Freie Universität Berlin

Metagenomic analysis of viruses present in faeces of domestic pigs and wild urban rats

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by:

Jana Sachsenröder

From: Quedlinburg

year of submission: 2016

Bundesinstitut für Risikobewertung

Bearbeitungszeit: 13.04.2010 bis 29.02.16

Die Arbeit wurde am Bundesinstitut für Risikobewertung, Berlin

unter der Leitung von Herrn Prof. Dr. Reimar Johne durchgeführt.

Erster Gutachter:

Prof. Dr. Reimar Johne

Bundesinstitut für Risikobewertung

Abteilung 4 „Biologische Sicherheit“

Diedersdorfer Weg 1

12277 Berlin

Zweiter Gutachter:

Prof. Dr. Rupert Mutzel

Freie Universität Berlin

Institut für Biologie (Mikrobiologie)

Königin-Luise-Str. 12-16

14195 Berlin

Disputation am: 03.06.2016

DANKSAGUNG

Ich möchte diese Chance nutzen und mich an dieser Stelle bei all denjenigen zu bedanken, die mich auf dem Weg zu meiner Promotion und während dieser Zeit so tatkräftig unterstützt haben.

Mein größter Dank gilt dabei ganz klar Herrn Prof. Dr. Reimar Johne, meinem Betreuer am BfR. Seine stetige Unterstützung im Laufe der gesamten Promotionszeit sowie beim Schreiben von Publikationen, sein Vertrauen und seine Geduld unter anderem bei der Abschlussarbeit machten meine Promotion und ihr Gelingen erst möglich.

Ein herzlicher Dank gehört auch Herrn Prof. Dr. Rupert Mutzel für seinen unermüdlichen Einsatz als Betreuer und Begutachter, nicht nur von meiner Promotion. Trotz der unzähligen Anfragen treffen Studenten bei ihm immer eine offene Tür an.

Des Weiteren möchte ich der gesamten Arbeitsgruppe der Virologie danken und dabei vor allem meiner geliebten Bürogemeinschaft. Durch ihre Unterstützung und die unermüdliche Aufbauarbeit, sei es nur in Form eines offenen Ohres, habe ich auch in Zeit von Misserfolgen oder Ratlosigkeit doch immer gern mein Promotionsthema bearbeitet und nie Freude am Lösen der Probleme verloren. Ich danke euch beiden, Eva und Anika!

Aber auch außerhalb meiner Arbeitsgruppe haben mich zahlreiche Menschen am BfR durch vielfältige Hilfe stets tatkräftig unterstützt und mein Arbeiten hier positiv gestaltet. Besonders zu nennen sind da Britta, mein erster Supervisorin am BfR, Annette, Jens, Nora und Alex und all die Heinzelmännchen, die im Hintergrund unser Abreiten am BfR ermöglichen. Danken möchte ich an dieser Stelle auch Dr. Beatriz Guerra Román, die mich stets ermuntert hatte überhaupt den Weg zu gehen, eine Dissertation anzufertigen und einer der größten Fans meines Promotionsthemas war.

Ein ganz besonderer Dank gehört aber auch Dr. Sven Twardziok (FU Berlin) und Matthias Scheuch (FLI, Insel Riems). Ohne die beiden wären die bioinformatischen Auswertungen nicht möglich gewesen. Vor allem Sven hat nie

die Geduld verloren, mit mir auch die komplexesten Sachverhalte zu erarbeiten und hat mir immer gern seinen Arbeitsplatz für meine Auswertungen überlassen.

Eine ganz besonders wertvolle Erfahrung verdanke ich Prof. Ph.D. Eric L. Delwart (BSRI, San Francisco, USA), der mir nicht nur die Möglichkeit eines fünfmonatigen Praktikums in seiner Arbeitsgruppe ermöglicht hat, sondern mich auch sehr herzlich aufgenommen hat und all sein „Know-how“ zur Verfügung gestellt hat. Durch seinen Mitarbeiter Ph.D. Terry F. F. Ng habe ich während dieser Zeit unschätzbar viel gelernt und mich ganz neu in mein Promotionsthema verliebt!

Vielen Dank auch all den anderen Kooperationspartnern wie Prof. Dr. Rainer G. Ulrich (FLI) und den SFB 852, in dessen Rahmen meine Promotion erst ermöglicht wurde. Auch hier habe ich tolle Menschen kennengelernt, die mich für immer mitprägen werden und mit denen ich zusammen die Höhen und Tiefen das SFB 852 teilen durfte. Ich danke euch Fanny, Lena und Katha.

Last but not least gebührt natürlich meiner lieben Familie und meinen geliebten Freunden ein ganz dickes „Ich danke euch!“ für eure Verständnis, eure an mich Glauben, eure stetigen Ermunterungen und vor allem dafür, dass ihr mich immer aufgebaut habt, wenn ich gestresst und frustriert war! Besonders möchte ich dabei meinem Freund Niko danken, der ganz großartig das Meiste davon kompensiert hat. Danke Niko, dass du mich stets bei allem uneingeschränkt unterstützt hast!

TABLE OF CONTENT

ZUSAMMENFASSUNG

Im Darm von Menschen und Tieren lebt eine Vielzahl von Mikroorganismen, die wichtige Funktionen bei der Verdauung und bei der Abwehr von Infektionen besitzen. Einige der Mikroorganismen können allerdings auch zu Erkrankungen führen. Die Darmflora setzt sich aus Bakterien, Pilzen und Protozoen zusammen, aber auch eine Vielzahl von Viren ist im Darm zu finden. Viren können ebenfalls als Krankheitserreger fungieren. Die Viren der Bakterien – die Bakteriophagen – können darüber hinaus das Wachstum von Bakterien regulieren und Gene übertragen. Während die Zusammensetzung der bakteriellen Darmflora in der letzten Zeit durch Metagenom-Analysen schon gut untersucht wurde, ist über die Zusammensetzung der Viren (das sogenannte Virom) im Darm bisher nur wenig bekannt. Dies ist vor allem auf methodische Schwierigkeiten zurückzuführen, die große Vielfalt der in einer Probe enthaltenen Viren zu identifizieren. Die Entwicklung neuer Technologien zur Hochdurchsatz-Sequenzierung von DNA - zusammenfassend als Next Generation Sequencing (NGS) bezeichnet – hat aber in den letzten Jahren Metagenomanalysen stark vereinfacht und damit auch die Analyse von Viromen möglich gemacht. Während erste Daten zur Zusammensetzung von Viromen des Menschen vorliegen, sind die Virome von Tieren bisher nur wenig untersucht worden.

Ziel der hier vorgestellten Untersuchungen sollte deshalb die Analyse von Viromen von Tieren sein. Da die Darmflora besonders komplex aufgebaut ist und viele krankmachende Viren über den Kot übertragen werden, sollten die Metagenomanalysen an Kotproben durchgeführt werden. Hierbei sollte sowohl die generelle Zusammensetzung des fäkalen Viroms ermittelt werden, als auch der mögliche Einfluss von Faktoren wie Alter und Ernährung auf diese berücksichtigt werden. Zunächst sollte hierfür eine NGS-basierte Methode zur sicheren und reproduzierbaren Analyse des fäkalen Viroms entwickelt werden. Danach sollte die Methode an zwei ausgewählten Tierarten angewendet werden. Schweine wurden ausgewählt, weil diese Tierart ein wichtiges landwirtschaftliches Nutztier in Deutschland darstellt, deren Darmgesundheit große ökonomische Bedeutung hat. Zunächst sollte die Zusammensetzung der Viren in Fäzes von Schweinen ermittelt werden. Danach sollte der Einfluss einer Fütterung mit probiotischen Bakterien

und des Alters der Schweine untersucht werden. Als zweite Tierart wurden wildlebende Ratten ausgewählt, weil diese als Reservoire von Krankheitserregern des Menschen bekannt sind. Die in den Fäzes erhaltenen Viren wurden vor allem hinsichtlich ihrer möglichen zoonotischen Übertragbarkeit auf den Menschen analysiert.

Die entwickelte Analyse-Methode beruht auf Filtrationen und Zentrifugationen, die alle Viruspartikel aus den Fäzes reinigen und konzentrieren. Anschließend wurde die Gesamt-Nukleinsäure aus den Viruspartikeln extrahiert, vermehrt und mittels NGS sequenziert. Bioinformatische Analysen ermittelten aus den Sequenzen die Zusammensetzung der Viren in der Probe. Ein Kontrollsystem, bestehend aus drei Bakteriophagen, die der Probe zugegeben wurden, wurde zur Optimierung der Methode und als Qualitätskontrolle entwickelt und benutzt.

Im Ergebnis konnten in acht Pool-Proben aus Fäzes von jeweils sechs Ferkeln beziehungsweise jungen Schweinen etwa 205 unterschiedliche Virusarten aus 36 verschiedenen Virusfamilien nachgewiesen werden. Die hauptsächlichen Virusgruppen waren Schweineviren und Bakteriophagen. Pflanzenviren, die im menschlichen fäkalen Virom häufig detektiert wurden und die wahrscheinlich aus Gewürzen aus der Nahrung stammen, wurden in den Schweineproben kaum vorgefunden. Mit Hilfe der Metagenom-Analyse konnte auch ein bisher unbekanntes Schweinevirus (als PigSCV bezeichnet) identifiziert werden.

Bei vergleichenden Untersuchungen von fäkalen Viromen aus einem Fütterungsversuch von Sauen und ihren Ferkeln mit dem probiotischen Bakterium *Enterococcus faecium* NCIMB 10415 konnte generell eine große Variabilität ihrer Zusammensetzungen festgestellt werden. Das fäkale Virom wurde allerdings kaum von der probiotischen Fütterung beeinflusst, sondern war vor allem stark vom Alter der Tiere abhängig. Während die jüngsten Ferkel einen hohen Anteil von Schweineviren und einen geringen Anteil von Bakteriophagen aufwiesen, war dies bei den Sauen genau umgekehrt. Darüber hinaus nahm die Diversität der Zusammensetzung der enthaltenen Viren mit dem Alter deutlich zu.

Bakteriophagen ignorierend, bestanden die fäkalen Virome von 20 wildlebenden Ratten, die aus der Innenstadt von Berlin stammten, hauptsächlich aus bekannten Rattenviren aus den Virusfamilien *Parvoviridae* und *Picobirnaviridae*. Es wurden

jedoch auch bisher unbekannte Picorna-, Bocaparvo- und Sapoviren sowie neue zirkuläre DNA-Viren identifiziert. Erstmals wurde ein Gruppe A-Rotavirus in Ratten nachgewiesen. Rotaviren sind als Gastroenteritis-Erreger bei Tieren und Menschen bekannt. Die Analyse des Gesamtgenoms des Ratten-Rotavirus zeigt, dass es eng mit Rotaviren von Menschen und Tieren verwandt ist und deshalb möglicherweise auf diese übertragen werden kann.

Zusammenfassend zeigen die Untersuchungen, dass die entwickelte Methode gut für die Analyse von fäkalen Viromen geeignet ist. Die Anwendung der Methode offenbart eine hohe Variabilität der Virome, die vor allem vom Alter der Tiere abhängig ist. Mit Hilfe der Methode konnten sowohl virale Gemeinschaften beschrieben werden als auch pathogene Viren detektiert und neue Viren identifiziert werden. Weiterführende Untersuchungen sollten einerseits die Methode vereinfachen, um größere Probenzahlen untersuchen zu können. Andererseits sollte insbesondere die bioinformatische Analyse der Sequenzdaten weiterentwickelt werden, um Virome in Zukunft noch umfassender und genauer bestimmen zu können und die Analyse von Wechselwirkungen mit deren Wirt und anderen Mikroorganismen zu ermöglichen.

SUMMARY

The gut of humans and animals comprises a high number of microorganisms, which exert important functions during food digestion and defense of infections. Some of the microorganisms can cause diseases. The gut flora is comprised of bacteria, fungi and protozoa, but a large number of viruses is also present in the gut. Viruses can also be agents of disease. In addition, the viruses of bacteria – the bacteriophages – can regulate the growth of bacteria and transfer genes. The composition of the bacterial gut flora has recently been analysed using metagenomic methods. In contrast, only little is known about the composition of viruses (the so-called virome) in the gut. This is mainly due to methodological problems to identify the large variety of viruses present in a sample. However, the development of high-throughput sequencing techniques – summarized as Next Generation Sequencing (NGS) – has simplified the metagenome analyses during the last years and enables the analysis of viromes. Although first data on human viromes are available now, the viromes of animals have been only scarcely analysed.

The aim of the investigations presented here is therefore the analysis of viromes from animals. As the enteric flora is complex and many pathogenic viruses are transmitted by faeces, the metagenomic analyses was conducted using faecal samples. By this, the general composition of the faecal viromes was assessed, but also the possible influence of factors like age and diet was considered. To this end, an NGS-based method for the reliable and reproducible analysis of the faecal virome was developed first. Thereafter, the method was applied to two animal species. Pigs were selected as this animal species represents an important farm animal in Germany and its gut health has a high economic impact. The composition of the viruses in the faeces of pigs should be analysed first. The influence of feeding probiotic bacteria und the age of pigs was investigated. The second selected animal species were wild rats, because these animals are known as reservoirs for pathogens for humans. The viruses detected in these faeces should be mainly analysed regarding their potential for zoonotic transmission to humans.

The developed analytic method is based on filtrations and centrifugations to purify and concentrate all virus particles from the faeces. Whole nucleic acid was extracted from the virus particles, amplified and sequenced using NGS. Bioinformatic analyses were used to determine the composition of viruses in the sample based of the sequence data. A control system, which consisted of three bacteriophages added to the samples, was developed and used for optimization of the method and as quality control.

In total, 205 different virus species from 36 different virus families were detected in eight pooled faecal sample each derived from six piglets or six young pigs. The main virus groups were pig viruses and bacteriophages. Plant viruses, which are frequently detected in human faecal viromes and which presumably originate from spices of the diet, comprised only a very small fraction in the pig samples. Using the metagenome analysis, a so far unknown pig virus (designated PigSCV) was also identified.

In comparative investigations of faecal viromes from a feeding trail of sows and their piglets with the probiotic bacterium *Enterococcus faecium* NCIMB 10415 a large variability of their composition could generally be identified. However, the faecal virome was nearly not influenced by the probiotic feeding, but strongly by the age of the animals. In contrast to the youngest piglets, which showed a high proportion of pig viruses and a lower proportion of bacteriophages, the sows showed the opposite picture. In addition, the diversity of viral communities increased significantly with the age.

Despite bacteriophages, the faecal viromes of 20 wild rats, which originated from the city of Berlin, mainly consisted of known rat viruses from the virus families *Parvoviridae* and *Picobirnaviridae*. However, so far unkown picorna-, bocaparvo- and sapoviruses as well as novel circular DNA viruses were also identified. Group A rotaviruses were detected for the first time in rats. Rotaviruses are known as pathogens causing gastroenteritis in animals and humans. The analysis of the whole genome of the rat rotavirus showed that it is closely related to rotaviruses from humans and animals and therefore it may possibly transmissible to them.

In conclusion, the investigations show that the developed method is well suited for the analysis of faecal viromes. The application of the method shows the high

variability of the virome, which is mainly dependent on the age of the animals. Using the method, we were able to describe viral communities and to detect pathogenic viruses as well as novel viruses identified. Further investigations should simplify the method in order to enable the analysis of higher sample numbers. In addition, the bioinformatic analysis of sequence data should be further developed in order to determine the viromes in more broadness and detail and to enable the analysis of interactions with other microorganisms and the host.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A.D.	Anno Domini
BLAST	Basic Local Alignment Search Tool
BoSCV	bovine stool associated circular virus
bp	base pairs
C++	programming language C
CAE	capillary arrays electrophoresis
cap	capsid protein
ChiSCV	chimpanzee stool associated circular virus
CsCl	cesium chloride
CV-like	circovirus-like
DDBJ	DNA Data Bank of Japan
DNA	desoyribonucleic acid
ds	double stranded
E. faecium	*Enterococcus faecium*
e.g.	exempli gratia for for example
EBI	European Bioinformatics Institute
ENA	European Nucleotide Archive
E-value	expect-value
FSfaCV	fur seal feces-associated circular DNA virus
g	gramme
GSP	Genome Sequencing Program
HAV	hepatitis A virus
HBV	hepatitis B virus
HCV	hepatitis C virus
HEV	hepatitis E virus
HGP	Human Genome Project
HIV	human immunodeficiency virus

ICTV	International Committee on Taxonomy of Viruses
IDIR	infectious diarrhea of infant rats
INSDC	International Nucleotide Sequence Database Collaboration
kb	kilobase
Mb	megabase
Mbp	megabasepair
NCBI	National Centre of Bioinforamtics
NCIMB	National Collection of Industrial, Food and Marine Bacteria
NGS	next generation sequencing
NHGRI	National Human Genome Research Institute
NIG	National Institute of Genetics
nm	nanometer
NoV	norovirus
nt	nucleotide
PCR	polymerase chain reaction
PEDV	porcine epidemic diarrhea virus
pH	pondus Hydrogenii
PigSCV	pig stool associated circular virus
PoSCV	porcine stool-associated circular ssDNA viruses
qPCR	quantitative real-time PCR
R	programming language R
ratSCV	rat stool associated circular virus
RAV	rotavirus group A
RBV	rotavirus group B
rep	replication initiator protein
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid;

RT	reverse transcription
RT-PCR	reverse transcription polymerase chain reaction
RT-qPCR	quantitative reverse transcription PCR
S	svedberg units
SCV	stool associated circular virus
SDA	strand displacement amplification
SIA	sequence-independent amplification
SIB	Swiss Institute of Bioinformatics
ss	single stranded
TFF	tangential flow filtration
TGS	Third Generation Sequencing
TTSuV	torque teno sus virus
TuSCV	turkey stool associated circular virus
USA	United States of America
µm	micrometer

1 GENERAL INTRODUCTION

The field of virology established around the turn of the 20th century when viruses were first discovered and recognised as a separate class of microbes. The focus of interest at that time was based on the assumption that viruses are the origin of infectious diseases (Levine and Enquist, 2007). Human pathogenic viruses and viruses, which are transmitted from animals to human (so-called zoonotic viruses), are of particular interest for human health. The idea of viruses as infecting agents is very old; the first use of the term virus can be traced to Cornelius Aulus in 50 A.D. However, the meaning of the term “virus” was quite different compared to today. Derived from Latin, the word “virus” was associated with venom or similar poisonous fluids (Lwoff, 1957). The first non-bacterial infectious agent was the Tobacco mosaic virus described in 1886 by Adolf Mayer and later verified by Beijerinck in 1898 (Bos, 1999). Since that time around 3200 virus species have been defined by the International Committee on Taxonomy of Viruses (ICTV, 2014) with the assumption that this is only a small number of the existing virus species.

Nowadays we know that viruses can be found in almost every ecosystem. It is assumed that viruses are the most abundant group of entities on earth (Edwards and Rohwer, 2005; Crawford, 2011). Nevertheless, our knowledge about viruses is limited. When compared to all postulated virus species, very few viruses are detected and most of the viral communities remain uncharacterized. Viruses are important beyond their role as agents of diseases. Some researchers assume that viruses have the ability to colonize organisms as commensal organisms, as seen in commensal bacteria (Blaser and Valentine, 2008). Viruses, especially bacteriophages (so-called phages) and retroviruses, also play a significant role for the horizontal gene transfer and are therefore very important for the processes of evolution (Canchaya et al., 2003a). Because viruses have a relatively simple molecular biological system, they are used to investigate and manipulate the function of cells (Valdez-Cruz et al., 2010). As intracellular parasites, viruses are important influencers of global nutrient cycles and are thus responsible for shaping ecosystems (Suttle, 2005).

In the past, the main scientific issue in virology was to detect and characterize single virus species. However, the development of new sequencing technologies in the mid-2000s enabled the simultaneous characterization of whole virus communities, which contain many different virus species, as well as the detection of entirely new viruses. The application of the techniques to samples of animals and humans may help to get insights into interaction of different viruses with their environment and their hosts. In addition, zoonotic viruses, which can be transmitted between animals and humans, may be more easily identified. It is expected, that the application of the new sequencing technologies will contribute to our understanding of virus transmission and the epidemiology of diseases.

1.1 Viral communities

Most habitats or organisms are colonized by different viruses creating so-called viral communities. Viruses interact with each other via different ways, e.g. competition for resources provided by host cells. They can be characterised as “invaders” and “predators” with respect to the host organism (Wooley et al., 2010). The composition of virus communities differs from habitat to habitat and every habitat can host its own viral community. A habitat could be any ecosystem, including sewage depots, bodies of water, soil, ocean or a segment of a host organism.

One possibility to characterize viral communities is to characterize their genomes, the so-called viral metagenome. A metagenome describes the whole genetic information of all organisms of interest living in one defined habitat. Metagenomics can therefore be understood as the study of all genomes present at a defined time point within one habitat. The term “**metagenome**” is also used to indicate that a nucleic acid is obtained directly from uncultured microorganisms found in environmental samples (Wooley et al., 2010) and includes the description of viruses without prior knowledge of them (DeLong, 2002; Chen and Pachter, 2005; Edwards and Rohwer, 2005). The entirety of the genetic information of all viruses in a defined habitat, the “viral metagenome”, is also called “**virome”.** In the case of complex animals like mammals, it is not the aim to determine the virome for the whole body but rather for particular organs like the intestine, lung or skin.

To get an idea of how many virus species make up a virome, Anthony and his team (2013) performed a study with samples from an Indian Flying Fox (*Pteropus giganteus*). By using statistical methods and extrapolating numbers of detected viruses related to nine virus families, they estimated a minimum of 320,000 mostly new mammalian viruses present in the sample. Breitbart and Rohwer (2005a) estimated that there are 10^{31} virus particles on Earth and that there are millions of different viral genotypes in one kilogram of marine sediment (Breitbart and Rohwer, 2005a).

The scientific field of viral metagenomics is quite young and has been technically possible only for about twelve years. The first viral metagenome projects were mainly focused on environmental samples, especially on water samples. One of the first studies was conducted by Breitbart and her colleagues in 2002, where they determined the virome of surface seawater. In the meantime, a large variety of different habitats has been analysed for the diversity of viral communities. Table 1 provides an overview of selected virome studies, including the pioneer virome studies.

One common observation of metagenomic studies was that the virome varied remarkably regarding the number of identified virus species within one virus community. For instance, between 374 and 7114 virus species could be identified in different seawater samples (Breitbart et al., 2002). In marine sediments, between 10^3 and 10^6 virus species could be identified (Breitbart et al., 2004).

Table 1: Overview on virome studies of different habitats

Material	Reference
Water	
fresh water	Djikeng, 2009
lake (Antartic)	Lopez-Bueno et al., 2009
sewage	Ng et al., 2012
surface water	Cole et al., 2003
ocean	Breitbart et al., 2002; Gitig, 2010; Labonte and Suttle, 2013
Others	
blood/ plasma	Breitbart et al., 2005a; Li et al., 2013
grapevine	Coetzee et al., 2010
insects: mosquito	Ng et al., 2012
rice paddy soil	Kim et al., 2008
soil	Williamson et al., 2003
tissue: brain; adrenal gland	Honkavuori et al., 2008
tissue: crop	Kister et al., 2008
urin	Santiago-Rodriguez et al., 2015

1.1.1 The gut virome and its composition

In the last five years, the intestinal tract of mammalians was of particular interest for viral metagenomics and a systematic analysis of mammalian faecal viromes has begun. Research interest in the gut virome is based on the significance of this habitat for health as well as the complexity of its composition (Bailey et al., 2005; Foca et al., 2015).

The intestinal tract of mammalians is colonized by diverse microbiota. This microbiota includes different organisms such as bacteria, archaea, fungi, parasites and viruses. It is known that the enteric flora of bacteria of different mammalian species and individual organisms is very diverse. The microbial diversity of the gut is based in part on the different functions of the gut. These functions range from the digestion of foodstuff and the synthesis of nutrients, like vitamins, to structural functions, such as the formation of an intestinal barrier for providing defence

against pathogens, and regulation of immunity (Berkhout, 2015; Foca et al., 2015; Hur and Lee, 2015; Gangarapu et al., 2014; Willing and Van Kessel, 2009). Due to a lack of suitable methods, very little was known about the compositions of faecal viromes until the availability of the Next Generation Sequencing (NGS) method (see chapter 1.3.3.3).

Nowadays, we know that the faecal virome represents a conglomeration of many kinds of different viruses. Many viruses replicate within the cells of the intestinal tract, others replicate in enteric bacteria and other microorganisms. Furthermore, multiple viruses pass through the intestinal tract after their replication within the cells of different organs. (Christensen, 1989; Goodgame, 1999; Zhang et al., 2006). In addition, some viruses are passively ingested with the food. The faecal virome is therefore dominantly composed of bacteriophages, host specific viruses and transient viruses such as plant viruses or other viruses derived from foodstuffs.

The gut viromes of mammalian animals have been of special interest for several reasons. The investigation of mammalian faecal samples allows for relatively easy identification of enteric viruses in general, which may include pathogenic viruses and zoonotic viruses. Especially some of the viruses present in faeces of mammals living in close proximity to human and those farm animals can potentially pose a risk to human health (Christou, 2011; Firth et al., 2014). Enteric zoonotic viruses are described in more detail in paragraph 1.1.2.2. Furthermore, viruses may also play significant roles as animal pathogens. In farm animals, such as pigs, infectious enteric diseases can cause high economic losses due to deaths of severely affected piglets or reduced animal growth (Zhang et al., 2014a). In addition, the determination of mammalian gut viromes gives insights in the stability and dynamics of viromes under certain conditions as well as different influencing factors and the identification of new viruses and pathogens.

The earliest faecal virome studies used massive cloning and conventional sequencing, which was very time-consuming and expensive. The first faecal virome study was done by Breitbart et al. (2003) and determined the human gut virome. In 2005 a study on the faecal virome of horses followed (Cann et al., 2007).Many faecal viromes of other animal species have been investigated,

whereby most studies focus largely on mammals and are based on NGS techniques. When the investigations of this PhD thesis started, no studies on faecal viromes of farm animals or rodents had been published. By the end of the investigations, four porcine viromes (Belak et al., 2013a; Lager et al., 2012; Shan et al., 2011; Zhang et al., 2014a), one wild rodent virome (Phan et al., 2011) and one wild rat virome (Firth et al., 2014) were available in addition to the own publications. The increased interest is also reflected by the list of other mammalian faecal viromes determined thus far (table 2).

Table 2: Overview on faecal virome studies of different host

Host animal	Reference
bats	Donaldson et al., 2010; Li et al., 2010a;
bovine	Kim et al., 2012
cats	Zhang et al. 2014b
dogs	Li et al., 2011a
dromedary	Woo et al., 2014
feline	Ng et al., 2014
ferret	Fehér et al., 2014; Smits et al., 2013
fur seal	Sikorski et al., 2013a
horse	Cann et al., 2007; Li et al., 2015
human	Holtz et al., 2014; Kim et al., 2011; Minot et al., 2011; Minot et al., 2013; Ogilvie et al., 2015; Reyes et al., 2010; Sasaki et al., 2015; Wylie et al., 2014; Zhang et al., 2006
pig	Belag et., 2013a; Lager et al., 2012; Shan et al., 2011; Zhang et al., 2014a
pigeon	Phan et al., 2013a
rabbit	Stenglein et al., 2012
sea lion	Li et al., 2011b
turkey	Day et al., 2010
wild rat	Firth et al., 2014
wild rodent	Phan et al., 2011

The number of detected virus species derived from mammalian faecal samples is very different. For example, up to 1200 virus species could be detected in human faeces (Breitbart and Rohwer, 2005a), whereby in equine faeces only 223 virus species were detected (Cann et al., 2007). Reyes et al. (2010) investigated human twins and their mothers and could detect between 52 and 2773 virotypes per faecal sample. These significant differences are evident between different starting materials, but also within the same starting material.

1.1.1.1 Bacteriophages

The term "**bacteriophages**" (so-called phages) summarizes all viruses inflecting bacteria (Breitbart and Rohwer, 2005a). According to most viral metagenome studies, it currently seems that most faecal viromes are dominated by bacteriophages.

A high abundance of bacteriophages was already postulated by Wilcox and Fuhrman (1994), which extrapolated from the number of prokaryotes that bacteriophages are the most abundant biological entities in the world. This assumption was supported by Wommack and Colwell (2000) who estimated that there are around ten bacteriophages per bacteria. In total, there might exist up to 10^{14} bacteria in the human intestinal tract (Kim and Ho, 2010), leading to the assumption that 10^{15} bacteriophages can be found in the human intestinal tract. Breitbart and Rohwer (2005a) estimated that there are 10^{31} viruses on earth, whereby most of them infecting bacteria.

Phages also have an ecological role in regulating bacterial growth (Kim et al., 2008). As predators of bacteria and due to their enormous numbers, bacteriophages have a very strong impact on the microbial biomass and therefore play an important role for several ecosystem functions (Angly et al., 2005; Engelhardt et al., 2014). They have a strong impact as a controlling factor for bacteria and archaea mortality. Due to killing of prokaryotes they are also responsible for the release of many nutrients, which are thereafter available for other organisms, thus further influencing the ecological systems (Clokie et al., 2011; Engelhardt et al., 2014; Suttler, 2007).

1.1.1.2 Enteric viruses and their zoonotic potential

The term "**enteric viruses**" summarizes viruses from different mammalian virus families, which can be found in the intestinal tract (Kotwal and Cannon, 2014). Analysis of viral metagenomes revealed enteric viruses to be the second most abundant group of viruses within faecal viromes.

Amongst the enteric viruses, several are the cause of serious diseases and are thus of considerable concern worldwide. Many of the known enteric viruses are associated with gastroenteritis and diarrhoea, but can also show other organ manifestation. Enteric viruses are predominantly transmitted by the faecal-oral route (FAO/ WHO, 2008; WHO, 2008 and 2011). One of the first detected enteric viruses was the Norwalk virus, identified in 1972 by Kapikian (Kapikian et al., 1972). This virus was the first detected member of the genus later designated as *Norovirus* (Robilotti et al., 2015). The identification of the Norwalk virus represented a milestone as, prior to the discovery, the assumption persisted that only bacteria cause diarrhoea. Subsequent to this discovery, a large number of other enteric viruses was identified including astrovirus, calicivirus, encephalomyocarditis virus, enteric hepatitis viruses (hepatitis A virus [HAV] and hepatitis E virus [HEV]), enteric adenovirus, enterovirus, orthoreovirus, rotavirus group A and B [RAV, RBV], and sapovirus (Christensen, 1989; Glass et al., 2001; WHO, 2008 and 2011). Enteric viruses are the most common cause of viral gastroenteritis in humans. Noroviruses and rotaviruses together infect millions of humans woldwide and cause thousands of deaths every year (Eckardt and Baumgart, 2011; Lee et al., 2013).

Enteric viruses are not only important for humans, but also for mammalian animals, in which they also can cause several diseases like gastroenteritis. In farm animals, enteric viruses represent a notable problem for agriculture due to the economic impact caused by the increased cost of medical treatment for animals and/or loss of livestock (Halaihel et al., 2010; Koenen et al., 1999). In humans and animals, infections by the same virus do not necessarily lead to disease. While HEV infections in pigs seem to be subclinical, HEV infections in humans can lead to hepatitis (Chandler et al., 1999). In contrast, rotavirus infections in children and piglets may lead to serious diarrhoea with lethal consequence (Desselberger, 2014).

Usually individual virus species infect only a small range of hosts. In fact, many viruses are specific to certain species. The reason behind the host specificity of viruses is the complex interplay of the virus components with specific enzymes and other structures of the host cell such as cell surface receptors. These characteristics generally limit transmissibility to other hosts. For instance, the hepatitis C virus (HCV) only infects humans (King et al, 2012a). However, some viruses like HEV have acquired the ability to cross species barriers. HEV is known to infect several mammals, including humans, pigs, deer and monkeys (Meng, 2013), indicating an extension of the host range. Other viruses such as the human immunodeficiency virus (HIV) jumped from the original host species (monkeys) to another host species (human), losing the ability to infect the original host species (Faria et al., 2014; Sharp and Hahn, 2011). The ability of a virus to cross host species barriers is important to assess the risk for infection of humans. Pathogens, which are transmitted from animals to humans are called **zoonotic viruses**. Some of these zoonotic viruses are causing very serious infections in humans. Rabies virus and avian influenza are examples of important zoonotic viruses (Christou, 2011; Abolnik, 2014).

Some of the enteric viruses are considered zoonotic. The zoonotic infection of humans can occur by direct transmission from animals or ingestion of products from the food chain (Brugere-Picoux and Tessier, 2010; Machnowska et al., 2014). While the zoonotic transmission for some enteric viruses is clear, e.g. HEV from wild boar to humans (Aggarwal, 2013; Schielke et al., 2009) and RAV from rabbits to humans (Matthijnssens et al., 2006; Purcell and Emerson, 2008), it is so far unclear for many enteric viruses found in livestock. Examples are astrovirus (Kohl and Kurth, 2014, Tse et al., 2011), norovirus (Wang et al., 2005a), encephalomyocarditis virus (Deutz et al., 2003), kobuvirus and sapovirus (Dufkova et al., 2013; Kohl and Kurth, 2014; Meng, 2012) which are commonly found in pigs. For astrovirus and encephalomyocarditis virus, a zoonotic transmission route is assumed due to the close relationship of certain animal and human strains.

1.1.2 The gut virome of pigs and rats

Pigs are one of the most important farm animals worldwide. In Germany, about 12.4 million pigs are kept every year (BMELV, 2015). Farmed pigs are known to

harbour zoonotic viruses like swine influenza virus (Abe et al., 2015), or HEV (Meng, 2013) which result in increased risk of zoonotic infection through the high number of animals living in close proximity to humans. In addition, many porcine enteric viruses may pose a risk for humans as they have the potential to enter the food chain by contamination of meat and other food products during slaughter etc. (Leblanc et al., 2014; Schielke et al., 2009). A further reason for an increased interest in porcine gut viromes is the high economic impact due to financial losses from infected or deceased animals. Swine pathogens such rotavirus cause serious diarrhoea, especially in piglets after weaning, resulting in significant economic losses for farmers and the food production industry (Cooper, 2000; Papp et al., 2013; Zhang et al., 2014a).

The **Norway or Brown rat** (*Rattus norvegicus*) is broadly distributed through Europe and is mainly found in a synanthropic habitat often linked to waste and wastewater or to crop fields near water (Amori and Cristaldi, 1999). Because of the very close proximity to humans, there is an increased risk of human infection by rat-born zoonotic pathogens. The list of pathogens transmitted by rats is long and includes bacteria (e.g. *Yersinia pestis* causing human plaque) and viruses, which can cause serious diseases. Furthermore, the list includes viruses with both clear zoonotic potential such as the Seoul hantavirus and cowpoxvirus (Campe et al., 2009; Himsworth et al., 2013; Meerburg et al., 2009) and those where the zoonotic potential is so far unclear, e.g. ratHEV (Johne et al., 2010), herpesviruses (Ehlers et al., 2007) and papillomaviruses (Schulz et al., 2012). Several different ways of transmission are possible, ranging from direct contacts with rats to inhalation and ingestion of virus-containing excretions. In addition, vector-transmitted infections by small arthropods acting as vectors are known.

1.1.3 Dynamics of the faecal virome

Like most biological systems, the gut virome is not stable and can be influenced by several factors, which are assumed to influence viromes. These are abiotic and biotic factors, such as general diet including the application of nutritional factors, treatments with medicine, changes due to illness, age, sex, and genetic relationship. Moreover, global factors like geographical differences can influence the virome composition.

The possibility for viruses to replicate in the host is strongly dependent on cell surface receptors and cellular enzymes as well as activity of the host immune system (Bailey et al., 2005; Stevenson et al., 1997; van Drunen Littel-van den Hurk and Watkiss, 2012). However, both factors are affected by the age of the host and by the developmental status of cells. Depending on the development stage and the age of cells, receptors on the cell surface and the enzyme repertoire of the organs and tissues can be different. The potential of viruses to infect the cell is dependent on the receptors existing on cell surface. In addition, the status of the immune system is host age-dependent. In infantile mammalians the adaptive and the innate immune system is not well developed and it takes time until it achieves the full protective status.

Phages and enteric viruses could be affected by the age of the host animal. Composition of the bacterial flora, which serves as a host reservoir for bacteriophages, changes repeatedly within the lifespan. For example as part of the weaning process and thus changes in the piglets' diet, diversity of the bacterial flora is increased. (Actis, 2014; Zentek et al., 2013). As consequence of this, the composition of the bacteriophages can be suspected to also change along with the composition of the virome (Minot et al., 2011). Changes in the bacterial gut flora lead also to changes in the intestinal mucosa. This may affect the replication of enteric viruses via a changed access to intestinal epithelial cells.

In summary, it must be assumed that viromes are strongly influenced by the age of the host organism. However, the effect of host age on the composition of the virome has been only poorly investigated so far. At the beginning of this thesis, only one study investigating the influence of the age on the human virome composition had been published. This study gave indications of a correlation. Breitbart et al. (2008), could detect only eight different virus genotypes in faeces of babies, while they postulated up to 1900 different genotypes in those of adults (results based on mathematical modeling).

Nutritional factors may also influence composition of the faecal virome. Changes in the diet will create different environments for enteric bacteria, which may act as hosts for bacteriophages and as targets of the immune system. In addition, direct effects on the virus stability or on the activity of immune cells are conceivable

(Kato and Ishiwa, 2015). Some specialized diets use probiotic bacteria for modulation of the gut health. **Probiotics** are defined as living organisms conferring a health benefit to the host when they were delivered in a sufficient amount (FAO/ WHO, 2001.).Probiotics have to fulfil certain criteria (Fontana et al., 2013): (i) They may not cause pathogenic-, toxic- and adverse side effects. (ii) They shall be able to survive and colonize the gastrointestinal tract. (iii) They shall be present in an adequate number of viable cells in the products in which they were delivered to the recipients.

The mode of action of probiotics is unique for each individual strain and the effects mainly depend on the delivered dose, but also on the application route and the intake frequency (Power et al., 2014). Several mechanisms have been postulated for the potential influence of probiotics on the composition of the faecal virome. First, they could create a direct effect by changing the physiological conditions through the production of distinct metabolisms. Second, they could affect the bacterial flora and therefore the number of hosts for bacteriophages. Probiotic strains compete for limited resources with the commensal microflora; they could produce antimicrobial agents or they could block the adherence of pathogens. Finally, there could be a modulation of the immune response through interactions with the gut-associated immune system (Kotzampassi and Giamarellos-Bourboulis, 2012). Probiotic strains can increase the production of B and T cells, natural killer cells and they can regulate the production of pro- and anti-inflammatory cytokines (Power et al., 2014; Fontana et al., 2013). In this manner, they affect the immune response against virus infection and also against bacteria, consequently leading to a changed enteric virus and phage population.

Although, different effects of probiotic treatment on the composition of faecal viromes have been postulated, no metagenomic studies addressing this question have been published. So far, only a few studies using other techniques were published. They are showing effects against viral replication by blocking of viral attachment, an increased resistance of epithelial cells against virus-induced lysis or the secretion of different compounds that protect epithelial cells from viral infection (Colbere-Garapin et al., 2007; Freitas et al., 2003). Most of the studies focused on rotavirus infection. A positive effect of different probiotic strains on rotavirus-induced gastroenteritis has been reported by several *in vitro* and *in vivo*

studies (Fang et al., 2009; Munoz et al., 2011; Zhang et al., 2013). Studies on the effect of probiotics on further enteric viruses are scarce. The beneficial effects of probiotic strains on liver diseases caused by viruses has been investigated for HCV and hepatitis B virus (HBV) (Imani-Fooladi et al., 2013; Lee et al., 2013; Loguercio et al., 2005). An inhibitory effect against transmissible gastroenteritis coronavirus has been also demonstrated (Kumar et al., 2010). Scarce are also studies focused on other porcine pathogenic viruses like porcine reproductive and respiratory syndrome virus (Kritas and Morrison, 2005).

1.2 Virus diversity as challenge for investigation of viromes

To analyse viral communities, the aim would be to develop one protocol suitable for simultaneous detection of all viruses independent of virus characteristics like genome type or virus structure. The difficulty in establishing such a method is reflected by the high number of different protocols published for metagenomic analysis of viruses as reviewed by Hall et al. (2014). The major challenge is the high diversity of the different viruses and the absence of any genomic sequence common to all viruses. The general characteristics of viruses compared to other types of organisms are summarized in table 3.

Table 3: Comparison of properties of viruses, bacteria and eukaryotic cells

	Viruses	Bacteria	Eukaroyota
Size	15 - 1000 nm	500 nm - 10 µm	1 - 30 µm
Growth	dependent on host cell machinery	autonomous	autonomous
Ribosomes	no	yes	yes
Genome size	0.5 - 1000 kb	0.5 - 10 Mb	10 - 50 Mb
Genetic material	RNA an DNA	DNA	DNA
Conserved sequences among all families	no	yes	yes
Gene number*	2[1)] - 2500[2)]	500[3)] - 4500[4)]	up to 30 000

*as far as known: [1)]Porcine circovirus type 1; [2)]*Pandoravirus salinus*; [3)]*Mycoplasma genitalium* [4)]*E. coli*

The next paragraphs will discuss some of the specific properties of viruses and the consequences for metagenomic analysis of virus communities in more detail.

1.2.1 Parasitic lifeform

As previously mentioned, viruses are obligate intracellular parasites. They require the biochemical machinery of the host cell for their own reproduction, a circumstance, which is also reflected in the definition of viruses: “Viruses are small infectious particles which can infect almost all types of living cells including bacteria and archaea and can replicate only inside of living cells or organisms” (Yoon, 2012). Therefore, viruses do not belong to living organisms or living cells (Koonin et al., 2006). Whereas living organisms always possess both nucleic acid types: RNA and DNA, viruses contain only either DNA or RNA as genetic material.

The obligate parasitic lifeform represents challenges for the detection of viruses by molecular biological methods and for the cultivation of viruses. As viruses do not encode own ribosoms, they do not have the highly conserved ribosomal RNA genes like the 16S rRNA in bacteria and archaea and the 18S rRNA in eukaryotes. As a result, those target genes cannot be used to detect poorly characterized viruses or for phylogenetic analysis. An additional problem presented by the parasitic lifeform is that virus cultivation is only possible when the virus host is known and when the cultivation conditions of the host specific host cells are known, which is often not the case. Without cultivation possibilities, viruses cannot be propagated from a sample, but must be analysed directly, which often calls for very sensitive detection methods.

1.2.2 Virus diversity: particle shape, size and genome type

As mentioned above, viruses are very diverse with regard to serval properties. To analyse virus communities, the simultaneous identification of all viruses present in a sample is necessary. The applied protocols are always a compromise regarding the different characteristics.

Virus particles generally consist of two parts: i) the genetic material and ii) a protein capsid to protect the genome. In some cases iii) an additional lipid envelope surrounding the protein capsid (Harrison, 2007; Yoon, 2012).

Protocols of virus particle separation often use physical characteristics like the size or density of the virus particles. However, virus particle shape varies from simple helical or icosahedral forms to very complex forms with several different

components, and some viruses have filamentous forms. In particular, bacteriophages like Enterobacteria phage T4 have a very complex shape. Enveloped viruses appear to be spherical. Some examples of virus particle shapes are given in figure 1, which shows an overview of the most common human viruses with their relative particle size. However, particle morphology alone is not sufficient to identify a virus species because there may be almost no differences in particle shape and morphology between viruses of different genera or families.

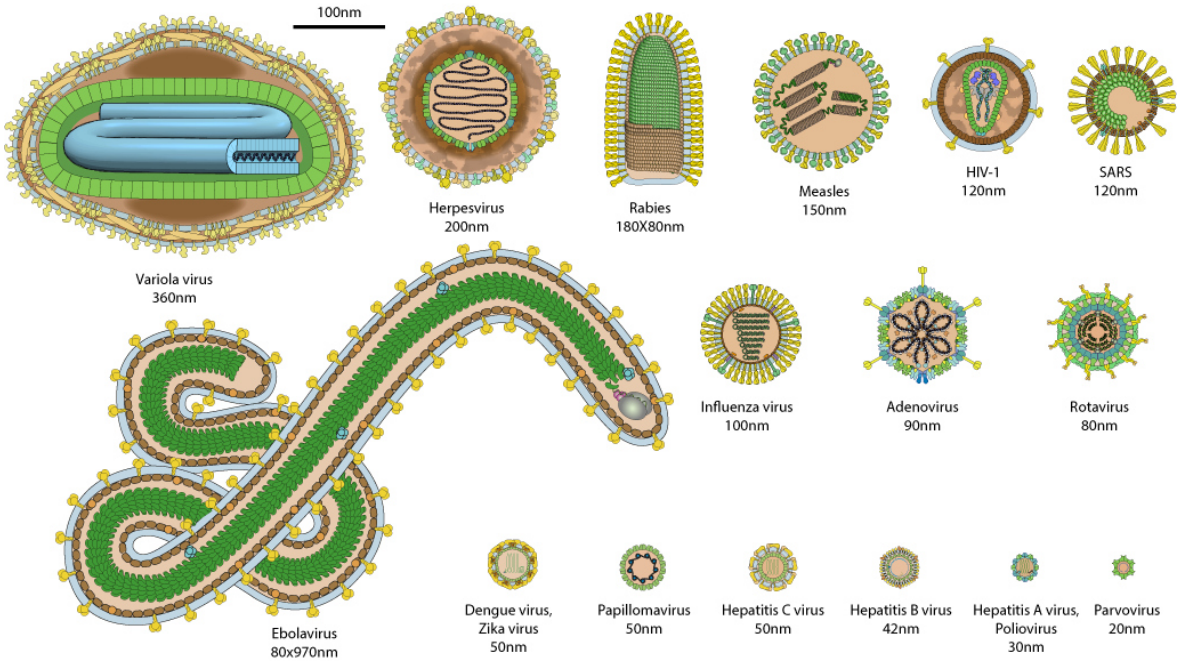


Figure 1: Examples of virus particle shapes and their relative size (viralzone 2015, SIB Swiss Institute of Bioinformatics; www.expasy.org/viralzone).

The fact that virus particles – with a few exceptions – are smaller than all other organisms (see table 3) could be used to separate most viruses from bacteria or host cells by using ultrafiltration. However, within all viruses the range in particle size is quite large, spanning from around 20 nm to 300 nm (Carstens, 2012). The range of particle size should be considered when applying filtration and centrifugation based purification methods, particularly because some viruses are showing extreme sizes. For example, some filamentous *Filoviridae* like ebola virus have a total length of up to 14000 nm with a diameter of around 80 nm (Sanchez et al., 2007). In contrast, one of the smallest viruses is the porcine circovirus with a

17 nm diameter (Ellis 2014). One of the largest known viruses is the recently discovered virus *Pithovirus sibericum* with a 1.5 µm diameter related to *Megavirales* (Legendre et al., 2014; Sharma et al., 2015). An overview of sizes of some viruses in relation to the size of bacteria and eukaryotic cells is given in figure 2.

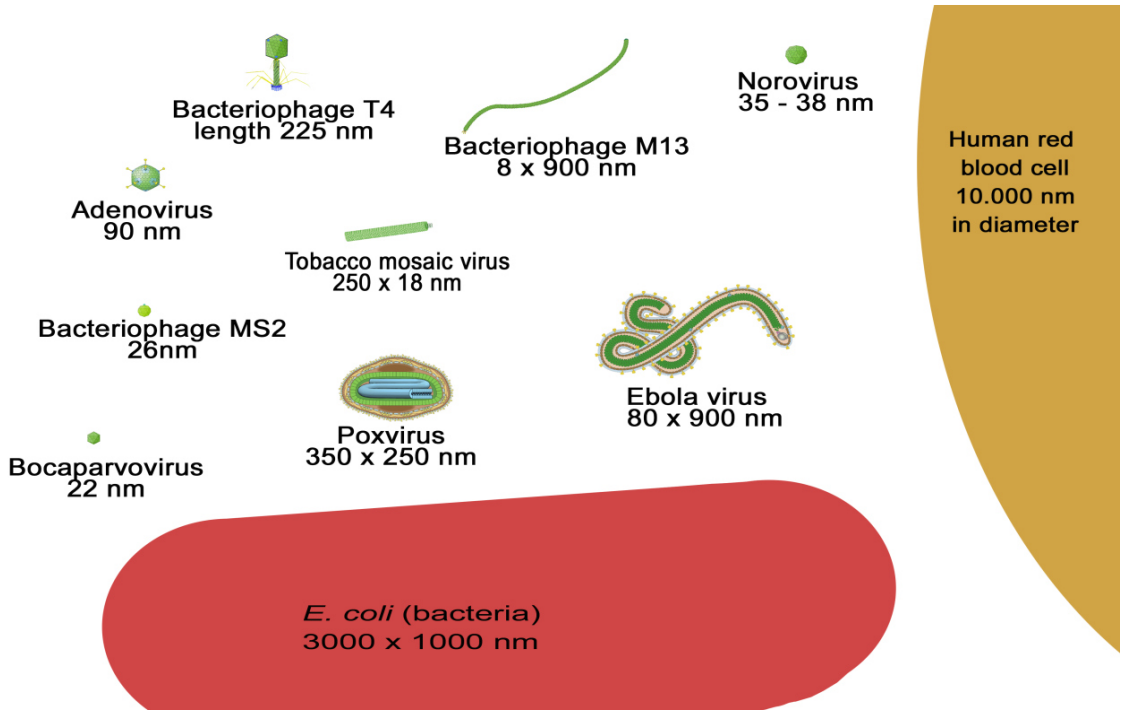


Figure 2: Relation of different viruses to bacteria and eukaryotic cells (Parts of the figure are provided from ViralZone, Swiss Institute of Bioinformatics, www.expasy.org/viralzone)

The **genomes of viruses** display also a large variety and much more variability than any other group of organisms. The genome size of different viruses may vary by a factor of 1,500 and ranges from 1,759 nt (Porcine circovirus type 1), encoding only two genes, to 2.5 Mbp in *Pandoravirus salinus,* encoding around 2500 genes (Yutin and Koonin, 2013). Further differences on genome types are presented in table 4.

Table 4: Schematic overview on viral genome types, with examples from different virus families

							
Host	**dsDNA**	**ssDNA**	**ssRNA+**	**ssRNA-**	**dsRNA**	**dsDNA-RT**	**ssRNA-RT**
Vertebrates	Herpesviridae 1 Adenoviridae 1 Polyoma-viridae 1 Papilloma-viridae 1 Alloherpes-viridae 1 Asfarviridae 1 Iridoviridae 1 Poxviridae 2	Parvoviridae 4 Circoviridae 5 Anelloviridae 5	Astroviridae 6 Caliciviridae 6 Picorna-viridae 6 Hepeviridae 6 Togaviridae 6 Corona-viridae 6 Flaviviridae 6 Arteriviridae 6 Nodaviridae 7	Bornaviridae 8 Rhabdo-viridae 8 Filoviridae 8 Paramyxo-viridae 8 Nyamaviridae 8 Bunyaviridae 9* Arenaviridae 9* Orthomyxo-viridae 10 Deltavirus 11	Birnaviridae 12 Picobirna-viridae 12 Reoviridae 14	Hepadna-viridae 15	Retroviridae
Bacteria	Myoviridae 1 Siphoviridae 1 Podoviridae 1 Tectiviridae 1 Corticoviridae 3 Sphaerolipo-viridae 1+2	Inoviridae 5 Microviridae 5	Leviviridae 6	/	Cystoviridae 13	/	/

1) one linear segment dsDNA
2) linear dsDNA, covalently closed
3) circular dsDNA
4) one linear segment ssDNA
5) one circular segment ssDNA
6) one linear segment ssRNA + sense
7) two linear segment ssRNA + sense
8) one linear segment ssRNA - sense
9) three linear segment ssRNA - sense
10) 6-8 linear segment ssRNA - sense
11) circular ssRNA
12) two segments dsRNA
13) three segments dsRNA
14) 9-12 segments dsRNA
15) circular partially dsDNA

* +/- orientation: + posivitiv sense;
- negativ sense
ds double stranded
ss singel stranded
DNA Deoxyribonucleic acid
RNA Ribonucleic acid

The genomes of viruses can differ regarding the nucleic acid type (RNA and DNA, single stranded- (ss-) or double stranded- (ds-) genomes) and genome topologies: circular, linear, linear with covalently closed ends and segmented genomes (Carstens, 2012). *Hepadnaviridae* present an additional complexity through the mixture of circular dsDNA genome with regions of single-stranded DNA. In the case of ss-genomes, the orientation of the coding genes also plays a role as both positive-sense (3' to 5' direction) and negative-sense strand virus genomes (5' to 3' direction) exist. This genomic variety complicates analysis of viromes using one

protocol for all genome types and could lead to overrepresentation or underrepresentation of particular genome types.

1.3 Methods to identify and characterise viruses and viromes

To characterise and identify individual virus species, traditional methods are very useful. This may include virus isolation, electron microscopy, polymerase chain reaction (PCR) techniques and Sanger sequencing, complemented by quantitative (qPCR) or reverse transcription PCR (RT-PCR). However, these methods are not very appropriate to characterise whole viromes, because they are not able to simultaneously identify multiple different viruses present in one sample. Therefore, traditional methods were partially replaced by modern sequencing technologies in combination with bioinformatics-based data analyses.

1.3.1 Traditional detection methods for viruses

The most commonly used methods for detection and characterisation of viruses can be generally divided into three categories: i) those that detect the viral genome (so-called “molecular biological method”); ii) those that detect antigenic virus components (so-called “serological method”), and; iii) those that study viral infectivity, using virus isolation in cells, bacterial cultures or experimental animals. Additional important tools to detect and characterise viruses are imaging techniques, like electron microscopy (Flint et al., 2015). However, electron microscopy is a method with low sensitivity and requires large amounts of virus particles for identification. Nevertheless, this method has the significant advantage that it is unbiased and allows the detection of unexpected and so far unknown viruses. Due to the limited numbers of different morphological shape types of viruses, however, the discrimination power of this method is comparatively low.

Virus isolation has been used for decades to identify novel viruses. Although this technique is widely used if the specific cell type and growth conditions of the virus are known, it is not efficient for screening virus communities due to the fact that each of the viruses would need its own specialized growth conditions. For a larger number of different viruses, cell culture systems remain unavailable.

Serological methods are based on the specific interaction of antibodies with antigenic structures of the viruses. As the availability of specific antibodies is a prerequisite for detection of a single virus species or genera, they are mostly used for specific virus detection and are not suitable for identification of currently unknown viruses or virus communities. Further difficulties include dealing with limited antigenic/serological cross-reactions.

Molecular biological methods are based on the detection of the viral genome. They comprise very powerful techniques, which are used for broad applications ranging from specific detection of virus strains to discovery of currently unknown viruses. Mostly, molecular biological methods work in two steps: amplification of nucleic acids followed by identification of nucleic acid sequence through sequencing techniques. For the analysis of viral communities, a combination of sequence-unspecific amplification of viral nucleic acid combined with DNA sequencing of all amplified products has been shown to represent a very efficient technique (Delwart, 2007; Kristensen et al., 2010). In the following paragraphs, these methods will be described in more detail as they have also been chosen for faecal virome analysis in the study presented in this thesis.

1.3.2 Nucleic acid amplification techniques

Polymerase chain reaction (PCR) was developed by Mullis and colleagues in 1983 (Mullis et al., 1986). The possibility to amplify DNA revolutionized the study of genomes, resulting in an enormous impact for almost all fields of biology. Ten years later, Walker and colleagues (Walker et al., 1992) developed another amplification method, the strand displacement amplification (SDA). In both methods, exponential DNA amplification results from coupling sense and antisense reactions based on strand displacement (Walker et al., 1992). Due to **sequence specificity** of primers used in PCR techniques, they are only suitable to identify known viruses, which contain the primer binding site in their genome (Hosono et al., 2003). Classical PCR assays are therefore not suitable to identify unrelated or new viruses from environmental samples.

Sequence-independent amplifications (SIA) were developed by using a mixture of primers with highly variable (random) sequences. They have been shown to be

suitable for metagenomic studies and are in nowadays wildly used. For analysis of viral communities, all virus particles present in a sample are first purified by centrifugation and filtration steps and then the viral nucleic acid is amplified by SIA. In contrast to PCR, SIA is not suitable to identify directly any species, because many different viruses will lead to detectable amplicons. Additional steps including sequencing of the amplicons have to be applied subsequently. The major advantage of SIA is to provide enough starting material for further analysis techniques such as DNA sequencing (Delwart, 2007; Schoenfeld et al., 2010).

1.3.3 DNA Sequencing

Due to the importance of DNA sequencing for virome studies, the following paragraph will include also a historical overview of sequencing techniques. DNA sequencing means any method or technology that is able to determine the desoxynucleotide (short: nucleotide) order in a DNA molecule. The possibility to determine the DNA sequence originated in late 70s and subsequently led to a revolution within many subjects of biological science, opening up the field of genomics. Nowadays, DNA sequencing is a very important key method for genetic and molecular biology. It also plays an important role in virology as it is used for identification of viruses. The historical development of sequencing techniques is summarized in table 5. It is divided into several major periods designated as first, second and third generation sequencing.

Table 5: Overview on the historical development of sequencing techniques

Time	Technical milestones	Use	Prinicple	Generation
1953	Watson and Crick discovered the DNA structure			
1972	First DNA sequencing method developed by Sanger	Sequencing short DNA fragments	Chain termination method	**First Generation Sequencing**
1972	First complete gene sequenced (from phage MS2)			
1976	First sequenced genome (from phage φX174)			
1977	Maxam and Gilbert developed chemical sequencing	Sequencing short DNA fragments	Chemical sequencing method	
1980's	Laser-based fluorescence dye detection combined with capillary electrophoresis	Whole-genome sequencing of single reference strains	Automated fluorescence-based sequencing	
1990's	Developing of pyrosequencing	Whole-genome sequencing of multiple strains	Automated pyrophosphate-based sequencing	
2000's	Start of the NGS era with 454 Pyrosequencer	First sequencing of metagenomes	Massive parallel sequencing	**Second Generation Sequencing (NGS)**
2010's	Several NGS platforms are available	NGS is used routinely in many fields		
open	Further development of NGS to TGS	Sequencing of single genome of single cells	Sequencing of single molecules	**Third Generation Sequencing (TGS)**

1.3.3.1 First Generation Sequencing

Methods for DNA sequencing were first developed by two independent groups led by Frederick Sanger in 1975 and by Allan Maxam and Walter Gilbert in 1977. The "dideoxy method" developed by Sanger used selective incorporation of chain-terminating dideoxynucleotides by a DNA polymerase during *in vitro* DNA replication (Sanger and Coulson, 1975; Sanger et al., 1977a). In recognition of Sanger, this method is still called Sanger sequencing. The Maxam and Gilbert method is also called chemical sequencing and is based on chemical modification of DNA and subsequent base-specific cleavage of the DNA (Maxam and Gilbert, 1977). To separate and visualize the fragments, both methods used gel electrophoresis and radioactive labelling. In recognition of their fundamental and

revolutionary contributions, Frederick Sanger and Walter Gilbert received the Noble Prize for Chemistry in 1980 (Royal Swedish Academy of Science, 1980).

For 25 years, the Sanger sequencing was the most widely used method. Compared to the Maxam and Gilbert method, the Sanger method was easier to use and required fewer hazardous reagents. It was also the Sanger sequencing technology, which was later further developed constituting the fundament for modern sequencing technologies. The two most important steps in the further development were the general automatization of the sequencing reactions and the development of base detection systems independent from electrophoresis and radiography. This was later realized by automated laser-based fluorescence dye detection in combination with capillary electrophoresis (Smith et al., 1986).

In 1992 Huang et al. presented a method using 25-capillary array electrophoresis (CAE), which was subsequently extended to 96-lane CAE by Paegel and his colleagues (Medintz et al., 2001). In 2005 Aborn et al. presented a system based on microfluidic plates with 768 DNA sequencing lanes. With these more efficient technologies, the lengths of the sequenced fragments were longer and researchers could determine whole small genomes. Nevertheless, using Sanger technologies, the sequencing of whole viromes or bigger genomes like that of mammals remained very time-consuming and extremely expensive. An additional problem was the analysis of the generated data. Further developments in bioinformatic tools were necessary to assemble sequencing reads to longer fragments, also called contigs (Luckey et al., 1990, Fancello et al., 2012; Kunin et al., 2008).

1.3.3.2 Shotgun sequencing

A general limitation of DNA sequencing is the generation of only short DNA fragments, designated as reads, with a theoretical length of up to 1000 bp. One solution for sequencing larger genomes was the so-called “primer walking”. After finishing a fragment, new primers were designed with binding sites on the previously determined sequence. Primer walking was first used in 1977 by Sanger to sequence the bacteriophage phi X174 (Sanger et al., 1977b). However, this

approach is quite time-consuming and therefore other strategies are needed for larger genomes.

One strategy to overcome this problem is the so-called shotgun sequencing. Genomes or longer DNA molecules are broken up randomly into many short fragments, which could be individually sequenced. By this, the generated fragments can be sequenced in parallel, which speeds up the process. The individual fragments are assembled to create the large original sequence using bioinformatic tools (Staden, 1979).

The combination of primer walking and whole-genome shotgun sequencing strategy enabled sequencing of the first mammalian genome, the human genome project (HGP). This project started in 1990 under the leadership of James Watson and was the world's largest collaborative biological project so far. It took 13 years to complete the first human genome sequence (Tripp and Grueber, 2011). In 1998, the commercial institute Celera started a second trial to sequence a human genome. They used a whole-genome shotgun sequencing strategy, which is faster, but also more risky for unbridgeable sequence gaps (Venter et al., 1996; Venter et al., 1998). To achieve the goal, Celera developed special algorithms for sequence assembly, which are commonly used for sequence assemblies until now (Huson et al., 2001; Venter, 2011).

1.3.3.3 The second generation of sequencing (NGS)

The combination of Sanger sequencing techniques and more efficient fluorescence detection methods has opened the way to the second generation sequencing method summarised to next generation sequencing (NGS). The term NGS summarises different novel approaches of DNA sequencing. They all share the high throughput, which means that up to 1.2 million single sequence reactions are performed at the same time. This allows a very deep insight into the sequenced sample. Therefore, these methods are also called “Deep sequencing”. The “deepness” means how many sequence reads per genome or fragment of interest are generated. The high throughput of the NGS platforms is achieved by miniaturization in instrumentation and microfluidic separation technologies but is still based on the sequencing principle developed by Sanger. The NGS platforms

differ in regard to DNA fragmentation, separation of the fragments, amplification strategies of the generated clones and the detection of incorporated nucleotides.

The most common NGS platforms are the 454 pyrosequencing, the Illumina technology and the Ion Torrent sequencing as summarized in table 6. Depending on the application, they have different advantages and disadvantages (Scholz et al., 2012). Pyrosequencing relies on the detection of pyrophosphate, which is released during the nucleotide incorporation of DNA strand extension (Ronaghi et al., 1996 and 1998). The Illumina technology follows the principle of sequencing by synthesis and reversible termination (Bentley et al., 2008), and Ion Torrent sequencing is based on ion semiconductor technology, whereby pH changes are detected (Grada and Weinbrecht, 2013).

Table 6: Comparison of commercially available next generation sequencing platforms

Platform-Name	Detection method	Read lengh	Applications	Comments	References
First Generation Sequencing					
Sanger-based capillary-based	fluorescent, dideoxy terminator	750-base reads	small molecules, fragments	most costly method, low throughput, high accuracy	Smith et al., 1986
Second Generation Sequencing					
Roche–454	light emitted from secondary reactions initiated by release of pyrophosphate	400-base reads	small genome, targeted sequencing, transcriptome, metagenomics, de novo-sequencing	cost limits, coverage is good	Ronaghi et al., 1996; Margulies, 2005
Illumina HiSeq	fluorescent emission from ligated dye-labelled nucleotides	100–150-base reads	large genomes, exome, targeted sequencing, transcriptome, metagenomics	very high coverage owing to high instrument output and low cost	Bentley et al., 2008
Ion Torrent	proton detection	> 200-base reads	large genomes, exome, targeted sequencing, transcriptome, metagenomics	high coverage, but longer reads than Illumina, low cost	Liu et., 2012
Third Generation Sequencing					
PacBio	fluorescent, single-molecule sequencing	Up to 10-kb- reads	de novo assemblies of genomes including long repetitive sequences	attractive for long reads, but low accuracy, limits applications	Karlsson et al., 2015
Oxford Nanopore	electronic signal as DNA passes through pore	no upper limit to the read length	singel cell analysis, single-molecule sequencing	not yet available	Karlsson et al., 2015

Compared to the first generation sequencing methods, NGS are less expensive, less laborious and less time-consuming. Therefore, it increased the speed and decreased the cost of genome sequencing. One main advantage is that NGS enables the sequencing of many genomes without any prior sequence information and makes it possible to get the genome information without any cultivation steps. (Delwart, 2007).

However, the technical further development was only one important step toward viral metagenomics. Similarly important was the price decline of NGS. The graph in figure 3 shows the price decline for sequencing only based on the used reagents and the raw sequencing process. However, additional costs occur through the time-consuming bioinformatics data analysis. In addition, costs have to be calculated for sample preparation, for staff and for equipment (Wetterstrand, 2015).

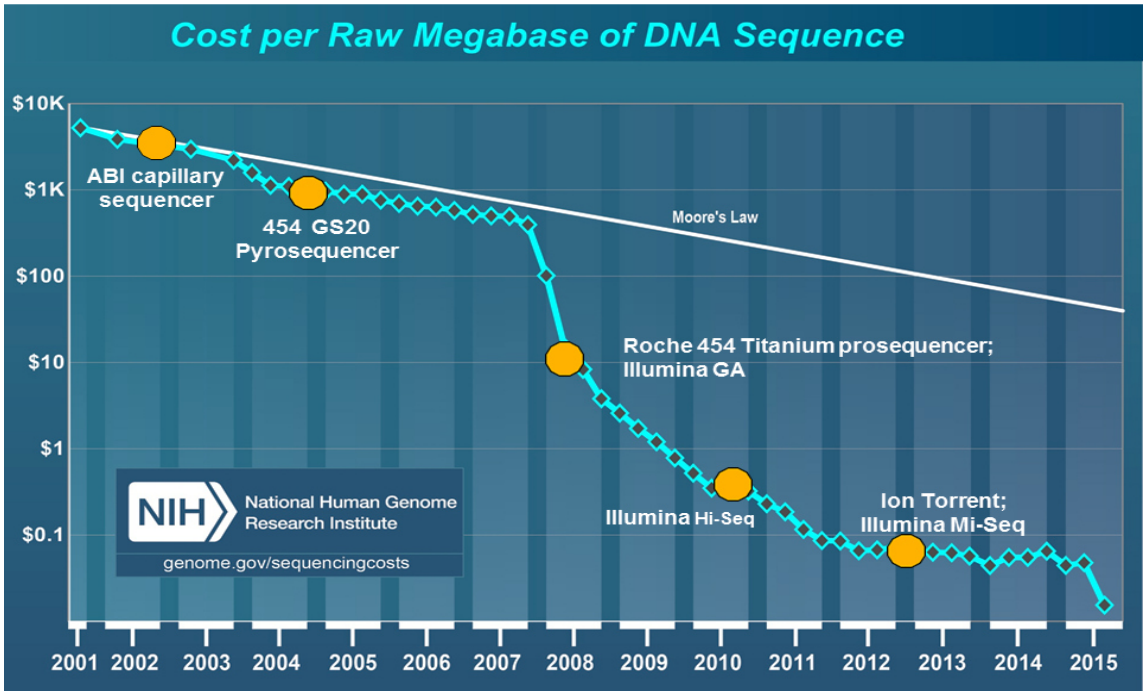


Figure 3: Development of DNA sequencing costs. From the NHGRI Genome Sequencing Program (Wetterstrand, 2015), available at: www.genome.gov/sequencingcosts.

1.3.3.4 Third Generation Sequencing

The Third Generation Sequencing (TGS) will make it possible to sequence directly single nucleic acid molecules in the future (Ozsolak, 2012; Diaz-Sanchez et al., 2013). Currently, such systems are not ready for the market, but several companies are working on it (Gut, 2013). In contrast to NGS using the Sanger sequencing principle, TGS will not need any prior amplification steps. The main advantage of these techniques would be the decreased amount of input DNA. Therefore, TGS will allow the study of single cells or single viruses. Further advantages will include an increased sequencing rate, longer read lengths, easier sample preparation, and further reduction of the costs (Schadt et al., 2010; Diaz-Sanchez et al., 2013).

1.3.4 Analysis of NGS data

Nowadays, the most time-consuming step in NGS is the analysis of the generated data. Therefore, supercomputing technologies and software solutions, which can handle the high amount of data are currently broadly developed (Yang et al., 2009). These challenges for the computing foster new fields within informatics and biology and the overlapping of these subjects mark the rebirth of the field of bioinformatics (Ueno et al., 2014).

The first step in NGS data analysis is the processing of the generated raw data, which is often an integral part of NGS platforms. This includes trimming of sequencing adaptors, quality control and removal of dubious reads with less quality, identification of contaminations and selection for a minimum read length (Karlsson et al., 2013). Especially for metagenomics, the identification and removal of systematic artefacts is essential. Mistakes, which may occur in this step, are in later phase not detectable and could lead to a shifted final picture of the metagenome (Gomez-Alvarez et al., 2009).

The second step is the DNA sequence analysis, which is often the most time-consuming step (Bzhalava and Dillner, 2013). How to analyse the data is dependent on the particular project. A metagenomic data analysis includes two main workflows: first, assembly of reads to contigs and second, the search for sequence homologies by comparison with appropriate databases to identify the

detected species or genes (Gut, 2013). A summary of the typical steps for NGS data analysis for identification of viromes are given in figure 4.

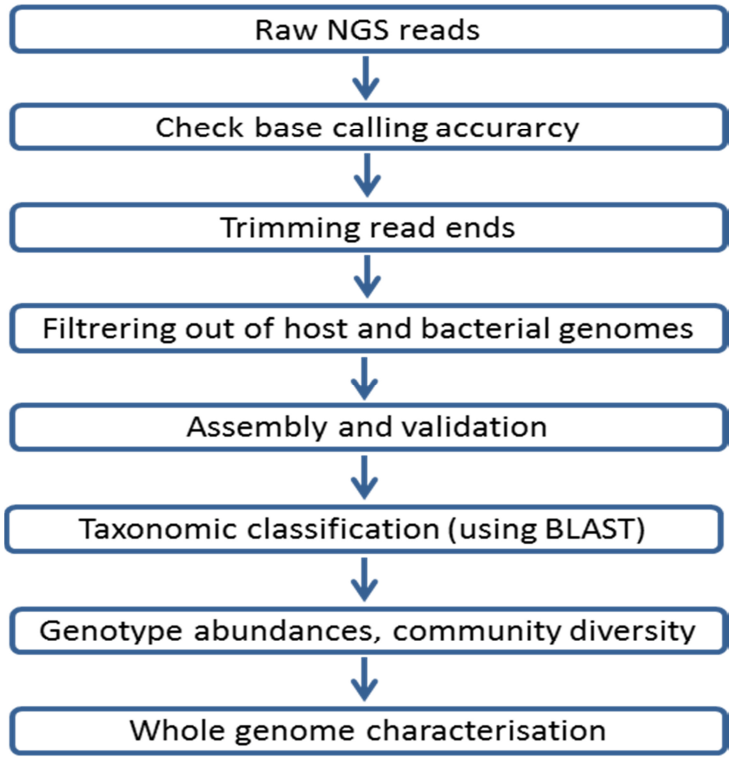


Figure 4: Schematic flowchart of viral metagenomic data analysis

The most important **databases** for metagenomics are three international databases for nucleotide sequences: GenBank; the European Nucleotide Archive (ENA) and the DNA Data Bank of Japan (DDBJ). GenBank is maintained by the National Centre of Bioinformatics (NCBI), ENA by the European Bioinformatics Institute (EBI) and the DDBJ by National Institute of Genetics (NIG) of Japan. All three databases are members of the International Nucleotide Sequence Database Collaboration (INSDC). They also provide several smaller databases, customized for special questions and applications.

Due to the huge number of available sequence data, they can only be analysed automatically by computer-assisted software algorithms. One of the mostly used tools for sequence homology search is the **BLAST** algorithm (Basic Local Alignment Search Tool) provided by NCBI. BLAST enables the comparison of

nucleic acid sequence and amino acid sequences with sequences of within the database (Altschul et al., 1990). Important software for sequence **assembly** are the Newbler software (Miller et al., 2010), Celera (Huson et al., 2001) or Bowtie (Langmead et al., 2012). Although several software tools have been developed during the last ten years, they cannot solve all questions and problems and additional programming skills are often necessary.

1.3.5 Application of NGS within virology

Within the field of virology, the most important applications of NGS are viral metagenomics, discovery of new viruses and identification of viral pathogens responsible for specific diseases. By the determination of the composition of viromes, the stability and dynamics of viral communities under certain conditions and influence of different factors may also be studied. Also in zoonosis research, outbreak investigations, monitoring of genetic variations, epigenetics and whole genome sequencing of viral genomes, NGS techniques are more and more widely used.

An increasing number of NGS studies investigate the totality of viruses contained in a specific sample, so-called “viral metagenomes” (see chapter 1.1). One important pioneer in general metagenomics was Norman R. Pace. In early works, he and his colleagues used PCR to explore the diversity of bacteria using their 16S rRNA genes (Lane et al., 1985; Schmidt et al., 1991). The results of this work led Pace to propose the idea to investigate DNA isolated directly from environmental samples, which allows the investigation of complex communities of unexplored microbial species. As mentioned before, however, viruses do not have universal phylogenetic markers and conserved regions like the rRNA genes. Accordingly, in the absence of those sequences, viral communities could only hardly been analysed in the past. The development of NGS techniques represented the key method to analyse viral communities. The culture-independent nature of NGS is an additional advantage of metagenomic approaches and makes this approach particularly suitable for the study of viral populations.

Due to the “open view”-nature of the applied NGS-based methods, many of those experiments led to the detection of **new viruses** or to identification of human and veterinary pathogens (Delwart, 2007; Minot et al., 2011; Li and Delwart, 2011c). One famous example is the Schmallenberg Virus (a new virus species related to known pathogenic genus *Orthobunyaviruses*) during a disease outbreak in cows and sheep with novel disease symptoms in 2011. It was discovered using NGS approaches (Hoffmann et al., 2012). However, not only pathogenic viruses are discovered using the NGS-based metagenomic approaches. Many so far unknown viruses with small circular ssDNA genome have been detected, which are not capable of replication in cell culture. Although no disease could be assigned to these viruses so far, the example demonstrates the potential of NGS technologies to discover new virus genera or families (Ng et al., 2014; Zhao et al., 2013).

2 AIM OF THE THESIS

The intestinal tract of animals contains a large community of different microorganisms, which can have important functions, e.g. during digestion of nutrients. Some of these microorganisms can also act as pathogens. The analysis of the composition of the bacterial intestinal microbiota in the past has given important insights into its functions and interactions with pathogens as well as the host immune system. However, the composition of the viral intestine community, its function and interaction with bacteria and the host, has only poorly studied so far. Novel NGS-based methods may enable the analysis of intestinal viromes.

The general aim of this thesis was therefore to analyse and characterise the faecal viromes of selected animal species. To this end, NGS-based method for the reliable and reproducible analysis of viruses present in faecal samples should be developed and applied to faecal samples derived from animals. Two different animal species have been selected for analysis: domestic pigs and urban wild rats.

Pigs represent one of the most important farm animal species in Germany. Factors influencing the porcine gut health are therefore of high economic impact. The faecal virome of pigs should first be determined in this thesis and thereafter, the effects of nutritional factors on it should be determined. In detail, a feeding trial with the probiotic bacterium *Enterococcus faecium (E. faecium)* should help to answer the question if probiotics could influence the virome composition. Rats are known as reservoir for several zoonotically transmitted diseases, some of them with a high impact on human health. Therefore, the faecal viromes of wild urban rats, which live in close proximity to humans, should be determined and potentially zoonotic viruses analysed.

The results should address five major questions:

i) What is the general composition of porcine and rat faecal viromes?

ii) How stable are faecal viromes and could they be influenced by nutritional factors?

iii) Which pathogenic viruses could be detected within faecal viromes?

iv) Have some of the viruses the potential to be zoonotically transmitted to humans?

v) Is the established method suitable to detect so far unknown viruses?

3 OWN PUBLICATIONS

3.1 First paper: Simultaneous Identification of DNA and RNA Viruses Present in Pig Faeces Using Process-Controlled Deep Sequencing

Jana Sachsenröder, Sven Twardziok, Jens A. Hammerl, Pawel Janczyk, Paul Wrede, Stefan Hertwig, Reimar Johne

PLoS One. 2012, 7(4): e34631.

Summary of paper 1

Animal faeces are comprised of millions of different viruses leading to very complex and diverse virus communities (viromes) in faeces. Due to the lack of suitable methods in the past, the porcine viral gut community is almost completely unknown.

In the current publication, we describe the development and application of a protocol for determination of the pig faecal viromes. A pooled sample of faeces from young pigs was used and analysed. The optimized protocol is based on the purification of the entire fraction of virus particles from pig faeces and the simultaneous extraction of the entire nucleic acid from the particles, which was thereafter randomly amplified by PCR and subjected to NGS (454 pyrosequencing). We also have developed a pipeline for subsequent bioinformatics analysis. Based on control systems used in quantitative PCR we have established a process control system to monitor the performance of the method using three different bacteriophages (T4, M13 and MS2). These bacteriophages showing different morphologies and genome types should reflect the wide range of virus properties. Defined amounts of the bacteriophages were added to the pooled pig faecal sample during the preparation and their abundance after the procedure was assessed by quantitative PCR. By this, the method was optimized. Later on, the bacteriophages were used as a quality control measuring the performance of the method of each analysed sample.

The virus community found in the pooled faecal sample from the pigs mainly consisted of viruses, which replicate in mammalian cells or in bacteria like pig

viruses or bacteriophages. Secondly, transiently passaged viruses from food or insects were identified. Using BLAST comparison, 7.7% of all generated reads showed a significant sequence similarity to known viruses. Mainly, bacteriophages were identified with 74% followed by mammalian viruses with 24%. The most abundant pig-specific viruses were: kobuvirus, rotavirus C, astrovirus, enterovirus B, sapovirus and picobirnavirus. Additional reads with sequence similarity to chimpanzee stool-associated circular ssDNA virus (ChiSCV) were identified. Whole-genome amplification and analysis of this virus genome indicated that this virus could represent a new pig virus named pig stool-associated circular ssDNA virus (PigSCV).

The results indicate that the method is suitable for simultaneous detection of DNA and RNA viruses in pig faeces including the identification of so far unknown viruses. It could be demonstrated that the developed process control consisting of three different bacteriophages could be useful for optimization of the method and for use as a quality control in metagenomic virus analyses.

Key messages of paper 1

- A new protocol for the simultaneous purification of DNA and RNA viruses present in faeces was developed.
- A process control consisting of three bacteriophages for use as a quality control of metagenomic analysis was established.
- The pig virome mainly consists of bacteriophages and pig viruses, whereas transient plant and insect viruses are only rarely detected.
- A novel small circular DNA virus, designated as PigSCV was identified.

Own contribution to paper 1

In this study, I developed the purification protocol for virus-like particles from a large amount of faeces. The phage process control was generated and optimized by me in collaboration with J. Hammerl and S. Hertwig (both BfR, Berlin). Furthermore, I developed a method for the simultaneously extraction and amplification of DNA and RNA from virus particles, which was subsequently used for the preparation of 454 pyrosequencing libraries. In this study, the process of 454 pyrosequencing was done by a commercial company. I applied the developed protocols to a pooled faecal sample derived from pigs and I did the main bioinformatics analysis of the generated processed reads in collaboration with S. Twardziok (FU Berlin). I also participated in the whole genome analysis of the novel pig virus PigSCV. Additionally, I wrote the major part of the manuscript and was engaged in critical reading and revision of the whole manuscript.

OPEN ACCESS Freely available online

PLoS one

Simultaneous Identification of DNA and RNA Viruses Present in Pig Faeces Using Process-Controlled Deep Sequencing

Jana Sachsenröder[1], Sven Twardziok[2], Jens A. Hammerl[1], Pawel Janczyk[1], Paul Wrede[2], Stefan Hertwig[1], Reimar Johne[1]*

1 Federal Institute for Risk Assessment, Berlin, Germany, **2** Institute for Molecular Biology and Bioinformatic, Charite, Berlin, Germany

Abstract

Background: Animal faeces comprise a community of many different microorganisms including bacteria and viruses. Only scarce information is available about the diversity of viruses present in the faeces of pigs. Here we describe a protocol, which was optimized for the purification of the total fraction of viral particles from pig faeces. The genomes of the purified DNA and RNA viruses were simultaneously amplified by PCR and subjected to deep sequencing followed by bioinformatic analyses. The efficiency of the method was monitored using a process control consisting of three bacteriophages (T4, M13 and MS2) with different morphology and genome types. Defined amounts of the bacteriophages were added to the sample and their abundance was assessed by quantitative PCR during the preparation procedure.

Results: The procedure was applied to a pooled faecal sample of five pigs. From this sample, 69,613 sequence reads were generated. All of the added bacteriophages were identified by sequence analysis of the reads. In total, 7.7% of the reads showed significant sequence identities with published viral sequences. They mainly originated from bacteriophages (73.9%) and mammalian viruses (23.9%); 0.8% of the sequences showed identities to plant viruses. The most abundant detected porcine viruses were kobuvirus, rotavirus C, astrovirus, enterovirus B, sapovirus and picobirnavirus. In addition, sequences with identities to the chimpanzee stool-associated circular ssDNA virus were identified. Whole genome analysis indicates that this virus, tentatively designated as pig stool-associated circular ssDNA virus (PigSCV), represents a novel pig virus.

Conclusion: The established protocol enables the simultaneous detection of DNA and RNA viruses in pig faeces including the identification of so far unknown viruses. It may be applied in studies investigating aetiology, epidemiology and ecology of diseases. The implemented process control serves as quality control, ensures comparability of the method and may be used for further method optimization.

Citation: Sachsenröder J, Twardziok S, Hammerl JA, Janczyk P, Wrede P, et al. (2012) Simultaneous Identification of DNA and RNA Viruses Present in Pig Faeces Using Process-Controlled Deep Sequencing. PLoS ONE 7(4): e34631. doi:10.1371/journal.pone.0034631

Editor: Houssam Attoui, Institute for Animal Health, United Kingdom

Received January 4, 2012; **Accepted** March 2, 2012; **Published** April 13, 2012

Funding: This study was supported by the German Research Foundation (DFG) through grant SFB 852/1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Reimar.Johne@bfr.bund.de

Introduction

About 191 million pigs are kept in Europe as farm animals [1]. As they provide a crucial source of food worldwide, viruses directly affecting the pig health have a great impact on food production in general. In addition, pigs are a natural habitat for zoonotic viruses, which can cause infectious diseases in humans, such as influenza virus [2], rotavirus [3] and hepatitis E virus [4]. Other viruses present in pigs such as bacteriophages or viruses causing subclinical infections may affect the pig health indirectly, e.g. by modification of the bacterial population or by modulation of the immune system as also shown for other hosts [5,6]. In order to understand the complex interaction between the different viruses, their host cells and the immune system, a simultaneous analysis of a broad range of virus species in a distinct compartment is desirable. For a long time, such an analysis was hampered by technical limitations [7]. Recently, this problem was solved by the development of novel molecular approaches utilizing deep sequencing techniques, e.g. 454 pyrosequencing [8]. This technique enables the simultaneous analysis of thousands of sequences present in a DNA sample. Deep sequencing techniques have been frequently used to study the composition of microbial communities in different kinds of environmental samples including faeces of humans and animals [9–11]. They have also been applied to the detection of single known and unknown viruses in various kinds of samples [12–16].

Some studies have applied deep sequencing methods to the analysis of the composition of viral communities in human faeces [17–21], but only one study focussed on pigs [22]. In most cases, the majority of detected viruses were bacteriophages. Among the RNA viruses, transitory plant viruses, which most probably originated from feed, were often identified. In addition, several human/animal viruses were detected.

The described protocols used for the analysis of the composition of the intestinal viral community generally consist of four basic steps: (i) purification and concentration of the virus particles present in the faeces, (ii) extraction of nucleic acids, (iii) deep sequencing of the nucleic acid and (iv) bioinformatic analysis of the sequence data. Despite this common backbone, the protocols differ from each other in several details. For example, different methods are used for purification and concentration of virus particles. As the virus particles are very heterogeneous in shape and size [23], the applied filtration method will strongly influence the result of the analysis. Moreover, the genomes of viruses consist of either DNA or RNA, which additionally may appear in different topologies [23]. Most of the protocols analyse DNA and RNA separately [13,17,21,22], which makes the comparison between both groups of viruses difficult. In many of the protocols, various amplification steps, which may affect the distribution of the different genome types, are implemented. In addition, either primary sequence reads or contigs assembled from these sequence reads were used for analysis. All of these variations in the applied protocols may lead to different results. Indeed, the reported composition of the human viral gut community varies remarkably, which - beside other factors - may also be caused by application of different protocols for its analysis [17–19,21].

Here, a protocol was established for the simultaneous analysis of DNA and RNA viruses present in pig faeces. A process control consisting of a mixture of bacteriophages with different morphology and genome type was added to the sample and used to assess the efficiency of the method. The use of a process control may enable the optimization of the method as well as a comparison to other published protocols. The optimized protocol was thereafter tested with a pooled pig faecal sample in order to analyse the composition of the viruses present in pig faeces.

Results

Establishment of a process control

To monitor the performance of the method and to optimize the distinct steps of the purification/concentration method, we used three bacteriophages (T4, M13, MS2) with different morphology and genome type [23] as process control (Tab. 1). Bacteriophages were chosen as they are easy to propagate and do not need extensive safety containment. T4 belongs to the family *Myoviridae* and has a large, double-stranded DNA. M13 is a filamentous phage and contains a small genome of single-stranded DNA. MS2 has a small icosahedral capsid containing a genome of single-stranded RNA. Using these three bacteriophages, the most common genome types were covered by the process control. All phages were added to the pig faecal suspension at a defined concentration. Quantitative real-time RT-PCR (qRT-PCR) protocols were established for these bacteriophages (Tab. 2), which enabled detection and quantification of the viruses at each step of purification. By comparison of the determined genome copy number at the end of the purification process with the number at the beginning, the efficacy of the purification protocol was calculated.

Optimization of the protocol for the detection of the viral community in pig faeces

Using the established process control, the method for purification of virus particles from pig faeces was gradually optimized. Several techniques were tested and various conditions were compared to each other, e.g. magnetic stirring vs. stomaching of the faecal sample, filtration vs. centrifugation for removal of larger debris, classical filtration vs. tangential flow filtration (TFF) for removal of smaller debris and bacteria, TFF vs. ultrafiltration for concentration of virus particles and caesium chloride gradient ultracentrifugation vs. sucrose gradient centrifugation for final purification and concentration of virus particles (data not shown). For example, a prolonged centrifugation time for removal of larger debris resulted in a marked loss of bacteriophage T4, whereas bacteriophages M13 and MS2 were less affected. Figure 1 shows the original qRT-PCR plots of the three bacteriophages from this purification step before and after centrifugation at 17,000× g for 5 hours. However, further processing of the sample by TFF without prior centrifugation was not practicable because of the appearance of filter occlusions. Therefore, a shorter centrifugation of three hours was used, although this might result in some loss of virus particles with high densities.

After the purification step had been optimized, several parameters were tested for efficient removal of free nucleic acids from the purified virus particle suspension. While DNase I digestion had only little effects on the control bacteriophages,

Table 1. Properties of the bacteriophages T4, M13 and MS2 used as process control.

	T4	M13	MS2
family	*Myoviridae*	*Inoviridae*	*Leviviridae*
particle shape	Icosahedral head+ contractile tail	filamentous	icosahedral
particle size	Ø 100 nm+ 100 nm tail	1000 nm long	Ø 30 nm
genome[1]	dsDNA	ssDNA	ssRNA
genome size[kb]	169	6.5 (circular)	3.6
density in CsCl [g/ml]	1.5	1.4	1.4
reads	29	137	175
contigs	/	4	5
average contig size		647	694

[1]ssDNA – single-stranded DNA; dsDNA – double-stranded DNA; ssRNA – single-stranded RNA

doi:10.1371/journal.pone.0034631.t001

Table 2. Primers and probes used for qPCR detection of the three bacteriophage genomes.

	sequence (5′–3′)	reference
	probe: FAM-TTG GGC GCG GTA ATG ATT CCT ACG-TAMRA	
M13	for: ACG CCT CGC GTT CTT AGA ATA CC	this study
	rev: ACC GCA CTC ATC GAG AAC AAG C	
	probe: FAM-CCT TTT TAG CTG CTT TAG TTT CTG C-TAMRA	
T4	for: GTA TCA GCA TCT TTA CCG CA	this study
	rev: GCT TTG GCT CGT AAA TTG GC	
	probe: FAM-ACC TCG GGT TTC CTG CTT GCT CGT-TAMRA	
MS2	for: GGC TGC TCG CGG ATA CCC	[28]
	rev: TGA GGG AAT GTG GGA ACC	

doi:10.1371/journal.pone.0034631.t002

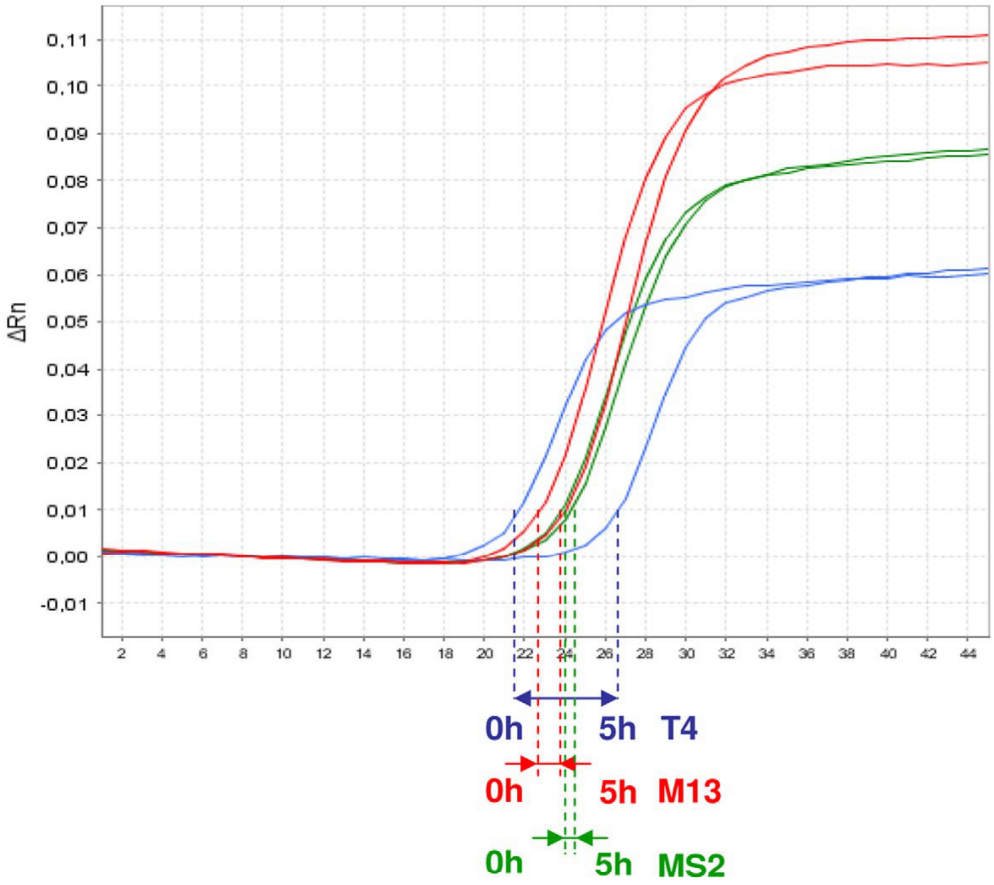


Figure 1. Optimization of the purification protocol using qRT-PCR detection of bacteriophages used as process control. The amplification curves determined by qRT-PCR for the phages T4 (blue), M13 (red) and MS2 (green) were determined before and after centrifugation for 17,000× g for 5 hours. The shift between the curves, which corresponds to the degree of virus loss, is indicated by arrows coloured according to the bacteriophage type.
doi:10.1371/journal.pone.0034631.g001

treatment with RNase A led to a significant decrease in the detectable genome number of bacteriophage MS2, which may be explained by the presence of residual RNase activity in the final nucleic acid preparation. Therefore, treatment with RNase was omitted in the final protocol. Tests with naked RNA prepared from bacteriophage MS2 that was added to pig faecal samples showed that the RNA was no longer detectable by qRT-PCR after 30 minutes of incubation at room temperature, suggesting that RNase treatment is not necessary at later steps.

Although nucleic acid amplification should be omitted to avoid errors caused by PCR, the optimized method contains an amplification step at the end of the procedure. This step had to be introduced, because the yield of nucleic acids was too low for direct use in deep sequencing. In order to minimize the amplification steps, aliquots were taken from different PCR cycles to determine the minimum amplification needed. A general scheme of the working steps in the optimized protocol is presented in Fig. 2. Applying this protocol, recovery rates of 4%, 125% and 105% were obtained for the genomes of T4, M13 and MS2, respectively, by comparing qRT-PCR results after addition of the phages with those from the final nucleic acid preparation. Recovery rates higher than 100% can be explained by concentration of the virus particles and by amplification of their genomes during the last step of the protocol. The low recovery rate of T4 may be explained by loss of this high density virus during clearing centrifugation as shown above.

Application of the optimized method to a pooled pig faecal sample

The faeces of five 35 day-old male pigs were pooled and subjected to purification of virus particles using the optimized protocol. By deep sequencing of the extracted nucleic acid, 69,613 reads with an average sequence length of 250 nucleotides were generated. After trimming of the sequence ends and exclusion of reads shorter than 50 nucleotides, 66,129 reads were included in the analyses. From these reads, 1,482 contigs could be assembled with a length between 100 bp and 1998 bp. The primary sequencing data are summarized in Table 3.

As a quality control of the method, the sequence reads were analysed for the presence of nucleotide sequences originating from the added bacteriophages T4, M13 and MS2. Sequences of all three phages could be identified. In detail, 175 reads originated

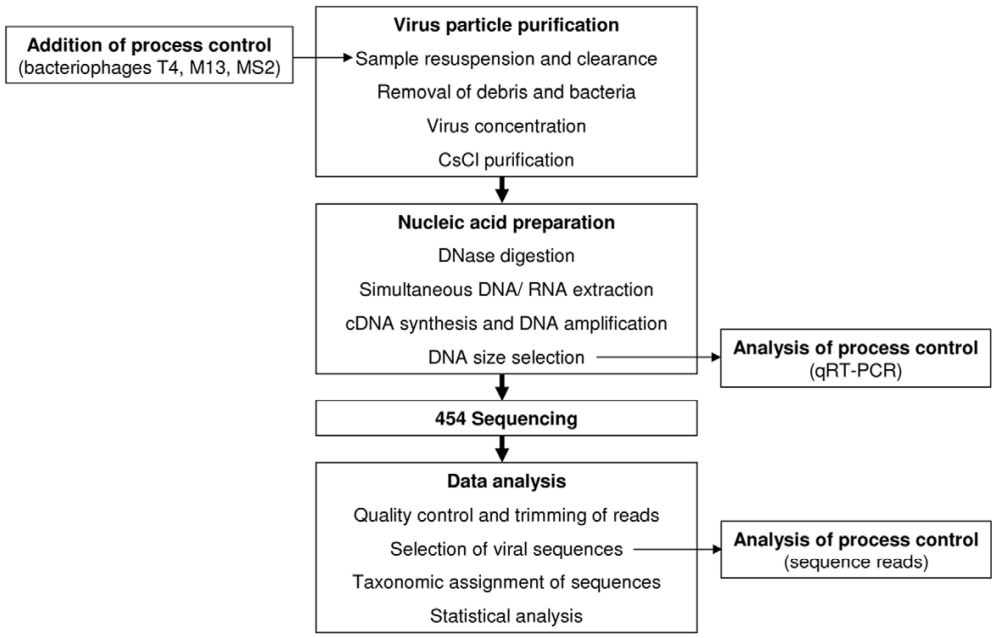


Figure 2. Schematic presentation of the general steps for analysis of the total fraction of virus particles from pig faeces by deep sequencing. The addition of the process control (bacteriophages T4, M13, MS2) and the subsequent points of analysis are indicated.
doi:10.1371/journal.pone.0034631.g002

from MS2, 137 reads from M13 and 29 reads from T4 (Tab. 1), thus correlating with the numbers of the bacteriophage genomes detected by qRT-PCR in the final preparation. These reads were excluded from the following data analysis.

Data analysis

Primary reads and contigs were checked for sequence identities with a viral genome database including the calculation of tBLASTx E scores. Only sequences with a tBLASTx score E $<= 10^{-4}$ were suspected to represent sequences of known and closely related viruses. By this, 5,366 primary reads (7.7%) and 315 contigs (21.3%) showed significant identities to known viral sequences and were therefore included in the further data analysis.

Out of the selected 5,366 primary reads, 3,965 reads (73.9%) showed significant sequence identities to bacteriophages, 1,282 reads (23.9%) were similar to mammalian viruses, 43 reads (0.8%) were identified as plant virus sequences and 56 reads (1.4%) could not be classified into one of these groups (Fig. 3A). An assignment of the primary reads to different types of viral genomes identified

Table 3. Primary data obtained for the pig faecal sample after deep sequencing.

total number of reads	69,613
number of reads included in analysis	66,129
average read length (bp)	250
total number of bases (Mb)	17,5
number of contigs	1,482
contigs >500 bp	321

doi:10.1371/journal.pone.0034631.t003

49.8% double-stranded (ds) DNA viruses, 26.4% single-stranded (ss) DNA viruses, 6.0% dsRNA viruses, 13.5% ssRNA viruses and 4.2% unknown genome types (Tab. 4, Fig. 3B). By analysis of the contigs, 39.0% dsDNA viruses, 36.2% ssDNA viruses, 6.3% dsRNA viruses, 14.3% ssRNA viruses and 4.1% unknown genome types were detected. An assignment of the sequences to known viral families was performed by application three different calculations (Tab. 5). Using the primary reads only, the most abundant virus families were *Siphoviridae* (30.1%), *Microviridae* (21.7 %) and *Myoviridae* (11.3%), all of them representing bacteriophages. Using the contigs only, three bacteriophage families were again most prominent, but with different percentages: 33.1% *Siphoviridae*, 14.9% *Myoviridae* and 9.1% *Podoviridae*. The maximum identities between the deduced amino acid sequences of the contigs with the lowest tBLASTx E scores and that of known members of the respective virus families ranged from 35% to 100% (Tab. 5). The lengths of the viral genomes are remarkably different between different virus families [23] and it can be assumed that longer genomes are represented by more primary reads per genome than smaller ones. Therefore, a correction factor for the genome length was introduced to give estimation for the particle number. After this correction, the most abundant viral families were *Microviridae* (bacteriophages, 45.9%) followed by the Chimpanzee stool associated circular ssDNA virus (ChiSCV, see below, 12.0%) and *Picornaviridae* (mammalian viruses, 10.8%) (Fig. 3C). The Shannon index was used to assess the diversity of the sequences detected in the sample. Shannon indexes of 4.7777 and 4.252 were calculated for the primary reads and contigs, respectively.

A closer inspection of the sequences revealed that a total of 15 different viruses with sequence identities to mammalian viruses were identified (Tab. 6). These included the RNA viruses: kobuvirus, rotavirus group A and C, astrovirus, enterovirus B, sapovirus, picobirnavirus, teschovirus, picornavirus, and the

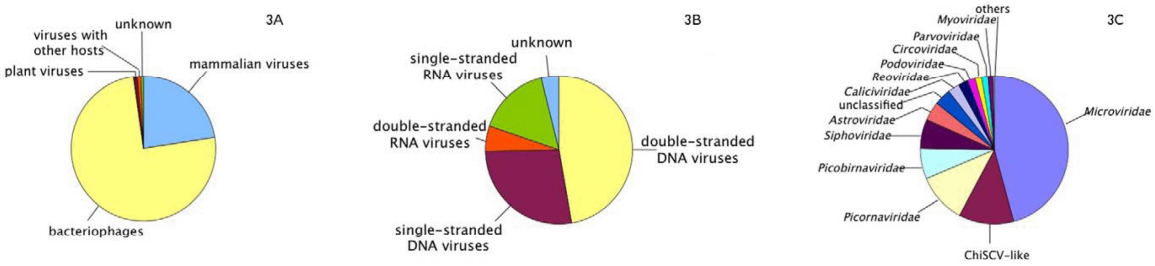


Figure 3. Summary of results for the detection of viruses in a pooled faecal sample derived from five 35 day-old male pigs. The numbers of generated viral sequences are analysed according to virus host (A), viral genome type (B) or assignment to a specific virus family (C). In the latter case, the numbers have been corrected according to the genome length of the respective virus species (see Text for details).
doi:10.1371/journal.pone.0034631.g003

DNA viruses: circovirus, bocavirus, pox virus, parvovirus and herpesvirus. In addition to these virus families, which had already been detected in pigs previously, a novel virus showing sequence identity to ChiSCV was detected and analysed in detail.

Whole genome analysis of a novel circular ssDNA virus

The whole genome of the suspected ChiSCV-related virus was amplified from the nucleic acid preparation in two fragments using PCR. Primers used in PCR were designed on the basis of the sequences derived from the deep sequencing analysis. After sequencing of the overlapping PCR products, the whole genome sequence was assembled and ORFs were identified. The genome of this virus has a length of 2,459 nucleotides and a circular topology (Fig. 4A). A total of 6 ORFs encoding proteins consisting of more than 100 amino acids were identified. The protein encoded by ORF1 shows the highest amino acid sequence identity (27.0%) to the replicase protein (Rep) of ChiSCV (Gen-Bank Acc.-No. ABD24829.1). ORF2 shows the highest amino acid sequence identity (49.2%) to the capsid protein (Cap) of ChiSCV (Gen-Bank Acc.-No. ADB24798.1). No significant similarities to known proteins could be determined when the amino acid sequences encoded by the other ORFs were analysed. The genome organization of PigSCV is different from that of ChiSCV: while Rep and Cap are encoded by the same DNA strand in ChiSCV, the respective genes are oriented in opposite direction in the PigSCV genome (Fig. 4A). A search for repetitive sequences revealed a short stemloop sequence between nucleotide positions 743 and 759 on the PigSCV genome (Fig. 4B). This stemloop also contains a conserved sequence with identitiy to stemloop sequences of several ChiSCV isolates. A phylogenetic analysis of the amino acid sequence of Rep together with related viruses from animals and plants revealed the closest relationship of PigSCV to ChiSCV (Fig. 4C). However, within this group, the PigSCV sequence forms a separate branch indicating a membership of this virus to a separate species.

In order to analyze PigSCV infection in the individual pigs, stored faecal and serum samples from the animals were tested by PCR for the presence of the PigSCV genome. As evident from Table 7, specific PCR products were detected in faecal samples of each of the tested piglets; however, at different time-points and with different intensity (Tab. 7). The highest detection rate was found in the pigs at 33 and 35 days of age. The PigSCV genome was not detected in serum samples derived from the pigs at 42 days of age.

Discussion

The viral community present in animal faeces is complex in function and composition. So far, only little is known about the diversity and distribution of viruses in this ecosystem. The analysis of the composition of viruses in faeces was hampered for a long time by the lack of suitable methods. However, the recent development of deep sequencing techniques enabled the analysis of the viral community in faeces from different species like human [18,19,21], horse [24], turkey [25], sea lion [26] and rodents [27]. For pigs, only one study was published very recently [22]. The applied protocols are diverse and not standardized; therefore, comparison of the generated data is difficult.

To overcome the problem of comparability, we here suggest the use of a process control for monitoring the efficiency of the

Table 4. Numbers of viral sequences detected in the pig faecal sample, according to genome types.

	Reads		contigs	
genome type[1]	number of reads (% of reads)	number of different virus species	number of contigs (% of contigs)	number of different virus species
dsDNA	2674 (49.8)	379	123 (39.0)	85
ssDNA	1419 (26.4)	29	114 (36.2)	20
dsRNA	323 (6.0)	21	20 (6.3)	3
ssRNA	726 (13.5)	17	45 (14.3)	9
unknown	224 (4.2)	20	13 (4.1)	5
Total	5366 (100.0)	466	315 (100.0)	121

[1]dsDNA – double-stranded DNA; ssDNA – single-stranded DNA; dsRNA – double-stranded RNA; ssRNA – single-stranded RNA
doi:10.1371/journal.pone.0034631.t004

Table 5. Number of viral sequences detected in the pig faecal sample, according to virus family.

contigs			Reads		reads/genome size	
	% of contigs	sequence identity* (length of compared sequence)		% of reads		% families (corrected)
Siphoviridae	33.06	75% (99aa)	*Siphoviridae*	30.06	*Microviridae*	45.86
Myoviridae	14.88	62% (102aa)	*Microviridae*	21.71	ChiSCV-like	12.01
Podoviridae	9.09	54% (125 aa)	*Myoviridae*	11.33	*Picornaviridae*	10.76
unclassified**	9.09	45% (80aa)	*Picornaviridae*	8.83	*Picobirnaviridae*	6.59
Microviridae	6.61	54% (206 aa)	unclassified**	6.78	*Siphoviridae*	6.38
Astroviridae	3.31	52% (182 aa)	*Podoviridae*	4.71	*Astroviridae*	4.17
Circoviridae	3.31	45% (58 aa)	*Reoviridae*	4.38	unclassified**	3.71
Parvoviridae	3.31	55% (77aa)	ChiSCV-like	3.24	*Caliciviridae*	2.39
Phycodnaviridae	2.48	42% (45aa)	*Astroviridae*	2.83	*Reoviridae*	2.28
Picornaviridae	1.65	87% (200aa)	*Caliciviridae*	1.83	*Podoviridae*	1.77
Reoviridae	1.65	93% (271aa)	*Picobirnaviridae*	1.62	*Circoviridae*	1.33
Herpesviridae	1.65	45% (29aa)	*Parvoviridae*	0.65	*Parvoviridae*	1.21
Poxviridae	1.65	100% (42aa)	*Phycodnaviridae*	0.58	*Myoviridae*	1.20
ChiSCV-like	0.83	68% (105aa)	*Poxviridae*	0.52	*Inoviridae*	0.07
Caliciviridae	0.83	92% (99aa)	*Circoviridae*	0.26	*Tectiviridae*	0.06
Picobirnaviridae	0.83	42% (86aa)	*Mimiviridae*	0.24	*Bromoviridae*	0.05
Mimiviridae	0.83	48% (42aa)	*Herpesviridae*	0.11	*Luteoviridae*	0.03
Potyviridae	0.83	50% (28aa)	*Tectiviridae*	0.09	*Phycodnaviridae*	0.02
Phaeovirus	0.83	42% (57aa)	*Iridoviridae*	0.09	*Poxviridae*	0.02
Inoviridae	0.83	29% (52aa)	*Inoviridae*	0.06	*Herpesviridae*	0.006
Iridoviridae	0.83	35% (78aa)	*Bromoviridae*	0.02	*Iridoviridae*	0.005
			Luteoviridae	0.02	*Bicaudaviridae*	0.002
			Bicaudaviridae	0.02	*Mimiviridae*	0.002

*sequence identity: Contigs were compared to known sequences of the indicated virus families and the respective contig with the lowest tBLASTx E score was used for calculation of deduced amino acid sequence identities.

**unclassified: The highest identities of these sequences were found to virus sequences, which have not been assigned to a virus family according to the data available in the GenBank database.

doi:10.1371/journal.pone.0034631.t005

protocol applied for the analysis. The process control consists of a set of previously characterized viruses, of which a defined number is added to the sample and then followed during the process. Similar process controls are well established in other analytical applications, such as PCR analysis of food and environmental samples [28–30]. For the application in analysis of viral communities, the process control should reflect the variety of viruses with different morphologies and genome types, which can be suspected to be present in the sample. The three selected bacteriophages used here are diverse in these aspects and represent a large proportion of known viruses. However, some morphologies, e.g. enveloped spherical viruses, as well as some genome types, e.g. double-stranded RNA, are not present so far and may be added in future in order to improve the process control. As a proof of principle, we could demonstrate that all of the bacteriophages added to the sample could be identified in the sequences generated by deep sequencing. This indicates that the bacteriophages could be monitored throughout the whole process until its end, which includes sequence analysis. Using the recovery rates or sequence read numbers of the control bacteriophages, a comparison of results from studies performed with different techniques should be possible in the future. Although in principle the three used bacteriophages seem to be appropriate for monitoring the efficiency of the process, the low recovery rate of bacteriophage T4 demand further developments for a robust dsDNA virus control. Since the majority of viral sequences identified in our analysis of the pig faecal sample originate from dsDNA viruses, the bacteriophage T4 control may not sufficiently reflect the behaviour of many dsDNA viruses. Therefore, changes in the preparation protocol for bacteriophage T4 (e.g. by adding steps preventing aggregation of virus particles) or a replacement by another virus should be considered in the future.

The process control was also used as a read-out for optimization of the distinct steps of the method. By this, it could be confirmed that some of the techniques applied in former studies, e.g. tangential flow filtration [31,32], are more effective than others. Besides minor variations of parameters in order to assess the optimal conditions for each step, some more basic decisions had to be made during the optimization process. One of them refers to the use of an RNase digestion step to remove contaminating free RNA from the virus particle suspension. As high amounts of RNase are difficult to inactivate or to remove from the preparation, residual RNase activity can destroy the viral RNA in the final nucleic acid preparation [32]. We show that free bacteriophage RNA is rapidly degraded in the faecal suspension within the first 30 minutes of the applied protocol; therefore,

Table 6. Data on mammalian viruses detected in the pig faecal sample.

		Reads		contigs		
virus	**genome type**	**number of reads (% of reads)**	**% of mammalian viruses**	**number of contigs (% of contigs)**	**% of mammalian viruses**	**sequence identity* (length of compared sequence)**
Kobuvirus	ssRNA	330 (5.8)	25.7	19 (15.4)	21.6	87% (200aa)
Rotavirus C	dsRNA	207 (3.6)	16.1	16 (13.0)	18.2	93% (271aa)
ChiSCV	ssDNA	174 (3.0)	13.6	11 (8.9)	12.5	68% (271aa)
Astrovirus	ssRNA	152 (2.7)	11.9	13 (10.6)	14.8	52% (182aa)
Enterovirus	ssRNA	134 (2.3)	10.5	9 (7.3)	10.2	94% (135aa)
Sapovirus	ssRNA	98 (1.7)	7.6	3 (2.4)	3.4	87% (158aa)
Picobirnavirus	dsRNA	87 (1.5)	6.8	3 (2.4)	3.4	49% (49aa)
Rotavirus A	dsRNA	28 (0.5)	2.2	1 (0.8)	1.1	79% (76aa)
Bocavirus	ssDNA	27 (0.5)	2.1	3 (2.4)	3.4	61% (46aa)
Poxvirus	dsDNA	15 (0.3)	1.2	2 (1.6)	2.3	100% (42aa)
Parvovirus	ssDNA	8 (0.1)	0.6	1 (0.8)	1.1	55% (77aa)
Teschovirus	ssRNA	7 (0.1)	0.5	/	/	/
Circovirus	ssDNA	7 (0.1)	0.5	5 (4.1)	5.7	41% (85aa)
Herpesvirus	dsDNA	5 (0.1)	0.4	2 (1.6)	2.3	22% (95aa)
Picornavirus	ssRNA	3 (0.1)	0.2	/	/	/
total		1282 (22.5)	100.0	88 (7.5)	100.0	/

*Contigs were compared to known sequences of the indicated virus families and the respective contig with the lowest tBLASTx E score was used for calculation of deduced amino acid sequence identities.

doi:10.1371/journal.pone.0034631.t006

RNase treatment seems not necessary at later steps. A second decision refers to the use of amplification steps after preparation of the nucleic acids. It is known that some techniques induce changes in the relative distribution of viral genome types, e.g. several isothermal amplification methods preferentially amplify small circular DNA molecules [33]. To avoid extensive amplification steps, a relatively high amount of starting material (100 g faeces) was used here. However, the yield of purified virus nucleic acid was too low for direct use in deep sequencing. Therefore, amplification was implemented in the protocol; however, the cycles were limited to the lowest possible number. The identification of all known viral genome types in the primary sequence reads of the pig faecal sample analyzed here indicate that the optimized protocol is suitable for simultaneous identification of DNA viruses and RNA viruses in pig faeces.

By application of the optimized protocol to the pooled pig faecal sample, we detected 7.7% sequences with a high sequence identity to known virus genomes. This value is similar to that of recently published viral metagenome studies, which range from 2% [18] to 13% [22]. The slight difference between our study and that of Shan et al. [22], which both analyzed pig faecal samples, may be explained by the use of different databases and different sequence identity cut-offs. In other aspects, the results of both studies are similar. For example, Shan et al. identified 99% of the mammalian viruses as RNA viruses, which is similar to the value of 96% obtained in our study. Kobuvirus, enterovirus, sapovirus, teschovirus, astrovirus and bocavirus were identified in both studies. The similarity of these results is quite surprising taking into consideration that the samples analyzed in both studies originated from pigs from different continents. However, some differences still exist as coronavirus was only found by Shan et al. [22] and rotavirus was only detected in our study. The detection of rotaviruses, which are known as causative agents of gastroenteritis in pigs [34,35], correlates with the fact that some of the pigs included in our study suffered from watery diarrhoea. In the study of Shan et al. [22], a correlation was found between the detection of bocavirus or coronavirus and the appearance of diarrhoea. Additionally, we detected a higher number of different virus families, compared to the study of Shan et al. [22], which may be explained by the use of a higher amount of faeces as starting material enabling detection of less represented viruses.

An additional factor with marked influence on the results of metagenomic analyses is the applied method for data analysis. First, a reliable cut-off for selection of sequences with significant sequence identities to known virus sequences has to be defined. Here, a tBLASTx E score $<= 10^{-4}$ was used, according to other studies applying tBLASTx E scores between 10^{-3} [20,21] and 10^{-5} [16,18]. However, using this cut-off, many sequences with relative low deduced amino acid sequence identities were included in the analysis (Tab. 5). In addition, it has to be considered that the analysis of deduced amino acid sequences may not be appropriate in the case of sequences representing non-coding regions of viral genomes. Second, it should be decided, whether primary reads or contigs are used. In our study, the overall distribution of viral genome types or virus families was rather similar when using primary reads or contigs for analysis. However, analysis of the contigs only may lead to a reduction of detected virus diversity, as the rare sequences of less represented viruses cannot be assembled into contigs. Indeed, by using the primary reads we could identify 447 different virus species, compared to 121 viruses identified by the use of contigs. The Shannon index of diversity determined using the primary reads was 4.777 compared to 4.252 using contigs. As additionally contig assembly is strongly dependent on the used assembly algorithm [36,37], it may be generally advisable to use primary reads for metagenomic data analysis. A third factor influencing results of data analysis refers to the highly diverse viral

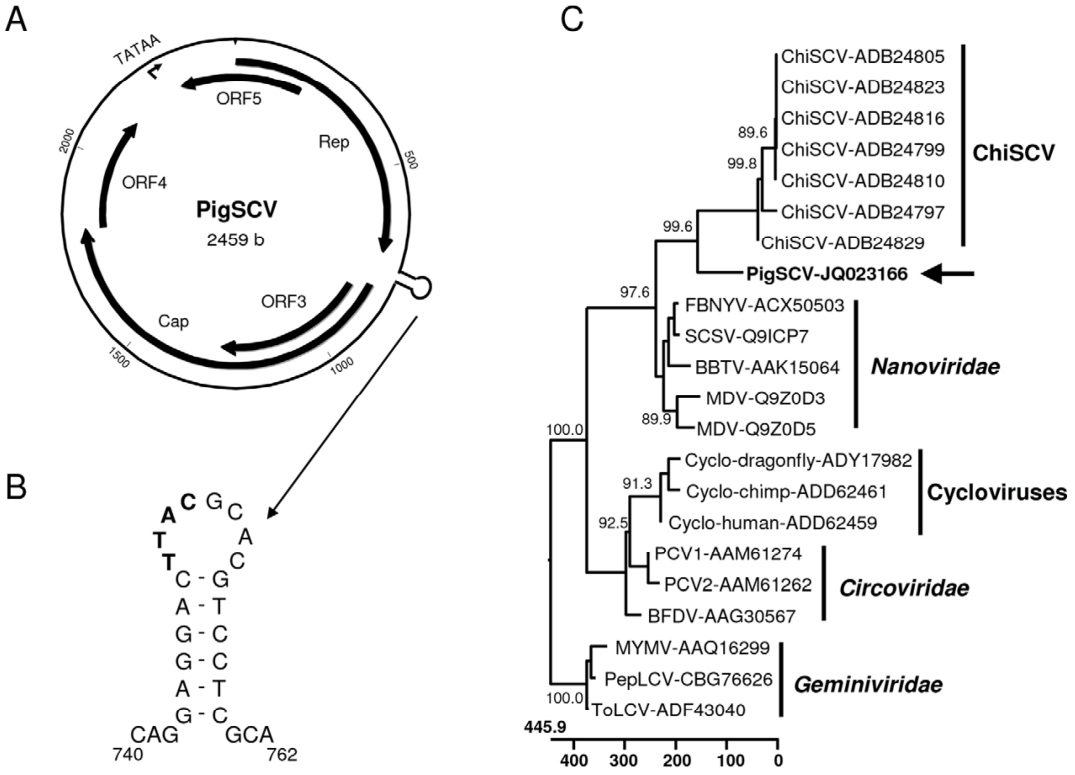


Figure 4. Genome analysis of the novel pig stool-associated single-stranded DNA virus (PigSCV) detected in the faecal sample. (A) Genome map: ORFs encoding the replication-associated protein (Rep), the capsid protein (Cap) and several unknown proteins (ORF3-5), the position of a putative TATA-box and the stemloop structure are indicated. (B) Proposed secondary structure of the stem-loop structure; nucleotides conserved in PigSCV and ChiSCV are shown in bold face. (C) Phylogenetic tree constructed on the basis of the deduced amino acid sequences of the Rep protein of selected circular ssDNA viruses of animals and plants. The position of PigSCV is indicated by an arrow.
doi:10.1371/journal.pone.0034631.g004

genome length. As a larger genome type is likely to be represented by more primary reads per genome than a smaller one, correction factors should be developed and included. We could show that the use of such correction factors can dramatically change the calculated distribution of viral families leading to a much higher abundance of mammalian viruses (with small genomes) than calculated without consideration of genome length.

Table 7. Detection of the PigSCV genome in samples from pigs using PCR.

age (days)	29	33	35	42	
sample	**faeces**	**Faeces**	**faeces**	**faeces**	**serum**
piglet 1	−	+	−	+	−
piglet 2	−	n.d.	++	−	−
piglet 3	−	+	++	n.d.	−
piglet 4	−	++	++	−	−
piglet 5	++	−	+	−	−

− no band visible after electrophoresis of PCR product
+ faint band visible after electrophoresis of PCR product
++ strong band visible after electrophoresis of PCR product
n.d. – not done
doi:10.1371/journal.pone.0034631.t007

By application of the optimized protocol and subsequent data analysis, sequences with similarities to a chimpanzee virus have been identified in the pig faeces. Subsequent amplification and analysis of the whole genome sequence indicate that this virus is closely related to the chimpanzee virus [38]. The two largest ORFs clearly show sequence similarities to the encoded replicase and capsid proteins of the chimpanzee virus, but their orientation in the genome is different and the percentages of the sequence identity are relatively low. As such virus sequences have not yet been described in pigs, it is concluded that it represents a novel pig virus. Analysis of samples from the pigs used in our experiment indicated that all of the piglets were infected with this virus, but the amounts and time-points of virus excretion in the faeces differed between the individuals. The different pattern of virus excretion in pigs receiving the same feed argues against an origin of the virus from infected plants that were ingested and excreted after intestinal passage, as originally suspected as one possibility for ChiSCV [38], although the feed could not be directly tested. No PigSCV genome was detected in the serum of the pigs; however, only one time-point could be analyzed because of the limited availability of the sera. Further studies are necessary in order to assess the origin, dissemination, organ distribution and clinical significance of this novel virus in pigs. PigSCV belongs to the

single-stranded circular DNA viruses, which are increasingly detected in animals, plants and environmental samples [39–41]. The reason for this increased detection is unknown so far; however, it may be explained by the increased use of specific amplification methods in the detection protocols, which are known to amplify circular DNA molecules more efficiently than others [33]. Shan et al. [22] also found several single-stranded circular DNA virus sequences in pig faeces, though, with identities to circo- and cycloviruses. Although the detection of the novel virus indicates the suitability of the protocol for detection of so far unknown viruses, it has to be clearly stated, that only those viruses can be identified, for which sequence data of related viruses are present in the databases. For the identification of members of totally new virus families, novel analytical tools have to be developed using "deep bioinformatics", e.g. functional predictions of putative proteins.

In summary, the results show that the established protocol enables the efficient analysis of the virus content in pig faeces including the identification of new viruses. The high percentage of bacteriophages detected in the faecal sample suggests an important function of these viruses in the modulation of the bacterial intestinal community, which should be analyzed in more detail in future. In addition, the clinical significance of the detected pig viruses - separate and in combination with each other – should be investigated further. The method may generally be useful in studies investigating aetiology, epidemiology and ecology of diseases. The implemented process control serves as quality control ensuring comparability of the method and may be used for further method optimization in future.

Materials and Methods

Ethics statement

The animal study was approved by the local ethic commitee of the state Berlin (Landesamt für Gesundheit und Soziales, Berlin, Germany) under the accession number G 0350/09.

Faecal samples

Faecal samples were derived from pigs kept in the experimental animal facility of the Federal Institute for Risk Assessment (Berlin, Germany) and used in preliminary studies for establishment and optimization of the methods. For final testing of the established protocol, a fresh pooled faecal sample (approximately 100 g) was derived from five 35 day-old male pigs housed in one pen of the experimental facility. The animals were in good overall clinical condition. However, two of the five pigs suffered from watery diarrhoea at the time of the collection of the faeces. After collection, the faeces were stored at 4°C for 2 days until analysis. Individual rectal faecal samples and serum samples were collected at different time-points and stored at −80°C.

Bacteriophages

The bacteriophages T4 (German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany), M13 (kindly provided by E. Lanka, Max Planck Institute for Molecular Genetics, Berlin, Germany) and MS2 (kindly provided by J. Dreier, Ruhr University of Bochum, Bad Oeynhausen, Germany) were used as process control to assess the efficiency of the method. Table 1 summarizes the properties of the phages. M13 and T4 were propagated on E. coli JM110 (New England BioLabs GmbH, Frankfurt, Germany), MS2 on E. coli strain Top10 F+ (Invitrogen, Leek, Netherlands). For quantification, the phage titer of lysates was determined as previously described [42]. For the detection of the bacteriophage genomes at distinct steps of purification and amplification, quantitative real-time RT-PCR (qRT-PCR) assays were established. Primers and probes used for the detection of the phages are listed in Table 2. All qRT-PCRs were performed using the QIAamp probe RT-PCR kit (Qiagen, Hilden, Germany) in an ABI 7500 cycler (Applied Biosystems; Darmstadt, Germany). After reverse transcription at 50°C for 30 min and activation of the polymerase at 95°C for 15 min, 45 cycles were performed, each consisting of 94°C for 15 sec, 56°C for 1 min and 72°C for 1 min. Although the bacteriophages T4 and M13 have DNA genomes, their detection was also carried out using this RT-PCR protocol in order to enable parallel analysis of all three bacteriophage genomes using the same cycling conditions. Ten-fold diluted nucleic acid preparations of the titrated phages were used as standards for quantification.

Purification and concentration of virus particles

Virus particles were purified and concentrated by a combination of tangential flow filtration (TFF), ultrafiltration and caesium chloride (CsCl) density gradient ultracentrifugation [17,21,31,32,43]. Briefly, approximately 100 g of pig faeces were resuspended in 1,000 ml SM-buffer by magnetic stirring. Thereafter, 1×10^7 plaque-forming units of each bacteriophage were added to the suspension. Aliquots of the faecal suspension were taken before and after adding of phages and tested at a later date by qRT-PCR. The sample was centrifuged at 17,000× g for 30 min to remove the large particulate debris and the supernatant was collected. The procedure was repeated by a three hour centrifugation to remove smaller particular structures. Afterwards, a first TFF was performed using a 0.22 µm filter (PALL Corporation, Middleton; MA, USA) to remove bacterial and eukaryotic cells and debris. The remaining filtrate was subjected to a second TFF with a 50 kDa filter (PALL Corporation, Middleton; MA, USA) to concentrate the virus particles. This viral preparation was further concentrated by centrifugation through Vivaspin 50,000 MWCO concentrators (Sartorius Stedim Biotech GmbH, Goetting, Germany) at 1,500× g resulting in a final volume of 36 ml. Volumes of 18 ml each were loaded onto a stepwise CsCl density gradient with density layers of 1.7, 1.5, 1.35 and 1.2 g ml^{-1} (each 5 ml) and centrifuged at 65,000× g for 14 hours at 10°C. The 1.35–1.5 g ml^{-1} layers were collected from the gradients using a syringe and pooled resulting in a final volume of 9 ml.

Nucleic acid preparation and deep sequencing

To eliminate free DNA present in the virus concentrate, 500 µl CsCl purified virus suspension were treated with 25 units DNase I (2,000 U/mg, bovine pancreas grad II; Roche Diagnostics GmbH, Mannheim, Germany) for 45 min at 37°C, followed by heat inactivation for 10 min at 65°C. Thereafter, DNA and RNA were simultaneously extracted from 200 µl using the RTP® Pathogen Kit (Invitek; Berlin, Germany). The extracted nucleic acids (75 ng per reaction) were randomly primed for cDNA synthesis using the TransPlex® Complete Whole Transcriptom Amplification Kit (WTA2, Sigma-Aldrich, St. Louis, MO, USA) according to the protocol recommended by the supplier; however, the annealing temperature was decreased to 40°C (cycles 1 and 2) and 45°C (cycles 3 and 4) to enable the simultaneous amplification of DNA and RNA. Aliquots of 75 µl each were removed from the WTA reaction at different cycle numbers, and purified as well as size-selected using MobiSpin S-400 Columns (MoBiTec, Goetting, Germany). The DNA concentration of the preparations was measured using a nanodrop spectrometer (PEQLAB Biotechnologie GMBH; Erlangen Germany) and the preparation derived from a minimum of amplification cycles with a DNA concentration higher than 50 ng/µl was used for deep sequencing. A total of 1 µg DNA was applied to deep sequencing on a 1/8 plate of the GS-FLX sequencer 454

Titanium (GS Titanium SV emPCR Kit (Lib-L) v2; GS Titanium PicoTiterPlate Kit 70×75; GS Titanium Sequencing Kit XLR70t; Life Sciences, Roche, Branford, USA) according to the manufacturer's protocol.

Data analysis

Primary sequence analysis was applied to raw sequence reads, which were subjected to amplification primer/adaptor trimming using SeqMan (DNASTAR, Lasergene, Madison, USA) and selection for a minimum length of 50 nt. In parallel, all primary reads were subjected to *de novo* contig assembly using the 454 Newbler Assembler [44] software (http://www.my454.com/), with criteria of 90% minimum overlap identity and a minimum overlap length of 40 nt. Homology searches for primary reads and assembled contigs were performed with tBLASTx [45] and CLC Main Workbench 6.2 (http://www.clcbio.com/index.php?id = 532) against a local database. This local database was created using EXCEL (2003). It included the viral genome non-redundant reference sequence nucleotide database (RefSeq, NCBI, ftp://ftp.ncbi.nih.gov/refseq/release/viral/ 30.08.2011 download) and additional sequences from recently discovered viruses, which had been manually added. BLAST results with an E-value $<= 10^{-4}$ were selected and used for further grouping analysis, which was performed using a script manually written with R 2.13 [46]. This analysis included counting of detected species and determination of their taxonomy, which was also used to determine the virus hosts. Correction of the different genome length of the virus families was done by using the formula: formal size factor = read number/genome size (the genome size for the respective virus species was derived from the above mentioned local database, which included these data from the RefSeq database, NCBI). The Shannon index [47] was calculated to compare the diversity of the species identified by primary reads and by assembled contigs.

Whole genome analysis of PigSCV

Two overlapping fragments covering the whole genome sequence of PigSCV were amplified by PCR using nucleic acid derived from the concentrated virus particles as template. The Long range PCR kit (Qiagen, Hilden, Germany) was used with primers PigSCV-1s (5′-CCA ATC AGA TTC ACG CTT ACC G-3′) and PigSCV-1as (5 ′AAC ATC GTC AAC CGT ATC ATG G-3′) or primers PigSCV-2s (5′-GGG CCA CGC ATG AAC CTT CC-3′) and PigSCV-2as (5′-ACC ATT GAA ATC ATC TGG GAT G-3′) for amplification of overlapping fragments with lengths of 1,849 bp or 1,123 bp, respectively. The PCR products were directly sequenced in an ABI 3730 DNA Analyser (Applied Biosystems). The whole genome sequence was assembled from the sequenced fragments, submitted to the GenBank database under the accession number JQ023166. ORFs were predicted using the SeqBuilder module of the DNASTAR software package (Lasergene, Madison, USA). Sequence distances were calculated and phylogenetic trees were constructed with the MegAlign module implemented in the DNASTAR software package.

Detection of PigSCV genome in pig samples

Faecal samples derived from the rectum of the pigs at d29, d33, d35 and d42 and serum samples of d42 were stored at −80°C. The nucleic acid was extracted from 200 µl faecal suspension or 100 µl serum using NucliSens Magnetic Extraction in a NucliSens EasyMag device (bioMerieux Deutschland GmbH, Nürtingen, Germany). PigSCV was detected by PCR using the TaKaRa ExTaq PCR Kit (TaKaRa Bio Europe, Göttingen, Germany) with primers PigSCV-2s and PigSCV-1as amplifying a 361 bp fragment of the PigSCV genome. PCR was performed by an initial denaturation step at 95°C for 5 min followed by 40 cycles each including 95°C for 30 sec, 60°C for 30 sec and 72°C for 45 sec, and a final elongation at 72°C for 5 min. PCR products were analysed by electrophoresis on ethidium bromide-stained agarose gels.

Acknowledgments

We thank E. Lanka (Max Planck Institute for Molecular Genetics, Berlin, Germany) and J. Dreier (Ruhr University of Bochum, Bad Oeynhausen, Germany) for kindly providing us with bacteriophages M13 and MS2, respectively, and Rose Whelan (Humboldt University Berlin, Berlin, Germany) for providing the individual pig samples.

Author Contributions

Conceived and designed the experiments: RJ. Performed the experiments: JS JH. Analyzed the data: JS ST JH SH RJ. Contributed reagents/materials/analysis tools: JH PJ. Wrote the paper: JS PJ PW SH RJ.

References

1. BMELV, Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz; Federal Ministry of Food, Agriculture and Consumer Protection. Available: http://berichte.bmelv-statistik.de/SJT-8031700-0000.pdf.
2. Sencer DJ (2011) Perspective: Swine-origin influenza: 1976 and 2009. Clin Infect Dis 52 Suppl. 1: 4–7.
3. Midgley SE, Bányai K, Buesa J, Halaihel N, Hjulsager CK, et al. (2011) Diversity and zoonotic potential of rotaviruses in swine and cattle across Europe. Vet Microbiol, Epub ahead of print.
4. van der Poel WH, Verschoor F, van der Heide R, Herrera MI, Vivo A, et al. (2001) Hepatitis E virus sequences in swine related to sequences in humans, The Netherlands. Emerg Infect Dis 7(6): 970–976.
5. Smith HW, Huggins MB (1983) Effectiveness of phages in treating experimental Escherichia coli diarrhoea in calves, piglets and lambs. J Gen Microbiol 129;8): 2659–2675.
6. Maidana-Giret MT, Silva TM, Sauer MM, et al. (2009) GB virus type C infection modulates T-cell activation independently of HIV-1 viral load. AIDS 23(17): 2277–2287.
7. Roossinck MJ, Saha P, Wiley GB, Quan J, White JD, et al. (2010) Ecogenomics: using massively parallel pyrosequencing to understand virus ecology. Mol Ecol 1: 81–88.
8. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, et al. (2005) Genome Sequencing in Open Microfabricated High Density Picoliter Reactors. Nature 437;7057): 376–380.
9. Riesenfeld CS, Schloss PD, Handelsman J (2004) Metagenomics: Genomic Analysis of Microbial Communities. Annu Rev Genet 38: 525–552.
10. Lamendella R, Domingo JWS, Ghosh S, Martinson J, Oerther DB (2011) Comparative fecal metagenomics unveils unique functional capacity of the swine gut. BMC Microbiol 11: 103.
11. Arumugan M, Raes J, Pelletier E, Paslier DL, Yamada T, et al. (2011) Enterotypes of the human gut microbiome. Nature 473;7346): 174–180.
12. Tang P, Chiu C (2010) Metagenomics for the discovery of novel human viruses. Future Microbiol 5(2): 177–189.
13. Djikeng A, Kuzmickas R, Anderson NG, Spiro DJ (2009) Metagenomic analysis of RNA viruses in a fresh water lake. PLoS One 4(9): e7264.
14. Li L, Pesavento PA, Shan T, Leutenegger CM, Wang C (2011) Viruses in diarrhetic dogs include novel kobuviruses and sapoviruses. J Gen Virol 92 (Pt 11): 2534–2541.
15. Jere KC, Mlera L, O'Neill HG, Potgieter AC, Page NA, et al. (2011) Whole genome analyses of African G2, G8, G9, and G12 rotavirus strains using sequence-independent amplification and 454 pyrosequencing. J Med Virol 83(11): 2018–2042.
16. Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, et al. (2008) Metagenomic analysis of human diarrhea: viral detection and discovery. PLoS Pathog 29; 4(2): e1000011.
17. Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton, et al. (2003) Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol 185(20): 6220–6223.
18. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, et al. (2011) The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res 21: 1616–1625.

19. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, et al. (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466(7304): 334–338.
20. Victoria JG, Kapoor, Li L, Blinkowa O, Slikas B, et al. (2009) Metagenomic Analyses of Viruses in Stool Samples from Children with Acute Flaccid Paralysis. J Virol 83(9): 4642–4651.
21. Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, et al. (2006) RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. PLoS Biol 4 (1): e3: 10.1371.
22. Shan T, Li L, Simmonds P, Wang C, Moeser A, et al. (2011) The fecal virome of pigs on a high-density farm. J Virol 85(22): 11697–11708.
23. Palmenberg AC, Sgro JY (2005) Virus Particle Structures. In: Fauquet CM, Mayo MA, Maniloff J, Desselberger U, Ball LA (eds) Virus Taxonomy: Eight Report of the International Committee on Taxonomy of Viruses Elsevier Academic Press, Amsterdam, Holland. pp 12–22.
24. Cann AJ, Fandrich SE Heaphy S (2005) Analysis of the Virus Population Present in Equine Faeces Indicates the Presence of Hundreds of Uncharacterized Virus Genomes. Virus Genes 30(2): 151–156.
25. Day JM, Ballard LL, Duke MV, Scheffler BE Zsak L (2010) Metagenomic analysis of the turkey gut RNA virus community. Virol J 7: 313.
26. Li L, Shan T, Wang C, Côté C, Kolman J, et al. (2011) The Fecal Viral Flora of California Sea Lions. J Virol 85(19): 9909–9917.
27. Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, et al. (2011) The Fecal Viral Flora of Wild Rodents. PLoS Pathog 7(9): e1002218.
28. Dreier J, Störmer M Kleesiek K (2005) Use of Bacteriophage MS2 as an Internal Control in Viral Reverse Transcription-PCR Assays. J Clin Microbiol 43(9): 4551–4557.
29. Costafreda MI, Bosch A, Pintó RM (2006) Development, evaluation, and standardization of a real-time TaqMan reverse transcription-PCR assay for quantification of hepatitis A virus in clinical and shellfish samples. Appl Environ Microbiol 72 72(6): 3846–3855.
30. Ward P, Poitras E, Leblanc D, Letellier A, Brassard J, et al. (2009) Comparative analysis of different TaqMan real-time RT-PCR assays for the detection of swine Hepatitis E virus and integration of Feline calicivirus as internal control. J Appl Microbiol 106(4): 1360–1369.
31. Breitbart M, Salamon P, Andresen B, Mahaffy JM, Segall AM, et al. (2002) Genomic analysis of uncultured marine viral communities. Proc Natl Acad Sci U S A 99(22): 14250–14255.
32. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F (2009) Laboratory procedures to generate viral metagenomes. Nature Prot 4(4): 470–483.
33. Johne R, Müller H, Rector A, van Ranst M, Stevens H (2009) Rolling-circle amplification of viral DNA genomes using phi29 polymerase. Trends Microbiol 17(5): 205–211.
34. Kim Y, Chang KO, Straw B, Saif LJ (1999) Characterization of group C rotaviruses associated with diarrhea outbreaks in feeder pigs. J Clin Microbiol 37(5): 1484–1488.
35. Halaihel N, Masía RM, Fernández-Jiménez M, Ribes JM, Montava R, et al. (2010) Enteric calicivirus and rotavirus infections in domestic pigs. Epidemiol Infect 138(4): 542–548.
36. Pignatelli M, Moya A (2011) Evaluating the fidelity of de novo short read metagenomic assembly using simulated data. PLoS One 6(5): e19984.
37. Lin Y, Li J, Shen H, Zhang L, Papasian CJ, Deng HW (2011) Comparative studies of de novo assembly tools for next-generation sequencing technologies. Bioinformatics 27(15): 2031–2037.
38. Blinkova O, Victoria J, Li Y, Keele BF, Sanz C, et al. (2010) Novel circular DNA viruses in stool samples of wild-living chimpanzees. J Gen Virol 91 (Pt 1): 74–86.
39. Kim HK, Park SJ, Nguyen VG, Song DS, Moon HJ, et al. (2011) Identification of a novel single stranded circular DNA virus from bovine stool. J Gen Virol, Epub ahead of print.
40. Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, et al. (2010) Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. J Virol 84(4): 1674–1682.
41. Ge X, Li J, Peng C, Wu L, Yang X, et al. (2011) Genetic diversity of novel circular ssDNA viruses in bats in China. J Gen Virol 92 (Pt 11): 2646–2653.
42. Sambrook J, Russel D (2001) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
43. Breitbart M, Fekts B, Kelley S, Mahaffy JM, Nulton J, et al. (2004) Diversity and population structure of a near-shore marine-sediment viral community. Proc Biol Sci 271(1539): 565–74.
44. Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics 95(6): 315–327.
45. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17): 3389–3402.
46. R Development Core Team (2011) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
47. Shannon CE (1997) A mathematical theory of communication. MD Comput. 14(4): 306–317.

3.2 Second paper: The general composition of the faecal virome of pigs depends on age, but not on feeding with a probiotic bacterium

Jana Sachsenröder, Sven O. Twardziok, Matthias Scheuch, Reimar Johne

PLoS One. 2014, 9(2):e88888.

Summary of paper 2

The composition of pig viromes has been only scarcely investigated so far. Especially factors, which could influence this composition are largely unexplored.

In this study, the effect of feeding a probiotic bacterial strain and the effect of the pig age on the faecal virome were investigated. To this end, an experimental pig feeding trial with sows and their piglets, which either received the probiotic bacterium *Enterococcus faecium* NCIMB 10415 (probiotics group) or not (control group), was performed. Pooled faecal samples of both groups were analysed at different time-points. For this purpose, the entire virus particles were purified, the nucleic acid was extracted, subjected to process controlled Next Generation Sequencing (454 pyrosequencing) and bioinformatically analysed.

Using BLAST comparison an average of 14% of the generated sequence reads showed significant sequence similarities to known viruses. The proportion of mammalian viruses compared to all identified viruses continuously decreased from the youngest piglets (12 days old, 55-77%) through the 54 days old piglets (24-30%) to the sows (8-10%). In contrast, the proportion of bacteriophages increased constantly from the youngest piglets (22-44%) through in the older piglets (68-72%) to the sows (90%). A clear effect of age on the composition of the virome showed also the Shannon index, which reflects the degree of diversity and which continuously increased with age. The application of a probiotic strain did not change the virome composition as no consistent differences between the viromes of probiotic treaded and control group animals were evident.

Altogether, the analysis underlines the high variability of pig faecal viromes and demonstrates a main dependence of the virome by the age of the pigs. A general

effect of the probiotic treatment on the virome composition was not evident by using the *E. faecium* NCIMB 10415 strain and the applied metagenomic method.

Key messages of paper 2

- The composition of the pig virome is highly variable.
- It is mainly influenced by the age of the pig, showing a decrease of the percentage of pig viruses and an increase of the percentage of bacteriophages by increasing age.
- The degree of diversity of the viruses contained in faeces of pigs also increases by age.
- The general composition of the pig virome is not influenced by feeding with the probiotic bacterium *E. faecium* NCIMB 10415.

Own contribution to paper 2

In this study, I collected the different pig samples and performed the purification of the virus particles from eight faecal samples. The whole protocol for metagenomic analysis as described in paper 1 was applied by me to these samples. Again, the 454 sequencing was done by a commercial company. I performed the main part of the bioinformatics analysis of the generated processed reads in collaboration with S. Twardziok (FU Berlin) and M. Scheuch (FLI, Insel Riems). Additionally, I wrote the major part of the manuscript and was engaged in critical reading and revision of the whole manuscript.

OPEN ACCESS Freely available online

PLOS ONE

The General Composition of the Faecal Virome of Pigs Depends on Age, but Not on Feeding with a Probiotic Bacterium

Jana Sachsenröder[1], Sven O. Twardziok[2], Matthias Scheuch[3], Reimar Johne[1]*

1 Federal Institute for Risk Assessment, Berlin, Germany, **2** Institute for Molecular Biology and Bioinformatic, Charité Berlin, Berlin, Germany, **3** Friedrich-Loeffler-Institute, Greifswald - Insel Riems, Germany

Abstract

Background: The pig faecal virome, which comprises the community of viruses present in pig faeces, is complex and consists of pig viruses, bacteriophages, transiently passaged plant viruses and other minor virus species. Only little is known about factors influencing its general composition. Here, the effect of the probiotic bacterium *Enterococcus faecium* (*E. faecium*) NCIMB 10415 on the pig faecal virome composition was analysed in a pig feeding trial with sows and their piglets, which received either the probiotic bacterium or not.

Results: From 8 pooled faecal samples derived from the feeding trial, DNA and RNA virus particles were prepared and subjected to process-controlled Next Generation Sequencing resulting in 390,650 sequence reads. In average, 14% of the reads showed significant sequence identities to known viruses. The percentage of detected mammalian virus sequences was highest (55–77%) in the samples of the youngest piglets and lowest (8–10%) in the samples of the sows. In contrast, the percentage of bacteriophage sequences increased from 22–44% in the youngest piglets to approximately 90% in the sows. The dominating mammalian viruses differed remarkably among 12 day-old piglets (kobuvirus), 54 day-old piglets (boca-, dependo- and pig stool-associated small circular DNA virus [PigSCV]) and the sows (PigSCV, circovirus and ''circovirus-like'' viruses CB-A and RW-A). In addition, the Shannon index, which reflects the diversity of sequences present in a sample, was generally higher for the sows as compared to the piglets. No consistent differences in the virome composition could be identified between the viromes of the probiotic bacterium-treated group and the control group.

Conclusion: The analysis indicates that the pig faecal virome shows a high variability and that its general composition is mainly dependent on the age of the pigs. Changes caused by feeding with the probiotic bacterium *E. faecium* could not be demonstrated using the applied metagenomics method.

Citation: Sachsenröder J, Twardziok SO, Scheuch M, Johne R (2014) The General Composition of the Faecal Virome of Pigs Depends on Age, but Not on Feeding with a Probiotic Bacterium. PLoS ONE 9(2): e88888. doi:10.1371/journal.pone.0088888

Editor: Xiang-Jin Meng, Virginia Polytechnic Institute and State University, United States of America

Received November 12, 2013; **Accepted** January 13, 2014; **Published** February 19, 2014

Funding: The study was funded by the Deutsche Forschungsgemeinschaft (DFG) through grant SFB852/1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Reimar.Johne@bfr.bund.de

Introduction

The viral community present in faeces is composed of a variety of viruses originating from the gut tissue, from intestinal microorganisms or from ingested food. The totality of viruses present in faeces has also been frequently designated as the faecal virome [1,2]. The functions of the faecal virome are supposed to be manifold, which include roles for the viruses as pathogens, regulators of bacterial growth, gene-transfer vehicles and modulators of the immune system [3–6]. Early insights into the composition of the human faecal virome were provided by random cloning strategies [7,8]. Later on, the availability of deep sequencing methods lead to more comprehensive analyses of faecal viromes [1,9], including the development of process-controlled techniques enabling comparison of different analyses [10].

The composition of the faecal virome of pigs has been studied recently [2,10,11]. Although samples derived from different continents had been analysed in these studies, the general composition was found to be similar. The majority of the detected virus sequences belonged to bacteriophages and pig viruses. Only a few sequences belonged to plant viruses as well as other viruses. Most of the bacteriophage sequences originated from viruses belonging to the families *Siphoviridae*, *Microviridae* and *Myoviridae* [10]. The most abundant porcine viruses were kobuvirus, rotavirus, pig stool-associated small circular DNA virus (PigSCV), astrovirus, sapovirus and enterovirus B. Most of them represent widely distributed enteric viruses of pigs [2,10,11]. Whereas rotaviruses are well-known pathogens of piglets, which may lead to diarrhoea [12,13,14], the clinical importance of the other viruses is a subject of controversy [10,15–22].

Only little is known about the stability and dynamics of the faecal virome under different conditions. For the human faecal virome, Reyes et al. [9] investigated the intra- and interpersonal variation by analysing faeces of monozygotic twins and their

mothers at different time-points. By this, it was found that the viromes were unique to the individuals regardless of their degree of genetic relatedness. Minot et al. [1] analysed the inter-individual variation of the human faecal virome and its dynamic response to diet. It was shown that the largest source of variance among the viromes was caused by interpersonal variations and not by the diet. A high interpersonal diversity of gut bacteriophages was also described in two humans which were monitored over a 2.5 year period [23]. In another study, a much lower diversity of the virus community was found in infants as compared to adults [24]. Although this study has been conducted by cloning followed by classical sequencing, mathematical modelling of the derived sequence data indicated that the virome of adults was composed of approximately 2000 genotypes as compared to only 8 genotypes in one week-old infants.

The observed beneficial effects of probiotic bacteria on enteric virus infections have been recently reviewed by Colbere-Garapin et al. [6]. This includes clinical studies showing beneficial effects of probiotic bacteria in children with rotavirus-caused diarrhoea [25–27]. Feeding with probiotic microorganisms such as *Lactobacillus rhamnosus GG, Saccharomyces boulardii* or *Bifidobacterium lactis* resulted in milder clinical symptoms, reduced virus shedding and shortened the duration of diarrhoea in children [27–29]. In pigs, *Enterococcus (E.) faecium* NCIMB 10415 is a commonly used probiotic bacterium [35,37]. It has been shown recently, that feeding of pigs with this probiotic bacterium affected shedding of enteric viruses dependent on the virus species [30]. Especially, rotavirus was shed later and in lower amounts in the group of piglets that received *E. faecium* NCIMB 10415 as compared to the control group. The specific mechanisms responsible for this effect are not known so far. However, changes in the mucosal and systemic immunity due to feeding with *E. faecium* NCIMB 10415 have been described [31–34]. In addition, direct interactions of this bacterium with enteric virus particles have been observed in *in vitro* studies [35]. However, it is not known so far, whether probiotic bacteria can also influence the general composition of the faecal virome, e.g. by changing the composition of the bacterial community, which represents the host population for bacteriophages, or by direct interactions with specific viruses.

Table 1. Detection of *E. faecium* NCIMB 10415 in faeces of sows and their piglets.

	sows		piglets	
	28 d ap	**14 d pp**	**12 d**	**54 d**
group C (control)	nd[1]	nd[1]	nd[1]	nd[1]
group P (*E. faecium*)	nd[1]	6.97[2]	nd[1]	3.75[2]

[1]nd = not detected.
[2]quantitative real-time PCR results are expressed as decadic logarithmic numbers of cells per gram faeces according to Starke et al. [37].
d = days; ap = ante partum; pp = post partum.
doi:10.1371/journal.pone.0088888.t001

The primary aim of the presented study was to analyze the effect of the probiotic bacterium *E. faecium* NCIMB 10415 on the general composition of the faecal virome in pigs. Faecal samples from sows and their piglets experimentally fed with or without the probiotic bacterium were analyzed using a process-controlled deep sequencing method. The populations of the detected virus sequences were compared between the feeding groups as well as the age groups and general insights into the stability and dynamics of the pig faecal virome under different age-related and feeding conditions were generated.

Materials and Methods

Ethic Statement

The animal experiment (pig feeding trial) was approved by the local state office of occupational health and technical safety "Landesamt für Gesundheit und Soziales Berlin" (LaGeSo Reg. Nr. 0347/09).

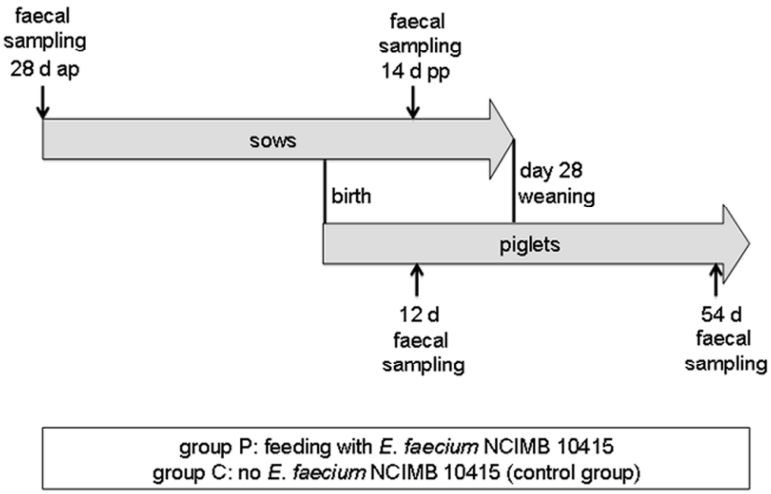


Figure 1. Schematic view of the experimental feeding trial. Sows and their piglets were fed with (group P) or without (group C) supplementation of the probiotic bacterium *Enterococcus faecium* NCIMB 10415. Pooled faecal samples were derived from the sows and the piglets of both groups at the indicated time-points.
doi:10.1371/journal.pone.0088888.g001

Animal Experiment and Sampling Scheme

The design of the pig feeding trial has been described in detail by Martin et al. [36] and is schematically shown in Figure 1. Briefly, sows and their piglets received either no probiotic bacterium or approximately 5×10^6 cfu/g *E. faecium* NCIMB 10415, which was fed with their diet starting at 28 days ante partum. The sows received a commercial diet (UNA-HAKRA, Hamburg, Germany). Additional feeding of the piglets started at 12 days of age with a non-medicated non-commercial pre-starter diet [36]. After weaning at 28 days of age, they were fed with a non-commercial mash starter diet [37]. The homogenous distribution of the probiotic in feed has been previously demonstrated by a colony hybridization assay [37].

The faeces of 6 sows of each group were sampled at day 28 ante partum (before *E. faecium* diet) and at day 14 post partum. Faeces of their piglets (6 from each group) were collected at day 12 and at day 54 of age (end of the experiment). Piglets were euthanized at the end of the experiment by intracardial injection of a lethal dose of tetracaine hydrochloride, mebezonium iodide and embutramide (T61, Intervet, Unterschleißheim, Germany). Although the whole experiment included a larger number of animals [30,37], only faeces of piglets were analyzed, for which the faeces of their mother sows had also been analysed. The faeces of each group and time-point were pooled. The samples were stored at −20°C until analysis.

Quantification of *E. faecium* NCIMB 10415 in Faeces

DNA was extracted from faecal samples and subsequently analyzed by real-time PCR specific for *E. faecium* NCIMB 10415 as previously described [38]. The standards used for quantification were prepared from negative pig faecal samples spiked with known amounts of cultured *E. faecium* NCIMB 10415 cells as described by Starke et al. [37] Results are expressed as log of cell numbers per g faeces.

Process Control

Three different bacteriophages (M13, MS2, T4) were grown, titrated and used as process controls for monitoring the efficiency of the virome analysis procedure as described previously [10]. A total of 10 µl of the bacteriophage mixture containing approximately 10^5 plaque-forming units of each bacteriophage was added per 1 g faeces.

Purification and Concentration of Virus Particles

Virus particles were purified from the faecal samples by a combination of tangential flow filtration (TFF) and caesium chloride (CsCl) density gradient ultracentrifugation, and concentrated by centrifugal filtration and TFF as described [10,39]. A total of 17 g of pooled faecal samples from the sows were used. Due to limited availability of faeces in the youngest age group, 1.7 g of the pooled faecal samples from the piglets was used. The samples were spiked with test-phages and resuspended 1:10 in SM-buffer (100 mM NaCl, 8 mM $MgSO_4$, 50 mM Tris-HCl pH 7.5) by magnetic stirring. The sample was centrifuged at 10,000 g for 30 min in order to remove the large particulate debris and the supernatant was collected. The procedure was repeated by centrifugation for 3 hours at 10,000 g to remove smaller particular structures. Afterwards, a first TFF was performed using a 0.22 µm filter (PALL Corporation, Middleton; MA, USA) to remove bacterial and eukaryotic cells and debris. The remaining filtrate was subjected to a second TFF with a 50 kDa filter (PALL Corporation, Middleton; MA, USA) in order to concentrate the virus particles. The viral preparations were further concentrated by centrifugation through Vivaspin 50,000 MWCO concentrators (Sartorius Stedim Biotech GmbH, Götting, Germany) at 3,500 g resulting in a final volume of 36 ml. The preparation was divided into two fractions of 18 ml, which were added separately onto preformed stepwise caesium chloride (CsCl) density gradients with density layers of 1.7, 1.5, 1.35 and 1.2 g/ml (5 ml each) and ultracentrifuged at 20,000 g for 14 hours at 10°C. The 1.35–1.5 g/ml layers were collected from the gradients using a syringe.

Nucleic Acid Preparation and Deep Sequencing

To eliminate free DNA present in the virus concentrate, an aliquot of 1 mL CsCl purified virus solution was treated with 50 units DNase I (2,000 U/mg, bovine pancreas grad II; Roche Diagnostics GmbH, Mannheim, Germany) for 45 min at 37°C, followed by heat inactivation for 10 min at 65°C. Thereafter, DNA and RNA were extracted simultaneously using NucliSENS magnetic extraction (bioMerieux, Nürtingen, Germany). The extracted nucleic acids (75 ng per reaction) were randomly primed for cDNA synthesis using the TransPlex® Complete Whole Transcriptom Amplification Kit (WTA2, Sigma-Aldrich, St. Louis, MO, USA) according to the protocol recommended by the supplier; however, the annealing temperature was decreased to 40°C (2 cycles) and 45°C (2 cycles) in order to enable the simultaneous amplification of DNA and RNA [10]. Aliquots of 75 µl each were removed from the WTA2 reaction at different cycle numbers, purified and size-selected using MobiSpin S-400 Columns (MoBiTec, Göttingen, Germany). The DNA concentration was measured from the preparations using a nanodrop spectrometer (Analytic Jena, Jena, Germany) and the preparation derived from a minimum of amplification cycles with a DNA concentration above 50 ng/µl was chosen for deep sequencing. A total of 1 µg DNA was used for deep sequencing on a 1/8 plate of the GS-FLX sequencer 454 Titanium (GS Titanium SV emPCR kit (Lib-L) v2; GS Titanium PicoTiterPlate Kit 70×75; GS Titanium Sequencing Kit XLR70t; Life Sciences, Roche, Branford, USA) according to the manufacturer's protocol. The raw sequence data have been submitted to the Sequence Read Archive (SRA) at GenBank as BioProject PRJNA232620 with SRA accession numbers SRP034937 (SRX396427–SRX396434).

Data Analysis

Primary sequence analysis was performed in two steps: identification of all virus species included in the samples and analysis of species abundances regarding selected sets of species. Raw sequence reads were subjected to primer/adaptor trimming using SeqMan (DNASTAR, Lasergene, USA) and selection for a minimum length of 50 nt. In parallel, all primary reads were subjected to de novo contig assemble using the 454 Newbler Assembler [40] software, with criteria of 90% minimum overlap identity and a minimum overlap length of 40 nt.

In order to create a local database containing all virus sequences with significant homologies to the sequence reads, homology searches for all primary reads were performed with BLASTx [41] against the non-redundant nucleotide database of NCBI [42]. In parallel, homology searches for the contigs were performed with CLC Main Workbench 6.2 [43] against the viral genome non-redundant reference sequence nucleotide database [44] and additional sequences from recently discovered viruses using the tBLASTx algorithms [39]. From both approaches, all BLAST results with an E-value $<=10^{-4}$ were selected and used for creation of the local sequence database.

Using this database, abundances of species were calculated. For Bray Curtis dissimilarity (see below), specific subsets, which consisted of mammalian viruses, bacteriophages or Enterococcus

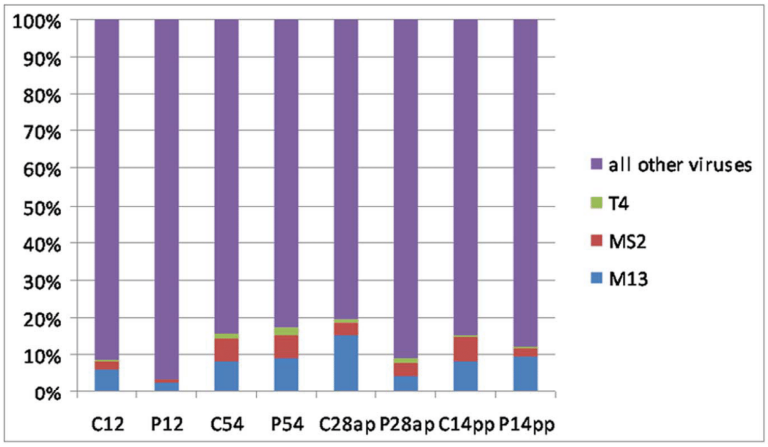


Figure 2. Detection of bacteriophages used as process control in the analysed samples. Equal amounts of the bacteriophages T4, MS2 and M13 were added to the pooled faecal samples prior to analysis and the generated reads were screened for the recovered genomic sequence reads of these bacteriophages. The percentage of the number of reads from these bacteriophages in relation to all detected virus reads is indicated. The samples are designated with the group letter (C – control, P – probiotic) and the day number (ap – ante partum, pp – post partum).
doi:10.1371/journal.pone.0088888.g002

phages, were used. In all cases, trimmed reads were mapped against the sequences of the local database to calculate species abundances with the readmapper Bowtie 2 2.0.5 [45]. Thereafter, numbers of mapped reads were corrected for multiple read assignments. The reads of the bacteriophages used as process control were subtracted from the number of the virus reads in subsequent analyses. Shannon index [46] was calculated to compare the diversity of the species identified by primary reads. The Shannon index is maximal for a sample with a balanced species distribution and it has a low value for a sample with an uneven species distribution; e.g. if some single species are a highly abundant. The maximal value depends on the number of species in a sample. Bray Curtis dissimilarity [47] was calculated for pairwise comparisons of samples and dendrograms were constructed by hierarchical clustering with the average linkage method. This analysis included counting of detected species and determination of their taxonomy, which was also used to determine the virus hosts (bacteria, vertebrates, plants etc.).

Results

Detection of *E. faecium* NCIMB 10415 in Faeces

A total of 8 pooled faecal samples were derived from sows and their piglets from an experimental feeding trial with the probiotic bacterium *E. faecium* NCIMB 10415. Four of the samples were derived from animals receiving the probiotic bacterium (group P) and four samples originated from the control group that did not receive probiotics (group C). A detailed scheme of the feeding trial is presented in Figure 1.

Table 2. Numbers and relative abundance of viral sequences and process control phage (test-phage) sequences in the analyzed samples.

			test-phages					without test-phages					
	sample name[1]	all reads	M13	MS2	T4	% test-phage of all reads	% test-phage of all viruses	viurses incl. phages	%_viuses incl. phages	viruses without phages	%_viruses without phages	only bacterio-phages	%_only bacterio-phages
piglets	C12	39670	482	168	18	1.7	9.4	7117	17.9	3981	10.0	3136	7.9
	P12	26115	221	67	11	1.1	3.8	7911	30.3	6115	23.4	1796	6.9
	C54	51534	745	552	114	2.7	18.6	7573	14.7	2387	4.6	5186	10.1
	P54	58796	1046	716	214	3.4	20.7	9535	16.2	2584	4.4	6951	11.8
sows	C28ap	64227	1085	262	63	2.2	24.4	5774	9.0	553	0.9	5221	8.1
	P28ap	55502	231	172	70	0.9	10.2	4628	8.3	429	0.8	4199	7.6
	C14pp	55851	495	378	41	1.6	18.1	5038	9.0	672	1.2	4366	7.8
	P14pp	38955	337	81	24	1.1	14.1	3143	8.1	338	0.9	2805	7.2
	average	48831	580	300	69	1.8	14.9	6340	14.2	2132	5.8	4208	8.4
	total	390650	4642	2396	555			50719		17059		33660	

[1]Sample names: C–control group; P–probiotic group; 12/54 days old; 28 ap: 28 days ante partum; 14 pp: 14 days post partum.
doi:10.1371/journal.pone.0088888.t002

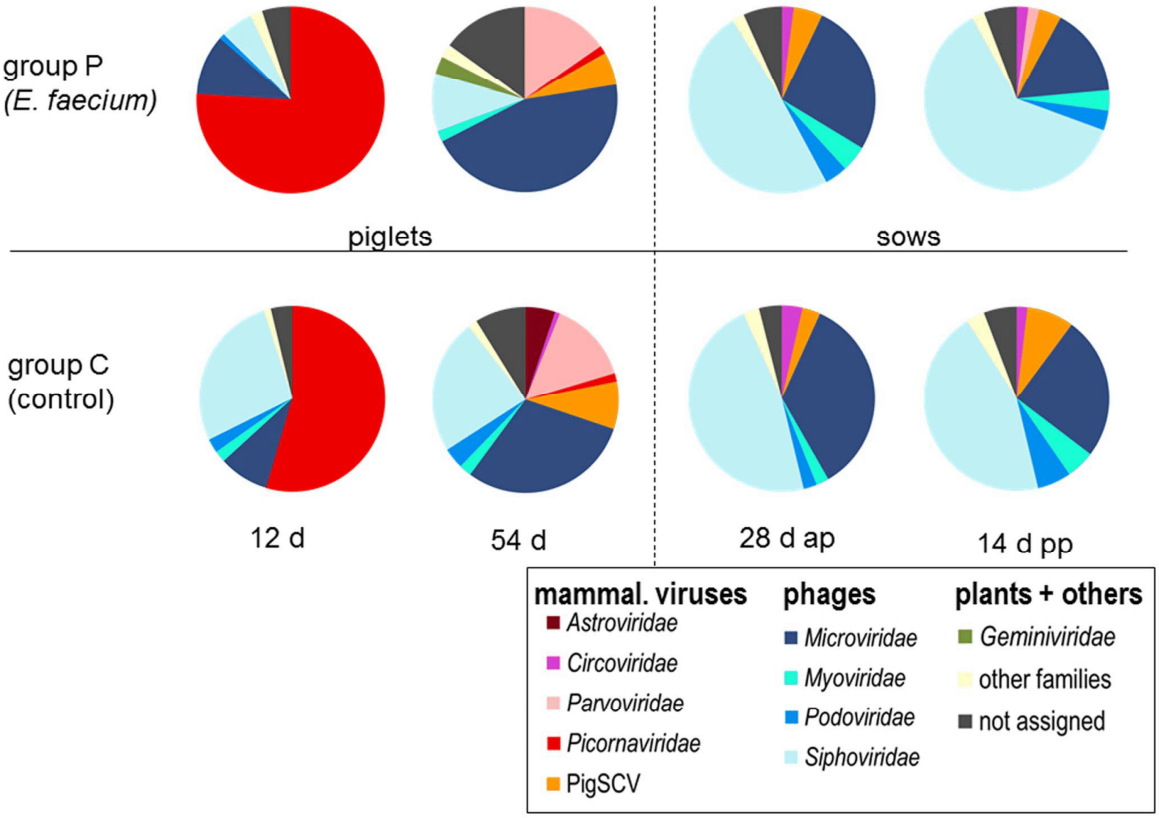


Figure 3. Relative abundance of virus families in the analyzed faecal viromes. The diagrams show the number of reads with sequence identities to a certain virus family in relation to all virus reads. Different colours were used for different virus families (see Legend). Virus families containing mammalian viruses are shown in shades of red, whereas those families containing bacteriophages are shown in shades of blue. The group receiving the probiotic bacterium *E. faecium* NCIMB 10415 (group P) is shown in the upper row; the control group (group C) is in the lower row. Samples derived from piglets are shown left and those from the sows are shown right. The time-points of sampling are indicated below.
doi:10.1371/journal.pone.0088888.g003

The presence of *E. faecium* NCIMB 10415 in the faeces of sows and their piglets was analyzed by quantitative real-time PCR. As shown in Table 1, *E. faecium* NCIMB 10415 was not detected in the samples of the control group. Also, no *E. faecium* NCIMB was detectable in the sample taken from the sows of the probiotic group immediately at the beginning of the experiment (28 day ante partum) as well as in the samples from the 12 day-old piglets of this group, which were still suckled at this time-point. Considerable amounts of *E. faecium* NCIMB 10415 were demonstrated in the samples taken from the sows at 14 day post partum and from the 54 day-old piglets, both belonging to the probiotic group.

Process-controlled Deep Sequencing of Virus Genomes in Faecal Samples

The 8 pooled faecal samples were analyzed by process-controlled deep sequencing. In total, 390,650 reads were generated, with an average of 48,831 reads per sample. The efficiency of the whole method was monitored by a process control consisting of three bacteriophages, which were added in constant amounts to the samples. In all samples the three test-phages could be detected representing 0.9% to 3.4% of all generated reads. The numbers of totally generated reads, test-phage reads and other virus reads is summarized for the individual samples in Table 2. The number of the test-phage reads in relation to the total virus reads ranged from 3.8% to 24.4% and is shown in Figure 2.

Analysis of Detected Virus Families and Respective Virus Hosts

Using a cut-off E-value of $<= 10^{-4}$ for the BLASTx homology search of the sequences, the viral reads could be assigned to 36 known virus families. Only 10 of these families dominated the faecal viromes representing more than 1% in at least one of the samples. As shown in Figure 3 and Table S1, the composition of the faecal viromes according to virus families varied remarkably among the samples. A grouping of the virus families according to the taxonomic kingdom of hosts of the contained viruses revealed that the main detected groups were mammalian viruses (colored red in Fig. 3) and bacteriophages (colored blue in Fig. 3). In contrast, viruses from other hosts (insects, plants, amphibians and fungi) ranked together between 0.2% and 3.4% only.

A closer inspection of the proportion of the read numbers from mammalian viruses compared to that from bacteriophages revealed marked differences between the samples derived from different age groups. In the youngest piglet group (12 days of age), the main fraction consisted of mammalian viruses with 55%

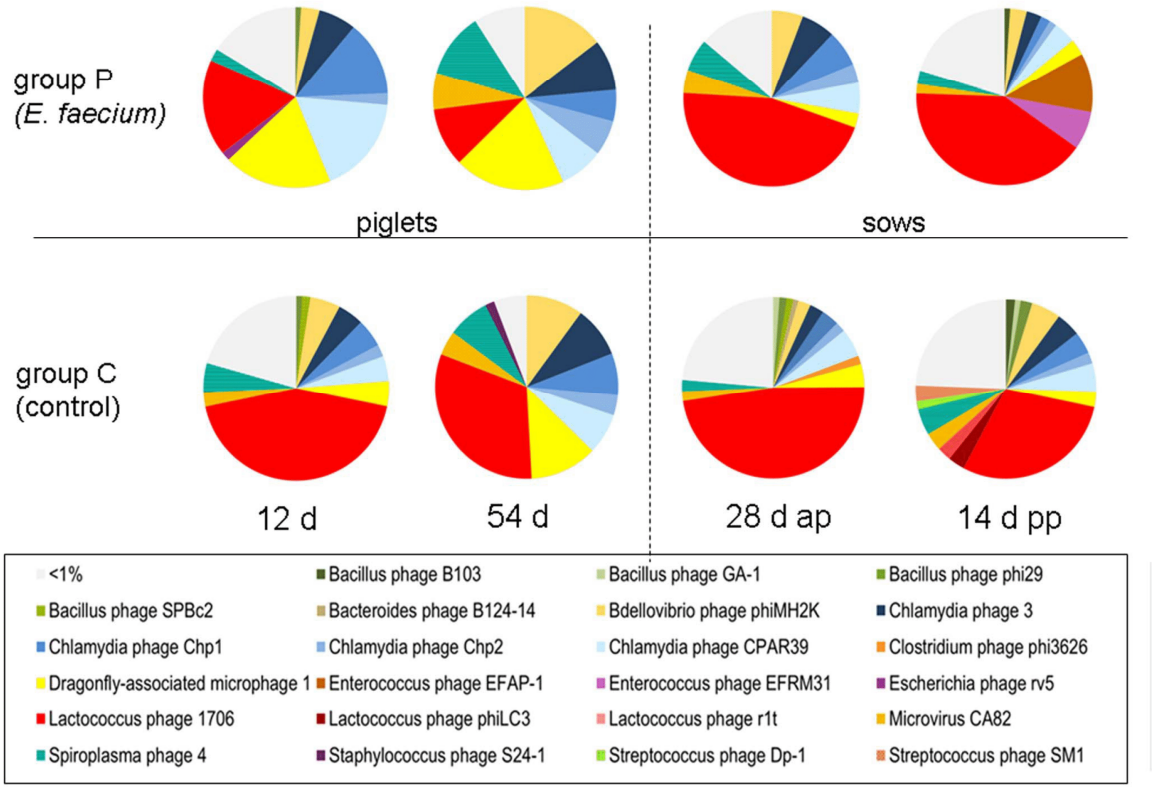


Figure 4. Relative abundance of bacteriophage species among all bacteriophages detected in the analyzed faecal viromes. The diagrams show the number of reads with sequence identities to a certain bacteriophage species in relation to all bacteriophage reads. Different colours were used for different bacteriophage species (see Legend). Bacteriophage species showing an abundance of less than 1% in a distinct faecal virome are subsumed in light grey colour (<1%). The group receiving the probiotic bacterium *E. faecium* NCIMB 10415 (group P) is shown in the upper row; the control group (group C) is in the lower row. Samples derived from piglets are shown left and those from the sows are shown right. The time-points of sampling are indicated below.
doi:10.1371/journal.pone.0088888.g004

(control group) and 77% (probiotic group). In the group of 54 day-old piglets, the proportion of mammalian viruses was reduced to 24% (control group) and 30% (probiotic group). Within the four groups of the sows (one year old) the amount of mammalian viruses ranged from 8% to 12%. In contrast to those findings, the proportion of bacteriophages increased with the age of the pigs. In the 12 day-old piglets, 44% (control group) and 22% (probiotic group) of the reads relate to bacteriophages. The percentage of bacteriophages increases in the 54 day-old piglets to 68% (control group) and 72% (probiotic group), whereas approximately 90% of the virus reads belong to bacteriophages in the four sow groups. No differences in the general composition of virus families or the respective hosts were evident, when the probiotic group was compared to the control group.

Analysis of Bacteriophages

In overall, sequences with significant identities to 524 known bacteriophage species were detected. The bacteriophages most abundant in the eight samples are shown in Figure 4 and Table S2. In all cases, the bacteriophage population is dominated by 9 to 16 species, which represent 76–90% of all bacteriophage reads of the respective sample. The most abundant phages as identified by the highest number of reads with sequence identities to known bacteriophage genomes are Lactococcus phage 1706, Dragonfly-associated microphage 1, Chlamydia phages 4, Chp1 and Chp2, Bdellovibrio phage phiMH2K, Spiroplasma phage 4, Microvirus CA82 as well as Enterococcous phages EFAP-1 and EFRM31. A comparison between the bacteriophage populations of the specific samples indicated that many of the most abundant bacteriophage species are present in all samples, however, with different relative frequency. Apart from that, the composition of the faecal virome with regard to bacteriophage species was relatively variable between the samples and every sample contained its unique collection of bacteriophages. No consistent differences between age groups and feeding groups were obvious when the abundance of bacteriophage species was analyzed. Interestingly, the sample taken at day 14 post partum from the *Enterococcus faecium*-fed sows contained relatively high amounts of the Enterococcous phages EFAP-1 (10.8%) and EFRM31 (7.0%), which were only sporadically detected in the other groups (0.02% to 0.6%).

Analysis of Animal Viruses

By analysis of all virus reads excluding the bacteriophages, sequences with significant identities to 205 known virus species were detected. Among that, 92.9% of the sequences belonged to viruses infecting mammalian animals. This percentage of mam-

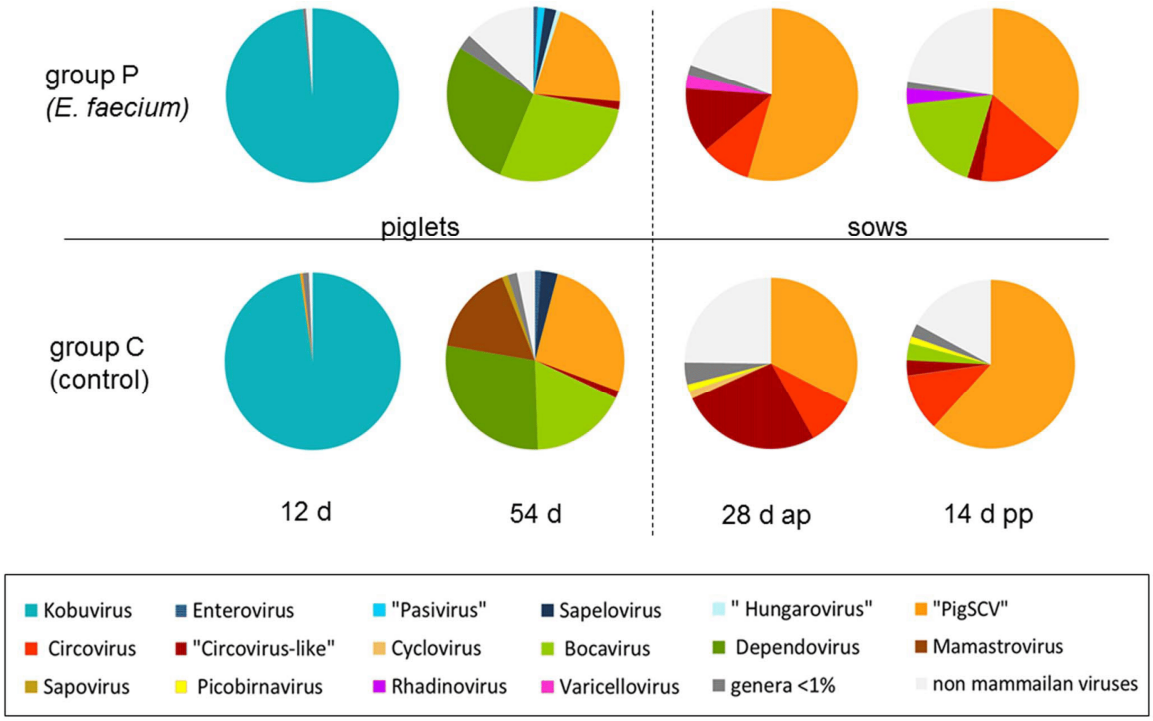


Figure 5. Relative abundance of mammalian virus genera among all animal viruses detected in the analyzed faecal viromes. The diagrams show the number of reads with sequence identities to a certain mammalian virus genus in relation to all animal virus reads. Different colours were used for different mammalian virus genera (see Legend). Mammalian viruses, which are so far not assigned to a certain genus, are indicated in apostrophes. Mammalian virus genera showing an abundance of less than 1% in a distinct faecal virome are subsumed in dark light grey colour (<1%). Viruses from non-mammalian hosts are subsumed in light grey colour. The group receiving the probiotic bacterium *E. faecium* NCIMB 10415 (group P) is shown in the upper row; the control group (group C) is in the lower row. Samples derived from piglets are shown left and those from the sows are shown right. The time-points of sampling are indicated below.
doi:10.1371/journal.pone.0088888.g005

malian viruses decreased with age, with an average of 99.1% in the 12 day-old piglet group, 91.8% in the 54 day-old piglet group and 79.0% in the sows. The relative percentage of the most abundant mammalian virus genera in the eight samples is shown in Figure 5 and Table S3, indicating that remarkable differences exist between the age groups. In the samples from 12 day-old piglets, almost all virus sequences (97.9%) belonged to porcine kobuvirus. In the samples from the 54 day-old piglets, dependovirus (27.9%), bocavirus (22.9%) and PigSCV (23.9%) dominated the virus sequence reads. In the samples from the sows, several different small circular DNA viruses such as PigSCV (46.3%) and circovirus including "circovirus-like" viruses (22.4%) were the most abun-

Table 3. Calculated Shannon indexes reflecting the diversity of the analyzed faecal viromes.

	Shannon index mammalia viruses		phages	
Samples[1]	**minimal value**	**maximal value**	**minimal value**	**maximal value**
C12	1,64898	3,295837	1,445446	3,091042
P12	2,158605	4,343805	1,167458	5,913503
C54	1,874457	4,634729	1,389295	6,226537
P54	2,066851	4,564348	1,300899	6,062785
C28ap	3,207631	4,812184	1,350963	6,267201
P28ap	3,127225	4,820282	1,754871	6,267201
C14pp	3,041299	4,682131	1,878738	6,23637
P14pp	2,644238	4,70048	1,732825	6,257668

[1]Sample designations: C–control group; P–probiotic group; 12/54 days old (piglets); 28 ap: 28 days ante partum (sows); 14 pp: 14 days post partum (sows).
doi:10.1371/journal.pone.0088888.t003

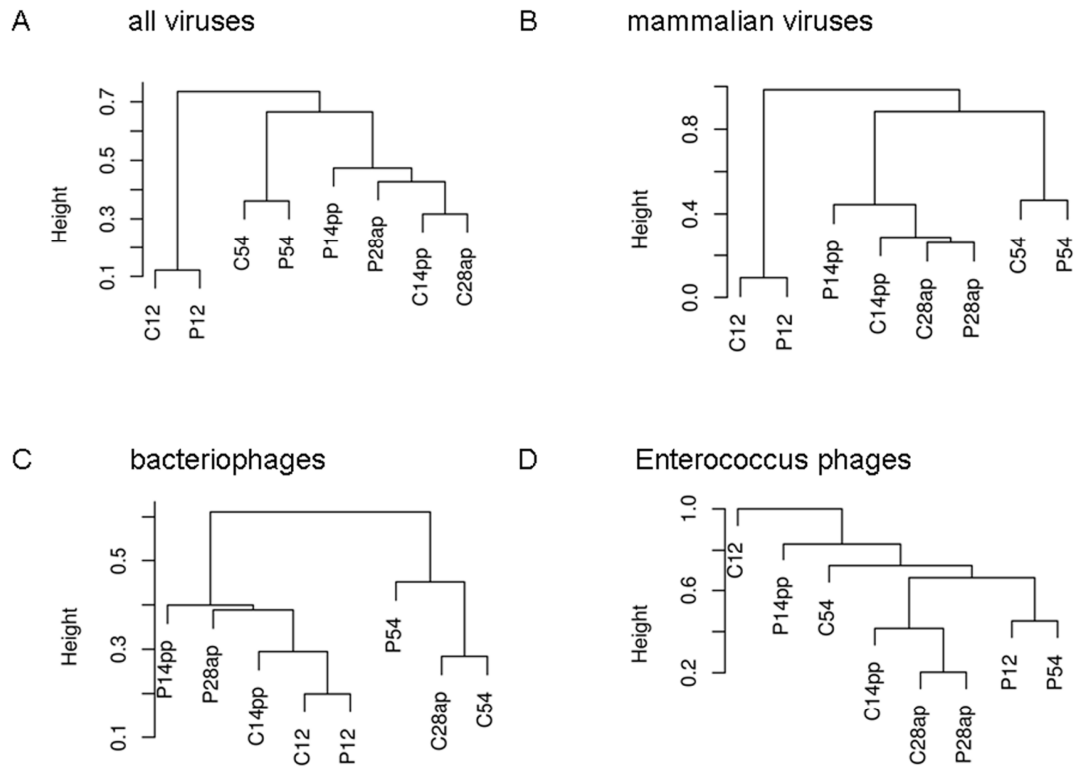


Figure 6. Cluster dendrograms showing the similarity of the composition of the analysed faecal viromes. The dendrograms were calculated using Bray Curtis dissimilarity. The presented dendrograms are based on a comparison of reads from all detected viruses (A), mammalian viruses (B), bacteriophages or Enterococcus phages (D). The samples are designated with the group letter (C – control, P – probiotic) and the day number (12/54 days old piglets; 28 days ante partum or 14 days post partum for sows). The dendrograms are scaled in Bray–Curtis dissimilarity units (BC; from 0 to 1).
doi:10.1371/journal.pone.0088888.g006

dant mammalian viruses. Among the "circovirus-like" viruses, sequences with highest identities to the viruses CB-A and RW-A were most often detected. No consistent differences were obvious between the group fed with the probiotic bacterium and the control group. However, a relatively high proportion of mamastrovirus sequences (16.3%) was detected in the sample derived from the 54 day-old piglets; while this virus was not detected in the other groups (less than 2 reads per sample).

Analysis of Diversity and Similarity of Faecal Viromes

The calculation of the Shannon index was used to assess the diversity of the sequences detected in the specific samples (Table 3). Generally, comparison of Shannon indexes between piglets and sows indicated that the diversity increased with age. When only the bacteriophage sequences were analysed, the average Shannon index of the piglet groups was 1.3 and that of the sows 1.7. For the mammalian virus sequences, the average Shannon index for the piglets was 1.9 and that for the sows 3.0.

Calculation of Bray-Curtis distances determined similarities of the faecal viromes detected in the specific samples. Figure 6 illustrates clustering of samples on the basis of Bray Curtis distance calculated by abundances of species-specific subsets. As shown in the dendrogram based on abundances of all virus species (including bacteriophages), a grouping according to age is evident (Fig. 6A). The two samples taken from the 12 day-old piglets cluster closely together and are separated from the other samples. Among these other samples, the two samples taken from the 54 day-old piglets form one separate branch, whereas the four samples of the sows are all contained in the other branch. A branching according to the feeding group is not evident from this dendrogram. The same grouping is evident, when only the mammalian virus sequences are analysed (Fig. 6B). The analysis of the bacteriophages shows no evident grouping according to age or feeding group (Fig. 6C). Also, no grouping according to age or feeding group was evident, when only the sequences of the Enterococcus phages were used for the analysis (Fig. 6D).

Discussion

Comparisons of the composition of intestinal viromes from different samples have been only scarcely described so far. A few studies investigated individual differences of faecal viromes and the influence of diet and age in humans [1,9,24], whereas similar studies on faecal viromes of pigs are almost missing. Technical problems with the use of deep sequencing methods for comparative virome analyses may represent one major problem in this context [10,48]. Here, we tried to overcome some of these problems by using a process-controlled deep sequencing approach [10]. By this, the efficiency of the analysis can be estimated for each sample, thereby enabling identification of major differences

due to different performances of the method. We could show here, that all types of the bacteriophages used as process control could be detected in the final data sets of all samples. This indicates that the method has a reproducible performance and the generated data can generally be used for comparative analyses. However, the detection rates of the process control bacteriophages varied between the samples from 0.9% to 3.4%. As the detection rate of the bacteriophages is – besides technical factors - also dependent on the amount of viruses initially present in the analyzed sample, improved deep sequencing methods enabling quantitative analyses should be developed in future for comparative virome investigations.

In the eight investigated pooled samples, the overall composition of the virus community was similar to that described for other pig faecal viromes [2,10,11]. The two major virus groups were bacteriophages and porcine viruses, whereas plant viruses and viruses with other hosts were only rarely detected. However, large differences were detected in the ratio between bacteriophages and mammalian viruses in the distinct samples; in addition, the diversity of detected virus species varied between the analysed viromes. These data indicate that the faecal virome of pigs is not uniform and static, but shows a remarkable variability. For human faecal viromes, a high variability even between the analyzed individuals has been described [1,9]. As only pooled faecal samples have been analyzed in the study presented here, future investigations are necessary in order to assess the inter-individual variability of faecal viromes of pigs.

The most obvious factor influencing the composition of the pig faecal virome was the age. The percentage of porcine viruses, which comprised the most abundant group in the youngest piglet samples, decreased dramatically in the samples from the older pigs. In parallel, the percentage of bacteriophages as well as the diversity of detected virus species increased by age. Interestingly, porcine kobuvirus and pig SCV, which both had been discovered only recently [10,49], were among the most frequently detected viruses in the faecal viromes of the youngest and oldest age groups, respectively. This underlines the importance of unfocused detection systems in order to get an undistorted picture of the abundance of viruses in a sample. As all samples analyzed here originated from an experimental feeding trial, the detected virus composition may vary in comparison to field-origin samples. However, the age-specific effect was strong and very similar in both analyzed groups, which were held completely separate during the whole period of the experiment. The differences may be explained by an age-related susceptibility to specific virus infections as well as by an increasing immunity to porcine viruses due to completed virus infections with higher age. In addition, the progressive diversification of the bacterial enteric flora, which serves as the host pool for bacteriophages, would also explain the increasing diversification of the virus flora by age. An increasing diversity of the virus species in faeces of humans has already been described [24].

In contrast to the age-related effect, no clear differences could be detected in the composition of the faecal viromes according to feeding with the probiotic bacterium *E. faecium* NCIMB 10415. A relatively high percentage of Enterococcus phages in the sample derived from an *E. faecium*-feeded group may indicate multiplication of the phage due to the application of its host. This explanation may indicate that a larger amount of the probiotic bacteria may be lysed by the bacteriophages and are therefore not available for the probiotic therapy; however, this interpretation is questionable as the bacteriophages were only found in one of the samples. Also, a relative high proportion of astrovirus was found in one of the samples of the control group. Interestingly, real-time RT-PCR analyses of samples derived from the same feeding trial confirmed the presence of astrovirus exclusively in the control group [30]. However, the same study indicated later shedding of rotaviruses with lower amounts in the probiotic group as compared to the control group, which was not detected by our virome analysis. A closer inspection of the data shows that up to 10^7 astroviruses per gram faeces were present in the samples, whereas only up to 10^5 rotaviruses per gram were detected [30]. Therefore, a lower sensitivity of the virome analysis method may explain the discrepancies and still deeper sequencings may be necessary in future to detect more subtle changes in the faecal virome composition due to probiotic feeding.

The composition of the identified bacteriophage species in the different samples revealed no consistent pattern. However, most of the detected sequence reads showed only moderate identities to the known bacteriophage sequences present in the database. Therefore, it has to be considered that the majority of the detected sequences belong to so far unknown bacteriophages and that the identified bacteriophage species represent only their next relatives. A definitive assignment of a host to these sequences is therefore currently not possible. The quality of the database with regard to genomic sequences of bacteriophages is crucial for virome analyses. For example, the high proportion of the detected Lactococcus phage 1706 may reflect the disproportionately high abundance of those phage sequences in the database as a consequence of intensive research on these bacteriophages, which are problematic agents for the diary cheese product industry [50]. In contrast, for another highly abundant bacteriophage, the dragonfly-associated microphage, the specific bacterial host is still unknown [51]. An increase of annotated bacteriophage sequences in the databases is therefore a prerequisite for studies on the interactions between bacteriophages and their host bacteria in the gut in future. Alternatively, a deeper sequencing may enable the assembly of complete bacteriophage genome sequences from the metagenomic data set. By this, an assignment to their hosts was possible by identification of inserted host-related sequences as recently described [3].

In summary the data show a high variability of the pig faecal virome. Most obvious are age-related differences in the proportion between pig viruses and bacteriophages as well as an increasing diversification of virus species by age. Consistent differences due to probiotic feeding could not be identified by our metagenomic analysis. The results of comparative pig virome analyses may help to understand the complex interactions between viruses, bacteria and the pig within the intestinal tract. Future research should focus on the optimization of the method in order to increase its sensitivity and on the improvement of the sequence databases, especially regarding annotated bacteriophage sequences. It will be interesting to apply the optimized techniques to analyse the diversity of faecal viromes in individuals and to identify further factors like geographical origin or disease-related changes influencing its composition.

Supporting Information

Table S1 Relative abundance of virus families in the analyzed faecal viromes. The table shows the number of reads with sequence identities to a certain virus family in relation to all virus reads (in %). Families showing an abundance of less than 1% in a distinct faecal virome are subsumed to other families. Families which were not classified by the International Committee on Taxonomy of Viruses (ICTV) are subsumed to not assigned. The group P received the probiotic bacterium E. faecium NCIMB 10415 (P) and the group C (C) received no probiotic.

(PDF)

Table S2 Relative abundance of bacteriophage species among all bacteriophages detected in the analyzed faecal viromes. The table shows the number of reads with sequence identities to a certain bacteriophage species in relation to all bacteriophage reads (in %). Bacteriophage species showing an abundance of less than 1% in a distinct faecal virome are subsumed (<1%). The group P received the probiotic bacterium E. faecium NCIMB 10415 (P) and the group C (C) received no probiotic.
(PDF)

Table S3 Relative abundance of mammalian virus genera among all animal viruses detected in the analyzed faecal viromes. The table shows the number of reads with sequence identities to a certain mammalian virus genus in relation to all animal virus reads (in %). Mammalian viruses, which are so far not assigned to a certain genus, are indicated in apostrophes. Mammalian virus genera showing an abundance of less than 1% in a distinct faecal virome are subsumed (genera < 1%). Viruses from non-mammalian hosts are subsumed to non mammalian viruses. The group P received the probiotic bacterium E. faecium NCIMB 10415 (P) and the group C (C) received no probiotic.
(PDF)

Acknowledgments

We thank Robert Pieper and his colleagues at the Institute for Animal Nutrition (Free University Berlin, Germany) for the coordination of the animal experiment and Paul Wrede (Charité, Berlin, Germany) for helpful discussion. We also thank Wilfried Vahjen and Ingo Starke (Institute for Animal Nutrition, Free University Berlin, Germany) for their support in detection of *E. faecium* NCIMB 10415 in the samples.

Author Contributions

Conceived and designed the experiments: RJ. Performed the experiments: JS. Analyzed the data: ST MS. Wrote the paper: JS RJ.

References

1. Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, et al. (2011) The human gut virome: Inter-individual variation and dynamic response to diet. Genome Res 21: 1616–1625.
2. Shan T, Li L, Simmonds P, Wang C, Moeser A, et al. (2011) The fecal virome of pigs on a high-density farm. J Virol. 85 (22): 11697–708.
3. Modi SR, Lee HH, Spina CS, Collins JJ (2013) Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome. Nature. Jul 11; 499(7457): 219–22. doi: 10.1038/nature12212.
4. Ogilvie LA, Bowler LD, Caplin J, Dedi C, Diston D, et al. (2013) Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences. Nat Commun. Sep 16; 4: 2420. doi: 10.1038/ncomms3420.
5. Steward GF, Preston CM (2011) Analysis of a viral metagenomic library from 200 m depth in Monterey Bay, California constructed by direct shotgun cloning. Virol J. Jun 9; 8: 287. doi: 10.1186/1743-422X-8-287.
6. Colbère-Garapin F, Martin-Latil S, Blondel B, Mousson L, Pelletier I, et al. (2007) Prevention and treatment of enteric viral infections: possible benefits of probiotic bacteria. Microbes Infect. Nov-Dec; 9(14–15): 1623–31.
7. Zhang T, Breitbart M, Lee WH, Run J-Q, Wei CL, et al. (2006) RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses. PLOS Biology 4 (1): e3: 10.1371.
8. Breitbart M, Hewson I, Felts B, Mahaffy JM, Nulton J, et al. (2003) Metagenomic analyses of an uncultured viral community from human feces. J Bacteriol 185(20): 6220–6223.
9. Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, et al. (2010) Viruses in the faecal microbiota of monozygotic twins and their mothers. Nature 466(7304): 334–338.
10. Sachsenröder J, Twardziok S, Hammerl JA, Janczyk P, Wrede P, et al. (2012) Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing. PLoS One. 2012; 7(4):e34631. doi: 10.1371/journal.pone. 0034631.
11. Lager KM, Ng TF, Bayles DO, Alt DP, Delwart EL, et al. (2012) Diversity of viruses detected by deep sequencing in pigs from a common background. J Vet Diagn Invest. Nov; 24(6): 1177–9. doi: 10.1177/1040638712463212.
12. Park JG, Kim HJ, Matthijnssens J, Alfajaro MM, Kim DS, et al. (2013) Different virulence of porcine and porcine-like bovine rotavirus strains with genetically nearly identical genomes in piglets and calves. Vet Res. 2013 Oct 1; 44(1): 88.
13. Midgley SE, Bányai K, Buesa J, Halaihel N, Hjulsager CK, et al. (2012) Diversity and zoonotic potential of rotaviruses in swine and cattle across Europe. Vet Microbiol. May 4; 156(3–4): 238–45. doi: 10.1016/j.vetmic.2011.10.027.
14. Martella V, Bányai K, Matthijnssens J, Buonavoglia C, Ciarlet M, et al. (2010) Zoonotic aspects of rotaviruses. Vet Microbiol. Jan 27; 140(3–4): 246–55. doi: 10.1016/j.vetmic.2009.08.028.
15. Zhang Q, Hu R, Tang X, Wu C, He Q, et al. (2013) Occurrence and investigation of enteric viral infections in pigs with diarrhea in China. Arch Virol. Aug; 158(8): 1631–6. doi: 10.1007/s00705-013-1659.
16. Di Profio F, Ceci C, Di Felice E, Marsilio F, Di Martino B, et al. (2013) Molecular detection of porcine kobuviruses in Italian swine. Res Vet Sci. Oct; 95(2): 782–5. doi: 10.1016/j.rvsc. 2013.06.020.
17. Verma H, Mor SK, Abdel-Glil MY, Goyal SM. (2013) Identification and molecular characterization of porcine kobuvirus in U. S. swine. Virus Genes. Jun; 46(3): 551–3. doi: 10.1007/s11262-013-0879-1.
18. Cheung AK, Ng TF, Lager KM, Bayles DO, Alt DP, et al. (2013) A divergent clade of circular single-stranded DNA viruses from pig feces. Arch Virol. Oct; 158(10): 2157–62. doi: 10.1007/s00705-013-1701-z.
19. Meng XJ (2012) Emerging and Re-emerging Swine Viruses. Transbound Emerg Dis. Jan 9. doi: 10.1111/j.1865-1682.2011.01291.
20. Moon HJ, Song D, Seon BH, Kim HK, Park SJ, et al. (2012) Complete genome analysis of porcine enterovirus B isolated in Korea. J Virol. Sep; 86(18): 10250. doi: 10.1128/JVI.01548-12.
21. Boros Á, Pankovics P, Reuter G (2011) Characterization of a novel porcine enterovirus in domestic pig in Hungary. Infect Genet Evol. Jul; 11(5): 1096–102. doi: 10.1016/j.meegid.2011.04.003.
22. De Benedictis P, Schultz-Cherry S, Burnham A, Cattoli G (2011) Astrovirus infections in humans and animals - molecular biology, genetic diversity, and interspecies transmissions. Infect Genet Evol. Oct; 11(7): 1529–44. doi: 10.1016/j.meegid.2011.07.024.
23. Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, et al. (2013) Rapid evolution of the human gut virome. Proc Natl Acad Sci U S A. Jul 23; 110 (30): 12450–5. doi: 10.1073/pnas.1300833110.
24. Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, et al (2008) Viral diversity and dynamics in an infant gut. Arch Virol. 2013 Apr; 158(4): 799–807. doi: 10.1007/s00705-012-1543-0.
25. Rowland I, Capurso L, Collins K, Cummings J, Delzenne N, et al. (2010) Current level of consensus on probiotic science-report of an expert meeting-London. Gut Microbes. Nov-Dec; 1(6): 436–9. doi: 10.4161/gmic.1.6.13610.
26. Szajewska H, Wanke M, Patro B (2011) Meta-analysis: the effects of *Lactobacillus rhamnosus* GG supplementation for the prevention of healthcare-associated diarrhoea in children. Aliment Pharmacol Ther. Nov; 34(9): 1079–87. doi: 10.1111/j.1365-2036.2011.04837.
27. Guandalini S (2011) Probiotics for prevention and treatment of diarrhea. J Clin Gastroenterol. Nov; 45 Suppl:S149–53. doi: 10.1097/MCG.0b013e3182257e98.
28. Erdoğan O, Tanyeri B, Torun E, Gönüllü E, Arslan H, et al. (2012) The comparition of the efficacy of two different probiotics in rotavirus gastroenteritis in children. J Trop Med. 2012: 787240. doi: 10.1155/2012/787240.
29. Fang SB, Lee HC, Hu JJ, Hou SY, Liu HL, et al. (2009) Dose-dependent effect of *Lactobacillus rhamnosus* on quantitative reduction of faecal rotavirus shedding in children. J Trop Pediatr. Oct; 55(5): 297–301. doi: 10.1093/tropej/fmp001.
30. Kreuzer S, Machnowska P, Aßmus J, Sieber M, Pieper R, et al. (2012) Feeding of the probiotic bacterium *Enterococcus faecium* NCIMB 10415 differentially affects shedding of enteric viruses in pigs. Vet Res. 2012 Jul 27; 43(1): 58. doi: 10.1186/1297-9716-43-58.
31. Siebert B, Reinhardt N, Kreuzer S, Bondzio A, Twardziok S, et al. (2013) *Enterococcus faecium* NCIMB 10415 supplementation affects intestinal immune-associated gene expression in post-weaning piglets. Manuscript Number: accepted but unpublished so far.VETIMM-D-13-742R3.
32. Scharek L, Guth J, Filter M, Schmidt MF (2007) Impact of the probiotic bacteria *Enterococcus faecium* NCIMB 10415 (SF68) *and Bacillus cereus* var. toyoi NCIMB 40112 on the development of serum IgG and faecal IgA of sows and their piglets. Arch Anim Nutr, 61: 223–234.
33. Scharek L, Guth J, Reiter K, Weyrauch KD, Taras D, et al. (2005) Influence of a probiotic *Enterococcus faecium* strain on development of the immune system of sows and piglets. Vet Immunol Immunopathol 105: 151–161.
34. Zhang J, Deng J, Wang Z, Che C, Li YF, et al. (2011) Modulatory effects of *Lactobacillus salivarius* on intestinal mucosal immunity of piglets. Curr Microbiol, 62: 1623–1631.
35. Chai W, Burwinkel M, Wang Z, Palissa C, Esch B, et al. (2013) Antiviral effects of a probiotic *Enterococcus faecium* strain against transmissible gastroenteritis coronavirus. Arch Virol. Apr; 158(4): 799–807. doi: 10.1007/s00705-012-1543-0.
36. Martin L, Pieper R, Kröger S, Goodarzi Boroojeni F, Vahjen W, et al. (2012) Influence of age and *Enterococcus faecium* NCIMB 10415 on development of small

intestinal digestive physiology in piglets. Anim Feed Sci Technol 2012, 175: 65–75.
37. Starke I, Pieper R, Neumann K, Zentek J, Vahjen W (2013) Individual response of mother sows to a probiotic *E. faecium* strain lead to different microbiota composition in their offspring accepted but unpublished so far.
38. Vahjen W, Taras D, Simon O (2007) Effect of the probiotic Enterococcus faecium NCIMB10415 on cell numbers of total Enterococcus spp., *E. faecium* and *E. faecalis* in the intestine of piglets. Curr Issues Intest Microbiol. Mar; 8(1): 1–7.
39. Thurber RV, Haynes M, Breitbart M, Wegley L, Rohwer F (2009) Laboratory procedures to generate viral metagenomes. Nature Prot 4(4): 470–483.
40. Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics. 2010 Jun; 95(6): 315–27. doi: 10.1016/j.ygeno.2010.03.001.
41. Altschuld SF, Gish W, Miller W, Myers E, Lipman DJ, et al. (1990) Basic Local Alignment Search Tool. J. Mol. Biol. 215: 403–410.
42. GenBank, NCBI, http://www.ncbi.nlm.nih.gov/genbank/(Accessed 2014 Jan 21).
43. CLC MainWorkbench 6.2; http://www.clcbio.com/index.php?id = 532 (Accessed 2014 Jan 21).
44. RefSeq, NCBI, ftp://ftp.ncbi.nih.gov/refseq/release/viral/01.08.2012.
45. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods. Mar 4; 9(4): 357–9. doi: 10.1038/nmeth.1923.
46. Shannon CE (1997) A mathematical theory of communication. MD Comput. 14(4): 306–317.
47. Bray JR, Curtis JT (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs. 27: 325–349. http://dx.doi.org/10.2307/1942268.
48. Baker KS, Leggett RM, Bexfield NH, Alston M, Daly G, et al. (2013) Metagenomic study of the viruses of African straw-coloured fruit bats: detection of a chiropteran poxvirus and isolation of a novel adenovirus. Virology. Jul 5; 441(2): 95–106. doi: 10.1016/j.virol.2013.03.014.
49. Reuter G, Boldizsár A, Kiss I, Pankovics P (2008) Candidate new species of Kobuvirus in porcine hosts. Emerg Infect Dis. Dec; 14(12): 1968–70. doi: 10.3201/eid1412.080797.
50. Garneau JE, Tremblay DM, Moineau S (2008) Characterization of 1706, a virulent phage from *Lactococcus lactis* with similarities to prophages from other Firmicutes. Virology. Apr 10; 373(2): 298–309. doi: 10.1016/j.virol.2007.12.002.
51. Rosario K, Dayaram A, Marinov M, Ware J, Kraberger S, et al. (2012) Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta). J Gen Virol. Dec; 93(Pt 12): 2668–81. doi: 10.1099/vir.0.045948-0.

3.3 Third paper: Metagenomic identification of novel enteric viruses in urban wild rats and genome characterization of a group A rotavirus

Jana Sachsenröder, Anne Braun, Patrycja Machnowska, Terry Fei Fan Ng, Xutao Deng, Sebastian Guenther, Samuel Bernstein, Rainer G. Ulrich, Eric Delwart, Reimar Johne

J Gen Virol. 2014 Dec;95(Pt 12):2734-47.

Summary of paper 3

Norway rats are well-known reservoirs of several zoonotic pathogens. So far, the composition of the virus community in faeces of urban wild rats has been only scarcely investigated and has been focused on human pathogens using virus specific RT-PCR amplification. Here the whole faecal viromes of 20 urban wild rats derived from Berlin, Germany were determined and selected virus sequences were characterised in more detail.

Virus particles were enriched from the faecal samples of the rats using filtrations and centrifugations. Thereafter, the total nucleic acid was purified, amplified and subjected to Next Generation Sequencing using the Illumina technology. Bioinformatic analysis methods were applied for identification and comparison of virus reads. Thereafter, (RT-)PCR assays combined with Sanger Sequencing were developed for further determination of virus distribution and characterization of selected viruses.

The analysis of the generated reads indicated that the faecal viromes of the rats were highly diverse and consisted of several known and unknown mammalian viruses (phage sequences were excluded from the analysis). The most abundant species belonged to the genera *Parvo*- and *Picornavirus*. Due to low sequence similarities to known representatives of described virus taxa, it was concluded that several novel rat viruses belonging to the genera *Picorna*-, *Bocaparvo*-, *Sapovirus* and “stool-associated circular ssDNA viruses” were identified.

Among others, several viruses potentially causing gastroenteritis, e.g. noro- and rotaviruses were identified. Reads of a norovirus and a rotavirus B could be identified as closely related to a recently detected rat norovirus and rat rotavirus strain IDIR, respectively. A zoonotic potential of these viruses was excluded because they clustered separately form respective human virus strains in phylogenetic trees. Nearly complete Sanger sequencing of the genome of a rotavirus A revealed a close relationship of some of the genome segments to human and animal rotavirus strains. The strain consists of three known genotypes (G3, P[3] and N2) and eight novel genotypes (I20-R11-C11-M10-A22-T14-E18-H13) and represents the first rotavirus A detected in rats.

Based on the results it could be concluded that the faecal rat viromes are highly variable and contain heterogenic enteric virus communities. Some of these viruses are new and others may pose a risk of zoonotic transmission to humans and animals.

Key messages of paper 3

- Faeces of wild urban rats harbour a large variety of known and unknown mammalian viruses.
- The rat faecal viromes are highly variable between the animals; however, parvo- and picornaviruses are most abundant.
- Novel rat viruses belonging to the genera *Rotavirus A, Picorna-, Bocaparvo-, Sapoviruses* and “stool-associated circular ssDNA viruses” were identified.
- A rotavirus A detected in some of the animals is closely related to other human and animal rotaviruses and may therefore pose a risk of zoonotic transmission.

Own contribution to paper 3

In this study, I prepared in collaboration with Terry F. F. Ng (Blood Systems Research Institute, BSRI; San Francisco, USA) all faecal samples using a novel purification method developed for small-sized samples. This included the purification of the virus-like particles by centrifugation and filtration, the extraction of the nucleic acid, preparation of the Illumina library for sequencing as well as the whole Illumina sequencing procedure. Subsequently I analysed the generated reads using a bioinformatics pipeline originally developed by Xutao Deng (BSRI). The (RT)-PCRs and Sanger sequencings of selected viruses were mainly done by Anne Braun and Reimar Johne (both BfR, Germany). I participated in the subsequent sequence analyses. Additionally, I wrote the respective parts of the manuscript and was engaged in critical reading and revision of the whole manuscript.

Metagenomic identification of novel enteric viruses in urban wild rats and genome characterization of a group A rotavirus

Jana Sachsenröder [1], Anne Braun [1], Patrycja Machnowska [1], Terry Fei Fan Ng [2], Xutao Deng [2], Sebastian Guenther [3], Samuel Bernstein [4], Rainer G. Ulrich [4], Eric Delwart [2], Reimar Johne [1*]

[1] Federal Institute for Risk Assessment, Max-Dohrn-Straße 8-10, 10589 Berlin, Germany

[2] Blood Systems Research Institute, 270 Masonic Ave., San Francisco, 94118 San Francisco, CA, USA

[3] Centre for Infection Medicine, Institute of Microbiology and Epizootics, Freie Universität Berlin, Germany Robert von Ostertagstrasse 7-13, 14163, Berlin, Germany

[4] Friedrich-Loeffler-Institut, Institute for Novel and Emerging Infectious Diseases, Südufer 10, 17493 Greifswald-Insel Riems, Germany

Journal's content category: Animal Viruses – Double-strand RNA

The GenBank/EMBL/DDBJ accession numbers are KJ879448- KJ879458 (rat rotavirus A), KM030273 (rat rotavirus B), KM030274 - KM030275 (rat picornavirus), KM030276 (rat bocavirus), KM030277 - KM030278 (rat SCV) and KM030279 (rat sapovirus).

*Corresponding author.
E-mail address: Reimar.Johne@bfr.bund.de (R. Johne).

Paper it to find under:

https://doi.org/10.1099/vir.0.070029-0

Abstract

Rats are known as reservoirs and vectors for several zoonotic pathogens. However, information on the viruses shed by urban wild rats that could pose zoonotic risk to human health is scare. Here, intestinal contents from 20 wild Norway rats (*Rattus norvegicus*) collected in the city of Berlin, Germany, were subjected to metagenomic analysis of viral nucleic acids. The determined faecal viromes of rats consisted of a variety of known and unknown viruses and were highly variable among the individuals. Members of the families *Parvoviridae* and *Picobirnaviridae* represented the most abundant species. Novel picorna-, boca-, sapo- and stool-associated circular ssDNA viruses (SCV) were identified, which showed only low sequence identities to known representatives of the corresponding taxa. In addition, noro- and rotaviruses were detected as potential zoonotic gastroenteritis viruses. However, partial genome sequence analyses indicated that the norovirus was closely related to the recently identified rat norovirus and the rotavirus B was closely related to the rat rotavirus strain IDIR; both viruses clustered separately from respective human virus strains in phylogenetic trees. In contrast, the rotavirus A sequences showed high identities to human and animal strains. Analysis of the nearly complete genome of this virus revealed the known genotypes G3, P[3] and N2 for three of the genome segments, whereas the remaining eight genome segments represented the novel genotypes I20-R11-C11-M10-A22-T14-E18-H13. In conclusion, the results indicate a high heterogeneity of enteric viruses present in urban wild rats; their ability to be transmitted to humans remains to be assessed in future.

Paper it to find under:

https://doi.org/10.1099/vir.0.070029-0

4 GENERAL DISCUSSION

4.1 Method for faecal virome analysis

4.1.1 Background

The study of animal viromes could help to identify zoonotic viruses, which may be transmitted to humans. Especially, viruses present in faeces of urban rats or farmed pigs may pose a zoonotic risk as close contacts to humans can be suspected (Christou, 2011; Firth et al., 2014). Additionally, viruses may also play significant roles as animal pathogens. Especially in farm animals like pigs, infectious enteric diseases can cause high economic losses due to deaths of severely affected piglets or reduced animal growth.

At the beginning of this study, no virome data were available from urban wild rats and farm animals with importance for food production. The only determined mammalian faecal viromes were derived from horses (Cann et al., 2007), humans (Breitbart et al., 2008; Reyes et al., 2010; Zhang et al., 2006) and bats (Donaldson et al., 2010; Li et al., 2010a). Until now, the list of investigated animal viromes is long including several porcine viromes (Belak et al., 2013a; Lager et al., 2012; Shan et al., 2011; Zhang et al., 2014a), and those of rats (Firth et al., 2014) cats (Ng et al., 2014; Zhang et al., 2014b), dromedary (Woo et al., 2014), ferrets (Smith et al., 2013), pigeon (Phan et al., 2013a) and shrew (Sasaki et al., 2015) (see table 2, introduction).

The analysis of viromes is difficult, because classical methods for virus detection like virus cultivation often fail to detect the broad range of virus types present in a sample. Only a small percentage of existing cell types and bacteria can be cultivated properly in order to serve as host organisms for the isolation and propagation of viruses or bacteriophages. Therefore, most of the existing viruses can be assumed not to be cultivable by standard laboratory procedures so far. Consequently, the existing genomic data of viruses can be suspected to be highly biased by the predominant identification of cultivable viruses, which does not represent a true picture of the genomic diversity of viruses (Wooley et al., 2010). Another limitation to analyse the entire virus diversity was due to technological

restrictions. Until the 2000s it was only possible to determine the genomes of some single viruses because the sequencing procedure was quite time-consuming and expensive.

With the development of the NGS technologies in the 2000s, a new area of sequencing has started. These NGS technologies allow researchers to simultaneously identify and characterize many virus genomes in a very short time and with comparably low costs. Furthermore, using NGS makes it possible to determine the genomes of whole virus communities taken directly from the environment. Genomes can be analysed without any prior cultivation step and sequenced without prior knowledge of the genome sequences.

However, due to the lack of standardized methods for NGS-based virome analysis, it was necessary to establish a reliable laboratory protocol. The application of this protocol should ideally result in an unbiased amplification and identification of all viral nucleic acids present in a sample. In order to optimize the method and to control the performance in each application, a process control system was introduced into the protocol.

4.1.2 Establishment of a process control

Viral metagenomic studies could be separated into three main parts: i) the purification of the virus-like particles and preparation/amplification of contained nucleic acids, ii) the sequencing process and iii) the data analyses (figure 5).

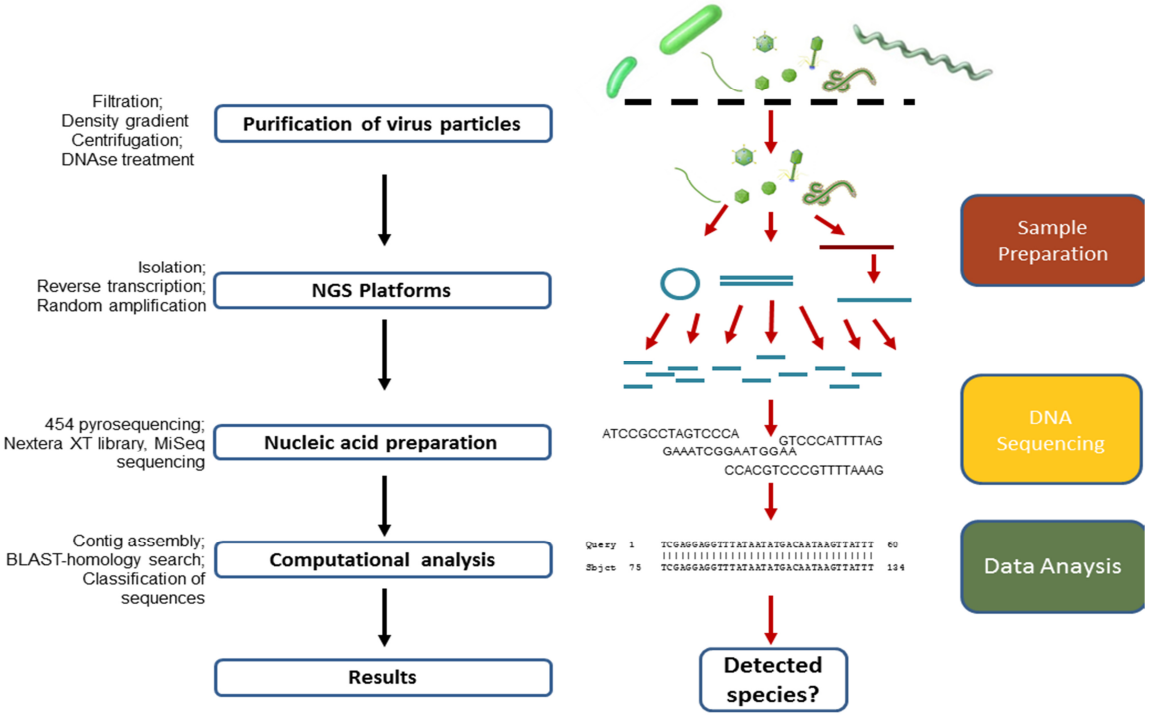


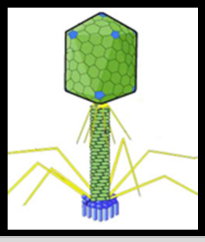
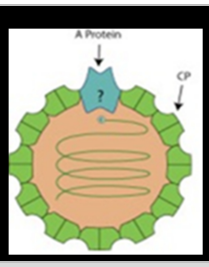
Figure 5: Schematic overview of the steps for analysis of the total fraction of viruses using NGS sequencing

Especially the first step of purification of the virus-like particles and amplification of contained nucleic acids is complex and the applied protocols are not standardized so far. In addition, all NGS technologies are still limited by barely reproducible steps during the whole procedure (Daber et al., 2013). Therefore, optimization and validation of a protocol is difficult, especially if samples are used for the development of the protocol, which contain an unknown composition of viruses, like stool samples.

These considerations led to the concept of an establishment of a process control for monitoring the efficiency of each step of the method as well as of the whole metagenomic analysis. The concept of process control was originally adapted from other analytical applications like quantitative reverse transcription PCR (RT-qPCR), which also uses similar controls to monitor the method performance (Dreier et al., 2005; Parshionikar et al., 2004; Rolfe et al., 2007; Stevenson et al., 2008). This process control for viral metagenome analyses should be composed of a defined amount and defined composition of known viruses, which are added at the start and monitored until the end of the analysis. One of the challenges was to

find a process control system, which reflects the high variability of viruses and is easy and safe to use. The process control developed by us consists of three different bacteriophages, which were added at the beginning to the faecal samples. For analysis of the abundance of the bacteriophages in each analytical step, specific RT-qPCRs were established (paper 1, table 2). The selected bacteriophages are T4, M13 and MS2, which differ remarkably regarding genome type as well as size and particle shape (see table 7), thus reflecting a large part of the viral diversity.

Table 7: Properties of the bacteriophages T4, M13 and MS2 used as process control

	T4 	M13 	MS2 
Family	*Myoviridae*	*Inoviridae*	*Leviviridae*
Particle shape	icosahedral head + contractile tail	filamentous	icosahedral
Particle size	Ø 100 nm + 100 nm tail	1000 nm long	Ø 30 nm
Genome	dsDNA	ssDNA	ssRNA
Genome size [kb]	169	6.5 (circular)	3,6
Density in CsCl [g/ml]	1,5	1,4	1,4

ssDNA – single-stranded DNA; dsDNA – double-stranded DNA; ssRNA – single-stranded
[1)]Pictures: with permission of ViralZone, Swiss Institute of Bioinformatics, www.expasy.org/viralzone

However, some morphologies like enveloped spherical viruses as well as some genome types like double-stranded RNA are not represented by them. The observation, that enveloped viruses are often underrepresented in faecal virome studies, may indicate that they should be included in the process control in future studies.

During the experiments, the developed process control turned out to be very valuable for the optimization of the metagenomic analysis method (paper 1). Furthermore, the process control was used later to monitor the method performance for each sample (paper 2, table 2). This quality control seems to be mostly necessary for comparative analyses of different samples.

4.1.3 Sample preparation

The detection of viruses in complex matrices using NGS is complicated due to several reasons. The first reason is the different genome sizes of organism. The average genome size of a mammalian host cell is 3.5×10^6 kb (Gregory et al., 2007), that of bacteria ranges from 0.16 – 13 Mb (Kuo et al., 2009; Nakabachi et al., 2006; Schneiker et al., 2007), but that of viruses ranges between only 4 – 1.2 Mb (Black and Thomas, 2012). With respect to this relationship, one bacterial genome accounts for approximately 500 virus genomes and one genome of a mammalian cell accounts for approximately 300,000 virus genomes if analysed by sequencing. Inefficient purification of virus particles leads to an abundance of genetic material of bacteria and host cells due to the small genome size of viruses in relation of the large genome sizes of bacteria and host cells. This may subsequently lead to their lack of detection within complex samples. The second reason for complication is that virus genomes are very diverse and there are no sequences conserved among all viruses. For analysis of bacteria or animal cells, the ribosomal RNA genes can be amplified and subsequently sequenced as they are present in all of these organisms (Mokili et al., 2012; Reyes et al., 2012; Weinstock, 2012). As viruses do not have such genes, a selective amplification of genomic material from all viruses is not possible.

For these reasons, the performance of an efficient purification method for virus particles from the samples is necessary as a first step of analysis. Otherwise, viruses will only be detectable by NGS if present in very large quantities. Thus, a protocol was developed to enrich the virus particles, which was thereafter successfully validated using the process control phages. Several purification systems for virus particles have been published; these systems are mainly based on their physical properties and include combinations of filtrations and (ultra)-centrifugation steps (Breitbart et al., 2003; Casas and Rohwer, 2007; Mokili et al., 2012; Thurber et al., 2009). In the first protocol (paper 1 and paper 2), the main steps included tangential flow filtration (TFF) and caesium-chloride density gradient ultracentrifugation. This enabled the use of large amounts (100 g) of faecal samples resulting in high amounts of nucleic acid for subsequent analysis.

The read number of the process control phages ranged between 0.5% (study 1, data not shown) and 3.4% of all reads (paper 2, table 2). However, TFF requires a minimum quantity of filtration material. For small sample sizes an alternative protocol based on syringe filtration for virus particle purification was used as this was thought to be more appropriate for the smaller sample sizes (0.5-1.0 g) of the rat faeces (paper 3).

The next step was the amplification of nucleic acids contained in the particles. Viruses show a large variety of genome types (DNA or RNA in different topologies) and genome size (see chapter 1.2.2, introduction), which complicates the isolation and purification of the majority of viruses by the use of only one universal protocol. This is also reflected in the large number of different published protocols to realize metagenomic virus discoveries reviewed by Hall et al. (2014). Most of them analyse DNA viruses separately from RNA viruses. However, we developed a protocol, which enables the simultaneous analysis of all virus genome types by a reverse transcription followed by a random amplification step. It could be shown that all types of viral genomes and all process control phages could be identified in pig faecal samples by this method (paper 1, table 4; paper 2, table 2).

Direct comparisons of our method with other published methods are difficult because no recovery rates for selected viruses (e.g. our process control phages) have been published so far. However, the average percentage of virus reads obtained with our TFF protocols ranged from 6.7%; (paper 1, table 4) and 17.9% (paper 2, table 2) and are comparable to published percentages (Minot et al., 2011; Shan et al., 2011; Zhang et al., 2014b). Only some insect virome studies showed a higher proportion (81%) of virus reads (Sasaki et al., 2015).

Generally, we aimed to use a large amount of sample material for the analysis to increase the chance to detect only minor abundant virus species. Additional DNA amplification steps should be kept to a minimum as they are also known to change the relative proportions of genomes, especially in the case of circular DNAs (Delwart, 2007; Johne et al., 2009). For the experiments with adult pigs, 100 g faeces were used as the starting material. In contrast to this, the majority of viral metagenome analyses was performed using only up to 5 g faeces as the starting material and involved additional DNA amplification steps to obtain a sufficient

amount of DNA for NGS analysis (Baker et al., 2013; Finkbeiner et al., 2008; Li et al., 2011a and 2011b; Masembe et al., 2012; Ng et al., 2011; Phan et al., 2011; Shan et al., 2011; Sikorski et al., 2013a; Smits et al., 2013; Victoria et al., 2009; Zhang et al., 2014a). Only two groups also purified virus particles from a larger amount (500 g) of faeces (Breitbart et al., 2003; Zhang et al., 2006). As only small amounts of faecal samples were available from the rats, we could only use 0.5 to 1.0 g faeces for our rat virome analysis (paper 3). Systematic investigations would be necessary to clarify the potential differences regarding the detection of viruses depending on the amount of starting material.

4.1.4 Sequencing

Several technical platforms for NGS are available. The pig viromes were sequenced using the 454 pyrosequenicng platform (Roche) by an external company. The rat virome was sequenced in-house using the MiSeq platform (Illumina). In both cases, the protocols from the manufacturers were followed and no additional optimization was necessary. The 454 pyrosequencing method is supposed to produce longer read lengths up to 400 nt, whereas the MiSeq method generates more but shorter reads (up to 250 nt), which may complicate the sequence identification via BLAST search (Liu et al., 2012; Loman et al., 2012; Mardis, 2011). In our experiments, both 454 pyrosequencing and Illumina reads achieved average read lengths of 250 bp to 300 bp. A disadvantage of 454 pyrosequencing is the high error rate for monobasic stretches. Additionally, at the time of our study, 454 pyrosequencing was more expensive with $10 per million bases compared to Illumina with $0.07 per million bases. We had chosen the 454 pyrosequencing for the pig viromes because of the expected longer read length. In the case of the rat viromes, however, it was the aim to sequence 20 samples, which was done using the MiSeq platform (due to economic reasons).

4.1.5 Data analysis

The analysis of NGS data is the most time-consuming and demanding step in NGS projects. To analyse the virus sequences, bioinformatic filtering pipelines are necessary, which process the data step by step until an assignment of the reads to taxonomic groups is possible. In the first steps, the read quality is assessed and sequences are trimmed by removal of primer sequences. These steps or data analysis can be done relatively easily with available computer programs developed for the NGS systems (Bzhalava and Dillner, 2013). A schematic flowchart of metagenomic data analysis is shown in the introduction, figure 4.

The next step is to assign the generated sequences to virus genomes, which is more complicated and remains unstandardized. The most important tool for identification of viral reads is a comparison of the generated reads with annotated sequences available in databases using BLAST algorithms (Altschul et al., 1990). The result of such a comparison is a list of the annotated sequences most similar to the read. The degree of sequence similarity is given with the BLAST E-value. The most common used cut-offs of the E-value in NGS studies range from an E-value ≤ 10^{-5} (Delwart, 2007; Hall et al., 2014) to an E-value ≤ 10^{-3} (Cann et al., 2007; McDaniel et al., 2013; Zhang et al., 2006) and in very rare cases where a very new viral flora was expected to an E-value ≤ 10^{-2} (Desnues et al., 2008). In general, the use of an insufficiently low cut-off leads to exclusive identification of known viruses with no possibility to detect new viruses. However, if the cut-off is too high, the assignment of a sequence to a specific virus is generally questionable as any (random) similarity is displayed. Moreover, particular E-values are not directly comparable due to the fact that they depend on the query sequence, the used database and the parameters which were used for the BLAST-algorithm (Mokili et al., 2012). As the genomes of viruses contain both conserved and highly variable regions, different cut-offs may be applied for the different regions. However, the automatic pipeline of viral metagenomic approaches requires the use of one cut-off for all generated sequences. We had chosen an E-value of the pig viromes of ≤ 10^{-5} for analysis as a sensitive read identification and a reduction of unspecific background noise was necessary to

compare the viromes. For the rat virome, we had chosen an E-value of ≤ 10^{-2} because we intended the detection of novel virus variants.

An additional restriction of virus identification is the availability of only very limited numbers of annotated viral genomes within the sequence databases. The identification of bacteriophages is particularly difficult as phage genomes are clearly underrepresented within sequence databases (Desnues et al., 2008; Edwards and Rohwer, 2005; Paul and Sullivan, 2005; Reyes et al., 2012). If the target virus genome is not present in the database, an identification of the read could be problematic. Minot et al. (2012) examined the human virome of 12 human volunteers to investigate the variability of viral genomes. They could demonstrate that DNA viruses of the human gut are rich in hypervariable regions and that up to 96% of NGS-generated reads from this bacteriophage were assigned as unique amino acid sequences using BLAST.

The third step is to determine the prevalence of species, genera or families based on the identified virus reads. Different approaches can be applied: i) comparison of the absolute numbers of primary reads; ii) comparison of single reads in relation to the virus genome length, or; iii) comparison of the number of assembled contigs. The first option is often used, but it does not take into account the different lengths of the viral genomes, which are remarkably variable between virus families (Carstens, 2012). However, longer genomes are represented by more reads per genome than smaller ones. Therefore, we used the second option with a correction factor for the genome length, which resulted in a remarkable shift of the abundance of virus genera and families (paper 1, table 5, 6). Nevertheless considering only single reads could lead to an unambiguous identification of viruses. One solution to overcome this problem is to first assemble the reads into longer contigs (option 3). Although the use of contigs lowers the risk of false BLAST results, assemblies of large datasets are quite complicated and contig assembly itself is strongly dependent on the used algorithm (Lin et al., 2011; Pignatelli and Moya, 2011). Those problems with contig assembly from NGS data are widely known and subject to intensive optimization (Cahais et al., 2012; Ji et al., 2011). This is also reflected by the vast number of assembly algorithms, which were developed in the last years. Between 2008 and 2010 alone more than 20 assembly softwares have been published; dozens of more are still waiting for

publishing (Li and Homer, 2010b). However, Kleffe and Hansen could demonstrate using several assembly softwares that even the initial order of reads has an influence on the assembly result (personal communication, 2013).

In our study, the overall distribution of virus genera and virus families was more similar when using primary reads or contigs for analysis than when considering genome length (paper 1, table 5, 6). Using contigs, a remarkable reduction of detected virus diversity was evident. It is likely that the sequences of less represented viruses could not be assembled into contigs, leading to a loss of such viruses. Indeed, by using the primary reads we could identify 466 different virus species, compared to 121 viruses identified by the use of contigs (paper 1, table 4).

Our results indicate that NGS data within one study cannot be easily compared with other studies if other analysis algorithms have been used. It may be advisable to generally use primary reads for metagenomic data analysis like we did for all following analyses in paper 2 and 3.

4.2 The porcine faecal virome

4.2.1 General composition

Two of our metagenomic studies investigated the composition of the porcine faecal virome. In both studies, significant diversity in the identified viruses could be determined. In the first study, the detected viruses assigned to 20 virus families (paper 1, table 5) and in the second study to 36 virus families (paper 2). In general, the porcine faecal viromes mainly consisted of bacteriophages and porcine viruses. Plant viruses and viruses with other hosts were only rarely detected (0.2% to 3.4%). This indicates that transient viruses such as food-derived plant viruses play only a minor role in the porcine faecal virome. This is in contrast to human faecal viromes, which have been shown to contain high quantities of plant viruses (Zhang et al., 2006). As the highest abundant RNA virus detected in the mentioned study was derived from pepper (Zhang et al., 2006), differences in the diet (e.g. regarding the use of spices) between humans and pigs may be responsible for the observed differences.

Further analyses indicated a high variability in the relative abundance of virus species in the pig faeces among the individual pigs analysed. A marked dependence of the porcine faecal virome composition on the age of the pigs was evident, which is discussed in more detail in paragraph 4.2.4. Nevertheless, a predominance of bacteriophages was evident in most of the analysed samples. Considering the importance of bacteriophages for regulation of bacterial growth and horizontal gene transfer (Clokie et al., 2011; McDaniel et al., 2013), the distinct composition and function of the bacteriophages present in the porcine intestine should be of special interest. However, none of the studies published so far on the porcine faecal virome included bacteriophages in their analysis (Belak et al., 2013a; Lager et al., 2012; Shan et al., 2011; Zhang et al., 2014a). Thus, a comparison of our data with that of published studies is not possible. A more detailed discussion of the bacteriophages in the porcine faecal virome is presented in paragraph 4.2.3.

There are four other published studies describing the analysis of porcine viruses in faecal samples (Belak et al., 2013a; Lager et al., 2012; Shan et al., 2011; Zhang et al., 2014a). All of these studies used samples from piglets and data on adult pigs were not included. Although these studies and our investigations analysed samples from piglets with different ages (7 days to 42 days old) and different geographical backgrounds (pigs from the USA, China, Sweden and Germany), it is a bit surprising that the general composition of the porcine viruses community was rather similar. All of the studies identified RNA viruses as the most abundant group of the mammalian viruses and kobuvirus, enterovirus, sapovirus, teschovirus, astrovirus and bocavirus could be detected in most of the pigs. However, some differences are evident, such as the detection of coronaviruses by Shan et al. (2011), Lager et al. (2012) and Zhang et al. (2014a), but not by Belak et al. (2013a) or in our studies. A comparison of all so far existing porcine faecal virome studies are given in table 8. Viruses, which are associated with causing diarrhoea, are highlighted in red. As coronaviruses are enveloped viruses they show different physicochemical behaviour. Differences in the sample preparation method may explain these differences (Delwart, 2007). Additionally, we detected a higher number of different virus families compared to the other studies, which may be explained by the use of a larger amount of starting material.

Table 8: Comparison of four porcine faecal virome studies Viruses which are associated with causing diarrhoea are highlight in **red**. By ICTV not accepted virus genera are highlight in " ".

.	Our study	Shan et a., 2011	Lager et al., 2012	Zahng et al., 2014	Belag et a., 2013a
sequencing method	454 pyro-sequencing, FLX Titanimum	454 pyro-sequencing, FLX Titanimum	455 pyro-sequencing, FLX Titanimum	Illumina, HiSeq 2000	454 pyro-sequencing FLX Titanimum
pooled or individual samples	pool of 6 animals	individual	individual	individual	pool of 21 animals
age	piglets: 12 and 54 days and sows	piglets: 19 - 30 days	piglets: 3 weeks	piglets: 20 - 30 days	piglets: 1 - 2 weeks
number animals	48	12 diarrhoea 24 healthy	2 healthy 2 diarrhoea	27 diarrhoea 29 healthy*	21 diarrhoea
	Kobuvirus	**diarrhoea pigs:**		Coronavirus PEDV	Porcine astrovirus
	Enterovirus	Coronavirus		Sapovirus	Kobuvirus
	"Pasivirus"	Kobuvirus		Porcine bocavirus	Calicivirus
	Sapelovirus	Enterovirus		Sapelovirus,	Rotavirus A
	"PigSCV"	Sapovirus		Torovirus	
	Circovirus and CV-like	Sapelovirus		Coronavirus	
	Bocaparvo-virus	Teschovirus		Porcine bocavirus	
	Mastadeno-virus/ Dependovirus	Astrovirus		"PoSCV"	
	Astrovirus	Bocavirus		Astrovirus	
	Sapovirus	CV-like		Kobuvirus	
	Picobirnavirus	**healthy pigs:**		Posavirus	
	Rotavirus	Kobuvirus		Porcine entervirus	
	Hungarovirus	Coronavirus		Po-circo-like virus	
		Enterovirus		Picobirnavirus	
		Sapovirus		TTSuV-2	
		Sapelovirus			
		Teschovirus			
		Astrovirus			
		Bocavirus			
		CV-like			

CV-like circovirus-like; **PEDV** porcine epidemic diarrhea virus; **PoSCV** porcine stool-associated circular ssDNA viruses; **TTSuV** torque teno sus virus
* Zhang et al., 2014: 29 healthy pigs were analysed by PCR

4.2.2 Mammalian viruses and pathogenic and zoonotic potential

The main mammalian viruses detected during the porcine virome analyses in this thesis (paper 1 and 2) are summarized in table 9 together with their associated disease or symptoms in pigs and/ or humans as far as known. In the different samples, mammalian viruses belonging to between one and ten different by the ICTV accepted genera were detected. Only PigSCV (see chapter 4.4.1, discussion) was found in all samples, albeit in different amounts. Bocaviruses, circoviruses and “circovirus-like” viruses could be found in most of samples. All other genera were present only in some of the individual samples.

Kobuvirus was found in most of the piglets. It is one of the newly emerging porcine viruses and was first detected in 2007 (Reuter et al., 2008, Reuter et al., 2009). Nowadays, kobuvirus seems to be ubiquitously distributed in the swine population worldwide (this study: paper 1 and 2; Belak et al., 2013a; Lager et al., 2012; Reuter et al., 2011; Shan et al., 2011; Zhang et al., 2014a). Biology, epidemiology and pathogenicity is only poorly understood and the involvement in porcine disease is unclear (Meng, 2012, Shan et al., 2011; Zhang et al., 2014a).

Astrovirus and kobuvirus were detected in our studies, partly with high titre (see paper 2; supplementary table 3), in animals without any symptoms of diarrhoea. Infections with astroviruses have also been previously described in healthy pigs. Zhang et al. (2014) found no difference in the prevalence of astroviruses in healthy pigs (21 of 29 pigs) compared to pigs with diarrhoea (20 of 27 pigs). However, it is still a matter of discussion whether astrovirus can cause gastroenteritis in pigs (Kreuzer et al., 2012; Luo et al., 2011; Mor et al., 2012; Xiao et al., 2013).

Table 9: Summary of mammalian virus found in porcine faecal viromes of paper 1 and paper 2 and their association with diseases

Genus	Family	Highest identity to:	Pooled sample first paper n = 1	Sows sample n = 4	Piglets sample n = 4	Enteic virus	Diseases in human	Diseases in pigs	Hosts	Zoonotic potential	References
Kobuvirus	*Picorna-viridae*	Porcine kobuvirus	x	0	2	x	astroenteritis	asymptomatic/ unlear: diarrhoea	pig, human, cattle, sheep	unclear	Reuter et ., 2011
Enterovirus	*Picorna-viridae*	Porcine enterovirus B	x	0	2	x	gastrointestinal and respiratory infection/ meningitis/ encephalitis	asymptomatic/ enteric diseases/ pneumonia/ polio-encephalomyelitis	human, mammals	unclear	Kaku et al., 2001; Tee et al., 2009
Sapelovirus	*Picorna-viridae*	Porcine sapelovirus 1		0	2	x	/	diarrhoea, pneumonia, polio encephalomyelitis	pig, simian, birds	no	Chen et al., 2012
"PigSCV"	not assigned		x	4	3	unclear	unclear	unclear	unclear	unlcear	Ng et a., 2015b
Circovirus + Circovirus-like	*Circo-viridae*	Porcine circovirus; Circovirus-like		4	0	unclear	/	postw eaning multisystemic w asting syndrome	pig, birds	no	Allan and Ellis, 2000
Bocavirus (=Bocapar-vovirus)	*Parvo-viridae*	Porcine bocavirus 3	x	2	2	x	respiratory tract	gastrointestinal tract andrespiratory symptoms	human, pig, cattle, dog	unlear	Manteufel and Truyen, 2008; Huang et al., 2014
Mastadeno-virus/ Dependovirus	*Adeno-viridae*	Porcine adenovirus A			2	x	respiratory, gastrointestinal infection	unclear	human, mammals	unclear	Maluquer de Motes et al., 2004
Astrovirus	*Astro-viridae*	Astrovirus w ild boar	x		1	x	asymptomatic/ unlear: enteritis	enteritis/ asymptom.	vertebrates	unlcear	Kapoor et al., 2009; Luo et al., 2011
Sapovirus	*Calic-iviridae*	Porcine enteric sapovirus	x			x	astroenteritis.	diarrhoea/ asymptomatic	pig, humans	unclear	Bank-Wolf et al., 2010
Picobirna-virus	*Picobirna-viridae*	Human picobirnavirus RNA	x	2		x	gastroenteritis in animals, humans, association to disease unclear	gastroenteritis in animals, humans, association to disease unclear	mammals	unlear	Ganesh et al., 2014; Wilhelmi et al., 2003
Rotavirus	*Reoviridae*	Porcine rotavirus C; Human rotavirus C	x			x	enteritis/ asymptom.	diarrhoea/ asymptomatic	human, vertebrates	proved	Martell et al., 2010; Papp et al., 2013
Hungarovirus	*Picorna-viridae*	Bovine hungarovirus			1	x	unlear	unlear	unlear/ found in: pig, cattle, sheep	unlear	Reuter et ., 2012

Two of the five pigs investigated in our first study suffered from watery diarrhoea (paper 1). One of the viruses detected in these samples was rotavirus. Rotaviruses are known agents for causing gastroenteritis in pigs (Kim et al., 1999; Halaihel et al., 2010). Metagenomic analysis may be useful for the identification of aetiological agents involved in certain diseases. Rotavirus was also detected in metagenomic studies of faeces of weaned and suckling pigs with diarrhoea carried out by several other groups (Theuns et al., 2016; Belak et al. (2013a and 2013b). Analogous to our studies, Belak et al. also found porcine astroviruses and kobuviruses.

Enteric viruses of pigs may also pose a risk of zoonotic transmission to humans. Indication for a zoonotic potential of the detected porcine enteric viruses may be provided by similarities of genome sequences and subsequent phylogenetic analyses as performed by Machnowska et al. (2014). They could show that the contained porcine astrovirus is only distantly related to human astrovirus strains, which indicates a separate evolution and no zoonotic transmission between pigs and humans. However, porcine rotavirus A grouped in the same branch together with human rotavirus A (Machnowska et al., 2014). Our results confirm the close relationship: we could detect reads with the highest similarity to human rotavirus A and some to porcine rotavirus A in the same sample (study 1; data in detail not published).

4.2.3 Bacteriophages

Phages play a key role in shaping the development and functional outputs of their host bacteria (Canchaya et al., 2003b; Ogilvie et al., 2013). They are of major significance for regulation of bacterial growth and horizontal gene transfer (Casas and Rohwer, 2007). However, their diversity is largely understudied due to the fact that their hosts are often unknown and/or the culture conditions are unidentified (Breitbart and Rohwer, 2005a; Casas and Rohwer, 2007; Wommack et al., 2009). Currently, the official taxonomy lists one bacteriophage order containing three families and seven additional phage families (King et al., 2012b).

The phage communities of the porcine faecal samples analysed in our studies were very diverse. The majority of detected bacteriophages could be grouped into

the four families *Microviridae, Myoviridae, Podoviridae, Siphoviridae* (paper 1, table 5; paper 2, supplementary table 1). A further consistent pattern of bacteriophages at the species level could not be identified by comparison of the samples. Only a few bacteriophages could be identified in all of the analysed samples: Bdellovirbrio phage phiMH2K, Chlamydia phage Chp1 and Chp2, Dragonfly-associated microphage 1, Spiroplasma phage 4 and Lactococcus phage 1706 (paper 2, figure 4; supplementary table 2).

However, most of the reads showed only minor similarities to known bacteriophage sequences present in the database. Therefore, it remains unclear whether the assignment of the generated sequence reads to a distinct bacteriophage species is correct, or if they represent related, but so far unknown bacteriophages. Currently, the databases contain only a low number of complete bacteriophage genomes making the identification of bacteriophages by metagenomic approaches and BLAST comparison generally difficult. As bacteriophages, which are closely related to each other, may infect different hosts (Whichard et al., 2010; Wongsuntornpoj et al., 2014), a reliable assignment of the bacteriophage host is also problematic (Casjens, 2005; Rodriguez-Valera et al., 2009; Wommack et al., 2009). This was also shown in a parallel analysis of the bacterial community of the samples, which was performed by a collaborating group of Dr. Vahjen (FU, Berlin; personal communication). Sequencing of the bacterial 16S rRNA genes from the samples revealed the presence of bacterial species, for which no phage was identified, and bacteriophages, for which the known host species was not present based on the bacterial 16S rRNA genes.

To study interactions between bacteriophages and their bacterial hosts within the gut of pigs, a future increase of annotated bacteriophage genomes within the databases is a prerequisite. Alternatively, Modi et al. (2013) solved the problem of a reliable association of bacteriophages and their host by the identification of host-related sequences inserted into the phage genome. However, this will require a very deep sequencing, which allows the assembly of complete phage genomes.

Some of the identified highly abundant bacteriophages may also be interpreted with care because of the quality of the database. For instance, Lactococcus phage 1706 has a high economic impact for the diary cheese product industry as it is

responsible for milk fermentation failures as predator of *Lactococcus lactis* (Garneau et al., 2008). Therefore, it is studied intensively and many genome sequences of this phage are in the database, which makes the identification of related sequences in metagenomic studies easy. Contrary, for the highly abundant dragonfly-associated microphage, which is as the name implies an insect-borne phage, the specific bacterial host is in reality not known (Rosario at al., 2012a).

4.2.4 Influence of age and probiotic feeding

Viral communities could be influenced by several also dynamic factors. Several abiotic and biotics factors could influence viromes, e.g. the application of nutritional factors, treatments with medicine, changes due to illness, age, sex and the intra- and intergenic relationship. Within this thesis, the influence of probiotics on the porcine faecal virome was analysed. To do so, it became evident that the most obvious factor influencing the composition of the faecal virome of pigs was not the probiotic feeding, but the age.

The **age-effect** was remarkable between piglets and adult pigs. By comparison of viromes from piglets and adult pigs, large differences in the ratio between bacteriophages and mammalian viruses could be observed. Additionally, the diversity of detected virus species varied between the analysed viromes derived from pigs with different ages. Within the youngest piglets, the most abundant virus group was porcine viruses, but the proportion of porcine viruses decreased dramatically in the samples from the older piglets to adult pigs. In parallel, the proportion of bacteriophages as well as the general diversity of detected virus species increased with age (paper 2, figure 3 and 6). Moreover, the composition of porcine virus species varied dramatically with the age of the pigs. While porcine kobuvirus was the most abundant virus species in piglets, the adult pigs viromes were dominated by PigSCV. Both viruses were exclusively detected either in the piglet or the adult pig groups (paper 2, figure 5). The Bray Curtis dissimilarity dendrograms, which show the compositional dissimilarity between two different groups, also clearly indicate age-specific effects for porcine viruses (paper 2, figure 6). Interestingly, the effect of age was very similar in both analysed groups,

despite the fact that they were held completely separate during the whole period of the experiment. Similar observations of increased virus diversity were also published for human faecal viromes (Breitbart et al., 2008). Whereas low diversity in an average of eight viral genotypes was found in babies, about 2000 viral genotypes were estimated for adults.

Although the reasons for the age-specific effect on the composition of the pig faecal virome are not known so far, several mechanisms may be hypothesised. First, the development of the immune system is subject to marked changes in the first month of life. The maternal immunity given by breast milk is only present in the first weeks of age (Holyoake et al., 1995; Muns et al., 2014) and confers protection against selected viruses, which disappears after a few weeks. Second, an age-related susceptibility to specific virus infections depending on development of organs/tissues with the presence of virus receptors may occur. Third, the adaptive immunity will confer resistance against specific virus infections after the first period of infection. In addition, there is the progressive development of the bacterial enteric flora, which represents the hosts for bacteriophages. The establishment of the intestinal microbial flora needs several weeks (Conroy et al., 2009; Huang et al., 2013) and the increasing presence of hosts for specific bacteriophages may explain the later observed increase in bacteriophages. When considering the influence of age on virome dynamics, the effect of general diet should not be disregarded as diet changes during the development. The most important change occurs with the weaning of breast milk; this is associated with several changes within the intestinal tract.

Probiotics may have beneficial effects on enteric virus infections (Colbere-Garapin et al., 2007). However, no studies on the general influence of probiotics on the composition of gut viromes have been published so far. Our study (paper 2) was done in a framework of investigations related to the effects of probiotics. Therefore, two feeding pig groups - one with probiotic bacterium and the other without - were analysed. In contrast to the marked influence of the age on the virus composition in pig faeces, no general effect could be observed regarding the probiotic bacterium *Enterococcus faecium* (*E. faecium*) NCIMB 10415 as feed additive (paper 2, figure 5 and 6).

Special attention was given to Enterococcus phages. These phages were detected in all samples, with a markedly large amount in one sample derived from the *E. faecium*-feeded group (paper 2, figure 6). As a predator of enterococci, these bacteriophages may also use the probiotic bacterium *E. faecium* NCIMB 10415 as a host. Thus, the large amount observed in the sample may be a result of high replication of this bacteriophage. This may also lead to a destruction of the probiotic bacterium, which may not be able to exert the probiotic function. However, this interpretation of the finding is questionable as the effect was only detected in one sample. Moreover, due to the lack of data in the GenBank database, the exact host for the detected bacteriophage could not be determined.

Slight differences could also be observed for porcine viruses. Astrovirus was detected only within the control group (paper 2, supplementary table 1). Interestingly, a parallel analysis of the samples by RT-qPCR confirmed the presence of astrovirus exclusively in the control group (Kreuzer et al., 2012). However, this finding is difficult to interpret as it simply may reflect a difference in astrovirus exposition between both groups at the beginning of the experiment, as already speculated by Kreuzer et al. (2012). A comparison of both studies (our metagenomic approach and the RT-qPCR approach) revealed that a higher sensitivity of the applied method may be a prerequisite in order to identify probiotics-induced effects. Using the real-time RT-PCR, differences could be identified for rotavirus which were recorded by Kreuzer et al. (2012), but not with our metagenomic method (paper 2). A much deeper sequencing may therefore be necessary for analysis of effects of probiotics on the faecal virome composition.

So far, the stability or dynamics of viromes was studied only for human faecal viromes. Minot et at. (2011) investigated the response of the faecal virome to diet and underlined a huge inter-individual variation. However, they also detected a minor reaction of the viral population to a specific diet. Analogous to this thesis, Breitbart et al. (2008) also identified a striking difference between the adult and infant viral communities regarding viral diversity. Viral diversity seems to be age-dependent and to increase in adults. Reyes et al. (2010) and Zhang et al. (2006) observed a minimal intrapersonal variation and very high interpersonal variation. Based on the repeated observation of high interpersonal variation, future studies should work with individual (instead of pooled) samples as seen in this study.

Furthermore, they should underline that the number of investigated samples is very important for the reliable assertions regarding dynamics of viromes.

4.3 The rat faecal virome

4.3.1 General composition

The Norway rat (*Rattus norvegicus*) often lives in close proximity to human quarters and is known as a reservoir for several zoonotic pathogens (Mitchell-Jones et al., 1999). These pathogens are different bacteria, but also viruses, e.g. Seoul hantavirus as a respiratory pathogen and cowpox virus as a skin pathogen (Himsworth et al., 2013; Meerburg et al., 2009). The presence of viruses in faeces of wild rats has been only scarcely analysed and most of the studies on enteric viruses of rats have only investigated pathogens of breeding colonies of laboratory rats (Baker, 1998). At the beginning of this study, only one paper (Phan et al., 2011) was published on a metagenomic analysis of rodent viromes; however, Norway rats were not included. Very shortly after publication of our study, Firth et al. (2014) published a similar metagenomic study for urban wild Norway rats collected in New York, whereby the overall virome composition of both studies was quite similar. A closer inspection of the comparison of the compositions follows later in this paragraph.

The results of our study - which included 20 faecal samples derived from wild Norway rats from Berlin - indicate a high heterogeneity of enteric viruses. The total number of reads per sample ranged from 29,923 to 910,647, indicating the presence of different amounts of viruses. The viral reads could be assigned to 34 known virus families and 75 virus genera. The most abundant viruses were grouped in 21 virus genera infecting mammalian hosts, whereas other virus genera infecting birds, invertebrates, insects, amoeba, fungi and plants were found at a markedly lower frequency (paper 3, table 2). This finding indicates that transient viruses, e.g. derived from the feed, contribute only marginally to the rat faecal virome, which is in contrast to the situation in humans as discussed above (Zhang et al., 2006). Bacteriophages were not analysed in this study.

Although the relative abundance of the different virus reads varied between rat samples, some of the virus genera were generally more abundant overall. Members of the families *Parvoviridae* and *Picobirnaviridae* represented the most abundant species; these both contain well-known viruses of rats (Fregolente et al., 2009; Jacoby et al., 1996). Less abundant viruses incuded known and unknown members of the genus *Bocavirus*, *Rotavirus*, stool-associated circular ssDNA viruses (SCV), *Dependovirus*, *Mastadenovirus*, circovirus-like viruses, *Cardiovirus* and *Enterovirus* (paper 3, table 2). Table 10 provides deeper insight into the composition of the faecal rat virome and gives an overview on diseases associated with the enteric viruses found in this study.

Table 10: Summary of mammalian virus found in rat faecal virome of paper 3 and their association to diseases

Genus	No. of rats	Familiy	Diseases in human	Diseases in rats	Hosts	Zoonotic potential	References
Bocaparvovirus	18	*Parvoviridae*	respiratory tract	unclear	human, pig, cattle, dog	unlear	Manteufel and Truyen, 2008
Cardiovirus	6	*Picornaviridae*	gastrointestinal tract myocarditis	labor inoculation leaded to neurologic paralysis	human, vertebrates	potential	Easterbrook et al., 2008; Rivera-Benitez et al. 014
Circovirus	9	*Circoviridae*	unlear	unlear	pigs, birds	unlear	Rosario et al. 2012b
"Circo-like"	16	*Circoviridae*	unlear	unlear	unlear	unlear	Rosario et al. 2012b
"Hungarovirus"	2	*Picornaviridae*	unlear	unlear	unlear; found so far in: pig, cattle, sheep	unlear	Reuter et al., 2012
Mamastrovirus	7	*Astroviridae*	enteritis/ asymptom	unlear	vertebrates	unlcear	Chu et al., 2010; De Benedictis et al., 2011
Mastadenovirus	16	*Adenoviridae*	intestinal, respiratory infection	unclear, mice: diarrhoea	human, mammals	unclear	Klempa et al., 2009
Protoparvovirus	20	*Parvoviridae*	no	asymptomatic infection, hepatic necrosis	vertebrates	not likly	Jacoby et al., 1987 and 1996; Wan et al., 2002
Rotavirus	12	*Reoviridae*	gastroenteritis	unlear; likly diarrheoa	human, vertebrates	RAV potential, RBV not likly	Ciarlet et al., 2002; Martella et al., 2010
Hepevirus	3	*Hepeviridae*	hepatitis	unclear	pig, rodent, chicken	for ratHEV not likly	Johne et al., 2012
Kobuvirus	2	*Picornaviridae*	gastroenteritis	unlear	human, pig, cattle, sheep	unclear	Reuter et ., 2011
Parechovirus	1	*Picornaviridae*	intestinal, respiratory infection	unlear	human, mice	potential	Hauffe et al., 2010; Niklasson et al., 2009
Picobirnavirus	16	*Picobirnaviridae*	gastroenteritis association to disease not proved	unlear	mammals	unlear	Ganesh et al., 2014; Wilhelmi et al., 2003
Pirconavirus	6	*Picornaviridae*	unlear	unlear	unlear	unlear	Ng et al., 2015
"Rosavirus"	4	*Picornaviridae*	unlear	unlear	unlear	unlear	Phan et al., 2013b
Sapovirus	1	*Picornaviridae*	gastroenteritis	unlear	human, pig	unclear, low potential	Hansman et al., 2007; Wang et al., 2005
Dependoviurs	11	*Adenoviridae*	unlear	unlear	unlear	unlear	Lochrie et al., 2006
Enterovirus	4	*Picornaviridae*	intestinal, respiratory infection, meningitis, encephalitis	unlear	human, mammals	unclear	Pfeifer, 2010
Sapelovirus	2	*Caliciviridae*	gastroenteritis	unlear	human, pig, mice	potential	Phan et al., 2011
Norovirus	2	*Caliciviridae*	gastroenteritis	unlear	human, mammals	unclear, human NoV detected in rats	Tse et al., 2012a; Wolf et al., 2013

A comparison of our results with that of the very recent study by Firth et al. (2014) shows similarities as well as some differences. Parvoviruses, bocaviruses, cardoviruses, picobirnaviruses and picornaviruses were identified in high quantities in both studies, indicating their wide distribution among urban rats in two continents. Also, the circular ssDNA virus (ratSCV) associated with rat faeces (as a member of the propagated new virus family "*Smacoviridae"*; see chapter 4.4.1*)* could be frequently found among the urban rats. Interestingly, this virus could not be found in the viromes of other wild rodent species (Phan et al., 2011). Differences were observed regarding the presence of astroviruses, rotaviruses and adenoviruses, which were detected in the rats from Berlin, but not in the rats from New York. In addition, in the study of Firth et al. (2014), a Seoul hantavirus strain was detected by specific PCR. This virus belongs to flaviviruses which may cause serious diseases in humans. Seoul virus could be repeatedly detected among other wild rats in the Netherlands (Verner-Carlsson et al., 2015). A difference in the geographical distribution of these viruses may be one reason for these findings. However, as the Seoul virus and the flavivirus have only been detected in additional PCR-based screenings of the samples, the different applied methodologies may also contribute to the observed differences. Table 11 compares the faecal rat virome from New York and Berlin.

Table 11: Comparison of two rat faecal viome studies of rats from New York and Berlin. The table reflects only viruses that were detected by unbiased high-throughput sequencing. Virus genera flagged with ” ” are not by the ICTV accepted genera

Rats New York				Rats Berlin		
Most similar species	**% of pools**	**Genus**	**Familiy**	**Most similar species2**	**New rat virus detected in Berlin**	**% of rats**
Porcine bocavirus	48	*Bocapar-vovirus*	*Parvoviridae*	Pig bocavirus Pine marten, bocavirus	Rat bocavirus	90
Calhevirus 1	30	*Calhevirus*	Picorna-like virus	/		/
Rat theilovirus 1	64	*Cardiovirus*	*Picorn-aviridae*	Theilovirus		30
Pig SCV	64	"SCV"	*Circoviridae*	Porcine SCV	RatSCV	50
Chicken anemia virus	64	*Circovirus*	*Circoviridae*	Bat circovirus		45
	/	"Circo-like"	*Circoviridae*	Pro-circo-like virus 41		80
Ovine hungarovirus	58	"Hungaro-virus"	*Picorna-viridae*	Ovine hungarovirus		10
Rat astrovirus	58	*Mamastro-virus*	*Astroviridae*	Rat astrovirus		35
Murine adenovirus A	55	*Mastadeno-virus*	*Adenoviridae*	Murine adenovirus B		80
Kilhamrat virus	73	*Protopar-vovirus*	*Parvoviridae*	Rat minute virus, Kilham rat virus		100
	/	*Rotavirus*	*Reoviridae*	Rotavirus A	Rat rotavirus A	25
Rat rotavirus B strain IDIR	9	*Rotavirus*	*Reoviridae*	Rat rotavirus Bstrain IDIR		40
Hepatitis E virus	3	*Hepevirus*	*Hepeviridae*	Hepatitis E virus		15
Mouse kobuvirus + Aichi virus	58	*Kobuvirus*	*Picorn-aviridae*	Kobuvirus, Sewage kathmandu		10
Ljungan virus Sebokele virus	15	*Parechovirus*	*Picorn-aviridae*	Human parechovirus		5
Human + Otarine picobirnavirus	67	*Picobirn-avirus*	*Picobirna-viridae*	Human picobirnavirus		80
Feline picornavirus	100	*Pircornavirus*	*Picorna-viridae*	unclassified	Rat-borne virus	30
Rosavirus	61	*Rosavirus*	*Picorna-viridae*	Mous rosavirus		20
Human + Porcine sapovirus	39	*Sapovirus*	*Picorn-aviridae*	Sapporo virus	Rat sapovirus	5
	/	*Dependo-viurs*	*Adenoviridae*	Rat adeno-associated virus		55
	/	*Enterovirus*	*Picorna-viridae*	Human enterovirus C		20
	/	*Sapelovirus*	*Caliciviridae*	Mouse sapelovirus		10
		Norovirus	*Caliciviridae*	Murine norovirus		10

4.3.2 Zoonotic potential of enteric rat viruses

Norway rats are well-known reservoirs of several zoonotic pathogens (Himsworth et al., 2013; Meerburg et al., 2009). In order to identify such zoonotic viruses in our samples, a more detailed sequence analysis including PCR-based screenings of some of the detected viruses was performed (in collaboration with Patrycja Machnowska, BfR). The first screening identified ratHEV, norovirus and rotavirus as possible zoonotic virus candidates.

HEV-like sequences were detected in three of the 20 rat faecal samples (paper 3, table 2). HEV is a pathogen, which may cause hepatitis in humans. At this point, the genotypes 1 to 4 of HEV are known to infect humans (Purcell et al., 2008; Scobie and Dalton, 2013). Recently, an HEV-related agent has been identified in wild Norway rats in Germany, designated as ratHEV, which is only distantly related to the human HEV genotypes 1 to 4 (Johne et al., 2010). Although the presence of ratHEV-specific antibodies has been described for a few forestry workers in Eastern Germany, the zoonotic potential of ratHEV seems to be low (Dremsek et al., 2012). A closer inspection of the HEV-like sequences from the rats in our study identified all of them as belonging to ratHEV. Therefore, the zoonotic potential of these HEV strains was considered low.

Norovirus-like sequences were detected in very low quantities in two of the 20 rat samples (paper 3, table 2). Norovirus infections are a leading cause of gastroenteritis in humans worldwide and in Germany (Atmar, 2010; Mäde et al., 2013). The norovirus genogroups I, II and IV are known to be pathogenic for humans. Currently, human noroviruses are not considered to be zoonotic infective. However, a genogroup I norovirus closely related to human strains has been recently detected in a Norway rat (Wolf et al., 2013). The very low number of virus sequences in our rat samples was problematic for analysis of the detected norovirus; nevertheless a close relationship to a recently discovered rat norovirus (Tse et al., 2012a) was evident (paper 3, supplementary figure 2). The rat norovirus was closely related to murine norovirus of genogroup V, which has not been detected in humans so far. Therefore, the risk for zoonotic transmission to humans seems to be low.

Rotavirus-like sequences have been detected in many of the analysed samples. Rotaviruses are agents of gastroenteritis in humans, which mainly cause severe disease in young children (Martella et al., 2010; Tate et al., 2012). Group A, B and C rotaviruses are known to infect humans. Rotaviruses are also found in several animal species and may cause gastroenteritis in young animals (Garmendia et al., 2015; Imagawa et al., 1991; Otto et al., 2015). Especially group A rotaviruses of animals are also considered as zoonotic agents (Martella et al., 2010). A closer investigation of our sequences indicated that six of the rats contained a rotavirus, which was most closely related to the rat rotavirus B strain IDIR (paper 3, supplementary figure 3). Interestingly, this is only the second description of a rotavirus B in rats since its first detection in laboratory rats in 1984 (Vonderfecht et al., 1984). Phylogenetic analysis showed that our rat sequences clustered separately from the human rotavirus B sequences, indicating a low potential for zoonotic transmission to human.

Sequences closely related to rotavirus A were detected in four of the 20 rats (paper 3, table 2) with one of them shedding a significant rotavirus amount (80.6% of all virus reads). This high virus amount enabled us to sequence the whole rotavirus genome by RT-PCR amplification and classic Sanger sequencing. The genome of rotaviruses is composed of eleven individual segments of double-stranded RNA, which encode one or two proteins each (Desselberger, 2014). From the genome sequence, it became evident that only three of the genome segments could be assigned to known genotypes (G3-P[3]-N2), whereas all the remaining eight genome segments present new genotypes which have never been detected to date. This might indicate that the virus is a rat-adapted strain and represents the first rotavirus A strain from rats. However, six of the genome segments showed the highest sequence identities to human rotaviruses (paper 3: table 3). The G3-P[3] genotype has mostly been detected in dogs and cats, although infections in rabbits, humans, rhesus monkeys and bats have also been documented (He et al., 2013; Martella et al., 2010), thus underlining a general zoonotic potential of G3-P[3] rotavirus strains. Consequently, zoonotic transmission of the rat strain to humans cannot be excluded.

4.4 Discovery of novel viruses

The NGS-based metagenomic analysis approach applied in our study is based on a comparison of the generated sequences with sequences of known viruses present in a database. Thus, identification of new viruses is clearly restricted as those sequences showing no or low similarities to known viruses are weeded out during the procedure. Nevertheless, novel viruses showing only moderate sequence similarities to known viruses can be identified by the method. Indeed, the NGS-based method has been repeatedly applied with success for discovery of novel virus species (Chiu, 2013; Delwart, 2007; Edwards and Rohwer, 2005; Li and Delwart, 2011c; Mokili et al., 2012; Reyes et al., 2012). Moreover, several novel viruses could be identified in our studies by using the method and were subsequently characterized further. This included novel stool-associated circular ssDNA viruses (SCV) of pigs and rats as well as novel sapovirus, bocavirus and picornavirus of rats. The detection of these new viruses underscores the high potential of the NGS-based metagenomic analysis approach for the non-targeted identification of novel viruses.

4.4.1 Pig and rat SCV

By analysis of the sequences derived from pig faeces (paper 1 and 2), similarities to the circular ssDNA virus associated with chimpanzee stool samples (Blinkova et al., 2010) were found. We named this new virus “pig stool-associated circular ssDNA virus” (pigSCV). At the point in time when the virus was identified, PigSCV and chimpanzee SCV were the only representatives of the SCV-like virus group. SCVs were later also detected in pig faecal samples from other countries and continents and were designated as porcineSCVs (Cheung et al., 2013; Kim et al., 2014; Shan et al., 2011; Sikorski et al., 2013b; Zhang et al., 2014a). These porcine SCVs show a close relationship to pigSCV, but have a different genome organization (see below). With the more common use of NGS, more related SCV species were found in faeces of cattle (Kim et al., 2012), fur seal (Sikorski et al., 2013a), gorilla, lemur black howler and humans (Ng et al., 2015b) and turkey (Reuter et al., 2014). In our study (paper 3), an SCV could also be detected for the first time in faeces of Norway rats, which was subsequently designated as ratSCV.

Although quite diverse in sequence and derived from the faeces of different animals (see figure 6), SCV genomes contain a set of conserved features. All SCV genomes are small and consist of circular DNA. They encode at least two proteins: a replication initiator protein (Rep) and a capsid protein (Cap). In addition, all SCVs contain a conserved canonical NANNNTTAC sequence downstream of the *rep* gene (Blinkova et al.; 2010; Cheung et al.; 2013; Kim et al., 2012, Kim et al., 2014; Reuter et al., 2014; Sikorski et al., 2013b). This sequence is considered to be responsible for the initiation of the rolling-circle replication of the virus genome (Rosario et al., 2012b). Other viruses containing Rep-like proteins can be found in the animal virus family *Circoviridae* and in the plant virus families *Geminiviridae* and *Nanoviridae*. However, in phylogenetic trees based on the Rep protein sequences, all SCVs cluster in the same branch, which is different from that of *Circoviridae, Geminiviridae* and *Nanoviridae* (figure 6). Also, the canonical sequence in these virus families is different compared to that of the SCVs. Therefore, it is currently proposed to place the SCVs into a novel virus family designated "*Smacoviridae*", small circular viruses (Ng et al., 2015b).

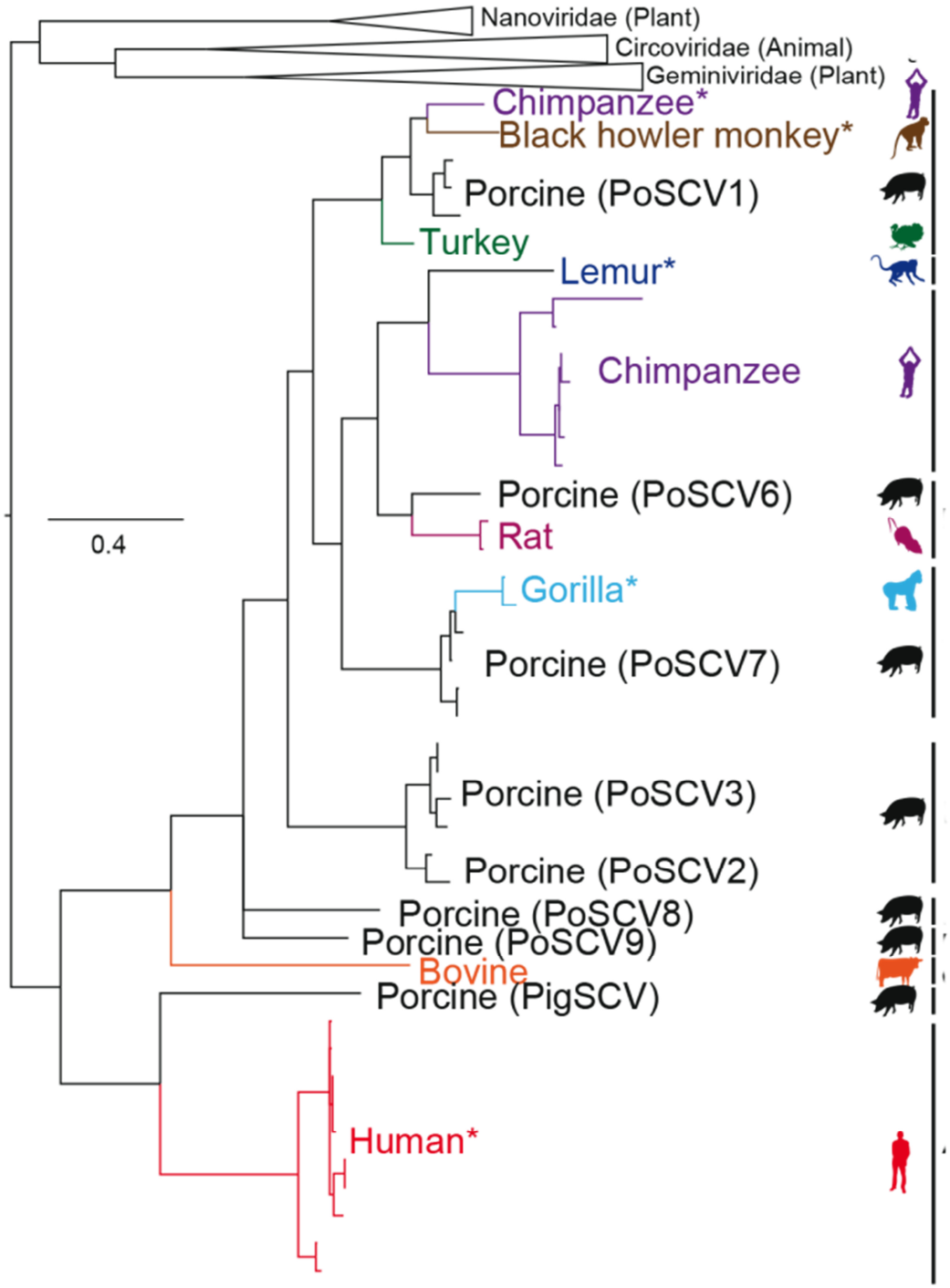


Figure 6: phylogenetic analyses of *rep* gene sequences of the postulated family *Smacoviridae*. Bayesian tree was generated using MrBayes, where 1,000,000 generations were sampled every 50 steps. Branches are coloured according to the viral animal sequences. (adopted from Ng et al., 2015b).

Regarding the orientation of the *rep* and *cap* genes, only PigSCV shows a head to tail configuration, whereas all other SCVs show a head to head configuration. Furthermore, the recently discovered porcine SCVs display a head to head

configuration (Cheung et al., 2014; Shan et al., 2011). A recombination event leading to a change in gene orientation in PigSCV may be speculated to have occurred in the past. Phylogenetically, PigSCV und the porcine SCVs are closely related; however, PigSCV branches together with humanSCV nearly the root of the porcine SCV group, which may indicate a beginning diversification of both SCVs and humanSCVs (figure 6).

The ratSCV was identified in 15 of the 20 investigated rat samples, indicating its broad distribution. It shall be noted that it was most closely related to a sequence annotated as a plasmid of an “uncultured bacterium” from rat caecum content (paper 3, figure 3; Jorgensen et al., 2014). Jorgensen investigated hundreds of circular novel plasmids and they did not recognize this sequence as a virus sequence. The genome of ratSCV and the rat caecum plasmid showed around 92% sequence identity, which indicate that they represent strains of the same virus from different geographical locations (paper 3, figure 3).

Indeed, the question of whether the detected SCV sequences represent virus genomes has not yet been sufficiently answered. All SCV strains have only been identified by genome detection in faeces, while neither virus particles nor cell culture isolated viruses or virus proteins have been isolated or demonstrated so far. Efforts to demonstrate the virus in other sample types like faeces were not successful. Neither in our study by testing serum samples (paper 1, table 7) nor by testing the food of SCV-positive tested animals (Kim et al., 2012) SCV could be detected outside of faeces by using PCR. Blinkova et al. (2010) assumed an origin of ChiSCV from infected plants that were ingested and excreted after intestinal passage. The results of our study argue against this hypothesis as all of the piglets excreted PigSCV, but the amounts and time-points of virus excretion in the faeces differed between the individuals (despite an identical diet). Further studies are required to assess the general organ distribution, origin, prevalence, and clinical significance of SCVs.

The clinical importance of SCVs is also not yet known. Recently, SCVs were repeatedly detected in unexplained human gastroenteritis outbreaks in France and in the USA. For the USA outbreak, no other known enteric pathogens causing gastroenteritis could be detected (Ng et al., 2016 publication in revision). Most of

the animal SCV sequences were obtained from faecal samples associated with diarrhoea (paper 1; Cheung et al., 2013; Reuter et al., 2014; Woo et al., 2014; Zhang, 2014a) and bovineSCV (Kim et al., 2012) was detected in faeces of calves with high fever and anorexia. In addition, some studies identified SCVs in samples of clinical healthy animals (paper 2; Kim et al., 2014; Sikorski et al., 2013b). Further investigations are necessary to determine whether SCV infection could cause diseases.

4.4.2 Rat Sapovirus

Two of the rats investigated in our study (paper 3) contained sequences with identities to sapoviruses. The detection of sapoviruses in rats had not yet been reported in the literature so far. Sapoviruses are known as causative agents of gastroenteritis in humans and distinct sapoviruses have also been detected in some animal species (Dufkova et al., 2013; Guo et al., 2001; Li et al., 2011a and 2011b; Tse et al., 2012b). Interestingly, the deduced amino acid sequences of a short genome segment showed the closest relationship of the rat sapovirus with human sapoviruses. However, the low percentage of identity and the phylogenetic branching between human and porcine sapoviruses (paper 3, figure 4) may indicate that the rat virus has a low potential of zoonotic transmission to humans. Nevertheless, human sapoviruses comprise a rather diverse group of viruses with high sequence variability (Hansman et al., 2007).Further characterization of the rat sapovirus, which was not possible here due to its low quantity within the samples (paper 3; table 2), should be performed in the future in order to assess the zoonotic potential of this virus.

4.4.3 Rat Bocavirus

A novel sequence of a rat bocavirus could be detected in most of the samples analysed in this study (paper 3). *Bocaparvoviruses* represent a genus of the family *Parvoviridae*, which has been described to be associated with respiratory and gastrointestinal diseases in humans and animals (Chow and Esper, 2009; Manteufel and Truyen, 2008); however they had not been detected in rats so far. Although the exact relationship of the rat bocavirus with other bocaviruses remains to be determined by whole-genome analysis, the data presented in this study

show only low sequence similarities of the analysed genome fragment to that of human viruses. Therefore, a low probability of its zoonotic transmission to humans has to be assumed.

4.4.4 Rat Picornavirus

Within approximately half of the rat samples several sequences could be detected, which showed similarities to various genes of different known picornaviruses. While the 3D gene sequences could be assigned to the genus *Sapelovirus* (paper 3, supplementary figure 1), that of the VP2 and VP3 gene showed insufficient sequence identities to known picornavirus genera (paper 3, figure 1). Moreover, the sequences also demonstrated few similarities to the rodent picornavirus genera *Cardiovirus* (Phan et al., 2011), *Mosa*- and *Rosaviruses* (Phan et al., 2011, 2013). These findings indicate that the rat picornavirus represents a novel genus of the family *Picornaviridae*, which is phylogenetically close to the root of the genera *Enterovirus* and *Sapelovirus* (paper 3, figure 1). Members of the genus *Enterovirus* are known to cause self-limiting diarrhoea and respiratory diseases. Several rhinoviruses, which are a major cause of the common cold in humans, are classified in the genus *Enterovirus*. Moreover, poliovirus, which can cause severe paralysis in humans, is a member of this genus (Pfeiffer et al., 2010). Therefore, it will be very interesting to analyse the pathogenic and zoonotic potential of the novel rat picornavirus.

In the study by Firth et al. (2014), sequences closely related to the novel rat picornavirus could also be detected, indicating a broad distribution of the virus in rats over two continents. Meanwhile, the whole genome of the virus was sequenced and analysed together with other representatives of the family *Picornaviridae* (Ng et al., 2015a). Hence, the distinctiveness of the virus was confirmed and a classification of the rat picornavirus as the prototype member of a novel *picornavirus* genus with the suggested designation "*Rabovirus*" (rat-borne virus) was indicated.

5 OUTLOOK

5.1.1 Problems to be solved in virome analysis

To date a direct **comparison of viromes** from different samples has only been attempted for human intestinal viromes from individuals with different diets or ages (Breitbart et al., 2008; Minot et al., 2011; Reyes et al., 2010), and from our studies on pigs. Generally, a comparison of metagenomic sequence data is difficult due to methodological and statistical reasons.

One major shortcoming of current metagenomic virome analyses is a low **reproducibility** or an unknown method performance due to missing validation or missing quality controls. The applied protocols for sample preparation, genome amplification and data analysis are quite diverse, error-prone and not standardized. To overcome this problem and to characterise the performance of the method, we developed here a process-controlled deep sequencing approach (see chapter 4.1.2: process contro). The detection of added bacteriophages was quantified and the calculated recovery rates provided an estimation of the method performance for each sample. However, some important virus morphologies such as enveloped viruses were not included in our process control. Further developments of reliable process controls should be attempted in future. In addition, the use of process controls should be mandatory in virome analysis in order to enable reliable comparisons of the data between different studies. Also, a more standardized data analysis process would contribute to a better ability to compare data from different studies.

So far, a **statistic validation** of metagenomic virome data is not possible. The high costs for analysis of an adequate number of samples and the work-intensive analysis of the results may be major restricting factors. However, it may also be necessary to further develop the statistical methods themselves as reliable algorithms for direct comparison of NGS sequence read data sets from different samples.

Another shortcoming of metagenomic virome analysis methods is the comparably low **sensitivity.** Comparison of the results of our study (paper 2) with that of an RT-PCR analysis of the same samples (Kreuzer et al., 2012) identified

considerable differences, such as the absence of rotavirus by NGS analysis and its presence by real-time RT-PCR. Up to 10^5 rotaviruses per gram faeces were present in the samples, which were not detected by NGS. In contrast, the 10^7 astroviruses per gram faeces were readily detected by NGS. Analogously Drexler et al. (2012) showed that when using NGS no paramyxovirus RNA was found, but using a consensus PCR its presence was readily demonstrated. To overcome this problem, the deepness of sequencing has to be increased. However, this will increase the costs and time of data analysis.

So far, viral metagenomic analyses seem to be not suitable for **absolute quantification** of viruses. Besides the shortcomings of the method reproducibility and sensitivity as mentioned above, some additional problems have to be solved. As the viruses do not contain genes, which are conserved in all viruses, the whole genomes have to be used for counting the viruses by NGS. However, as shown in our study (paper 1), the virus species prevalence differs greatly if NGS reads, contigs or genome size-corrected data are used for calculation. It has to be considered that longer genomes are represented by more reads per genome compared to shorter genomes. Due to the remarkable difference of the genome length between different virus families, the use of a general correction factor for the genome length should be introduced.

A major bottleneck to identify the viruses is the **quality of the used database.** As the identification of the viruses is done by sequence comparison with those present in the database, an increase of annotated complete viral genomes is desirable. If only short genome fragments are present in the database and the generated NGS read maps to another region of the viral genome, no identification of the virus is possible. Especially for phage genomes, there seems to be a significant lack of genomes due to missing research activities or activities very targeted to specific fields of bacteriophage research.

Last, it should be mentioned that it is not currently possible to **detect totally new viruses** using the applied metagenomic approaches. Although the genome of those viruses will be sequenced by NGS, only viruses showing at least moderate sequence similarity to related viruses in the databases can be detected via BLAST comparison. For the identification of totally new viruses, novel analytical tools have

to be developed. New software tools could be helpful, e.g. functional predictions of putative proteins provided by the bioinformatics.

5.1.2 Future prospects

Although several problems have to be solved, NGS-based metagenomic analysis provides a promising tool for future study of viromes and their dynamics. Several applications of the technique are possible and may be used to answer open questions in the interactions between viruses in their ecological niches and their host cells. The importance of this interplay for animal health and zoonotic virus transmission should be investigated. Also, the distinct role of bacteriophages for regulation of bacterial growth and change of their properties by horizontal gene transfer would be an interesting topic for future NGS-based studies. The influence of external factors such as new probiotics or therapeutics on the composition of the virome may provide new opportunities to treat diseases. Generally, the NGS-based methods also represent very useful tools for disease diagnostics. As they are not targeted to specific agents, they could ideally be used for identification of the etiological agents during disease outbreaks, which is currently only hampered by the high costs and the laborious handling.

After identification of novel viruses by the NGS-based metagenomic methods, it will be of importance to go ahead to characterize the viruses and possible diseases caused by them in more detail. For the SCVs, which were identified here for the first time in pigs and rats, it will be fundamental to clearly show the virus nature of these structures. Also, their function in the gut and a possible involvement in diseases have to be analysed. The same applies to the novel picornavirus found in our study in Norway rats. After completion of the genome sequence (Ng et al., 2015a), the host range, transmission dynamics and involvement of disease of the virus should be investigated.

As animals may serve as a source of zoonotic pathogens, the viruses detected here in pigs and rats should be further investigated regarding their zoonotic potential. In particular rotaviruses of pigs and rats have been identified by our studies as potentially pathogenic and zoonotic agents and should be further characterized. The association of rotavirus A with rats, which was shown here for

the first time, should be investigated in more detail in order to assess the risk of transmission of rotaviruses from urban rats to humans.

6 PUBLICATION LIST

1. Sachsenröder J., Twardziok S., Hammerl J.A., Janczyk P., Wrede P., Hertwig S., Johne R., 2012. **Simultaneous identification of DNA and RNA viruses present in pig faeces using process-controlled deep sequencing**. PLoS One. 7(4):e34631.
2. Sachsenröder J., Twardziok S.O., Scheuch M., Johne R., 2014. **The general composition of the faecal virome of pigs depends on age, but not on feeding with a probiotic bacterium**. PLoS One. 9(2):e88888.
3. Sachsenröder J., Braun A., Machnowska P., Ng T.F., Deng X., Guenther S., Bernstein S., Ulrich R.G., Delwart E., Johne R., 2014. **Metagenomic identification of novel enteric viruses in urban wild rats and genome characterization of a group A rotavirus**. The Journal of general virology 95, 2734-2747.
4. Ng T.F., Sachsenröder J., Reuter G., Knowles N.J., Delwart E., Johne R., 2015. **Rabovirus: a proposed new picornavirus genus that is phylogenetically basal to enteroviruses and sapeloviruses**. Archives of virology. Oct;160(10):2569-75.
5. Ng T.F.F., Zhang W., Sachsenröder J., Kondov N.d.-C., Antonio Charlys; , Vega E., Holtz L., Wu G., Wong D., Stine O.C., Antonio M., Mulvaney U., Muench M., Deng X., Ambert-Balay K., Pothier P., Vinje J., Delwart E., 2015. **A diverse group of small circular DNA viral genomes in human and non-human primate stools**. Virus Evolution. 1(1):vev017.
6. Ng T.F.F., Zhenga W., Sachsenröder J., Kondova N.O., Costaa C., Everardo Vegad U.S.M., Muencha M.O., Xutao Denga X., Ambert-Balaye K.,Pothierf P., Vinjég J., Delwart E., 2015. **Detection of small circular DNA viral genomes in unexplained gastroenteritis outbreaks**. Virus Evolution (publication in revision)
7. Trojnar E. and Sachsenröder J (shared first authorship), Twardziok S., Reetz J., Otto P.H., Johne R., 2013. **Identification of an avian group A rotavirus containing a novel VP4 gene with a close relationship to those of mammalian rotaviruses**. The Journal of general virology 94, 136-142.
8. Johne R., Reetz J., Ulrich R.G., Machnowska P., Sachsenröder J., Nickel P., Hofmann J., 2014. **An ORF1-rearranged hepatitis E virus derived from a chronically infected patient efficiently replicates in cell culture**. Journal of viral hepatitis 21, 447-456.

7 LITERATURE

1. Abe H., Mine J., Parchariyanon S., Takemae N., Boonpornprasert P., Ubonyaem N., Patcharasinghawut P., Nuansrichay B., Tanikawa T., Tsunekuni R., Saito T., 2015. **Co-infection of influenza A viruses of swine contributes to effective shuffling of gene segments in a naturally reared pig**. J Virol 484, 203-212.

2. Abolnik C., 2014. **A current review of avian influenza in pigeons and doves (Columbidae)**. Vet Microbiol 170, 181-196.

3. Aborn J.H., El-Difrawy S.A., Novotny M., Gismondi E.A., Lam R., Matsudaira P., McKenna B.K., O'Neil T., Streechon P., Ehrlich D.J., 2005. **A 768-lane microfabricated system for high-throughput DNA sequencing**. Lab on a chip 5, 669-674.

4. Actis G.C., 2014. **The gut microbiome**. Inflamm Allergy Drug Targets 13, 217-223.

5. Aggarwal R., 2013. **Hepatitis e: epidemiology and natural history**. J Clin Exp Hepatol 3, 125-133.

6. Allan G.M., Ellis J.A., 2000. **Porcine circoviruses: a review**. J Vet Diagn Invest, Inc 12, 3-14.

7. Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J., 1990. **Basic local alignment search tool**. J Mol Biol 215, 403-410.

8. Amori G., Cristaldi M., Edited by A. J. Mitchell-Jones G., Amori W., Bogdanowicz B., Krystufek P.J.H., Reijnder F. Spitzenberger M., Stubbe J.B.M., Thissen V., Vohralik V., Zima J., 1999. ***Rattus norvegicus*. In The Atlas of European Mammals** T & AD Poyser Natural History Academic Press. pp. 278–279.

9. Angly F., Rodriguez-Brito B., Bangor D., McNairnie P., Breitbart M., Salamon P., Felts B., Nulton J., Mahaffy J., Rohwer F., 2005. **PHACCS, an online tool for estimating the structure and diversity of uncultured viral communities using metagenomic information**. BMC bioinformatics 6, 41.

10. Anthony S.J., Epstein J.H., Murray K.A., Navarrete-Macias I., Zambrana-Torrelio C.M., Solovyov A., Ojeda-Flores R., Arrigo N.C., Islam A., Ali Khan S., Hosseini P., Bogich T.L., Olival K.J., Sanchez-Leon M.D., Karesh W.B., Goldstein T., Luby S.P., Morse S.S., Mazet J.A., Daszak P., Lipkin W.I., 2013. **A strategy to estimate unknown viral diversity in mammals**. MBio 4, e00598-00513.

11. Atmar R.L., 2010. **Noroviruses - State of the Art**. Food Environ Virol 2, 117-126.

12. Bailey M., Haverson K., Inman C., Harris C., Jones P., Corfield G., Miller B., Stokes C., 2005. **The development of the mucosal immune system pre-**

and post-weaning: balancing regulatory and effector function. Proc Nutr Soc 64, 451-457.

13. Baker D.G., 1998. **Natural pathogens of laboratory mice, rats, and rabbits and their effects on research**. Clin Microbiol Rev 11, 231-266.

14. Baker K.S., Leggett R.M., Bexfield N.H., Alston M., Daly G., Todd S., Tachedjian M., Holmes C.E., Crameri S., Wang L.F., Heeney J.L., Suu-Ire R., Kellam P., Cunningham A.A., Wood J.L., Caccamo M., Murcia P.R., 2013. **Metagenomic study of the viruses of African straw-coloured fruit bats: detection of a chiropteran poxvirus and isolation of a novel adenovirus**. Virology 441, 95-106.

15. Belak S., Karlsson O.E., Blomstrom A.L., Berg M., Granberg F., 2013a. **New viruses in veterinary medicine, detected by metagenomic approaches**. Vet Microbiol 165, 95-101.

16. S., Karlsson O.E., Leijon M., Granberg F., 2013b. **High-throughput sequencing in veterinary infection biology and diagnostics**. OIE Rev. Sci. Tech. 32, 893-915.

17. Bentley D.R., Balasubramanian S., Swerdlow H.P., Smith G.P., Milton J. and more, last author: Smith A.J., 2008. **Accurate whole human genome sequencing using reversible terminator chemistry**. Nature 456, 53-59.

18. Berkhout B., 2015. **With a Little Help from my Enteric Microbial Friends**. Front Med (Lausanne) 2, 1-4

19. Black L.W., Thomas J.A., 2012. **Condensed genome structure**. Adv Exp Med Biol 726, 469-487.

20. Bank-Wolf B.R., König M., Thiel H.J., 2010. **Zoonotic aspects of infections with noroviruses and sapoviruses**. Vet Microbiol 27, 204-12.

21. Blaser M.J., Valentine F.T., 2008. **Viral commensalism in humans?** J Infect Dis 198, 1-3.

22. Blinkova O., Victoria J., Li Y., Keele B.F., Sanz C., Ndjango J.B., Peeters M., Travis D., Lonsdorf E.V., Wilson M.L., Pusey A.E., Hahn B.H., Delwart E.L., 2010. **Novel circular DNA viruses in stool samples of wild-living chimpanzees**. J Gen Virol 91, 74-86.

23. BMEL, 2015. **Statistical data of the BMEL**. https://www.bmel.de/EN/Services/Statistics/_Texte/Statistik.html. Access October 2015.

24. Bos L., 1999. **Beijerinck's work on tobacco mosaic virus: historical context and legacy**. Philos Trans R Soc Lond B Biol Sci 354, 675-685.

25. 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 USA 99, 14250-14255.

26. 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, 6220-6223.

27. Breitbart M., Felts B., Kelley S., Mahaffy J.M., Nulton J., Salamon P., Rohwer F., 2004. **Diversity and population structure of a near-shore marine-sediment viral community**. Proc Biol Sci 271, 565-574.

28. Breitbart M., Rohwer F., 2005a. **Here a virus, there a virus, everywhere the same virus?** Trends Microbiol 13, 278-284.

29. Breitbart M., Rohwer F., 2005b. **Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing**. BioTechniques 39, 729-736.

30. Breitbart M., Haynes M., Kelley S., Angly F., Edwards R.A., Felts B., Mahaffy J.M., Mueller J., Nulton J., Rayhawk S., Rodriguez-Brito B., Salamon P., Rohwer F., 2008. **Viral diversity and dynamics in an infant gut**. Res Microbiol 159, 367-373.

31. Brugere-Picoux J., Tessier P., 2010. **Viral gastroenteritis in domestic animals and zoonoses**. Bull Acad Natl Med 194, 1439-1449.

32. Bzhalava D., Dillner J., 2013. **Bioinformatics for Viral Metagenomics**. J Data Mining Genomics Proteomics 4. 1000134

33. Cahais V., Gayral P., Tsagkogeorga G., Melo-Ferreira J., Ballenghien M., Weinert L., Chiari Y., Belkhir K., Ranwez V., Galtier N., 2012. **Reference-free transcriptome assembly in non-model animals from next-generation sequencing data**. Mol Ecol Resour 12, 834-845.

34. Campe H., Zimmermann P., Glos K., Bayer M., Bergemann H., Dreweck C., Graf P., Weber B.K., Meyer H., Buttner M., Busch U., Sing A., 2009. **Cowpox virus transmission from pet rats to humans, Germany**. Emerg Infect Dis 15, 777-780.

35. Canchaya C., Fournous G., Chibani-Chennoufi S., Dillmann M.L., Brussow H., 2003a. **Phage as agents of lateral gene transfer**. Curr Opin Microbiol 6, 417-424.

36. Canchaya C., Proux C., Fournous G., Bruttin A., Brussow H., 2003b. **Prophage genomics**. Microbiol Mol Biol Rev 67, 238-276.

37. Cann A.J., Fandrich S.E., Heaphy S., 2005. **Analysis of the virus population present in equine faeces indicates the presence of hundreds of uncharacterized virus genomes.** Virus genes 30, 151-156.

38. Carstens E.B., Editor: King A.M.Q., Adams M.J.,Carstens E.B., Lefkowitz E.J., 2012. **Introduction to virus taxonomy in 9th report of the ICTV. Virus taxonomy: classification and nomenclature of viruses**. Elsevier Academic Press, San Diego. Part I, p 3-20.

39. Casas V., Rohwer F., 2007. **Phage metagenomics**. Methods Enzymol 421, 259-268.

40. Casjens S.R., 2005. **Comparative genomics and evolution of the tailed-bacteriophages**. Curr Opin Microbiol 8, 451-458.

41. Chandler J.D., Riddell M.A., Li F., Love R.J., Anderson D.A., 1999. Serological evidence for swine hepatitis E virus infection in Australian pig herds. Vet Microbiol 68, 95-105.

42. Chen K., Pachter L., 2005. **Bioinformatics for whole-genome shotgun sequencing of microbial communities**. PLoS Comput Biol 1, 106-112.

43. Chen J., Chen F., Zhou Q., Li W., Song Y., Pan Y., Zhang X., Xue C., Bi Y., Cao Y., 2012. **Complete genome sequence of a novel porcine Sapelovirus strain YC2011 isolated from piglets with diarrhea**. J Virol 86, 10898.

44. Cheung A.K., Ng T.F., Lager K.M., Bayles D.O., Alt D.P., Delwart E.L., Pogranichniy R.M., Kehrli M.E., Jr., 2013. **A divergent clade of circular single-stranded DNA viruses from pig feces**. Arch Virol 158, 2157-2162.

45. Cheung A.K., Ng T.F., Lager K.M., Alt D.P., Delwart E.L., Pogranichniy R.M., 2014. **Unique circovirus-like genome detected in pig feces**. Genome Announc 2, e00251-14.

46. Chiu C.Y., 2013. **Viral pathogen discovery**. Curr Opin Microbiol 16, 468-478.

47. Chow B.D., Ou Z., Esper F.P., 2010. **Newly recognized bocaviruses (HBoV, HBoV2) in children and adults with gastrointestinal illness in the United States**. J Clin Virol 47, 143-147.

48. Christensen M.L., 1989. **Human viral gastroenteritis**. Clin Microbiol Rev 2, 51-89.

49. Christou L., 2011. **The global burden of bacterial and viral zoonotic infections**. Clin Microbiol Infect 17, 326-330.

50. Chu D.K., Chin A.W., Smith G.J., Chan K.H., Guan Y., Peiris J.S., Poon L.L., 2010. **Detection of novel astroviruses in urban brown rats and previously known astroviruses in humans**. J Gen Virol 91, 2457-2462.

51. Ciarlet M., Conner M.E., Finegold M.J., Estes M.K., 2002. **Group A rotavirus infection and age-dependent diarrheal disease in rats: a new animal model to study the pathophysiology of rotavirus infection**. J Virol 76, 41-57.

52. Clokie M.R., Millard A.D., Letarov A.V., Heaphy S., 2011. **Phages in nature**. Bacteriophage 1, 31-45.

53. Coetzee B., Freeborough M.J., Maree H.J., Celton J.M., Rees D.J., Burger J.T., 2010. **Deep sequencing analysis of viruses infecting grapevines: Virome of a vineyard. Virology** 10, 157-63.

54. Colbere-Garapin F., Martin-Latil S., Blondel B., Mousson L., Pelletier I., Autret A., Francois A., Niborski V., Grompone G., Catonnet G., van de Moer

A., 2007. **Prevention and treatment of enteric viral infections: possible benefits of probiotic bacteria**. Microbes Infect 9, 1623-1631.

55. Cole J.R., Myrold D.D, Nakatsu C.H., Owens, P.R., Kowalchuk G., Tebbe C., Tiedje J.M., 2010. **Development of Soil Metadata Standards for International DNA Sequence. 19th** World Congress of Soil Science. Published on DVD.

56. Conroy M.E., Shi H.N., Walker W.A., 2009. **The long-term health effects of neonatal microbial flora**. Curr Opin Allergy Clin Immunol 9, 197-201.

57. Cooper V.L., 2000. **Diagnosis of neonatal pig diarrhea**. Vet Clin North Am Food Anim Pract 16, 117-133.

58. Crawford D., 2011. **Viruses: A Very Short Introduction**. 1st edition. Oxford University Press, New York. Chapter 1, p 1-16.

59. Daber R., Sukhadia S., Morrissette J.J., 2013. **Understanding the limitations of next generation sequencing informatics, an approach to clinical pipeline validation using artificial data sets**. Cancer Genet 206, 441-448.

60. Day J.M., Ballard L.L., Duke M.V., Scheffler B.E., Zsak L., 2010. **Metagenomic analysis of the turkey gut RNA virus community.** Virol J 12, 7:313.

61. De Benedictis P., Schultz-Cherry S., Burnham A., Cattoli G., 2011. **Astrovirus infections in humans and animals - molecular biology, genetic diversity, and interspecies transmissions**. Infect Genet Evol 11, 1529-1544.

62. DeLong E.F., 2002. **Microbial population genomics and ecology**. Curr Opin Microbiol 5, 520-524.

63. Delwart E.L., 2007. **Viral metagenomics**. Rev Med Virol 17, 115-131.

64. Desnues C., Rodriguez-Brito B., Rayhawk S., Kelley S., Tran T., Haynes M., Liu H., Furlan M., Wegley L., Chau B., Ruan Y., Hall D., Angly F.E., Edwards R.A., Li L., Thurber R.V., Reid R.P., Siefert J., Souza V., Valentine D.L., Swan B.K., Breitbart M., Rohwer F., 2008. **Biodiversity and biogeography of phages in modern stromatolites and thrombolites**. Nature 452, 340-343.

65. Desselberger U., 2014. **Rotaviruses**. Virus Res. 190, 75-96.

66. Deutz A., Fuchs K., Nowotny N., Auer H., Schuller W., Stunzner D., Aspock H., Kerbl U., Kofer J., 2003. **Sero-epidemiological studies of zoonotic infections in hunters-comparative analysis with veterinarians, farmers, and abattoir workers**. Wien Klin Wochenschr 115 Suppl 3, 61-67.

67. Diaz-Sanchez S., Hanning I., Pendleton S., D'Souza D., 2013. **Next-generation sequencing: the future of molecular genetics in poultry production and food safety**. Poult Sci 92, 562-572.

68. Djikeng A., Kuzmickas R., Anderson N.G., Spiro D.J., 2009. **Metagenomic analysis of RNA viruses in a fresh water lake**. PloS one 4, e7264.

69. Donaldson E.F., Haskew A.N., Gates J.E., Huynh J., Moore C.J., Frieman M.B., 2010. **Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat**. J Virol 84, 13004-13018.

70. Dreier J., Stormer M., Kleesiek K., 2005. **Use of bacteriophage MS2 as an internal control in viral reverse transcription-PCR assays**. J Clin Microbiol 43, 4551-4557.

71. Dremsek P., Wenzel J.J., Johne R., Ziller M., Hofmann J., Groschup M.H., Werdermann S., Mohn U., Dorn S., Motz M., Mertens M., Jilg W., Ulrich R.G., 2012. **Seroprevalence study in forestry workers from eastern Germany using novel genotype 3- and rat hepatitis E virus-specific immunoglobulin G ELISAs**. Med Microbiol Immunol 201, 189-200.

72. Drexler J.F., Corman V.M., Muller M.A., Maganga G.D., Vallo P., Binger T., Gloza-Rausch F., Rasche A., Yordanov S., Seebens A., Oppong S., Adu Sarkodie Y., Pongombo C., Lukashev A.N., Schmidt-Chanasit J., Stocker A., Carneiro A.J., Erbar S., Maisner A., Fronhoffs F., Buettner R., Kalko E.K., Kruppa T., Franke C.R., Kallies R., Yandoko E.R., Herrler G., Reusken C., Hassanin A., Kruger D.H., Matthee S., Ulrich R.G., Leroy E.M., Drosten C., 2012. **Bats host major mammalian paramyxoviruses**. Nat Commun 3, 796.

73. Dufkova L., Scigalkova I., Moutelikova R., Malenovska H., Prodelalova J., 2013. **Genetic diversity of porcine sapoviruses, kobuviruses, and astroviruses in asymptomatic pigs: an emerging new sapovirus GIII genotype**. Arch Virol 158, 549-558.

74. Easterbrook J.D., Kaplan J.B., Glass G.E., Watson J., Klein S.L., 2008. **A survey of rodent-borne pathogens carried by wild-caught Norway rats: a potential threat to laboratory rodent colonies**. Lab Anim 42, 92-98.

75. Eckardt A.J., Baumgart D.C., 2011. **Viral gastroenteritis in adults**. Recent Pat Anti-infect Drug Discov 6, 54-63.

76. Edwards R.A., Rohwer F., 2005. **Viral metagenomics**. Nat Rev Microbiol 3, 504-510.

77. Ehlers B., Kuchler J., Yasmum N., Dural G., Voigt S., Schmidt-Chanasit J., Jakel T., Matuschka F.R., Richter D., Essbauer S., Hughes D.J., Summers C., Bennett M., Stewart J.P., Ulrich R.G., 2007. **Identification of novel rodent herpesviruses, including the first gammaherpesvirus of Mus musculus**. J Virol 81, 8091-8100.

78. Ellis J., 2014. **Porcine circovirus: a historical perspective**. Vet Pathol 51, 315-327.

79. Engelhardt T., Kallmeyer J., Cypionka H., Engelen B., 2014. **High virus-to-cell ratios indicate ongoing production of viruses in deep subsurface sediments**. ISME J 8, 1503-9

80. Fancello L., Raoult D., Desnues C., 2012. **Computational tools for viral metagenomics and their application in clinical research**. Virology 434, 162-174.

81. Fang S.B., Lee H.C., Hu J.J., Hou S.Y., Liu H.L., Fang H.W., 2009. **Dose-dependent effect of Lactobacillus rhamnosus on quantitative reduction of faecal rotavirus shedding in children**. J Trop Pediatr 55, 297-301.

82. FAO and WHO. 2001. **Health and nutrition properties of probiotics in food including powder milk with live lactic acid bacteria. Guidelines for the evaluation of probioticsin food**. FAO Food and Nutrition Paper 85. ISSN 0254-4725.

83. FAO and WHO, 2008. **Viruses in food: Scientific advice to support risk managment activities: meeting report**. Microbial Risk Assessment Series 13. ISSN1726-5274.

84. Faria N.R., Rambaut A., Suchard M.A., Baele G., Bedford T., Ward M.J., Tatem A.J., Sousa J.D., Arinaminpathy N., Pepin J., Posada D., Peeters M., Pybus O.G., Lemey P., 2014. **HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations**. Science 346, 56-61.

85. Fehér E., Pazar P., Kovacs E., Farkas S.L., Lengyel G., Jakab F., Martella V., Banyai K., 2014. **Molecular detection and characterization of human gyroviruses identified in the ferret fecal virome**. Arch Virol 159, 3401-3406.

86. Finkbeiner S.R., Allred A.F., Tarr P.I., Klein E.J., Kirkwood C.D., Wang D., 2008. **Metagenomic analysis of human diarrhea: viral detection and discovery**. PLoS pathogens 4, e1000011.

87. Firth C., Bhat M., Firth M.A., Williams S.H., Frye M.J., Simmonds P., Conte J.M., Ng J., Garcia J., Bhuva N.P., Lee B., Che X., Quan P.L., Lipkin W.I., 2014. **Detection of Zoonotic Pathogens and Characterization of Novel Viruses Carried by Commensal Rattus norvegicus in New York City**. MBio 5, e01933-01914.

88. Flint S.J., Racaniello V.R., Rall G.F., Skalka A.M., 2015. **Discovery of** Viruses in "**Principles of Virology, Volume 1: Molecular Biology"**. 4th Edition. ASM Press, Washington, DC. Part 1, p 10-12

89. Foca A., Liberto M.C., Quirino A., Marascio N., Zicca E., Pavia G., 2015. **Gut inflammation and immunity: what is the role of the human gut virome?** Mediators Inflamm, e:326032.

90. Fontana L., Bermudez-Brito M., Plaza-Diaz J., Munoz-Quezada S., Gil A., 2013. **Sources, isolation, characterisation and evaluation of probiotics**. Br J Nutr 109 Suppl 2, S35-50.

91. Fregolente M.C., de Castro-Dias E., Martins S.S., Spilki F.R., Allegretti S.M., Gatti M.S., 2009. **Molecular characterization of picobirnaviruses from new hosts**. Virus Res 143, 134-136.

92. Freitas M., Tavan E., Cayuela C., Diop L., Sapin C., Trugnan G., 2003. **Host-pathogens cross-talk. Indigenous bacteria and probiotics also play the game**. Biol Cell 95, 503-506.

93. Ganesh B., Masachessi G., Mladenova Z., 2014. **Animal picobirnavirus**. Virusdisease 25, 223-238.

94. Gangarapu V., Yildiz K., Ince A.T., Baysal B., 2014. **Role of gut microbiota: obesity and NAFLD**. Turk J Gastroenterol 25, 133-140.

95. Garmendia A.E., Lopez W., Ortega N., Chamorro M.J., 2015. **Molecular characterization of rotavirus isolated from alpaca (Vicugna pacos) crias with diarrhea in the Andean Region of Cusco, Peru**. Vet Microbiol 180, 109-112.

96. Garneau J.E., Tremblay D.M., Moineau S., 2008. **Characterization of 1706, a virulent phage from Lactococcus lactis with similarities to prophages from other Firmicutes**. Virology 373, 298-309.

97. Gitig D., 2010. **Marine metagenomics**. BioTechniques 48, 361-365.

98. Glass R.I., Bresee J., Jiang B., Gentsch J., Ando T., Fankhauser R., Noel J., Parashar U., Rosen B., Monroe S.S., 2001. **Gastroenteritis viruses: an overview**. Novartis Foundation symposium 238, 5-19; discussion 19-25.

99. Gomez-Alvarez V., Teal T.K., Schmidt T.M., 2009. **Systematic artifacts in metagenomes from complex microbial communities**. ISME J 3, 1314-1317.

100. Goodgame R.W., 1999. **Viral infections of the gastrointestinal tract**. Curr Gastroenterol Rep 1, 292-300.

101. Grada A., Weinbrecht K., 2013. **Next-generation sequencing: methodology and application.** J Invest Dermatol 133 (8):e11.

102. Gregory T.R., Nicol J.A., Tamm H., Kullman B., Kullman K., Leitch I.J., Murray B.G., Kapraun D.F., Greilhuber J., Bennett M.D., 2007. **Eukaryotic genome size databases**. Nucleic Acids Res 35, 332-338.

103. GSP, 2015. Data from the National Human Genome Research Institute (NHGRI). http://www.genome.gov/sequencingcosts. Access April 2015.

104. Guo M., Evermann J.F., Saif L.J., 2001. **Detection and molecular characterization of cultivable caliciviruses from clinically normal mink and enteric caliciviruses associated with diarrhea in mink**. Arch Virol 146, 479-493.

105. Gut I.G., 2013. **New sequencing technologies.** Clin Transl Oncol15, 879-881.

106. Halaihel N., Masia R.M., Fernandez-Jimenez M., Ribes J.M., Montava R., De Blas I., Girones O., Alonso J.L., Buesa J., 2010. **Enteric calicivirus and rotavirus infections in domestic pigs**. Epidemiol Infect 138, 542-548.

107. Hall R.J., Wang J., Todd A.K., Bissielo A.B., Yen S., Strydom H., Moore N.E., Ren X., Huang Q.S., Carter P.E., Peacey M., 2014. **Evaluation of rapid and simple techniques for the enrichment of viruses prior to metagenomic virus discovery**. J Virol Methods 195, 194-204.

108. Hansman G.S., Oka T., Katayama K., Takeda N., 2007. **Human sapoviruses: genetic diversity, recombination, and classification**. Rev Med Virol 17, 133-141.

109. Hauffe H.C., Niklasson B., Olsson T., Bianchi A., Rizzoli A., Klitz W., 2010. **Ljungan virus detected in bank voles (Myodes glareolus) and yellow-necked mice (Apodemus flavicollis) from Northern Italy**. J Wildl Dis 46, 262-266.

110. He B., Yang F., Yang W., Zhang Y., Feng Y., Zhou J., Xie J., Feng Y., Bao X., Guo H., Li Y., Xia L., Li N., Matthijnssens J., Zhang H., Tu C., 2013. **Characterization of a novel G3P[3] rotavirus isolated from a lesser horseshoe bat: a distant relative of feline/canine rotaviruses**. J Virol 87, 12357-12366.

111. Himsworth C.G., Parsons K.L., Jardine C., Patrick D.M., 2013. **Rats, cities, people, and pathogens: a systematic review and narrative synthesis of literature regarding the ecology of rat-associated zoonoses in urban centers**. Vector Borne Zoonotic Dis 13, 349-359.

112. Hoffmann B., Scheuch M., Hoper D., Jungblut R., Holsteg M., Schirrmeier H., Eschbaumer M., Goller K.V., Wernike K., Fischer M., Breithaupt A., Mettenleiter T.C., Beer M., 2012. **Novel orthobunyavirus in Cattle, Europe, 2011**. Emerg Infect Dis 18, 469-472.

113. Holtz L.R., Cao S., Zhao G., Bauer I.K., Denno D.M., Klein E.J., Antonio M., Stine O.C., Snelling T.L., Kirkwood C.D., Wang D., 2014. **Geographic variation in the eukaryotic virome of human diarrhea**. Virology 468-470, 556-564.

114. Holyoake P.K., Dial G.D., Trigg T., King V.L., 1995. **Reducing pig mortality through supervision during the perinatal period**. J Anim Sci 73, 3543-3551.

115. 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, 1883-1886.

116. Hosono S., Faruqi A.F., Dean F.B., Du Y., Sun Z., Wu X., Du J., Kingsmore S.F., Egholm M., Lasken R.S., 2003. **Unbiased whole-genome amplification directly from clinical samples**. Genome Res 13, 954-964.

117. Huang X.H.C., Quesada M.A., Mathies R.A., 1992. **DNA Sequencing Using Capillary Array Electrophoresis**. Anal Chem 64, 2149-2154.

118. Huang X.Z., Zhu L.B., Li Z.R., Lin J., 2013. **Bacterial colonization and intestinal mucosal barrier development**. World J Clin Pediatr 2, 46-53.

119. Huang J., Mor S.K., Erber J., Voss E., Goyal S.M., 2014. **Detection and characterization of porcine bocavirus in the United States**. Arch Virol 159, 1797-1801.

120. Huson D.H., Reinert K., Kravitz S.A., Remington K.A., Delcher A.L., Dew I.M., Flanigan M., Halpern A.L., Lai Z., Mobarry C.M., Sutton G.G., Myers E.W., 2001. **Design of a compartmentalized shotgun assembler for the human genome**. Bioinform.;17 Suppl 1:S132-9.

121. Harrison S.C., 2007. **History of virology in: "Fields virology"**. 5th edition. Editor: Field B., Knipe D.M., Howley P.M., Philadelphia : Wolters Kluwer Health/Lippincott Williams & Wilkins. Chapter 3, p 59-98.

122. Hur K.Y., Lee M.S., 2015. **Gut Microbiota and Metabolic Disorders**. Diabetes Metab J 39, 198-203.

123. ICTV, 2014. **Virus Taxonomy.** http://www.ictvonline.org/virusTaxonomy.asp. Access October 2015.

124. Imagawa H., Sekiguchi K., Anzai T., Fukunaga Y., Kanemaru T., Ohishi H., Higuchi T., Kamada M., 1991. **Epidemiology of equine rotavirus infection among foals in the breeding region**. J Vet Med Sci 53, 1079-1080.

125. Imani-Fooladi A.A., Mahmoodzadeh Hosseini H., Nourani M.R., Khani S., Alavian S.M., 2013. **Probiotic as a novel treatment strategy against liver disease**. Hepat Mon 13, e7521.

126. Jacoby R.O., Ball-Goodrich L.J., Besselsen D.G., McKisic M.D., Riley L.K., Smith A.L., 1996. **Rodent parvovirus infections**. Lab Anim Sci 46, 370-380.

127. Jacoby R.O., Bhatt P.N., Gaertner D.J., Smith A.L., Johnson E.A., 1987. **The pathogenesis of rat virus infection in infant and juvenile rats after oronasal inoculation**. Arch Virol 95, 251-270.

128. Ji Y., Shi Y., Ding G., Li Y., 2011. **A new strategy for better genome assembly from very short reads**. BMC bioinformatics 12, 493.

129. 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, 205-211.

130. Johne R., Plenge-Bonig A., Hess M., Ulrich R.G., Reetz J., Schielke A., 2010. **Detection of a novel hepatitis E-like virus in faeces of wild rats using a nested broad-spectrum RT-PCR**. J Gen Virol 91, 750-758.

131. Johne R., Dremsek P., Kindler E., Schielke A., Plenge-Bonig A., Gregersen H., Wessels U., Schmidt K., Rietschel W., Groschup M.H., Guenther S., Heckel G., Ulrich R.G., 2012. **Rat hepatitis E virus: geographical clustering within Germany and serological detection in wild Norway rats (Rattus norvegicus)**. Infect Genet Evol 12, 947-956.

132. Jorgensen T.S., Xu Z., Hansen M.A., Sorensen S.J., Hansen L.H., 2014. **Hundreds of circular novel plasmids and DNA elements identified in a rat cecum metamobilome**. PloS one 9, e87924.

133. Kaku Y., Sarai A., Murakami Y., 2001. **Genetic reclassification of porcine enteroviruses**. J Gen Virol 82, 417-424.

134. 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, 1075-1081.

135. Kapoor A., Li L., Victoria J., Oderinde B., Mason C., Pandey P., Zaidi S.Z., Delwart E., 2009. **Multiple novel astrovirus species in human stool**. J Gen Virol 90, 2965-2972.

136. Kapoor A., Simmonds P., Lipkin W.I., Zaidi S., Delwart E., 2010. **Use of nucleotide composition analysis to infer hosts for three novel picorna-like viruses**. J Virol 84, 10322-10328.

137. Karlsson O.E., Hansen T., Knutsson R., Lofstrom C., Granberg F., Berg M., 2013. **Metagenomic detection methods in biopreparedness outbreak scenarios**. Biosecur Bioterror 11, 146-157.

138. Karlsson E., Larkeryd A., Sjodin A., Forsman M., Stenberg P., 2015. **Scaffolding of a bacterial genome using MinION nanopore sequencing**. Sci Rep 5, 11996.

139. Kato K., Ishiwa A., 2015. **The role of carbohydrates in infection strategies of enteric pathogens**. Trop Med Health 43, 41-52.

140. Kim Y., Chang K.O., Straw B., Saif L.J., 1999 **Characterization of group C rotaviruses associated with diarrhea outbreaks in feeder pigs**. J Clin Microbiol. 37, 1484-8.

141. Kim K.H., Chang H.W., Nam Y.D., Roh S.W., Kim M.S., Sung Y., Jeon C.O., Oh H.M., Bae J.W., 2008. **Amplification of uncultured single-stranded DNA viruses from rice paddy soil**. Appl Environ Microbiol74, 5975-5985.

142. Kim Y.S., Ho S.B., 2010. **Intestinal goblet cells and mucins in health and disease: recent insights and progress**. Curr Gastroenterol Rep 12, 319-330.

143. Kim M.S., Park E.J., Roh S.W., Bae J.W., 2011. **Diversity and abundance of single-stranded DNA viruses in human feces**. Appl Environ Microbiol77, 8062-8070.

144. Kim H.K., Park S.J., Nguyen V.G., Song D.S., Moon H.J., Kang B.K., Park B.K., 2012. **Identification of a novel single-stranded, circular DNA virus from bovine stool**. J Gen Virol 93, 635-639.

145. Kim A.R., Chung H.C., Kim H.K., Kim E.O., Nguyen V.G., Choi M.G., Yang H.J., Kim J.A., Park B.K., 2014. **Characterization of a complete genome of a circular single-stranded DNA virus from porcine stools in Korea**. Virus genes 48, 81-88.

146. King A.M.Q., Adams M.J., Carstens E.B., Lefkowitz E.J., 2012a. **9th report of the ICTV. Virus taxonomy: classification and nomenclature of viruses**. Elsevier Academic Press, San Diego. Part II, p 1014-1020.

147. King A.M.Q., Adams M.J., Carstens E.B., Lefkowitz E.J., 2012b. **9th report of the ICTV. Virus taxonomy: classification and nomenclature of viruses**. Elsevier Academic Press, San Diego. Part I, p 3-20.

148. Klempa B., Kruger D.H., Auste B., Stanko M., Krawczyk A., Nickel K.F., Uberla K., Stang A., 2009. **A novel cardiotropic murine adenovirus representing a distinct species of mastadenoviruses**. J Virol 83, 5749-5759.

149. Koenen F., Vanderhallen H., Castryck F., Miry C., 1999. **Epidemiologic, pathogenic and molecular analysis of recent encephalomyocarditis outbreaks in Belgium**. Zentralbl Veterinarmed B 46, 217-231.

150. Kohl C., Kurth A., 2014. **European bats as carriers of viruses with zoonotic potential**. Viruses 6, 3110-3128.

151. Koonin E.V., Senkevich T.G., Dolja V.V., 2006. **The ancient Virus World and evolution of cells**. Biol Direct 1, 1-27.

152. Kotwal G., Cannon J.L., 2014. **Environmental persistence and transfer of enteric viruses**. Curr Opin Virol 4, 37-43.

153. Kotzampassi K., Giamarellos-Bourboulis E.J., 2012. **Probiotics for infectious diseases: more drugs, less dietary supplementation**. Int J Antimicrob Agents 40, 288-296.

154. Kreuzer S., Machnowska P., Assmus J., Sieber M., Pieper R., Schmidt M.F., Brockmann G.A., Scharek-Tedin L., Johne R., 2012. **Feeding of the probiotic bacterium Enterococcus faecium NCIMB 10415 differentially affects shedding of enteric viruses in pigs**. Vet Res 43, 58.

155. Kristensen D.M., Mushegian A.R., Dolja V.V., Koonin E.V., 2010. **New dimensions of the virus world discovered through metagenomics**. Trends Microbiol 18, 11-19.

156. Kritas S.K., Morrison R.B., 2007. **Effect of orally administered Lactobacillus casei on porcine reproductive and respiratory syndrome (PRRS) virus vaccination in pigs**. Vet Microbiol 119, 248-255.

157. Kruse H., kirkemo A.M., Handeland K., 2004. **Wildlife as source of zoonotic infections**. Emerg Infect Dis 10, 2067-2072.

158. Kumar R., Seo B.J., Mun M.R., Kim C.J., Lee I., Kim H., Park Y.H., 2010. **Putative probiotic Lactobacillus spp. from porcine gastrointestinal tract inhibit transmissible gastroenteritis coronavirus and enteric bacterial pathogens**. Trop Anim Health Prod 42, 1855-1860.

159. Kunin V., Copeland A., Lapidus A., Mavromatis K., Hugenholtz P., 2008. **A bioinformatician's guide to metagenomics**. Kunin 72, 557-578.

160. Kuo C.H., Moran N.A., Ochman H., 2009. **The consequences of genetic drift for bacterial genome complexity**. Genome Res 19, 1450-1454.

161. Labonte J.M., Suttle C.A., 2013. **Previously unknown and highly divergent ssDNA viruses populate the oceans**. ISME J 7, 2169-2177.

162. Lager K.M., Ng T.F., Bayles D.O., Alt D.P., Delwart E.L., Cheung A.K., 2012. **Diversity of viruses detected by deep sequencing in pigs from a common background**. J Vet Diagn Invest, Inc 24, 1177-1179.

163. Lane D.J., Pace B., Olsen G.J., Stahl D.A., Sogin M.L., Pace N.R., 1985. **Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses**. Proceedings of the National Academy of Sciences of the United States of America 82, 6955-6959.

164. Langmead B., Salzberg S.L., 2012. **Fast gapped-read alignment with Bowtie 2**. Nat methods 9, 357-359.

165. Leblanc D., Houde A., Gagne M.J., Plante D., Bellon-Gagnon P., Jones T.H., Muehlhauser V., Wilhelm B., Avery B., Janecko N., Brassard J., 2014. **Presence, viral load and characterization of Torque teno sus viruses in liver and pork chop samples at retail**. Int J Food Microbiol 178, 60-64.

166. Lee R.M., Lessler J., Lee R.A., Rudolph K.E., Reich N.G., Perl T.M., Cummings D.A., 2013. **Incubation periods of viral gastroenteritis: a systematic review**. BMC Infect Dis 13, 446.

167. Legendre M., Bartoli J., Shmakova L., Jeudy S., Labadie K., Adrait A., Lescot M., Poirot O., Bertaux L., Bruley C., Coute Y., Rivkina E., Abergel C., Claverie J.M., 2014. **Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology**. Proc Natl Acad Sci USA 111, 4274-4279.

168. Levine A.J., Enquist L.W., 2007. **History of virology in: "Fields virology"**. 5th edition. Editor: Field B., Knipe D.M., Howley P.M., Philadelphia : Wolters Kluwer Health/Lippincott Williams & Wilkins. Chapter 1, p 3-25.

169. Li L., Victoria J.G., Wang C., Jones M., Fellers G.M., Kunz T.H., Delwart E., 2010a. **Bat guano virome: predominance of dietary viruses**

from insects and plants plus novel mammalian viruses. J Virol 84, 6955-6965.

170. Li H., Homer N., 2010b. **A survey of sequence alignment algorithms for next-generation sequencing**. Brief Bioinform 11, 473-483.

171. Li L., Pesavento P.A., Shan T., Leutenegger C.M., Wang C., Delwart E., 2011a. **Viruses in diarrhoeic dogs include novel kobuviruses and sapoviruses**. J Gen Virol 92, 2534-2541.

172. Li L., Shan T., Wang C., Cote C., Kolman J., Onions D., Gulland F.M., Delwart E., 2011b. **The fecal viral flora of California sea lions**. J Virol 85, 9909-9917.

173. Li L., Delwart E., 2011c. **From orphan virus to pathogen: the path to the clinical lab**. Curr Opin Virol 1, 282-288.

174. Li L., Deng X., Linsuwanon P., Bangsberg D., Bwana M.B., Hunt P., Martin J.N., Deeks S.G., Delwart E., 2013. **AIDS alters the commensal plasma virome**. J Virol 87, 10912-10915.

175. Li L., Giannitti F., Low J., Keyes C., Ullmann L.S., Deng X., Aleman M., Pesavento P.A., Pusterla N., Delwart E., 2015. **Exploring the virome of diseased horses**. J Gen Virol 96, 2721-2733.

176. Lin Y., Li J., Shen H., Zhang L., Papasian C.J., Deng H.W., 2011. **Comparative studies of de novo assembly tools for next-generation sequencing technologies**. Bioinformatics 27, 2031-2037.

177. Liu L., Li Y., Li S., Hu N., He Y., Pong R., Lin D., Lu L., Law M., 2012. **Comparison of next-generation sequencing systems**. J Biomed Biotechnol 2012, 251364.

178. Lochrie M.A., Tatsuno G.P., Arbetman A.E., Jones K., Pater C., Smith P.H., McDonnell J.W., Zhou S.Z., Kachi S., Kachi M., Campochiaro P.A., Pierce G.F., Colosi P., 2006. **Adeno-associated virus (AAV) capsid genes isolated from rat and mouse liver genomic DNA define two new AAV species distantly related to AAV-5**. Virology 353, 68-82.

179. Loguercio C., Federico A., Tuccillo C., Terracciano F., D'Auria M.V., De Simone C., Del Vecchio Blanco C., 2005. **Beneficial effects of a probiotic VSL#3 on parameters of liver dysfunction in chronic liver diseases**. J Clin Gastroenterol 39, 540-543.

180. Loman N.J., Misra R.V., Dallman T.J., Constantinidou C., Gharbia S.E., Wain J., Pallen M.J., 2012. **Performance comparison of benchtop high-throughput sequencing platforms**. Nat Biotechnol 30, 434-439.

181. Lopez-Bueno A., Tamames J., Velazquez D., Moya A., Quesada A., Alcami A., 2009. **High diversity of the viral community from an Antarctic lake**. Science 326, 858-861.

182. Luckey J.A., Drossman H., Kostichka A.J., Mead D.A., D'Cunha J., Norris T.B., Smith L.M., 1990. **High speed DNA sequencing by capillary electrophoresis**. Nucleic Acids Res 18, 4417-4421.

183. Luo Z., Roi S., Dastor M., Gallice E., Laurin M.A., L'Homme Y., 2011. **Multiple novel and prevalent astroviruses in pigs**. Vet Microbiol 149, 316-323.

184. Lwoff A., 1957. **The concept of virus**. J Gen Microbiol 17, 239-253.

185. Machnowska P., Ellerbroek L., Johne R., 2014. **Detection and characterization of potentially zoonotic viruses in faeces of pigs at slaughter in Germany**. Vet Microbiol 168, 60-68.

186. Mäde D., Trubner K., Neubert E., Hohne M., Johne R., 2013. **Detection and Typing of Norovirus from Frozen Strawberries Involved in a Large-Scale Gastroenteritis Outbreak in Germany**. Food Environ Virol. 5, 162-168.

187. Maluquer de Motes C., Clemente-Casares P., Hundesa A., Martin M., Girones R., 2004. **Detection of bovine and porcine adenoviruses for tracing the source of fecal contamination**. Appl Environ Microbiol 70, 1448-1454.

188. Manteufel J., Truyen U., 2008. **Animal bocaviruses: a brief review**. Intervirology 51, 328-334.

189. Mardis E.R., 2011. **A decade's perspective on DNA sequencing technology**. Nature 470, 198-203.

190. 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, 376-380.

191. Martella V., Banyai K., Matthijnssens J., Buonavoglia C., Ciarlet M., 2010. **Zoonotic aspects of rotaviruses**. Vet Microbiol 140, 246-255.

192. Masembe C., Michuki G., Onyango M., Rumberia C., Norling M., Bishop R.P., Djikeng A., Kemp S.J., Orth A., Skilton R.A., Stahl K., Fischer A., 2012. **Viral metagenomics demonstrates that domestic pigs are a potential reservoir for Ndumu virus**. Virol J 9, 218.

193. Matthijnssens J., Rahman M., Martella V., Xuelei Y., De Vos S., De Leener K., Ciarlet M., Buonavoglia C., Van Ranst M., 2006. **Full genomic analysis of human rotavirus strain B4106 and lapine rotavirus strain**

30/96 provides evidence for interspecies transmission. J Virol 80, 3801-3810.

194. Maxam A., Gilbert W., 1977. **A new method for sequencing DNA**. Proc Natl Acad Sci USA. 74, 560-4.

195. McDaniel L.D., Rosario K., Breitbart M., Paul J.H., 2013. **Comparative metagenomics: natural populations of induced prophages demonstrate highly unique, lower diversity viral sequences**. Environ Microbiol 16, 570-85.

196. Medintz I.L., Paegel B.M., Blazej R.G., Emrich C.A., Berti L., Scherer J.R., Mathies R.A., 2001. **High-performance genetic analysis using microfabricated capillary array electrophoresis microplates**. Electrophoresis 22, 3845-56.

197. Meerburg B.G., Singleton G.R., Kijlstra A., 2009. **Rodent-borne diseases and their risks for public health**. Crit Rev Microbiol 35, 221-270.

198. Meng X.J., 2012. **Emerging and Re-emerging Swine Viruses**. Transbound Emerg Dis 59 Suppl 1, 85-102

199. Meng X.J., 2013. **Zoonotic and foodborne transmission of hepatitis E virus**. Semin Liver Dis 33, 41-49.

200. Miller J.R., Koren S., Sutton G., 2010. **Assembly algorithms for next-generation sequencing data**. Genomics 95, 315-327.

201. Minot S., Sinha R., Chen J., Li H., Keilbaugh S.A., Wu G.D., Lewis J.D., Bushman F.D., 2011. **The human gut virome: inter-individual variation and dynamic response to diet**. Genome Res 21, 1616-1625.

202. Minot S., Grunberg S., Wu G.D., Lewis J.D., Bushman F.D., 2012. **Hypervariable loci in the human gut virome**. Proceedings of the National Academy of Sciences of the United States of America 109, 3962-3966.

203. Minot S., Bryson A., Chehoud C., Wu G.D., Lewis J.D., Bushman F.D., 2013. **Rapid evolution of the human gut virome**. Proceedings of the National Academy of Sciences of the United States of America 110, 12450-12455.

204. Mitchell-Jones A.J., Bogdanowicz W., Krystufek B., Reijnders P.J.H., Spitzenberger F., Stubbe C., Thissen J.B.M., Vohralik V., Zima J., 1999. ***Rattus norvegius*. The Atlas of European Mammals**. London: Academic Press, T. & A. D. Poyser Natural History. pp278-279

205. Modi S.R., Lee H.H., Spina C.S., Collins J.J., 2013. **Antibiotic treatment expands the resistance reservoir and ecological network of the phage metagenome**. Nature 499, 219-222.

206. Mokili J.L., Rohwer F., Dutilh B.E., 2012. **Metagenomics and future perspectives in virus discovery**. Curr Opin Virol 2, 63-77.

207. Mor S.K., Chander Y., Marthaler D., Patnayak D.P., Goyal S.M., 2012. **Detection and molecular characterization of Porcine astrovirus strains associated with swine diarrhea**. J Vet Diagn Invest, Inc 24, 1064-1067.

208. Mullis K., Faloona F., Scharf S., Saiki R., Horn G., Erlich H., 1986. **Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction**. Cold Spring Harb Symp Quant Biol 51, 263-273.

209. Munoz J.A., Chenoll E., Casinos B., Bataller E., Ramon D., Genoves S., Montava R., Ribes J.M., Buesa J., Fabrega J., Rivero M., 2011. **Novel probiotic Bifidobacterium longum subsp. infantis CECT 7210 strain active against rotavirus infections**. Appl Environ Microbiol 77, 8775-8783.

210. Muns R., Silva C., Manteca X., Gasa J., 2014. **Effect of cross-fostering and oral supplementation with colostrums on performance of newborn piglets**. J Anim Sci 92, 1193-1199.

211. Nakabachi A., Yamashita A., Toh H., Ishikawa H., Dunbar H.E., Moran N.A., Hattori M., 2006. **The 160-kilobase genome of the bacterial endosymbiont Carsonella**. Science 314, 267.

212. Ng T.F., Willner D.L., Lim Y.W., Schmieder R., Chau B., Nilsson C., Anthony S., Ruan Y.J., Rohwer F., Breitbart M., 2011. **Broad Surveys of DNA Viral Diversity Obtained through Viral Metagenomics of Mosquitoes**. PloS one 6.

213. Ng T.F., Marine R., Wang C., Simmonds P., Kapusinszky B., Bodhidatta L., Oderinde B.S., Wommack K.E., Delwart E., 2012. **High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage**. J Virol 86, 12161-12175.

214. Ng T.F., Mesquita J.R., Nascimento M.S., Kondov N.O., Wong W., Reuter G., Knowles N.J., Vega E., Esona M.D., Deng X., Vinje J., Delwart E., 2014. **Feline fecal virome reveals novel and prevalent enteric viruses**. Vet Microbiol 171, 102-111.

215. Ng T.F., Sachsenroder J., Reuter G., Knowles N.J., Delwart E., Johne R., 2015a. **Rabovirus: a proposed new picornavirus genus that is phylogenetically basal to enteroviruses and sapeloviruses**. Arch Virol 160, 2569-2575.

216. Ng T.F., Zhang W., Sachsenröder J., Kondov N.O., da Costa A.C., Vega E., Holtz L.R., Wu G., Wang D., Stine C.O., Antonio M., Mulvaney U.S., Muench M.O., Deng X., Ambert-Balay K., Pothier P., Vinjé J., Delwart E., 2015b. **A diverse group of small circular ssDNA viral genomes in human and non-human primate stools**. Virus Evol 1, vev017.

217. Ng T.F., Zhenga W., Sachsenröder J., Kondova N.O., Costaa C., Everardo Vegad U.S.M., Marcus O. Muencha,b, Xutao Denga,b, Katia Ambert-Balaye, Pierre Pothierf, Jan Vinjég, Eric Delwart, 2016*. **Detection**

of small circular DNA viral genomes in unexplained gastroenteritis outbreaks. Virus Evol (*publication in revision).

218. Niklasson B., Samsioe A., Papadogiannakis N., Gustafsson S., Klitz W., 2009. **Zoonotic Ljungan virus associated with central nervous system malformations in terminated pregnancy**. Birth defects research. Part A, Clinical and molecular teratology 85, 542-545.

219. Ogilvie L.A., Bowler L.D., Caplin J., Dedi C., Diston D., Cheek E., Taylor H., Ebdon J.E., Jones B.V., 2013. **Genome signature-based dissection of human gut metagenomes to extract subliminal viral sequences**. Nat Commun 4, 2420.

220. Ogilvie L.A., Jones B.V., 2015. **The human gut virome: a multifaceted majority.** Front Microbiol 11; 6-918.

221. Otto P.H., Rosenhain S., Elschner M.C., Hotzel H., Machnowska P., Trojnar E., Hoffmann K., Johne R., 2015. **Detection of rotavirus species A, B and C in domestic mammalian animals with diarrhoea and genotyping of bovine species A rotavirus strains**. Vet Microbiol 179, 168-176.

222. Ozsolak F., 2012. **Third-generation sequencing techniques and applications to drug discovery**. Expert Opin Drug Discov 7, 231-243.

223. Papp H., Laszlo B., Jakab F., Ganesh B., De Grazia S., Matthijnssens J., Ciarlet M., Martella V., Banyai K., 2013. **Review of group A rotavirus strains reported in swine and cattle**. Vet Microbiol 165, 190-199.

224. Parshionikar S.U., Cashdollar J., Fout G.S., 2004. **Development of homologous viral internal controls for use in RT-PCR assays of waterborne enteric viruses**. J Virol Methods 121, 39-48.

225. Paul J.H., Sullivan M.B., 2005. **Marine phage genomics: what have we learned?** Curr Opin Biotechnol 16, 299-307.

226. Pfeiffer J.K., 2010. **Innate host barriers to viral trafficking and population diversity: lessons learned from poliovirus**. Adv Virus Res77, 85-118.

227. Phan T.G., Kapusinszky B., Wang C., Rose R.K., Lipton H.L., Delwart E.L., 2011. **The fecal viral flora of wild rodents**. PLoS Pathog 7, e1002218.

228. Phan T.G., Vo N.P., Boros A., Pankovics P., Reuter G., Li O.T., Wang C., Deng X., Poon L.L., Delwart E., 2013a. **The viruses of wild pigeon droppings**. PloS one 8, e72787.

229. Phan T.G., Vo N.P., Simmonds P., Samayoa E., Naccache S., Chiu C.Y., Delwart E., 2013b. **Rosavirus: the prototype of a proposed new genus of the Picornaviridae family**. Virus genes 47, 556-558.

230. Pignatelli M., Moya A., 2011. **Evaluating the Fidelity of De Novo Short Read Metagenomic Assembly Using Simulated Data**. PloS one 6.

231. Power S.E., O'Toole P.W., Stanton C., Ross R.P., Fitzgerald G.F., 2014. **Intestinal microbiota, diet and health**. Br J Nutr 111, 387-402.

232. Purcell R.H., Emerson S.U., 2008. **Hepatitis E: an emerging awareness of an old disease**. J Hepatol 48, 494-503.

233. Reuter G., Boldizsar A., Kiss I., Pankovics P., 2008. **Candidate new species of Kobuvirus in porcine hosts**. Emerg Infect Dis 14, 1968-1970.

234. Reuter G., Boldizsar A., Pankovics P., 2009. **Complete nucleotide and amino acid sequences and genetic organization of porcine kobuvirus, a member of a new species in the genus Kobuvirus, family Picornaviridae**. Arch Virol 154, 101-108.

235. Reuter G., Boros A., Pankovics P., 2011. **Kobuviruses - a comprehensive review**. Rev Med Virol 21, 32-41.

236. Reuter G., Pankovics P., Knowles N.J., Boros A., 2012. **Two closely related novel picornaviruses in cattle and sheep in Hungary from 2008 to 2009, proposed as members of a new genus in the family Picornaviridae**. J Virol 86, 13295-13302.

237. Reuter G., Boros A., Delwart E., Pankovics P., 2014. **Novel circular single-stranded DNA virus from turkey faeces**. Arch Virol 159, 2161-2164.

238. Reyes A., Haynes M., Hanson N., Angly F.E., Heath A.C., Rohwer F., Gordon J.I., 2010. **Viruses in the faecal microbiota of monozygotic twins and their mothers**. Nature 466, 334-338.

239. Reyes A., Semenkovich N.P., Whiteson K., Rohwer F., Gordon J.I., 2012. **Going viral: next-generation sequencing applied to phage populations in the human gut**. Nat Rev Microbiol 10, 607-617.

240. Rivera-Benitez J.F., Rosas-Estrada K., Pulido-Camarillo E., de la Pena-Moctezuma A., Castillo-Juarez H., Ramirez-Mendoza H., 2014. **Serological survey of veterinarians to assess the zoonotic potential of three emerging swine diseases in Mexico**. Zoonoses Public Health 61, 131-137.

241. Robilotti E., Deresinski S., Pinsky B.A., 2015. **Norovirus**. Clin Microbiol Rev 28, 134-164.

242. Rodriguez-Valera F., Martin-Cuadrado A.B., Rodriguez-Brito B., Pasic L., Thingstad T.F., Rohwer F., Mira A., 2009. **Explaining microbial population genomics through phage predation**. Nat Rev Microbiol 7, 828-836.

243. Rolfe K.J., Parmar S., Mururi D., Wreghitt T.G., Jalal H., Zhang H., Curran M.D., 2007. **An internally controlled, one-step, real-time RT-PCR**

assay for norovirus detection and genogrouping. J Clin Virol 39, 318-321.

244. Ronaghi M., Karamohamed S., Pettersson B., Uhlen M., Nyren P., 1996. **Real-time DNA sequencing using detection of pyrophosphate release**. Anal Biochem 242, 84-89.

245. Ronaghi M., Uhlen M., Nyren P., 1998. **A sequencing method based on real-time pyrophosphate**. Science 281, 363, 365.

246. Rosario K., Dayaram A., Marinov M., Ware J., Kraberger S., Stainton D., Breitbart M., Varsani A., 2012a. **Diverse circular ssDNA viruses discovered in dragonflies (Odonata: Epiprocta)**. J Gen Virol 93, 2668-2681.

247. Rosario K., Duffy S., Breitbart M., 2012b. **A field guide to eukaryotic circular single-stranded DNA viruses: insights gained from metagenomics**. Arch Virol 157, 1851-1871.

248. Royal Swedish Academy of Science: Nobel Prize, http://www.kva.se/en/Prizes/Nobel-prizes/Laureates-list/. Access October 2015.

249. Sanchez A., Geisbert T.W., Feldmann H., 2007. **Filoviridae: Marburg and Ebola virus in: "Fields virology"**. 5th edition. Editor: Field B., Knipe D.M., Howley P.M., Philadelphia : Wolters Kluwer Health/Lippincott Williams & Wilkins. Chapter 40, p 3-25.

250. Sanger F., Coulson A.R., 1975. **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase**. J Mol Biol 94, 441-448.

251. Sanger F., Nicklen S., Coulson A.R., 1977a. **DNA sequencing with chain-terminating inhibitors**. Proc Natl Acad Sci USA 74, 5463-7.

252. Sanger F., Air G.M., Barrell B.G., Brown N.L., Coulson A.R., Fiddes C.A., Hutchison C.A., Slocombe P.M., Smith M., 1977b. **Nucleotide sequence of bacteriophage phi X174 DNA**. Nature 265, 687-695.

253. Santiago-Rodriguez T.M., Ly M., Bonilla N., Pride D.T., 2015. **The human urine virome in association with urinary tract infections**. Front Microbiol 6, 14.

254. Sasaki M., Orba Y., Ueno K., Ishii A., Moonga L., Hang'ombe B.M., Mweene A.S., Ito K., Sawa H., 2015. **Metagenomic analysis of the shrew enteric virome reveals novel viruses related to human stool-associated viruses**. J Gen Virol 96, 440-452.

255. Schadt E.E., Turner S., Kasarskis A., 2010. **A window into third-generation sequencing**. Hum Mol Genet 19, R227-240.

256. Schielke A., Sachs K., Lierz M., Appel B., Jansen A., Johne R., 2009. **Detection of hepatitis E virus in wild boars of rural and urban regions**

in Germany and whole genome characterization of an endemic strain. Virol J 6, 58.

257. Schmidt T.M., DeLong E.F., Pace N.R., 1991. **Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing**. J Bacteriol 173, 4371-4378.

258. Schneiker S., Perlova O., Kaiser O., Gerth K., Alici A., Altmeyer M.O., Bartels D., Bekel T., Beyer S., Bode E., Bode H.B., Bolten C.J., Choudhuri J.V., Doss S., Elnakady Y.A., Frank B., Gaigalat L., Goesmann A., Groeger C., Gross F., Jelsbak L., Jelsbak L., Kalinowski J., Kegler C., Knauber T., Konietzny S., Kopp M., Krause L., Krug D., Linke B., Mahmud T., Martinez-Arias R., McHardy A.C., Merai M., Meyer F., Mormann S., Munoz-Dorado J., Perez J., Pradella S., Rachid S., Raddatz G., Rosenau F., Ruckert C., Sasse F., Scharfe M., Schuster S.C., Suen G., Treuner-Lange A., Velicer G.J., Vorholter F.J., Weissman K.J., Welch R.D., Wenzel S.C., Whitworth D.E., Wilhelm S., Wittmann C., Blocker H., Puhler A., Muller R., 2007. **Complete genome sequence of the myxobacterium Sorangium cellulosum**. Nat Biotechnol 25, 1281-1289.

259. Schoenfeld T., Liles M., Wommack K.E., Polson S.W., Godiska R., Mead D., 2010. **Functional viral metagenomics and the next generation of molecular tools**. Trends Microbiol 18, 20-29.

260. Scholz M.B., Lo C.C., Chain P.S., 2012. **Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis**. Curr Opin Biotechnol 23, 9-15.

261. Schulz E., Gottschling M., Ulrich R.G., Richter D., Stockfleth E., Nindl I., 2012. **Isolation of three novel rat and mouse papillomaviruses and their genomic characterization**. PloS one 7, e47164.

262. Scobie L., Dalton H.R., 2013. **Hepatitis E: source and route of infection, clinical manifestations and new developments**. J Viral Hepat 20, 1-11.

263. Shan T., Li L., Simmonds P., Wang C., Moeser A., Delwart E., 2011. **The fecal virome of pigs on a high-density farm**. J Virol 85, 11697-11708.

264. Sharma V., Colson P., Chabrol O., Pontarotti P., Raoult D., 2015. **Pithovirus sibericum, a new bona fide member of the "Fourth TRUC" club**. Front Microbiol 6, 722.

265. Sharp P.M., Hahn B.H., 2011. **Origins of HIV and the AIDS pandemic**. Cold Spring Harb Perspect Med 1, a006841.

266. Sikorski A., Dayaram A., Varsani A., 2013a. **Identification of a Novel Circular DNA Virus in New Zealand Fur Seal (Arctocephalus forsteri) Fecal Matter**. Genome Announc 1, e00558-13.

267. Sikorski A., Arguello-Astorga G.R., Dayaram A., Dobson R.C., Varsani A., 2013b. **Discovery of a novel circular single-stranded DNA virus from porcine faeces**. Arch Virol 158, 283-289.

268. Smith L.M., Sanders J.Z., Kaiser R.J., Hughes P., Dodd C., Connell C.R., Heiner C., Kent S.B., Hood L.E., 1986. **Fluorescence detection in automated DNA sequence analysis**. Nature 321, 674-679.

269. Smits S.L., Raj V.S., Oduber M.D., Schapendonk C.M., Bodewes R., Provacia L., Stittelaar K.J., Osterhaus A.D., Haagmans B.L., 2013. **Metagenomic analysis of the ferret fecal viral flora**. PloS one 8, e71595.

270. Staden R., 1979. **A strategy of DNA sequencing employing computer programs**. Nucleic Acids Res 6, 2601-2610.

271. Stenglein M.D., Velazquez E., Greenacre C., Wilkes R.P., Ruby J.G., Lankton J.S., Ganem D., Kennedy M.A., DeRisi J.L., 2012. **Complete genome sequence of an astrovirus identified in a domestic rabbit (Oryctolagus cuniculus) with gastroenteritis**. Virol J 9, 216.

272. Stevenson P.G., Hawke S., Sloan D.J., Bangham C.R., 1997. **The immunogenicity of intracerebral virus infection depends on anatomical site**. J Virol 71, 145-151.

273. Stevenson J., Hymas W., Hillyard D., 2008. **The use of Armored RNA as a multi-purpose internal control for RT-PCR**. J Virol Methods 150, 73-76.

274. Suttle C.A., 2005. **Viruses in the sea**. Nature 437, 356-361.

275. Suttle C.A., 2007. **Marine viruses--major players in the global ecosystem**. Nat Rev Microbiol 5, 801-812.

276. Tate J., Burton AH, Boschi-Pinto C, Steele AD, Duque J, Parashar U.D., 2012. **2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis.** Lancet Infect Dis. 12, 136-141.

277. Tee K.K., Takebe Y., Kamarulzaman A., 2009. **Emerging and re-emerging viruses in Malaysia, 1997-2007**. Int J Infect Dis 13, 307-318.

278. Theuns S., Vyt P., Desmarets L.M., Roukaerts I.D., Heylen E., Zeller M., Matthijnssens J., Nauwynck H.J., 2016. **Presence and characterization of pig group A and C rotaviruses in feces of Belgian diarrheic suckling piglets**. Virus research 213, 172-183.

279. Thurber R.V., Haynes M., Breitbart M., Wegley L., Rohwer F., 2009. **Laboratory procedures to generate viral metagenomes**. Nat protocols 4, 470-483.

280. Tripp S., Grueber M., 2011. **Economic Impact of the Human Genome Project**. Columbus: Battelle Memorial Institute. http://www.battelle.org/docs/default-document-

library/economic_impact_of_the_human_genome_project.pdf (Access December 2015)

281. Tse H., Chan W.M., Tsoi H.W., Fan R.Y., Lau C.C., Lau S.K., Woo P.C., Yuen K.Y., 2011. **Rediscovery and genomic characterization of bovine astroviruses**. J Gen Virol 92, 1888-1898.

282. Tse H., Chan W.M., Lam C.S., Lau S.K., Woo P.C., Yuen K.Y., 2012a. **Complete genome sequences of novel rat noroviruses in Hong Kong**. J Virol 86, 12435-12436.

283. Tse H., Chan W.M., Li K.S., Lau S.K., Woo P.C., Yuen K.Y., 2012b. **Discovery and genomic characterization of a novel bat sapovirus with unusual genomic features and phylogenetic position**. PloS one 7, e34987.

284. Ueno K., Ishii A., Ito K., 2014. **ELM: enhanced lowest common ancestor based method for detecting a pathogenic virus from a large sequence dataset**. BMC bioinformatics 15, 254.

285. Valdez-Cruz N.A., Caspeta L., Perez N.O., Ramirez O.T., Trujillo-Roldan M.A., 2010. **Production of recombinant proteins in E. coli by the heat inducible expression system based on the phage lambda pL and/or pR promoters**. Microb Cell Fact 9, 18.

286. van Drunen Littel-van den Hurk S., Watkiss E.R., 2012. **Pathogenesis of respiratory syncytial virus**. Curr Opin Virol 2, 300-305.

287. Venter J.C., Smith H.O., Hood L., 1996. **A new strategy for genome sequencing**. Nature 381, 364-366.

288. Venter J.C., Adams M.D., Sutton G.G., Kerlavage A.R., Smith H.O., Hunkapiller M., 1998. **Shotgun sequencing of the human genome**. Science 280, 1540-1542.

289. Venter J.C., 2011. **Genome-sequencing anniversary. The human genome at 10: successes and challenges**. Science 331, 546-547.

290. Verner-Carlsson J., Lohmus M., Sundstrom K., Strand T.M., Verkerk M., Reusken C., Yoshimatsu K., Arikawa J., van de Goot F., Lundkvist A., 2015. **First evidence of Seoul hantavirus in the wild rat population in the Netherlands**. Infect Ecol Epidemiol 5, 27215.

291. Victoria J.G., Kapoor A., Li L., Blinkova O., Slikas B., Wang C., Naeem A., Zaidi S., Delwart E., 2009. **Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis**. J Virol 83, 4642-4651.

292. Vonderfecht S.L., Huber A.C., Eiden J., Mader L.C., Yolken R.H., 1984. **Infectious diarrhea of infant rats produced by a rotavirus-like agent**. J Virol 52, 94-98.

293. Walker GT, Fraiser MS, Schram JL, Little MC, Nadeau JG, DP. M., 1992. **Strand displacement amplificatio**. Nucleic Acids Res. 11, 1691-6.

294. Wan C.H., Soderlund-Venermo M., Pintel D.J., Riley L.K., 2002. **Molecular characterization of three newly recognized rat parvoviruses**. J Gen Virol 83, 2075-2083.

295. Wang Q.H., Han M.G., Cheetham S., Souza M., Funk J.A., Saif L.J., 2005a. **Porcine noroviruses related to human noroviruses**. Emerg Infect Dis 11, 1874-1881.

296. Wang Q.H., Han M.G., Funk J.A., Bowman G., Janies D.A., Saif L.J., 2005b. **Genetic diversity and recombination of porcine sapoviruses**. J Clin Microbiol 43, 5963-5972.

297. Weinstock G.M., 2012. **Genomic approaches to studying the human microbiota**. Nature 489, 250-256.

298. Wetterstrand K.A., 2015. **DNA Sequencing Costs**. Data from the National Human Genome Res Institute (NHGRI) Genome Sequencing Program (GSP) Available at: www.genome.gov/sequencingcosts. Access December 2015.

299. Whichard J.M., Weigt L.A., Borris D.J., Li L.L., Zhang Q., Kapur V., Pierson F.W., Lingohr E.J., She Y.M., Kropinski A.M., Sriranganathan N., 2010. **Complete genomic sequence of bacteriophage felix o1**. Viruses 2, 710-730.

300. WHO, 2008. **Guidelines for drinking-water quality**, in: 3rd (Ed.). ISBN 978 92 4 154761 1

301. WHO, 2011. **Guidelines for drinking-water quality**. ISBN 978 92 4 154815 4th edition.

302. Wilcox R.M., Fuhrman J.A., 1994. **Bacterial viruses in coastal seawater- lytic rather than lysogenic production**. Mar Ecol Prog Ser 114, 35-45.

303. Wilhelmi I., Roman E., Sanchez-Fauquier A., 2003. **Viruses causing gastroenteritis**. Clin Microbiol Infect 9, 247-262.

304. Williamson K.E., Wommack K.E., Radosevich M., 2003. **Sampling natural viral communities from soil for culture-independent analyses**. Appl Environ Microbiol69, 6628-6633.

305. Willing B.P., Van Kessel A.G., 2009. **Intestinal microbiota differentially affect brush border enzyme activity and gene expression in the neonatal gnotobiotic pig**. J Anim Physiol Anim Nutr 93, 586-595.

306. Wolf S., Reetz J., Johne R., Heiberg A.C., Petri S., Kanig H., Ulrich R.G., 2013. **The simultaneous occurrence of human norovirus and hepatitis E virus in a Norway rat (Rattus norvegicus)**. Arch Virol 158, 1575-1578.

307. Wommack K.E., Colwell R.R., 2000. **Virioplankton: Viruses in aquatic ecosystems**. Microbiol Mol Biol Rev 64, 69-114.

308. Wommack K.E., Bench S.R., Bhavsar J., Mead D., Hanson T., 2009. **Isolation independent methods of characterizing phage communities 2: characterizing a metagenome**. Methods Mol Biol 502, 279-289.

309. Wongsuntornpoj S., Moreno Switt A.I., Bergholz P., Wiedmann M., Chaturongakul S., 2014. **Salmonella phages isolated from dairy farms in Thailand show wider host range than a comparable set of phages isolated from U.S. dairy farms**. Vet Microbiol 172, 345-352.

310. Woo P.C., Lau S.K., Teng J.L., Tsang A.K., Joseph M., Wong E.Y., Tang Y., Sivakumar S., Bai R., Wernery R., Wernery U., Yuen K.Y., 2014. **Metagenomic analysis of viromes of dromedary camel fecal samples reveals large number and high diversity of circoviruses and picobirnaviruses**. Virology 471-473, 117-125.

311. Wooley J.C., Godzik A., Friedberg I., 2010. **A primer on metagenomics**. PLoS Comput Biol 6, e1000667.

312. Wylie K.M., Mihindukulasuriya K.A., Zhou Y., Sodergren E., Storch G.A., Weinstock G.M., 2014. **Metagenomic analysis of double-stranded DNA viruses in healthy adults**. BMC Biol 12, 71.

313. Xiao C.T., Gimenez-Lirola L.G., Gerber P.F., Jiang Y.H., Halbur P.G., Opriessnig T., 2013. **Identification and characterization of novel porcine astroviruses (PAstVs) with high prevalence and frequent co-infection of individual pigs with multiple PAstV types**. J Gen Virol 94, 570-582.

314. Yang M.Q., Athey B.D., Arabnia H.R., Sung A.H., Liu Q., Yang J.Y., Mao J., Deng Y., 2009. **High-throughput next-generation sequencing technologies foster new cutting-edge computing techniques in bioinformatics**. BMC genomics 10 Suppl 1, I1.

315. Yoon K.Y., Editor: Jeffrey J. Zimmerman L.A.K., Alejandro Ramirez, Kent J. Schwartz, Gregory W. Stevenson 2012. **Overview of viruses in "Diseases of Swine",** 10th Edition. John Wiley & Sons. Chapter 23, p 383-634.

316. Yutin N., Koonin E.V., 2013. **Pandoraviruses are highly derived phycodnaviruses**. Biology direct 8, 25.

317. Zentek J., Ferrara F., Pieper R., Tedin L., Meyer W., Vahjen W., 2013. **Effects of dietary combinations of organic acids and medium chain fatty acids on the gastrointestinal microbial ecology and bacterial metabolites in the digestive tract of weaning piglets**. J Anim Sci 91, 3200-3210.

318. Zhang T., Breitbart M., 2006. **RNA Viral Community in Human Feces: Prevalence of Plant Pathogenic Viruses**.

319. Zhang Z., Xiang Y., Li N., Wang B., Ai H., Wang X., Huang L., Zheng Y., 2013. **Protective effects of Lactobacillus rhamnosus GG against human rotavirus-induced diarrhoea in a neonatal mouse model**. Pathog Dis 67, 184-191.

320. Zhang B., Tang C., Yue H., Ren Y., Song Z., 2014a. **Viral metagenomics analysis demonstrates the diversity of viral flora in piglet diarrhoeic faeces in China**. J Gen Virol 95, 1603-1611.

321. Zhang W., Li L., Deng X., Kapusinszky B., Pesavento P.A., Delwart E., 2014b. **Faecal virome of cats in an animal shelter**. J. Gen. Virol. 95, 2553-2564.

322. Zhao G., Krishnamurthy S., Cai Z., Popov V.L., Travassos da Rosa A.P., Guzman H., Cao S., Virgin H.W., Tesh R.B., Wang D., 2013. **Identification of novel viruses using VirusHunter--an automated data analysis pipeline**. PloS one 8, e78470.

8 APPENDIX

8.1 Eidesstattliche Erklärung:

Hiermit erkläre ich, dass ich die Arbeit selbständig und ausschließlich unter der Verwendung der angegebenen Mittel verfasst habe.

Die verwendeten Quellen und Hilfsmittel sind vollständig angegeben

Ort, Datum Unterschrift

8.2 Curriculum Vitae

For reasons of data protection, the curriculum vitae is not published in the electronic version.