

Virus Discovery and Characterization using Next-Generation Sequencing

Sander van Boheemen

© Copyright 2014, Sander van Boheemen

The research described in this thesis was performed at the Department of Viroscience, Erasmus MC, Rotterdam, The Netherlands and financially supported by the European Commission Seventh Framework Program for Research and Technological Development Project EMPERIE 223498. The research is within the framework of the Erasmus Postgraduate School Molecular Medicine.

Cover: First published diagrammatic picture on the double helix structure of DNA by Watson and Crick, as published in Nature in 1953.

Printing of this thesis was financially supported by Viroclinics Biosciences BV, Roche Diagnostics, and Greiner Bio-One.

ISBN: 978-90-8891-932-9

Book design & layout: Egon Sars & Sander van Boheemen

Printing: Proefschriftmaken.nl || Uitgeverij BOXPress

Virus Discovery and Characterization using Next-Generation Sequencing

Virus ontdekking en karakterisatie
met behulp van next-generation sequencing

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. H.A.P. Pols

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op
donderdag 11 september 2014 om 15:30 uur

door

Sander van Boheemen

geboren te Enkhuizen.



Promotiecommissie

Promotoren: Prof.dr. R.A.M. Fouchier
 Prof.dr. A.D.M.E. Osterhaus

Overige leden: Prof.dr. A. van Belkum
 Prof.dr. E.J. Snijder
 Prof.dr. M.D. de Jong

It is what it is

Table of contents

| | | |
|-----------|---|------------|
| Chapter 1 | General Introduction | 11 |
| Chapter 2 | Classical virus detection | |
| | 2.1 A Family-Wide RT-PCR Assay for Detection of Paramyxoviruses and Application to a Large-Scale Surveillance Study PLos One 2012 | 21 |
| Chapter 3 | Virus detection using next-generation sequencing | |
| | 3.1. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia New England Journal of Medicine 2012 | 37 |
| | 3.2 Cidofovir Gel as Treatment of Follicular Spicules in Multiple Myeloma JAMA Dermatology 2014 | 51 |
| Chapter 4 | Virus characterization using next-generation sequencing | |
| | 4.1 Identification, Characterization, and Natural Selection of Mutations Driving Airborne Transmission of A/H5N1 virus Cell 2014 | 59 |
| | 4.2 Genomic Characterization of a Newly Discovered Coronavirus Associated with Acute Respiratory Distress Syndrome in Humans MBio 2012 | 89 |
| | 4.3 Excessive Production and extreme ADAR editing of Human Metapneumovirus Defective Interfering RNA is Associated with Type I Interferon Induction Journal of General Virology 2014 | 107 |
| Chapter 5 | Summarizing Discussion | 123 |
| Chapter 6 | References | 137 |
| Chapter 7 | Nederlandse Samenvatting | 151 |
| Chapter 8 | Dankwoord | 157 |
| Chapter 9 | About the author | 163 |
| | Curriculum Vitae | |
| | PhD Portofolio | |
| | List of Publications | |

General Introduction

Emerging infectious diseases

Infectious diseases can be caused by a wide variety of pathogens, including viruses, bacteria, protozoa, fungi, helminths and prions. For numerous known pathogens, the disease incidence has increased over the last few decades. In addition, several previously unknown pathogens have recently been discovered, adding up to the impact of infectious diseases in humans and animals. The term "emerging infectious diseases" has been coined to describe infectious diseases caused by newly discovered pathogens and by known pathogens that have recently become more widespread.

Much emerging infectious disease research focuses on the importance of the interplay between humans, animals, and the environment in relation to health and disease. Because of industrialization, globalization, urbanization, deforestation, extensive commercialization of agriculture, and various other factors, the incidence of exposure of both animal and human populations to pathogens has increased [1]. This results in increased risks of a wide range of zoonotic diseases, i.e. diseases that can be passed from animals to humans. Animals are known to represent the origin of up to 75% of the emerging infectious diseases in humans. Examples of such zoonoses are the human immunodeficiency virus (HIV) and Ebola virus that were transmitted from non-human primates, and Lyme disease that is transmitted from animals by ticks [2-4]. On many occasions, humans are infected through an intermediate animal host rather than the original "reservoir" species. Hendra virus, Nipah virus, and SARS coronavirus (SARS-CoV) are pathogens that originate from bats, but infected humans through horses, pigs, and palm civets, respectively [5, 6]. The rate of contact between a pathogen reservoir and a novel host species is an important factor for successful interspecies transmission. Unfortunately, due to the changes described above, these contact rates have increased [7]. From 1980 onwards, the number of infectious outbreaks in both humans and animals have increased. Nearly 40 infectious diseases in humans have been identified that were unknown a few decades ago. In addition, many known infectious diseases that were thought to be under control have recently re-emerged [8].

According to the WHO, "It would be extremely naïve and complacent to assume that there will be no other disease like AIDS, Ebola, or SARS, sooner or later" [9]. In response to these words, several international initiatives were launched to counteract this threat, including a European consortium that leads the framework 7 program "EMPERIE" [10]. Many of the recent outbreaks of infectious disease in humans and animals were caused by RNA-viruses, as summarized in figure 1. RNA viruses are among the most variable and taxonomically diverse pathogens, with a wide variety of host species.

During the course of an outbreak, it is crucial to rapidly identify the causative agent. Pathogen identification enables containment strategies, specific diagnostic techniques, and treatments to be developed and implemented. The use of rapid, cost-effective molecular techniques that allow broad-scale screening of samples and subsequent pathogen discovery, is a strategy to respond to

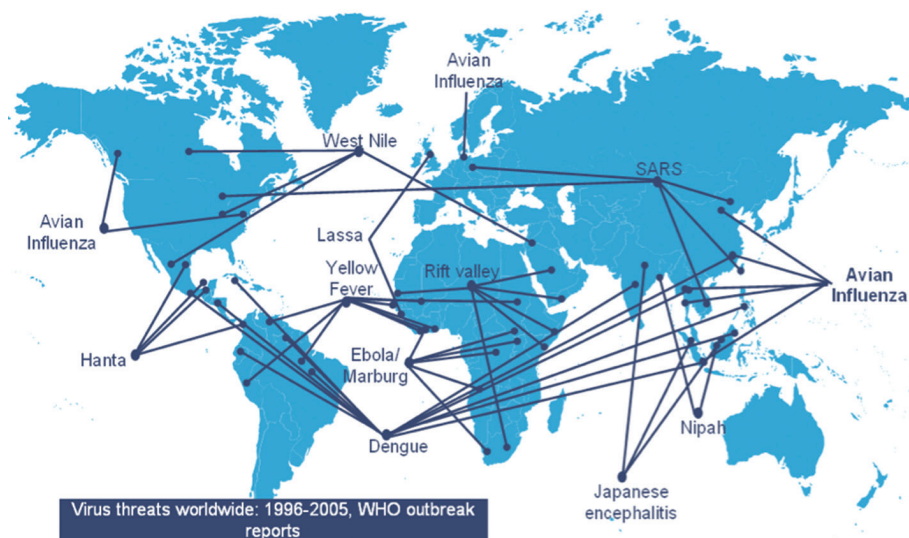


Figure 1. Worldwide RNA virus outbreaks from 1996-2005, as reported in outbreak reports of the WHO.

zoonotic disease emergence through early identification. In order to effectively counteract emerging viruses that threaten society, we need to have international collaborations of multidisciplinary groups integrated in the key national and international public health systems, combined with state-of-the-art technology for diagnostics.

Virus diversity

Viruses are very diverse. Even so, they can be grouped and placed into taxonomic systems. Morphology, nucleic acid type of the genome, sequence diversity mode of replication, host organisms, and the type of disease they cause, are the main phenotypic characteristics that form the basis of virus classification. However, because there are different ways to interpret these characteristics, virus classification is the subject of ongoing debate and proposals.

Two classification systems exist, that complement each other. Where the Baltimore classification system places viruses into one of seven groups based on the method of viral mRNA synthesis [11], the International Committee on Taxonomy of Viruses (ICTV) has placed viruses in seven orders, based on the inference that the virus families it contains have most likely evolved from a common ancestor [12]. Viruses are classified at different levels, starting with the order and continuing from family, subfamily, genus, until species. Within the seven orders, 103 families, 22 subfamilies, 455 genera, and 2,828 species of viruses have been defined by the ICTV [12]. Most of the viral families are not assigned to one of the seven orders.

Classical virus diagnostics

Classical virus diagnostics encompass a combination of techniques. The main repertoire of techniques used for the laboratory diagnosis of viral infections includes virus isolation, detection of viral antigens and nucleic acids, and detection of virus-specific antibodies. Other techniques are also occasionally used to complement the main diagnostics. Among those techniques are electron microscopy and histology.

Although time-consuming, cell culture provides a crucial element because this technique is capable of providing a viable virus isolate that can be used for further characterization. An additional advantage is that, in contrast to most antigen and nucleic acid detection methods, viral culture allows detection of multiple viruses, not all of which may have been suspected at the time the culture was started. Growth of virus in cell culture is usually detected by visualizing morphological changes in the cells, known as cytopathic effect (CPE). For example, one can observe the formation of syncytia, which are multinucleated cells that can result from multiple cell fusions of uninuclear cells. Cells infected by HIV and certain paramyxoviruses can cause these syncytia. Cell rounding, cell lysis, cell detachment, or production of inclusion bodies, are other examples of virus-induced cytopathic effects.

An alternative technique for virus detection is the detection of antigen via fluorescent antibody (FA) staining, immunoperoxidase staining or an enzyme-linked immunosorbent assay (ELISA). Of these techniques, FA staining is the most widely used in diagnostic virology. Rapid viral diagnosis by means of FA staining was first described in 1956 [13]. These techniques have two main advantages. First, the detection is fast, with results becoming available within hours of receiving the specimen in the laboratory. Second, there is no requirement for virus viability in the specimen, allowing greater flexibility in handling and transport of specimens.

Development of polymerase chain reaction (PCR) analysis in 1985 has greatly improved the diagnosis of virus infections through sensitive detection of specific viral nucleic acids [14]. To detect viral RNA, reverse transcriptase (RT) is used to copy the viral RNA into DNA. Real-time PCR (RT-PCR) assays are a major development that has greatly expanded the applicability of PCR analysis for diagnosis of virus infections. In these assays, a fluorescent signal is generated at the same time as PCR takes place. The assays are run on instruments that include optical systems to excite fluorescent dyes and detect fluorescent emissions. The combination of amplification and signal detection reduces the time required for nucleic acid detection significantly. When virus species-specific PCR assays do not reveal a pathogen, but a species from the same family is suspected, consensus-degenerate PCRs that detect whole genera or even whole families of viruses can give an indication of virus presence in a sample. However, these consensus-degenerate PCR assays are generally less sensitive as compared to species-specific assays.

Detection of specific antibodies raised upon virus infection is a traditional method to diagnose viral infections. Its clinical effectiveness is limited because of the need for comparison of acute and convalescent antibody titers. In particular chronic infections have been determined using this

serological method. For example, with HIV or HCV, presence of virus-specific antibodies is always (HIV) or usually (HCV) an indication of current infection.

It is important to note that the diagnosis of an infectious agent does not necessarily mean that this agent is also the cause of the disease under investigation. To prove association of the pathogen identified with the observed disease, the Koch postulates must be fulfilled. Koch's postulates are four criteria designed to establish a causal relationship between a microbe and a disease. The criteria are: (i) the microorganism or other pathogen must be present in all cases of the disease, (ii) the pathogen can be isolated from the diseased host and grown in pure culture, (iii) the pathogen from the pure culture must cause the disease when inoculated into a healthy, susceptible laboratory animal, and (iv) the pathogen must be reisolated from the new host and shown to be the same as the originally inoculated pathogen.

Virus discovery using next-generation sequencing

Approximately two decades after revealing the structure of DNA, a method for DNA sequencing with chain-terminating inhibitors was published, introducing Sanger sequencing [15]. Around the same time, Maxam–Gilbert sequencing was also developed. The latter method was based on base-specific partial chemical modification of DNA and subsequent cleavage of the DNA backbone at sites adjacent to the modified nucleotides [16]. However, Sanger sequencing would become the main method for sequencing for over 35 years. Limitations to this method are its limited scalability and the costs of sequencing large genomes. Because of this, only limited numbers of model organisms and human pathogens were selected for full-genome sequencing. A second, or next, generation of sequencers (NGS) was developed to overcome these limitations. Of these NGS technologies, the Illumina and Roche 454 sequencers have been the most frequently used NGS platforms. Illumina sequencing technology is based on reversible dye-terminators technology. Clonal amplification is achieved via bridge PCR, where fragments are amplified using primers that are attached to a solid surface to form DNA clusters. The main advantage of Illumina sequencing is that there is a high sequence yield and the cost per base is relatively low. Disadvantage of this technique is that Illumina produces characteristic systematic base calling biases [17]. In addition, different tiles of the sequencing plate can result in reads that contain a different quality [18], and higher sequencing error rates are observed at the 3' ends of the reads as compared to the 5' ends. 454 sequencing is a parallelized version of pyrosequencing, where the order of the nucleotides is determined based on sequencing-by-synthesis. The method amplifies DNA bound to a bead inside a water-in-oil emulsion, leading to clonal amplification of a single fragment. One bead is deposited per well of a plate, which can contain 500.000 or 1.000.000 wells. The main advantage of this technique is the fact that it produces long reads and that it is relatively fast in doing so. The disadvantages of 454 sequencing are that it has a high error rate in regions that contain homopolymers caused by indistinguishable light intensity [19, 20] and up to 15% of the produced reads are a consequence of artificial amplification [21]. As the 454 machines produce

long reads that lead to longer contigs in a de novo assembly, it makes the platform quite suitable for pathogen discovery.

Many viruses cannot be easily propagated in cell culture. As a consequence, detection of these viruses requires culture-independent methods. In early virus discovery programs, investigators turned to non-specific amplification methods, which might detect viruses that are genetically too distinct to be detected by degenerate or consensus PCR. However, this process of random amplification, cloning, transformation to bacteria and sequencing of bacterial colonies is very laborious and time consuming. With the arrival of NGS, this process has improved substantially, reducing valuable time and money. The required concentration of starting material is much less as compared to classical cloning strategies, because NGS technologies are relatively sensitive. The combination of random amplification [22] with NGS has proven to be a valuable tool in virus discovery [23].

A powerful tool for pathogen discovery, or to map the genetic microbial diversity within a given sample, is to use metagenomics. Metagenomics is the study of genetic material from microbial communities recovered directly from (environmental) samples, without the use of prior culturing [24]. The fact that this technique is independent of culturing facilitates the possibility to sequence large groups of microorganisms in many environments that are unculturable. In its early days, bacterial metagenomics was mostly based on sequencing of the hypervariable regions of 16S ribosomal RNA [25]. Because no gene exists that is common to every virus, viral metagenomics is based on sequencing of random amplified material. With the arrival of NGS platforms, metagenomics can quickly recover full genome sequences of viruses and bacteria. NGS technologies have decreased the amount of manual labor and increased the data output, resulting in increased sample diversity. Known and previously unknown viruses have been discovered in diverse samples by metagenomics. Examples of these viruses and samples are the discovery of multiple picornaviruses in the stool of children with acute flaccid paralysis [26], and the discovery of phages and viruses from algae in uncultured marine samples [27].

Bioinformatics

Once a sequence has been obtained, a search of this sequence against a database of nucleotide or protein sequences is needed to determine the origin of the sequence. This is mostly done using the Basic Local Alignment Search Tool (BLAST). BLAST is a computational algorithm that finds regions of local similarity between sequences [28]. Nucleotide or protein sequences are compared to sequences in the databases, and calculations are made on the statistical significance of possible matches.

The importance of sequencing is shown merely by the sheer dimension and expansion of international sequence databases. However, once a sample has been sequenced, it can still be a challenge to find new viruses in the dataset, especially when the viruses are highly divergent from sequences in current nucleotide and protein databases. In such cases, many reads amplified

by random amplification do not have significant nucleotide or amino acid sequence similarities to any sequence in the respective databases. Potentially new viruses might be amongst these sequences, making them of great importance to virus discovery. As the output of NGS platforms have increased and will increase further, the need for more efficient ways to analyze sequences from samples is needed.

Virus characterization

Viruses are capable of infecting all types of life forms [29]. It has been postulated that viruses have existed since living cells first evolved, and have shaped the evolution of higher organisms [30]. Virus evolution at much shorter time frames can also have significant impact. Virus diversity and virus adaptation may facilitate the infection of new hosts. Sometimes, vaccines and antiviral drugs become ineffective as a consequence of accumulation of (resistance) mutations. Therefore, there is an increasing interest in methods to study the accumulation of mutations during virus circulation.

Because RNA viruses have short replication cycles, and low-fidelity polymerases, their genomes have a remarkable potential for rapid evolution. This high evolution rate is the basis for the concept of viral quasispecies. A viral quasispecies is a group of viruses related by a similar mutation or mutations, competing within a highly mutagenic environment [31, 32].

Previously, analysis of quasispecies was done by PCR and Sanger sequencing, detecting mutations with a frequency of 10-20% [33, 34]. When it comes to Sanger sequencing, the clouds of diverse virus variants thus cause minority species to be overlooked, and the majority species will determine the “consensus sequence” within a sample. When studying the evolution of viruses, natural selection can induce these possible fitter minority species to evolve into a decisive majority species, and present vital information in terms of viral evolution. For known mutations, line probe assays have been used successfully [35]. However, the emergence of NGS techniques has dramatically lowered the detection frequency to 0,1% of sequenced reads, without prior knowledge of specific mutations [33].

The rapid identification and characterization of newly emerging pathogens of humans and animals is of key importance to prevent major public and animal health threats and economical and biodiversity disasters. SARS etiology was established within one month, which contributed significantly to the development of relevant intervention strategies that prevented the further global spread of the SARS coronavirus. It was the combination of classical virological and epidemiological tools on the one hand and advanced molecular sequencing and characterization techniques on the other hand, that allowed this rapid and successful response. In recent years, the application of high throughput sequencing technologies has had a major impact on the biological sciences, in particular also for the field of infectious diseases. These technologies will provide the basis for future rapid detection of newly emerging pathogens and for the rapid implementation of intervention strategies.

OUTLINE OF THIS THESIS

In this thesis, novel approaches for viral discovery and characterization were investigated, to reduce the time required from sample collection to pathogen identification, to increase the overall value of sequencing data, and to allow application of NGS approaches to large number of samples.

Chapter 2 describes the development of a primer set to detect all viruses within the Paramyxoviridae family using classical RT-PCR techniques. We implemented an option of high-throughput fragment analysis that can be applied to the RT-PCR assay, making it possible to test large numbers of human or animal samples for the presence of paramyxoviruses. This test was validated by screening a large collection of specimens that was collected from wild birds.

Chapter 3 describes the use of NGS for virus discovery. In chapter 3.1 we report on the clinical data, virus isolation, and molecular identification of a newly emerging human coronavirus using a pan-coronavirus RT-PCR assay and NGS. This virus had infected a 60-year-old Saudi man who was admitted to a private hospital in Jeddah, Saudi Arabia. The man developed acute pneumonia and subsequent renal failure with fatal outcome. The virus represents a novel *Betacoronavirus*, with bat coronaviruses HKU4 and HKU5 as closest known relatives, and has been dubbed Middle East Respiratory Syndrome coronavirus (MERS-CoV). The full genome sequence of the newly discovered MERS-CoV, along with extensive phylogenetic and taxonomic analyses, is described more elaborately in chapter 3.2.

In chapter 3.3, we used unbiased amplification and deep sequencing on a hair follicle of a 70-year-old male patient who had small, folliculocentric, coarse, yellowish spicules located on his face, chest, and lower arms. Based on further diagnostic testing, this patient was diagnosed with follicular spicules of multiple myeloma (FSMM). Facial treatment with Cidofovir cleared the yellowish spicules from the man's face, suggestive for a viral etiology of the disease. We demonstrated the presence of Merkel cell polyomavirus in this hair follicle by NGS.

Chapter 4 describes the use of NGS for detailed virus characterization. In chapter 4.1 we identified a minimal set of amino acid substitutions required for airborne virus transmission of an influenza A/H5N1 virus between ferrets. Only five amino acid substitutions were shown to confer the capacity to spread between ferrets via the airborne route. Substitutions E627K in PB2 and H99Y in PB1 were shown to collectively lead to enhanced transcription and virus replication. Substitution H103Y in HA stabilized the HA and decreased the pH of membrane fusion. Either Q222L or G224S in HA were required to alter the receptor-binding preference, and T156A in HA enhanced this binding avidity by loss of a glycosylation site. Using deep sequencing, we found that there was a strong natural selection on the substitutions responsible for airborne transmission; mutant viruses emerged very early in ferrets to become dominant virus variants rapidly.

Using an unbiased NGS virus discovery protocol, we investigated the presence of defective interfering particles (DIs) in human metapneumovirus (HMPV) samples in chapter 4.2. We show that DIs were only produced when the virus was passaged at high multiplicity of infection (m.o.i.).

DIs were shown to induce the type I interferon signaling cascade. Interestingly, the nucleotide sequences of these DIs were extensively edited - up to 70% -, and further bioinformatics and PCR assays suggested that editing was the result of the enzyme Adenosine Deaminase Acting on RNA (ADAR).

Finally, chapter 5 summarizes the findings that are described in this thesis.

A Family-Wide RT-PCR Assay for Detection of Paramyxoviruses and Application to a Large-Scale Surveillance Study

Sander van Boheemen, Theo M. Bestebroer, Josanne H. Verhagen,
Albert D. M. E. Osterhaus, Suzan D. Pas, Sander Herfst,
Ron A. M. Fouchier

PLoS One 2012; 7(4):e34961

ABSTRACT

Family-wide molecular diagnostic assays are valuable tools for initial identification of viruses during outbreaks and to limit costs of surveillance studies. Recent discoveries of paramyxoviruses have called for such assay that is able to detect all known and unknown paramyxoviruses in one round of PCR amplification. We have developed a RT-PCR assay consisting of a single degenerate primer set, able to detect all members of the *Paramyxoviridae* family including all virus genera within the subfamilies *Paramyxovirinae* and *Pneumovirinae*. Primers anneal to domain III of the polymerase gene, with the 3' end of the reverse primer annealing to the conserved motif GDNQ, which is proposed to be the active site for nucleotide polymerization. The assay was fully optimized and was shown to indeed detect all available paramyxoviruses tested. Clinical specimens from hospitalized patients that tested positive for known paramyxoviruses in conventional assays were also detected with the novel family-wide test. A high-throughput fluorescence-based RT-PCR version of the assay was developed for screening large numbers of specimens. A large number of samples collected from wild birds was tested, resulting in the detection of avian paramyxoviruses type 1 in both barnacle and white-fronted geese, and type 8 in barnacle geese. Avian metapneumovirus type C was found for the first time in Europe in mallards, greylag geese and common gulls. The single round family-wide RT-PCR assay described here is a useful tool for the detection of known and unknown paramyxoviruses, and screening of large sample collections from humans and animals.

INTRODUCTION

The *Paramyxoviridae* family within the order of *Mononegavirales* includes a large number of human and animal viruses that are responsible for a wide spectrum of diseases [36]. Measles virus (MV) is one of the most infectious human viruses known, and has been targeted by the World Health Organization for eradication through the use of vaccines. The paramyxovirus family includes several other viruses with high prevalence and public health impact in humans, like respiratory syncytial virus (RSV), human metapneumovirus (HMPV), mumps virus (MuV), and the parainfluenza viruses (PIV) [37]. In addition, newly emerging members of the *Paramyxoviridae* family – hendra and nipah virus – have caused fatal infections in humans upon zoonoses from animal reservoirs [38–40]. In animals, Newcastle disease virus (NDV) and Rinderpest virus (RPV) were among the viruses with the most devastating impact on animal husbandry. Members of the *Paramyxoviridae* family switch hosts at a higher rate than most other virus families [41] and infect a wide range of host species, including humans, non-human primates, horses, dogs, sheep, pigs, cats, mice, rats, dolphins, porpoises, fish, seals, whales, birds, bats, and cattle [12]. Thus, the impact of paramyxoviruses to general human and animal welfare is immense.

The *Paramyxoviridae* family consists of two subfamilies, the *Paramyxovirinae* and the *Pneumovirinae*. The subfamily *Paramyxovirinae* includes five genera: *Rubulavirus*, *Avulavirus*, *Respirovirus*, *Henipavirus* and *Morbillivirus*. The subfamily *Pneumovirinae* includes two genera: *Pneumovirus* and *Metapneumovirus* [12]. Classification of the *Paramyxoviridae* family is based on differences in the organization of the virus genome, the sequence relationship of the encoded proteins, the biological activity of the proteins, and morphological characteristics [36], [12]. Virions from this family are enveloped, pleomorphic, and have a single-stranded, non-segmented, negative-sense RNA genome. Complete genomic RNA sequences for known members of the family range from 13–19 kilobases in length. The RNA consists of six to ten tandemly linked genes, of which three form the minimal polymerase complex; nucleoprotein (N or NP), phosphoprotein (P) and large polymerase protein (L). Paramyxoviruses further uniformly encode the matrix (M) and fusion (F) proteins, and – depending on virus genus – encode additional surface glycoproteins such as the attachment protein (G), hemagglutinin or hemagglutinin-neuraminidase (H, HN), short-hydrophobic protein (SH) and regulatory proteins such as non-structural proteins 1 and 2 (NS1, NS2), matrix protein 2 (M2.1, M2.2), and C and V proteins [36], [12].

Routine diagnosis of paramyxovirus infections in humans and animals is generally performed by virus isolation in cell culture, molecular diagnostic tests such as reverse transcriptase polymerase chain reaction (RT-PCR) assays, and serological tests. Such tests are generally designed to be highly sensitive and specific for particular paramyxovirus species. However, to detect zoonotic, unknown, and newly emerging pathogens within the *Paramyxoviridae* family, these tests may be

less suitable. Development of virus family-wide PCR assays has greatly facilitated the detection of previously unknown and emerging viruses. Examples of such PCR assays are available for the flaviviruses [42], coronaviruses [43, 44] and adenoviruses [45]. For the *Paramyxoviridae*, Tong *et al.* described semi-nested or nested PCR assays to detect members of the *Paramyxovirinae* or *Pneumovirinae* subfamily or groups of genera within the *Paramyxovirinae* subfamily [46]. Although these tests are valuable for specific purposes, nesting of PCR assays and requirement for multiple primer-sets are sub-optimal for high-throughput diagnostic approaches, due to the higher risk of cross-contamination, higher cost, and being more laborious. Here, a PCR assay is described that detects all genera of the *Paramyxoviridae* with a single set of primers without the requirement of nesting. This assay was shown to detect all known viruses within the *Paramyxoviridae* family tested. As the assay is implemented in a high-throughput format of fragment analysis, the test will be useful for the rapid identification of zoonotic and newly emerging paramyxoviruses.

RESULTS

Design of oligonucleotides for PCR detection of paramyxoviruses

Thirty-three RNA-dependent RNA polymerase gene nucleotide sequences recorded in Genbank, representative for the species within the *Paramyxoviridae* family, were downloaded. BioEdit Sequence Alignment Editor was used to align the sequences, and to run Clustal W alignments. First, an amino acid sequence alignment was made of all species within the *Paramyxoviridae* family to locate conserved motifs. The most conserved motifs were found in domain III of the polymerase protein. A forced nucleotide alignment was made based on the motifs found in the amino acid alignment. On the basis of this alignment the consensus degenerate and inosine-containing forward and reverse primers, PMX1 and PMX2, respectively, were designed. Because there was ambiguity on several nucleotide positions in the primers, nucleotide degeneracy was applied to both the forward and reverse primer. The primers were designed to amplify a fragment with a total size of 121 base pairs. The variability between the genomic sequences of all paramyxovirus species at each nucleotide position in the PCR primers was calculated using the entropy algorithm available from the Bioedit software package [47, 48]. Entropy values were plotted for each nucleotide position in the primers (Figure 1). Considerable heterogeneity was still present for each of the oligonucleotides, but this was primarily restricted to the 5' ends of the oligonucleotides. The 3' ends of oligonucleotides are of greatest importance for the successful amplification by PCR. The 3' end of the reverse primer PMX2 was designed to anneal to the highly conserved motif GDNQ, which has been proposed to be the active site for nucleotide polymerization [49]. The entropy plots show high conservation of the 14 terminal nucleotides at the 3' ends of both primers. Forward primer PMX1 contains a thymine at position 20, where hendra virus, nipah virus and PIV-5 contain a cytosine at that position. Reverse primer PMX2

shows 2 mismatches within the 14 nucleotides at the 3' end of the primer. In this region, APMV-6 and henipaviruses have a single nucleotide mismatch to the primer.

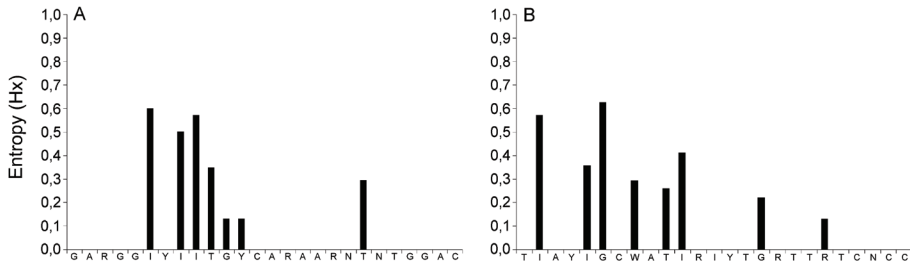


Figure 1. Entropy plots of primer-annealing sites in human and animal paramyxovirus sequences available from Genbank. Primer positions are given in the 5' to 3' direction. The sequences recognized by oligonucleotides PMX1 and PMX2 were compared to representative paramyxovirus genera and species sequences ($N = 33$), and their heterogeneities are displayed in panels A and B, respectively. The degree of heterogeneity was expressed as entropy as defined by Shannon: $H(1) = -\sum f(b, 1) \ln [f(b, 1)]$, where $H(1)$ is the uncertainty at position 1, b represents a residue out of the allowed choices for the sequence in question (A, C, G, T, -), and $f(b, 1)$ is the frequency at which residue b is found at position 1 [48].

Optimization of amplification

After primer design, various PCR reaction conditions were optimized, including primer concentrations (range 10 to 50 pmol), magnesium concentrations (range 0.5 to 4 mM), and cycling parameters. For each condition, serial virus dilutions of 10^0 to 10^7 were made and conditions were optimized to detect the highest dilution of input virus. Working concentrations of 50 pmol primer yielded the highest sensitivity. The maximum magnesium concentration recommended for AmpliTaq Gold DNA Polymerase of 4.0 mM was found to be optimal in this test. A gradient PCR with annealing temperatures ranging from 38°C to 46°C was performed, with 41°C determined as optimal annealing temperature (data not shown).

Validation of broad reactivity

To test the specificity of primers PMX-1 and PMX-2, RNA was isolated from the stocks of 28 paramyxoviruses from 6 genera (Table 1). Viruses from the *Henipavirus* genus were not tested, since these viruses are not available in our laboratory. RNA was isolated from high titer virus stocks in a 50 μ L volume, and 11 μ L of the RNA was used for cDNA synthesis, PCR amplification and gel electrophoresis to visualize the amplified fragments of 121 base pairs, including the primers (Figure 2).

L-gene fragments of all 28 paramyxovirus strains tested in this experiment were amplified, including members of the *Avula*-, *Morbilli*-, *Respiro*-, *Rubula*-, *Pneumo*- and *Metapneumovirus* genera. Additional experiments using RNA isolated from stocks of viruses from different virus families showed no cross reactivity of the assay with influenza A virus, influenza B virus, rhinovirus, human coronaviruses 229E, OC43, and NL63, and adenovirus (data not shown).

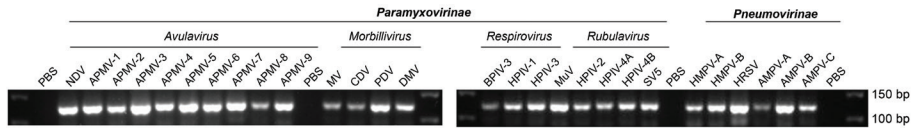


Figure 2. Detection of a wide range of paramyxoviruses by a single RT-PCR reaction. RNA was isolated from the indicated virus stocks and, after cDNA synthesis, used for PCR analysis and subsequent agarose gel electrophoresis. Genomic sequences of twenty-eight paramyxoviruses from six genera were amplified by PCR using the PMX1/PMX2 primer pair. PBS indicates phosphate-buffered saline used as negative control in the entire procedure.

Table 1. Virus stocks tested for evaluation of paramyxovirus detection by RT-PCR.

| Virus and classification | Abbreviation | Strain | Passage |
|-----------------------------------|--------------|-------------------------------|----------|
| <i>Paramyxovirinae</i> | | | |
| <i>Avulavirus</i> | | | |
| Avian paramyxovirus type 1 | APMV-1 | Chicken/N-Ireland/Ulster/2C | Eggs |
| Avian paramyxovirus type 2 | APMV-2 | Chicken/California/Yucaipa/56 | Eggs |
| Avian paramyxovirus type 3 | APMV-3 | Neophema/Holland/449/75 | Vero-118 |
| Avian paramyxovirus type 4 | APMV-4 | Duck/Hongkong/D3/75 | Eggs |
| Avian paramyxovirus type 5 | APMV-5 | strain Kunitachi | Vero-118 |
| Avian paramyxovirus type 6 | APMV-6 | Duck/Hongkong/199/77 | Eggs |
| Avian paramyxovirus type 7 | APMV-7 | Dove/Tennessee/4/75 | Eggs |
| Avian paramyxovirus type 8 | APMV-8 | Goose/Delaware/1053/76 | Eggs |
| Avian paramyxovirus type 9 | APMV-9 | Domestic Duck/New York/22/78 | Eggs |
| Newcastle disease virus | NDV | Nobilis ND Clone 30 | Eggs |
| <i>Morbillivirus</i> | | | |
| Measles virus | MV | Edmonston-Zagreb EZ 19 | Vero-118 |
| Mumps virus | MuV | unknown origin | NA |
| Canine distemper virus | CDV | Brussels | Vero-118 |
| Phocine distemper virus type 1 | PDV | NL/88 | Vero-118 |
| Dolphin morbillivirus | DMV | 16A | Vero-118 |
| <i>Respirovirus</i> | | | |
| Bovine parainfluenza virus type 3 | BPIV-3 | unknown origin | tMK |
| Human parainfluenza virus type 1 | HPIV-1 | C35 | tMK |
| Human parainfluenza virus type 3 | HPIV-3 | C243 | tMK |
| <i>Rubulavirus</i> | | | |
| Human parainfluenza virus type 2 | HPIV-2 | Greer, VR-92 | tMK |
| Human parainfluenza virus type 4a | HPIV-4A | M-25 | tMK |
| Human parainfluenza virus type 4b | HPIV-4B | CH 19503 | tMK |
| Simian parainfluenza virus 5 | SV5 | W3 | Eggs |
| <i>Pneumovirinae</i> | | | |
| <i>Pneumovirus</i> | | | |
| Human respiratory syncytial virus | HRSV | A2 | Hep-2 |
| <i>Metapneumovirus</i> | | | |
| Avian metapneumovirus type A | AMPV-A | Nobilis TRT | Vero-118 |
| Avian metapneumovirus type B | AMPV-B | Nobilis Rhino CV | Vero-118 |
| Avian metapneumovirus type C | AMPV-C | Colorado | Vero-118 |
| Human metapneumovirus A | HMPV-A | NL/1/00 | tMK |
| Human metapneumovirus B | HMPV-B | NL/1/99 | tMK |

PCR fragment detection using a Genetic Analyzer

Fragment analysis using a 3130x/ Genetic Analyzer was used to facilitate screening of larger numbers of samples without the requirement of running agarose gels (Figure 3). To this end, the forward primer PMX1 was labelled with the fluorescent dye 6-FAM. Samples were analyzed in a 96-well format. Seven tenfold serial dilutions of an APMV-3 virus stock were made and amplified using the pan-paramyxovirus RT-PCR assay. Upon agarose gel electrophoresis, 10^{-5} was the last dilution of virus still yielding a visible band. Upon fragment analysis, positive samples could be detected up to a dilution of 10^{-7} . When both forward and reverse primers were labelled, no significant increase in detection was observed (data not shown). Therefore, only the forward primer PMX1 was labelled.

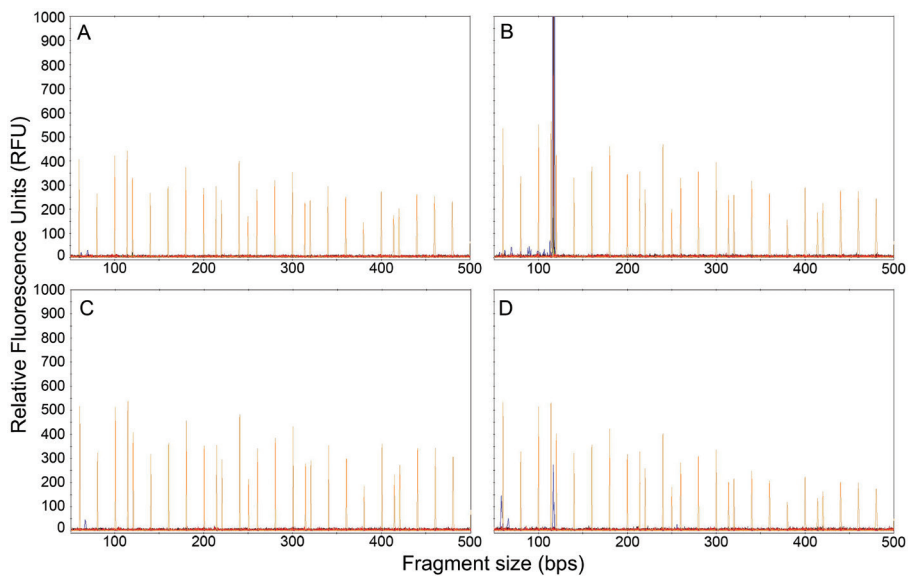


Figure 3. Fragment analysis plots of L gene fragments amplified by RT-PCR. The orange peaks represent LIZ-600 size standards. The size standard peaks are 60, 80, 100, 114, 120, 140, 160, 180, 200, 214, 220, 240, 250, 260, 280, 300, 314, 320, 340, 360, 380, 400, 414, 420, 440, 460, 480, 500, 514, 520, 540, 560, 580 and 600 bps. The blue peaks are PCR products amplified by PMX1 and PMX2, where PMX1 is labelled with the 6-FAM fluorescent dye. A and C represent avian samples in which no paramyxoviruses were found. B and D show peaks around 121 nucleotides. Blue peaks in B and D represent APMV-1 and APMV-C respectively. The relative fluorescence units are dependent on the amount of amplicon after RT-PCR.

Evaluation of specificity and sensitivity

To assess the sensitivity of the PMX1-PMX2-based assay as compared to standard diagnostic tests, we obtained anonymized human clinical samples that had tested positive for different paramyxoviruses using specific Taqman assays in the clinical virus diagnostic unit of Erasmus MC. Clinical samples were selected for diversity in virus species, type of clinical specimen, and virus load. The sample collection included clinical specimens positive for MuV, HMPV, MV, PIV-1, PIV-2, PIV-3, PIV-4, RSV-A, and RSV-B. Four different types of clinical specimens were used:

oral, (sputum, saliva, throat swab and mouth swab), nose (nose wash and nose swab), lung (broncho-alveolar lavage), and other (plasma and urine). Virus load, as measured by the cycle threshold (Ct) value in real-time Taqman assays, ranged from Ct 15 to Ct 38. Thirty-five samples were selected, RNA was extracted using the MagnaPure LC system, and the RT-PCR assay for detection of paramyxoviruses and fragment analysis was performed (Figure 4). Out of the 35 samples, the pan-paramyxovirus RT-PCR assay detected 27 human paramyxoviruses. The sample with the lowest Ct value (highest concentration target nucleic acid) that remained negative in the pan-paramyxovirus RT-PCR assay was a nose wash containing PIV-2 with a CT value of 30. The virus load in the 8 samples that remained negative in the pan-paramyxovirus RT-PCR assay (mean Ct 34.7, standard deviation 2.5) was lower than the load in the 27 samples yielding a positive reaction (mean Ct 24.5, standard deviation 5.2). Out of the virus specimens tested, 2/3 of the MuV specimens were positive, 4/4 for HMPV, 2/3 for MV, 3/5 for PIV-1, 4/5 for PIV-2, 3/3 for PIV-3, 3/5 for PIV-4, 4/5 for RSV-A, and 4/5 for RSV-B.

Fifty-four additional human throat samples were obtained from the clinical virus diagnostic unit of Erasmus MC. These samples were collected in 2000 and 2001 from patients with respiratory illnesses. The samples tested negative for the presence of RSV, influenza A, B and C, PIV-1 and -4, adenoviruses and rotaviruses by direct immunofluorescence on throat swabs and by immunofluorescence upon tissue culture. Of these 54 samples, one tested positive for PIV-1, two for PIV-4 and three for RSV with the pan-paramyxovirus RT-PCR assay followed by nucleotide sequencing of the PCR fragments (data not shown).

High-titer virus stocks of HMPV, HRSV, HPIV-1, HPIV-2, HPIV-3, HPIV-4, and MV were serially diluted to 10^{-10} . Dilutions were tested using agent-specific real-time PCR assays in our diagnostics department. The same RNA was used for the pan-paramyxovirus RT-PCR assay with subsequent fragment analysis. For HMPV, agent-specific real-time PCR assays detected positive samples up to a dilution of 10^{-6} , while fragment analysis detected positive samples up to a dilution of 10^{-5} . For the other viruses, these comparative dilutions were: RSV 10^{-7} and 10^{-5} , PIV-1 10^{-7} and 10^{-5} , PIV-2 10^{-5} and 10^{-4} , PIV-3 10^{-8} and 10^{-5} , PIV-4 10^{-6} and 10^{-5} , MV 10^{-8} and 10^{-5} . Thus, on average, the pan-paramyxovirus RT-PCR assay was 2-log less sensitive than agent-specific RT-PCR assays.

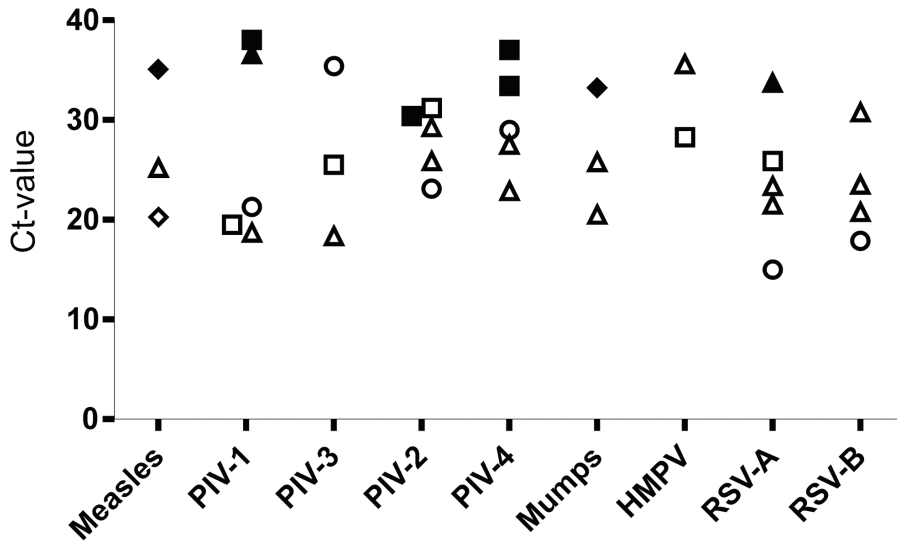


Figure 4. Evaluation of the pan-paramyxovirus RT-PCR assay for virus detection in 35 clinical samples with Taqman-confirmed human paramyxoviruses. Virus loads, as measured by Ct-values in virus specific Taqman assays, are plotted against paramyxovirus species (x-axis). Shapes of symbols indicate the type of clinical specimen (△; oral samples, □; nose samples, ◇; lung samples, ○; other samples). Non-filled shapes represent samples that tested positive in fragment analysis, while filled shapes represent samples that tested negative.

Phylogenetic analysis of amplified inserts

The pan-paramyxovirus RT-PCR assay amplifies a highly conserved region in the polymerase gene. To test whether the variability in the amplicon is sufficient to classify virus specimens to a virus subfamily or genus, a phylogenetic analysis of 33 different paramyxovirus L-gene fragments was performed (Figure 5).

As can be seen from the phylogenetic analysis based solely on the 70 base pair nucleotide sequences between primers PMX1 and PMX2 of the paramyxovirus species, the viruses clustered together as genera. These data indicate that the amplicon sequences obtained with PMX1-PMX2 can be used to (at least) roughly classify the viruses detected.

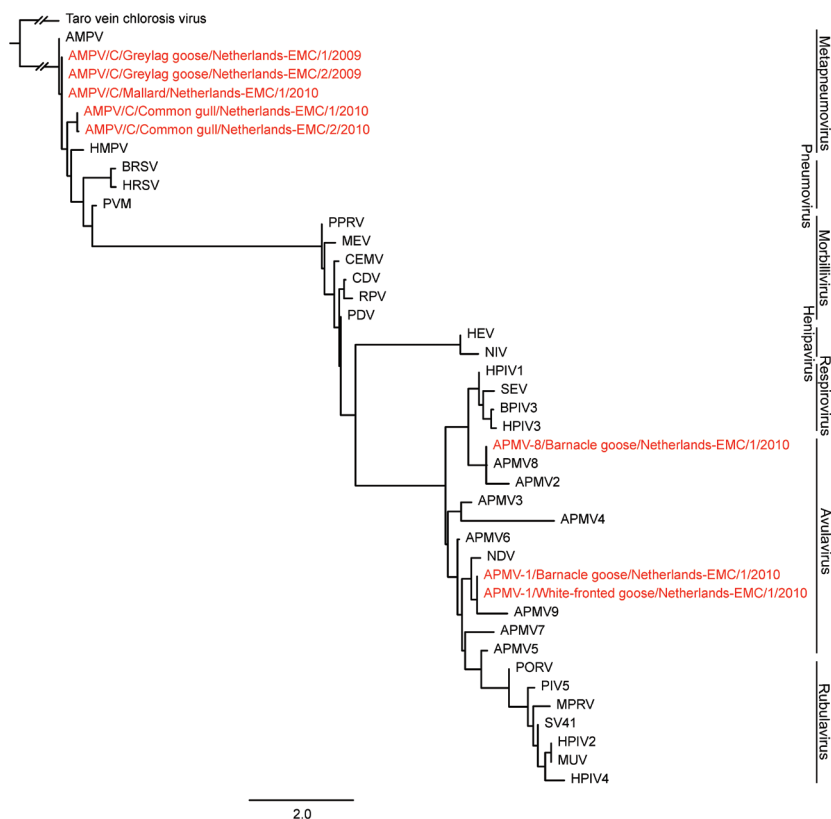


Figure 5. Maximum likelihood (ML) phylogenetic tree based on a 70 nt L-gene fragment for 33 different Paramyxovirus species. An L-gene fragment of the rhabdovirus Taro vein chlorosis virus (AY674964) was used as outgroup. Sequenced in black were obtained from Genbank while sequences in red were obtained from the avian surveillance study reported here. The length of the branch of Taro vein chlorosis virus as an outgroup is broken, and therefore does not represent its genetic distance to the rest of the tree. Accession numbers used for the phylogenetic tree were: NC003443, EU627591, NC009489, NC002200, NC009640, NC006430, NC006428, EU338414, EU403085, FJ177514, GU206351, NC003043, FJ231524, FJ215863, EU910942, NC002617, NC002161, C003461, NC001796, NC001552; NC001906, NC002728, NC001921, NC005283, NC001498, NC006383, Y09630, NC006296, NC001989, NC001781, NC006579, NC007652, NC004148.

Application of the pan-paramyxovirus RT-PCR assay to high-throughput detection of paramyxoviruses in avian samples

Samples from migratory birds were collected by expert ornithologists throughout the Netherlands and tested using a real-time RT-PCR assay targeting the influenza A virus matrix gene for an ongoing influenza surveillance program [50]. A selection of influenza virus negative samples was tested for the presence of members of the *Paramyxoviridae* family (Table 2). Samples were selected to span different species of birds, geographical locations, and multiple years. A total of 847 samples were tested for the presence of paramyxoviruses and yielded 8 positives. APMV-1 was found in a barnacle goose (*Branta leucopsis*) and a white-fronted goose (*Anser albifrons*), and APMV-8 in a barnacle goose. Avian metapneumovirus type C (AMPV-C) was found in 5 samples collected from 3 different bird species; mallard (*Anas platyrhynchos*), greylag goose (*Anser anser*) and common gull (*Larus canus*). To our knowledge, this is the first time that AMPV-C has been found in Europe.

Within the avian influenza virus surveillance program, numerous hemagglutinating agents were obtained upon inoculation of 11-day-old embryonated chicken eggs with bird specimens. While the vast majority of these agglutinating agents were identified as influenza A viruses in hemagglutination inhibition assays with antisera against all influenza virus subtypes, seventeen hemagglutinating agents could not be identified as influenza A virus. RNA was extracted from these unknown hemagglutinating agents and the RT-PCR assay for detection of paramyxoviruses was performed. This resulted in the identification of 12 isolates of APMV-1 obtained from mallards, barnacle geese, northern shovelers (*Anas clypeata*), common teals (*Anas crecca*), and white-fronted geese, one APMV-4 isolate from a mallard, two AMPV-6 isolates from mallards, and two APMV-9 isolates from gadwall and mallard.

Table 2. Avian samples tested for the presence of paramyxoviruses by RT-PCR.

| Order Family | Species | Tested (N) | Positive (N) | N (%) | Viruses detected |
|-----------------|---|---------------|-----------------|----------|------------------|
| Anseriformes | | | | | |
| Ducks | | | | | |
| | Mallard (<i>Anas platyrhynchos</i>) | 250 | 1 | 0,4 | AMPV-C |
| Geese | | | | | |
| | White-fronted goose (<i>Anser albifrons</i>) | 224 | 1 | 0,5 | APMV-1 |
| | Barnacle goose (<i>Branta leucopsis</i>) | 72 | 2 | 2,8 | APMV-1, APMV-8 |
| | Greylag goose (<i>Anser anser</i>) | 30 | 2 | 6,7 | AMPV-C (2) |
| | Bean goose (<i>Anser fabalis</i>) | 15 | 0 | 0 | |
| | Canada goose (<i>Branta canadensis</i>) | 4 | 0 | 0 | |
| Charadriiformes | | | | | |
| Gulls | | | | | |
| | Common gull (<i>Larus canus</i>) | 84 | 2 | 2,4 | AMPV-C (2) |
| Terns | | | | | |
| | Common tern (<i>Sterna hirundo</i>) | 2 | 0 | 0 | |
| Waders | | | | | |
| | Dunlin (<i>Calidris alpina</i>) | 84 | 0 | 0 | |
| | Turnstone (<i>Arenaria interpres</i>) | 30 | 0 | 0 | |
| | Oystercatcher (<i>Haematopus ostralegus</i>) | 18 | 0 | 0 | |
| | Common redshank (<i>Tringa totanus</i>) | 16 | 0 | 0 | |
| | Grey plover (<i>Pluvialis squatarola</i>) | 7 | 0 | 0 | |
| | Red knot (<i>Calidris canutus</i>) | 6 | 0 | 0 | |
| | Great ringed plover (<i>Charadrius hiaticula</i>) | 5 | 0 | 0 | |

DISCUSSION

Here, a high-throughput pan-paramyxovirus RT-PCR assay is described. The polymerase gene is the most conserved gene of the *Paramyxoviridae* family, and was therefore selected as the target for primer design. The primers were designed using an amino acid alignment of the RNA-dependent RNA polymerase protein that contained motifs conserved among all members of the seven genera composing the *Paramyxoviridae* family. The forward and reverse primers anneal at sites that are highly conserved. Sequence comparison of the polymerase gene of paramyxoviruses has revealed six highly conserved domains (I to VI), which are predicted to be essential for the key functions of RNA binding, RNA replication, and protein kinase activity [51, 52]. Domain III was proposed to function as the nucleotide polymerase. The primers used in this assay anneal to domain III. Most interestingly, the 3' end of the reverse primer PMX2 anneals to the conserved motif GDNQ, which is proposed to be the active site for nucleotide polymerization [49]. Because of this conservation, we can not exclude that this assay may detect other *Mononegavirales*, such as filo-, borna-, and rhabdoviruses. However, primer PMX1 shows a maximum of 3 mismatches with paramyxoviruses, whilst filo-, borna-, and rhabdoviruses show on average 6 or 7 mismatches with this primer. Likewise, primer PMX2 has a maximum of 4 mismatches with paramyxoviruses, whilst filo-, borna-, and rhabdoviruses show on average 7 to 9 mismatches. It is thus likely that detection of paramyxoviruses is most efficient. In the future, this conserved region may also be used as a target for the development of a pan-rhabdovirus RT-PCR assay or for other members of the order *Mononegavirales* as this region appears to be conserved within several virus families.

Consensus degenerate and inosine-containing primers were used to account for mismatches in strains from different paramyxovirus genera. Using these degenerate primers, all 28 paramyxoviruses that were present in our laboratory, encompassing six out of seven genera, were amplified successfully. Viruses within the *Henipavirus* genus were not tested because these viruses are not available in our facility. The thymine at position 20 in the forward primer does not match the cytosine at the corresponding location in the *Henipavirus* genus L-gene. This same mismatch was also seen in SV5. Since, the primers were able to detect SV5 (Figure 2), detection of henipaviruses should theoretically also be possible.

Application of the pan-paramyxovirus RT-PCR assay to human specimens in which the presence of paramyxoviruses was confirmed using routine species-specific diagnostic tests revealed that the assay detected paramyxoviruses in samples with Ct values up to 35 (Figure 3). Nine different clinical specimen types were tested. Paramyxoviruses were not detected in the two urine samples that were present in the sample set, which was likely due to the low copy numbers present in these samples (Ct 33 and 35). Paramyxoviruses were readily detected in all other clinical specimen types. Overall, the pan-paramyxovirus RT-PCR detected 27/35 paramyxovirus positive

specimens from humans and detected paramyxovirus in 6/54 samples that tested negative in routine diagnostic assays. These results appear to indicate satisfactory sensitivity to detect zoonotic, unknown, and newly emerging pathogens within the *Paramyxoviridae* family. It is important to note that the sensitivity of the pan-paramyxovirus RT-PCR assay is less than what can be achieved with more specific RT-PCR based tests.

The application of the pan-paramyxovirus RT-PCR assay was further evaluated using specimens from wild migratory birds. Screening of 847 wild birds in the Netherlands revealed 8 paramyxoviruses, detected using the fragment analysis method. Initially, a SYBR green melting curve-based assay was examined. However, this assay was not sensitive enough for screening of uncultured specimens. Fragment analysis was found to be a robust method for paramyxovirus screening in large sample collections. Three different paramyxovirus species were detected. Among the 8 positive samples, the AMPV-C was detected in three birds. To our knowledge, AMPV-C has not been detected previously in wild birds in Europe. It will be of interest to investigate the prevalence of the European AMPV-C in wild migratory birds further, and to test the genetic relationship with AMPV-C in the USA and HMPV.

Although the primer binding sites for the pan-paramyxovirus RT-PCR assay are highly conserved, the 70 nucleotide amplicon insert displays substantial variation. This variability facilitates the identification of amplified virus sequences to the species or genus level (Figure 4). Thus, the described test not only allows rapid detection of zoonotic, unknown, and newly emerging pathogens, but also gives initial hints to its classification that may aid in the further characterization.

Others have developed primer sets to detect *Paramyxoviridae* viruses previously that have already proven their value [12]. However, one primer set that detects all species of the *Paramyxoviridae* family without the need of nesting PCR reactions has thus far not been described. With the addition of fragment analysis applicable to the RT-PCR assay, paramyxovirus detection has now become possible in a high-throughput manner. The single round pan-paramyxovirus RT-PCR assay described here may thus be a useful tool for the detection of known and unknown paramyxoviruses, and screening of large sample collections from humans and animals.

MATERIALS AND METHODS

Oligonucleotide primer design

Primers were designed to anneal to conserved motifs in the RNA-dependent RNA polymerase gene as described in the results section. Consensus degenerate and inosine-containing primers were designed to account for variability among the different paramyxovirus species. The final

optimized forward and reverse oligonucleotide sequences are PMX1 (5'-GAR-GGI-YII-TGY-CAR-AAR-NTN-TGG-AC-3') and PMX2 (5'-TIA-YIG-CWA-TIR-IYT-GRT-TRT-CNC-C-3'), with G, A, C, T representing normal nucleotides, and I, Y, W, R, N, representing inosine, pyrimidine (C, T), weak (A, T), purine (A, G) and any (A, C, T, G) nucleotide respectively.

Viral nucleic acid isolation

For initial optimization of virus testing of tissue culture supernatants and allantoic fluids, RNA was isolated manually using the high pure RNA isolation kit (Roche Diagnostics, Almere, The Netherlands) according to instructions from the manufacturer. For high-throughput screening using fragment analysis, RNA was isolated using a MagnaPure LC system with a MagnaPure LC total nucleic acid isolation kit (Roche) according to instructions from the manufacturer.

Copy DNA (cDNA) synthesis and Polymerase Chain Reaction

A SuperScript III One-Step reverse transcription kit (Invitrogen, Bleiswijk, The Netherlands) was used to synthesize cDNA from extracted RNA. The optimized RT mixture contained 11 μL of RNA extract, 1 μL (500 $\mu\text{g}/\text{mL}$) Random Primers (Promega, Leiden, The Netherlands), 0.5 μL (40 U/ μL) Ribonuclease Inhibitor (Promega), and 1 μL (10 mM each) deoxynucleoside triphosphates (Roche) in a 13.5 μL volume. After a 5 min incubation at 65°C for optimal primer hybridisation to template, 4 μL (10 \times) First-Strand buffer, 1 μL (0.1 M) DTT, 0.5 μL (40 U/ μL) Ribonuclease Inhibitor (Promega) and 1 μL (200 U/ μL) SuperScript III Reverse Transcriptase was added to the mixture in a 20 μL volume. The RT mixture was sequentially incubated at 25°C for 5 min and 42°C for 1 hour to obtain cDNA.

PCR was optimized with respect to enzymes, primer sets, and concentrations of reagents as well as cycling parameters. The PCR mixture contained 50 pmol of each forward and reverse primer, 4 μL of cDNA, 1 μL (10 mM each) deoxynucleoside triphosphate, 5 μL 10 \times PCR Gold buffer, 8 μL (25 mM) MgCl_2 , and 0.5 μL (2.5 U/ μL) AmpliTaq Gold DNA Polymerase (Applied Biosystems, Bleiswijk, The Netherlands). Water was then added to achieve a final volume of 50 μL . The PCR mixture was incubated at 94°C for 10 min, then 35 cycles at 94°C for 15 s, 41°C for 30 s, 72°C for 30 s, and a final extension at 72°C for 7 min.

Amplicon detection

PCR products were visualized by blue light after electrophoresis on a 2.5% agarose gel containing 1 \times GelStar® Nucleic Acid Gel Stain (Lonza, Breda, The Netherlands) in 1 \times Tris-borate buffer (pH 8.0). A 50 bp DNA Ladder (Invitrogen) was used to estimate amplicon size.

For high-throughput testing, fragment analysis was used. To this end, oligonucleotide PMX1 was labelled with 6-carboxyfluorescein (6-FAM) and PMX2 was used unlabeled. PCR was performed as described above. Subsequently, 0.5 μL LIZ-600 Size Standard (GeneScan, Freiburg, Germany) was mixed with 9 μL formamide and 0.5 μL PCR product. Fragment analysis was performed

using the 3130x/Genetic Analyzer (Applied Biosystems) and data was analysed using GeneMapper software (Applied Biosystems). Classification of positive human and avian specimens was based on nucleotide sequencing of the PCR fragment and performing a NCBI BLASTn search on the 70 nt insert.

Virus stocks and specimens

High titer virus stocks of paramyxoviruses used for validation of broad reactivity of the assay are listed in table 1. Clinical specimens from humans to test for sensitivity of fragment analysis were obtained from the clinical diagnostic unit of the virology department, and were anonymized. Wild birds were trapped by expert ornithologists. Cloacal and/or oropharyngeal swab specimens were collected with sterile cotton swabs. All samples were stored in transport medium consisting of Hanks balanced salt solution containing 0.5% lactalbumin, 10% glycerol, 200 U/mL penicillin, 200 µg/mL streptomycin, 100 U/mL polymyxin B sulfate, 250 µg/mL gentamicin, and 50 U/mL nystatin (ICN, Zoetermeer, The Netherlands). All bird samples were stored at -80°C or at -20°C if rapid transport or storage at -80°C was not possible. Frozen samples were stored at -80°C in the laboratory upon arrival and were thawed no more than two times prior to analysis.

Phylogenetic tree analysis

Paramyxovirus L-gene fragments spanning the PMX1-PMX2 amplified region were downloaded from Genbank. Nucleotide sequence alignments of paramyxovirus L gene inserts was done using the Clustal W multiple alignment tool in BioEdit Sequence Alignment Editor [47]. Maximum likelihood (ML) trees were inferred using PAUP* (Phylogenetic Analysis Using Parsimony [53], version 4b10), by means of a full heuristic search and the tree bisection-reconnection (TBR) method based on the best-fit models of nucleotide substitution models determined by MODELTEST [54]. The preferred model of nucleotide substitution was GTR+G, based on the Akaike information criterion.

ACKNOWLEDGEMENTS

We thank the following ornithologists: J.G.B. van Dijk, D. Jonkers, F. Majoor, G.J.D.M. Müskens, E. van Oort, K. Oosterbeek, and T. de Vaal and their respective organisations, Alterra Wageningen, Eendenkooi Stichting, De Kooikersvereniging, SOVON Dutch Centre for Field Ornithology and the NIOO-KNAW Netherlands Institute of Ecology.

Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia

Ali M. Zaki, M.D., Ph.D., Sander van Boheemen, M.Sc., Theo M. Bestebroer, B.Sc.,
Albert D.M.E. Osterhaus, D.V.M., Ph.D., and Ron A.M. Fouchier, Ph.D.

N Engl J Med. 2012 Nov 8;367(19):1814-20

ABSTRACT

A previously unknown coronavirus was isolated from the sputum of a 60-year-old man who presented with acute pneumonia and subsequent renal failure with a fatal outcome in Saudi Arabia. The virus (called HCoV-EMC) replicated readily in cell culture, producing cytopathic effects of rounding, detachment, and syncytium formation. The virus represents a novel *Betacoronavirus* species. The closest known relatives are bat coronaviruses HKU4 and HKU5. Here, the clinical data, virus isolation, and molecular identification are presented. The clinical picture was remarkably similar to that of the severe acute respiratory syndrome (SARS) outbreak in 2003 and reminds us that animal coronaviruses can cause severe disease in humans.

INTRODUCTION

Coronaviruses are enveloped, single-stranded, positive-sense RNA viruses that are phenotypically and genotypically diverse [56]. Coronaviruses are widespread in bats around the world but can be found in many other species as well, including birds, cats, dogs, pigs, mice, horses, whales, and humans [56]. They may cause respiratory, enteric, hepatic, or neurologic diseases, with variable severity in various animal species. In humans, four respiratory coronaviruses -human coronaviruses (HCoV) 229E, OC43, NL63, and HKU1- are known to be endemic. In addition, in 2003 a previously unknown coronavirus caused an outbreak of SARS in humans [43, 57, 58]. The diversity of coronaviruses is facilitated by the infidelity of the RNA-dependent RNA polymerase, the high frequency of RNA recombination, and the unusually large genomes for RNA viruses [56, 59]. These factors not only have led to the diversity of known coronaviruses but also have facilitated the emergence of viruses with new traits that allow the organism to adapt to new hosts and ecologic niches, sometimes causing zoonotic events.

CASE REPORT

A 60-year-old Saudi man was admitted to a private hospital in Jeddah, Saudi Arabia, on June 13, 2012, with a 7-day history of fever, cough, expectoration, and shortness of breath. He had no history of cardiopulmonary or renal disease, was receiving no long-term medications, and did not smoke. The physical examination revealed a body-mass index (the weight in kilograms divided by the square of the height in meters) of 35.1, a blood pressure of 140/80 mm Hg, a pulse of 117 beats per minute, a temperature of 38.3°C, and a respiratory rate of 20 breaths per minute.

Chest radiography performed on admission showed low lung volumes, bilateral enhanced pulmonary hilar vascular shadows more prominent on the left, and accentuated bronchovascular lung markings. Multiple segmental, patchy, veiling opacities were present in the middle and lower lung fields; the costophrenic angles were not blunted. The cardiac silhouette was not enlarged, with a dilated, unfolded aorta (Figure 1A). On chest radiography performed 2 days later, the opacities had become more confluent and dense (Figure 1B). Computed tomography performed 4 days after admission revealed few subcentimetric mediastinal hilar lymph nodes, bilateral dependent airspace opacities with air bronchograms, scattered areas of ground-glass opacity, interstitial septal thickening, and nodularity in upper lobes, with minimal bilateral pleural effusions and no pericardial pleural effusion (Figure 1C). Collectively, these findings were consistent with an infection.

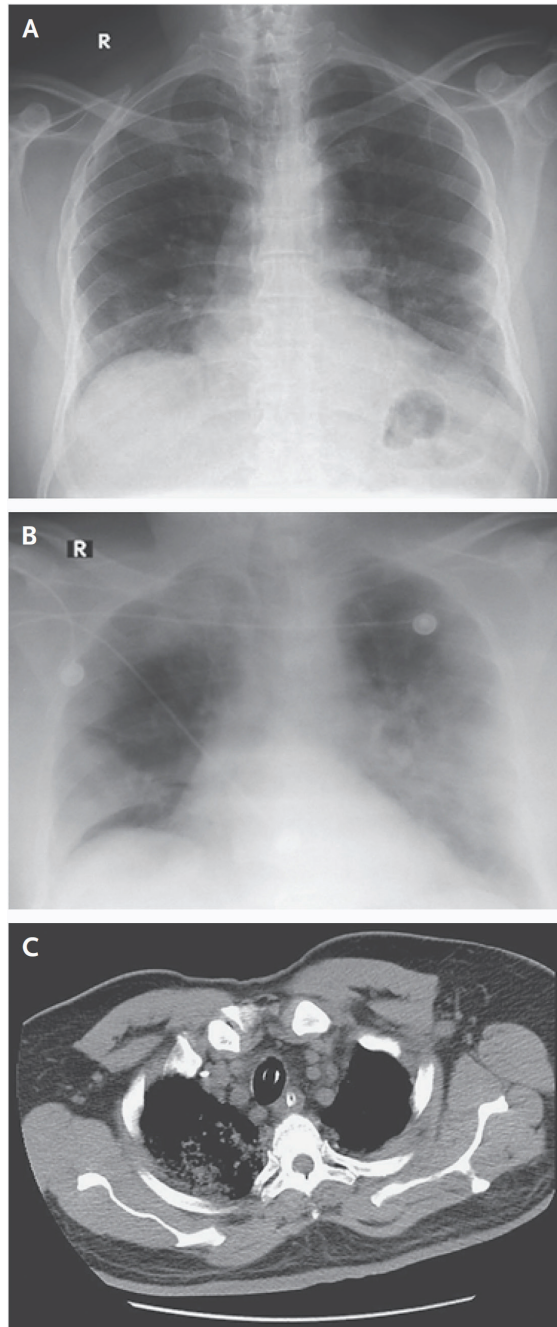


Figure 1. Abnormalities on Chest Imaging of the Patient. Shown are chest radiographs of the patient on the day of admission (Panel A) and 2 days later (Panel B) and computed tomography (CT) 4 days after admission (Panel C).

On day 1, treatment was started with oseltamivir, levofloxacin, piperacillin–tazobactam, and micafungin. On day 4, treatment with meropenem was started, since *Klebsiella pneumoniae* that was sensitive to meropenem was detected on bronchoscopy and tracheal lavage performed on day 2. *Staphylococcus aureus*, which was sensitive to a wide range of antimicrobials, was collected from a sputum sample collected on admission. *Acinetobacter* was detected in a tracheal aspirate sample collected on the day of death. No other pathogens were detected in respiratory specimens, and no bacterial growth was detected from blood samples.

On the day after admission, the patient was transferred to an intensive care unit, where he underwent intubation for mechanical ventilation. Laboratory findings obtained on admission showed normal white-cell counts, except for a relatively high percentage of neutrophils (92.5%) and a low percentage of lymphocytes (4.3%) (Table 1). Levels of liver enzymes, blood urea nitrogen, and creatinine were within the normal ranges. Somewhat elevated liver enzymes were detected on day 7 and later, with levels of alanine aminotransferase of 20 IU, 78 IU, and 47 IU per liter on days 1, 7, and 8, and levels of aspartate aminotransferase of 33 IU and 96 IU per liter on days 1 and 8, respectively. The patient tested negative for the human immunodeficiency virus; testing was not performed for pneumocystis pneumonia.

Starting on day 3 after admission, levels of blood urea nitrogen and creatinine progressively increased. Starting on day 8, the white-cell count began to rise, reaching a peak of 23,800 cells per cubic millimeter on day 10, with neutrophilia, persistent lymphopenia, and progressive thrombocytopenia. Arterial oxygen saturation ranged from 78 to 98% (Table 1). On day 11 after admission (June 24, 2012), the patient died of progressive respiratory and renal failure. A postmortem examination was not performed.

METHODS

Clinical Specimens and Viral Culture

Blood samples were collected in vacutainers with and without EDTA. Sputum samples were collected in sterile cups, after which virus transport medium was added; samples were stirred and centrifuged at 2000 rpm for 10 minutes. Supernatant was transferred to a new sterile tube and used to inoculate Vero and LLC-MK2 cells by adsorption for 1 hour at room temperature, after which 2% fetal bovine serum in minimal essential medium Eagle was added. Flasks were incubated in a carbon dioxide incubator at 37°C and observed daily for 15 days for cytopathic changes with change of medium every 3 days.

Table 1. Laboratory Data for the Patient.*

| Date | Laboratory test results | | | | | | |
|---------|-------------------------|------------------------------------|------|------|------|--------------|-----------------|
| | WBC | NEU | LYM | PLA | BUN | CRE | SO [†] |
| | | <i>x1.000 cells/mm³</i> | | | | <i>mg/dl</i> | <i>%</i> |
| June 13 | 9.3 | 8.9 | 0.4 | 168 | N.A. | 0.92 | N.A. |
| June 14 | 8.9 | 7.9 | 0.6 | 157 | N.A. | N.A. | N.A. |
| June 15 | 8.6 | 7.5 | 0.6 | 163 | 14.2 | 1.07 | 85.0 |
| June 16 | 9.6 | 8.6 | 0.6 | 149 | 27.2 | 2.14 | 84.8 |
| June 17 | 6.7 | 6.1 | 0.5 | 148 | 61.9 | 5.13 | 93.3 |
| June 18 | 8.4 | 7.5 | 0.4 | 171 | 98.1 | 7.79 | 98.8 |
| June 19 | N.A. | N.A. | N.A. | N.A. | 104 | 8.78 | N.A. |
| June 20 | 8.6 | 8.2 | 0.3 | 156 | N.A. | 9.67 | 78.1 |
| June 21 | 11.8 | 11.2 | 0.4 | 112 | 116 | 7.9 | 91.5 |
| June 22 | 21 | 20.1 | 0.4 | 85 | 94 | 7.26 | 83.7 |
| June 23 | 23.8 | 22.8 | 0.3 | 78 | 121 | 9.63 | 92.1 |
| June 24 | 22.4 | 21.3 | 0.5 | 69 | 123 | 7.3 | 87 |

* Reference ranges are as follows: white-cell count, 4×10^3 to 10×10^3 per cubic millimeter; neutrophil count, 2×10^3 to 7×10^3 per cubic millimeter; lymphocyte count, 1×10^3 to 4×10^3 per cubic millimeter; platelets, 140×10^3 to 400×10^3 per cubic millimeter; blood urea nitrogen, 8 to 26 mg per deciliter; creatinine, 0.7 to 1.2 mg per deciliter; and arterial oxygen saturation, 95 to 100%. To convert the values for blood urea nitrogen to millimoles per liter, multiply by 0.357. To convert the values for creatinine to micromoles per liter, multiply by 88.4. NA denotes not available. † All values for arterial oxygen saturation were obtained with the patient breathing supplemental oxygen; the exact values for the fraction of inspired oxygen are not known.

Viral Diagnostics

Respiratory epithelial cells from sputum were washed three times in phosphate-buffered saline (PBS), resuspended in 1 ml of PBS, and spotted on Teflon-coated slides. Slides were left to air-dry and then fixed for 10 minutes in chilled acetone. Slides were tested by indirect immunofluorescence for influenza A virus, influenza B virus, parainfluenza viruses types 1 to 3, adenovirus, and respiratory syncytial virus with the use of a Bartels Viral Respiratory Screening and Identification Kit, as described by the manufacturer (Trinity Biotech). The same procedure was used to detect viral antigens in inoculated cells after cytopathic effects had been observed. To this end, cells were scraped from tissue-culture flasks, and cells with media were transferred to a sterile centrifuge tube and prepared as described for respiratory epithelial cells from sputum. Supernatant from sputum as well as from experimentally inoculated cell cultures that displayed cytopathic effects (and uninfected cultures as negative controls) was extracted with the use of a High Pure Viral Nucleic Acid Kit, as described by the manufacturer (Roche). Extracted nucleic acids were tested by reverse-transcriptase polymerase chain reaction (RT-PCR) assay designed to detect all known paramyxoviruses [46, 60] coronaviruses [43, 61] and enteroviruses [62] and by real-time PCR for adenoviruses [63].

Viral Genome Sequencing

To sequence the PCR fragments of the pan-coronavirus PCR [43] amplicons were purified from the gel and sequenced with the use of a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and a 3130x/ Genetic Analyzer (Applied Biosystems), according to the manufacturer's instructions. To further characterize the virus genome, we used a random-amplification deep-sequencing approach. Supernatant was cleared from cellular debris by low-speed centrifugation, and virus was filtered through a 0.45- μ m centrifugal filter unit (Millipore) to minimize bacterial background. We used OmniCleave endonuclease (Epicenter) to remove free DNA and RNA, according to the manufacturer's protocol. Viral RNA was extracted from supernatants in infected cell cultures with the use of a High Pure RNA Isolation Kit (Roche). To remove mammalian ribosomal RNA, we used Ribo-Zero rRNA Removal Kit RZH110424 (Epicenter), according to the manufacturer's protocol. RNA underwent reverse transcription with the use of circular permuted primers [64] that were extended with random hexamer sequences. DNA was amplified by means of PCR with the circular permuted primers.

We sequenced the amplified fragments using the Roche 454 GS FLX sequencing platform. A fragment library was created according to the manufacturer's protocol without DNA fragmentation (GS FLX Titanium Rapid Library Preparation, Roche). The emulsion-based clonal amplification PCR (Amplification Method Lib-L) and GS junior sequencing run was performed according to the manufacturer's instructions (Roche). The sequence reads were trimmed at 30 nucleotides from the 3' and 5' ends to remove all primer sequences. Sequence reads from the GS FLX sequencing data were assembled into contig maps (a set of overlapping DNA segments) with the use of CLC Genomics software, version 4.6.1 (CLC Bio). Using the 454 sequencing platform, we obtained approximately 90% of the virus genome sequence. Subsequently, specific primers were designed to amplify overlapping fragments of approximately 800 bp by means of PCR. These PCR products were purified from the gel and sequenced with the use of a BigDye Terminator v3.1 Cycle Sequencing Kit and a 3130x/ Genetic Analyzer, according to the manufacturer's instructions.

RESULTS

Detection of a Coronavirus

The day 1 sputum sample tested negative by indirect immunofluorescence assays for influenza A and B viruses, parainfluenza viruses types 1 to 3, respiratory syncytial virus, and adenovirus. However, for a sputum sample obtained on admission, inoculation in LLC-MK2 and Vero cells resulted in cytopathic changes suggestive of virus replication (Figure 2A). Cytopathic changes consisted of syncytium formation in LLC-MK2 cells at low pH and rounding and detachment of cells at neutral or alkaline pH in Vero and LLC-MK2 cells. On passage of the culture supernatant to fresh cells, the same cytopathic effects were observed within 5 days. Virus was not isolated from a blood sample collected on admission or from a tracheal aspirate sample collected 4 days after admission.

Indirect immunofluorescence assays for the detection of influenza A and B viruses, parainfluenza viruses types 1 to 3, respiratory syncytial virus, and adenovirus were performed with the infected cell cultures, but again with negative results. In contrast, when these slides were incubated with serum samples collected from the patient 10 and 11 days after admission, the samples reacted strongly when dilutions of 1:20 were tested on immunofluorescence assay specific for IgG antibodies. No attempts were made to detect virus-specific IgM antibodies. In contrast, 2400 control serum samples collected from persons seeking medical attention at the Dr. Soliman Fakeeh Hospital in Jeddah from 2010 through 2012 remained negative in this assay. These data suggested that antibodies to an unknown virus had developed in the patient, although such antibodies were not detectable in the general population over the previous 2 years.

Real-time PCR assays specific for adenovirus, enterovirus, human metapneumovirus, and human herpesvirus types 1 to 3 yielded negative results with the use of nucleic acids extracted from the inoculated cell-culture supernatants. Furthermore, family-wide PCR assays that can detect all known paramyxoviruses [46, 60] also yielded negative results. However, family-wide PCR assays for the detection of coronaviruses [43, 61] yielded PCR fragments of the expected sizes.

Genetic Analysis of a Novel Coronavirus

The PCR fragments of the pan-coronavirus PCR [43] were sequenced. This sequence corresponded with a conserved region of open reading frame 1b of the replicase gene of a coronavirus. Reference coronavirus genome sequences were downloaded from GenBank and aligned with the amplified fragment of the newly discovered virus, hereafter called HCoV-EMC (for Erasmus Medical Center). A maximum-likelihood tree was constructed to infer the phylogenetic relationships (Figure 2B). This phylogenetic tree showed that HCoV-EMC belonged to lineage C of the genus *Betacoronavirus*, along with the bat coronaviruses HKU4 and HKU5. The *Betacoronavirus*

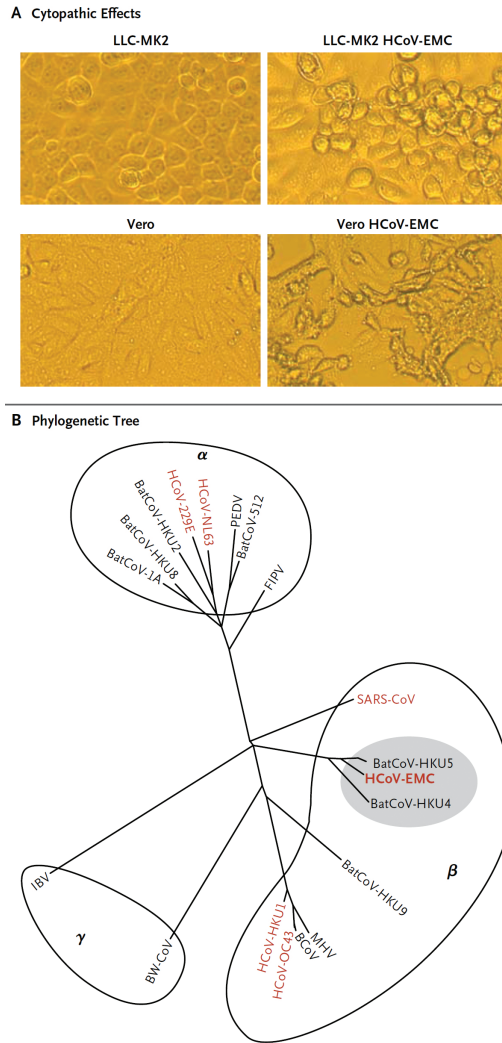


Figure 2. Cytopathic Effects and a Phylogenetic Tree of the Novel Coronavirus. Panel A shows cytopathic effects in LLC-MK2 and Vero cell cultures after inoculation with the novel coronavirus HCoV-EMC. Panel B shows the genetic relationship between HCoV-EMC and other coronaviruses in a maximum-likelihood phylogenetic tree. The tree is based on a 408-nucleotide fragment of the open reading frame 1b. The following viruses and accession numbers were used: feline infectious peritonitis virus (FCoV; NC007025), HCoV-229E (NC002645), porcine epidemic diarrhea virus (PEDV; NC003436), HCoV-NL63 (DQ445911), BatCoV-1A (NC010437), BatCoV-HKU8 (NC010438), BatCoV-HKU2 (NC009988), BatCoV-512 (DQ648858), bovine coronavirus (BCoV; NC003045), HCoV-OC43 (AY585228), HCoV-HKU1 (AY884001), murine hepatitis virus (MHV; NC006852), BatCoV-HKU5 (EF065509), BatCoV-HKU9 (EF065513), SARS-CoV (AY345988), BatCoV-HKU4 (EF065505), avian infectious bronchitis virus (IBV; NC001451), Beluga whale coronavirus (BWCoV; EU111742), and HCoV-EMC (JX869059). Alpha, beta, and gamma denote genera in the *Coronavirinae* subfamily. The five other coronaviruses that have been detected in humans are indicated in red. Lineage C of the *Betacoronavirus* genus containing HCoV-EMC is shaded in gray.

genus contains three additional lineages; A, B, and D [56]. HCoV-HKU1 and HCoV-OC43 belong to lineage A, whereas SARS-CoV belongs to lineage B. Lineage D does not contain any human pathogens and is represented in the tree by *Rousettus* bat coronavirus HKU9. Thus, HCoV-EMC is distinct from previously known human coronaviruses. HCoV-NL63 and HCoV-229E are even more distinct from HCoV-EMC, since these two human pathogens belong to a different genus, *Alphacoronavirus* (Figure 2B).

To further characterize the virus, approximately 90% of the virus genome sequence was obtained on sequence analysis with the use of the 454 platform. Subsequently, specific primers were designed to amplify overlapping PCR fragments of approximately 800 bp each for conventional Sanger sequencing. The nearly full-length sequence was obtained (GenBank accession number, JX869059), for which final annotation remained in progress at the time of this report. The HCoV-EMC virus genome encodes the open reading frames common to betacoronaviruses, including open reading frame 1ab, which encodes many enzymatic products, the spike-surface glycoprotein (S), the small-envelope (E) protein, the matrix (M) protein, and the nucleocapsid (N) protein, as well as several nonstructural genes. The genome does not encode a hemagglutinin-esterase protein, in contrast to some other betacoronaviruses.

We compared the open reading frame 1ab gene product of HCoV-EMC with those of the other betacoronaviruses, HKU4 and HKU5, to test whether HCoV-EMC might belong to one of these known species or whether it represents a new species within the genus. The International Committee on Taxonomy of Viruses (ICTV) considers that viruses sharing more than 90% of sequence identity in the conserved replicase domains belong to the same species [56]. This 90% identity threshold serves as the sole species demarcation criterion. Since the identity of amino acid sequences in these conserved domains of open reading frame 1ab between HCoV-EMC and HKU4 and HKU5 was less than 80%, we concluded that HCoV-EMC represented a novel *Betacoronavirus* species, although such classification requires formal ICTV approval.

DISCUSSION

The first decade of the 21st century has witnessed an increase in the number of coronaviruses that have been identified, along with a corresponding increase in the number of coronavirus genomes that have been sequenced. Such increases were due to the discovery of the SARS coronavirus, which resulted in a global outbreak of pneumonia in 2003 that affected persons in approximately 30 countries and resulted in about 800 deaths [65]. Before 2003, only two human coronaviruses were known, HCoV-229E and HCoV-OC43, both discovered in the 1960s [66, 67]. After the emergence of the SARS-CoV in 2003, two additional human coronaviruses were

discovered, HCoV-NL63 and HCoV-HKU1 [68-70]. Here we report the isolation and characterization of the sixth coronavirus that apparently may infect humans.

On the basis of genetic data, the ICTV has identified four virus clusters within the *Coronavirinae* subfamily, of which three represent approved genera; *Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus*. The five known human coronaviruses all belong to the genera *Alphacoronavirus* (HCoV-229E and HCoV-NL63) and *Betacoronavirus* (HCoV-OC43, HCoV-HKU1, and SARS-CoV) [43, 57, 58, 66-69]. HCoV-EMC is the first human coronavirus in lineage C of the *Betacoronavirus* genus. Its closest relatives are coronaviruses HKU4 and HKU5, isolated from *Tylosycteris pachypus* and *Pipistrellus abramus* bats, respectively [71].

As compared with other coronaviruses, HCoV-EMC was isolated and propagated relatively easily in Vero and LLC-MK2 cells. The only other human coronaviruses that replicate well in these monkey-cell lines are SARS-CoV and HCoV-NL63, which both use human angiotensin-converting enzyme 2 as their receptor. We hypothesize that one or more species of animals, possibly bats, were the reservoir host of this new coronavirus. Saudi Arabia harbors numerous bat species, including pipistrellus bats, which were found to carry BatCoV-HKU5 in Asia.

The patient's findings on chest radiography together with the clinical symptoms indicated acute respiratory distress syndrome (ARDS) with multiorgan dysfunction syndrome (MODS), similar to what has been described in severe cases of influenza and SARS [72-74]. These pneumonic changes did not respond to antibacterial treatment [75]. The patient was treated with oseltamivir for the possibility of infection with the H1N1 swine flu virus. Hematologic changes were evident in this patient in the form of lymphopenia, neutrophilia, and late thrombocytopenia. Abnormal hematologic variables were also quite common among patients with SARS. Lymphopenia was the most common finding in a cohort of 157 patients with SARS. In those patients, postmortem findings showed lymphopenia in various lymphoid organs with no features of bone marrow failure or reactive hemophagocytic syndrome [76]. The patient also had progressive impairment of renal function, similar to what had been described in some patients with SARS and possibly attributed to direct infection of renal tissue by the virus. The renal impairment in this case started on the 9th day of symptoms and progressed over the course of the patient's illness.

No symptoms were observed in the hospital among doctors and nurses caring for the patient, which suggests that the disease did not spread readily. However, staff members were not tested for antibodies against the virus for confirmation. Now that the genome sequence of HCoV-EMC has become available and rapid diagnostic tests specific for HCoV-EMC have been developed [77], thorough epidemiologic investigations are warranted. Such studies should initially focus on identifying the original source of the virus (including bats and other animal species) and

potential transmission events between the infected patient and direct contacts. The development of serologic assays for surveillance studies is important.

Three months after the hospitalization of the patient in Jeddah, it was reported that a second patient with a history of travel to Saudi Arabia who had been transferred from a hospital in Qatar to a hospital in London was infected with the same virus [78]. At present, links between the two infected patients or a potential common source of infection have not been identified. No additional cases have been identified, although several are still under investigation. Epidemiologic investigations, active case findings with the use of updated case definitions [78] and syndrome surveillance in combination with sensitive diagnostic tests will be key to monitoring the present situation and -if necessary- to intervene in a potential outbreak. It will be equally important to test whether HCoV-EMC fulfills Koch's postulates as the causative agent of severe respiratory disease.

This case is a reminder that although most infections with human coronaviruses are mild and associated with common colds, certain animal and human coronaviruses may cause severe and sometimes fatal infections in humans. Although HCoV-EMC does not have many of the worrisome characteristics of SARS-CoV, we should take notice of the valuable lessons learned during the 2003 SARS outbreak with respect to outbreak investigations and management.

ACKNOWLEDGEMENTS

Supported in part by the European Commission Seventh Framework Program for Research and Technological Development Project EMPERIE. We thank Dr. P.L.A. Fraaij for helpful discussions and comments.

Cidofovir Gel as Treatment of Follicular Spicules in Multiple Myeloma

Sander van Boheemen, MSc, Terry Jones, PhD, Barbara Muhlemann, BSc,
Mariet C. Feltkamp, PhD, Ron A.M. Fouchier, PhD and Enes Hajdarbegovic, MD.

JAMA Dermatology

ABSTRACT

Importance

The cause of follicular spicules in multiple myeloma is not known.

Observations

We present a case of follicular spicules in a multiple myeloma patient which is very reminiscent of trichodysplasia spinulosa caused by a polyomavirus. However, no trichodysplasia spinulosa-associated polyomavirus could be isolated from the skin lesions. The spicules were positive for Merkel cell carcinoma virus, which is also a polyomavirus.

Conclusions and Relevance

Follicular spicules in multiple myeloma are probably not caused by the trichodysplasia spinulosa-associated virus. Merkel cell polyomavirus could contribute to the origin of this dermatosis.

INTRODUCTION

Multiple myeloma has many cutaneous manifestations. Follicular spicules are relatively rare among them [79]. Previous papers on this dermatosis categorized it as paraneoplastic in origin. No specific treatment is available. The clinical picture is very reminiscent of trichodysplasia spinulosa which is known to be caused by the trichodysplasia spinulosa-associated polyoma virus (TSV). In this manuscript we describe a case of follicular spicules in a patient with multiple myeloma. We attempted to identify a possible viral cause and to classify this enigmatic entity.

CASE REPORT

A male patient in his seventies consulted us regarding coarse outgrowths on his face which had been present for several months (Figure 1). These lesions were asymptomatic but the patient was very concerned about his appearance. He had been treated with mid-potency topical corticosteroids for several weeks with no any apparent effect. The patient had no previous medical history but he complained of fatigue and malaise. Examination of his skin revealed small, folliculocentric, coarse, yellowish spicules located mainly on his face but also on his chest and lower arms. Ulcerated papules covered with hemorrhagic crusts were also visible on his torso. All his fingernail plates were white (Figure 2).



Figure 1. The effect of cidofovir treatment. A) Folliculocentric spicules on the face of the patient. B) Cidofovir treatment cleared the spicules from the face.



Figure 2. Leukonychia of the nails. A white discoloration appeared on the fingernails.

We considered the differential diagnosis of facial folliculocentric spicules comprising trichodysplasia spinulosa (TS) and follicular spicules of multiple myeloma (FSMM). The first is a well-characterized infection with TSV in immunosuppressed patients. The latter is seen in patients with multiple myeloma (MM) but its pathophysiology is not understood. Conventional hematoxylin and eosin staining of a skin biopsy taken on the forehead showed a widened hair follicle with a protruding spicule with hyperparakeratosis surrounded by multiple eosinophilic bodies (Figure 3). Electron microscopy of a follicular keratinocyte revealed multiple, round structures of approximately 1 μm in diameter arranged in a colon-like pattern (Figure 4). Further magnification showed these structures to consist of fibrillar bodies arranged in a paracrystalline configuration. No viral particles could be found. These findings are reminiscent of both TS and FSMM making differentiation based on these investigations alone impossible. Further serological testing showed increased serum protein of 101 g/L which was primarily composed of gamma-globulins 58.5 g/L. No antibodies against the human immunodeficiency virus were found and leukocyte levels were normal. A bone marrow biopsy revealed 50% infiltration by monoclonal plasmacytoid, CD138-positive and IgG producing cells. Fluorescence in-situ hybridization revealed an extra copy of chromosome 9 in 41% of the analyzed cells fitting the diagnosis of multiple myeloma (MM). With this information we diagnosed the patient with FSMM.

Because the etiology of FSMM is unknown, specific treatment is lacking. However, the clinical and histopathological findings in FSMM are reminiscent of TS. Identifying TSV in the lesions of our patient could provide basis for treatment. We therefore endeavored to identify a possible viral causative agent for FSMM. TS has been well characterized as a viral dermatosis caused by a polyomavirus but in case of FSMM no publications record attempts to identify the causative agent. A TSV-specific PCR [80], did not show presence of TSV in the spicules. To rule out another polyomavirus, rolling circle amplification was undertaken. However, no polyomavirus was detected using this technique.

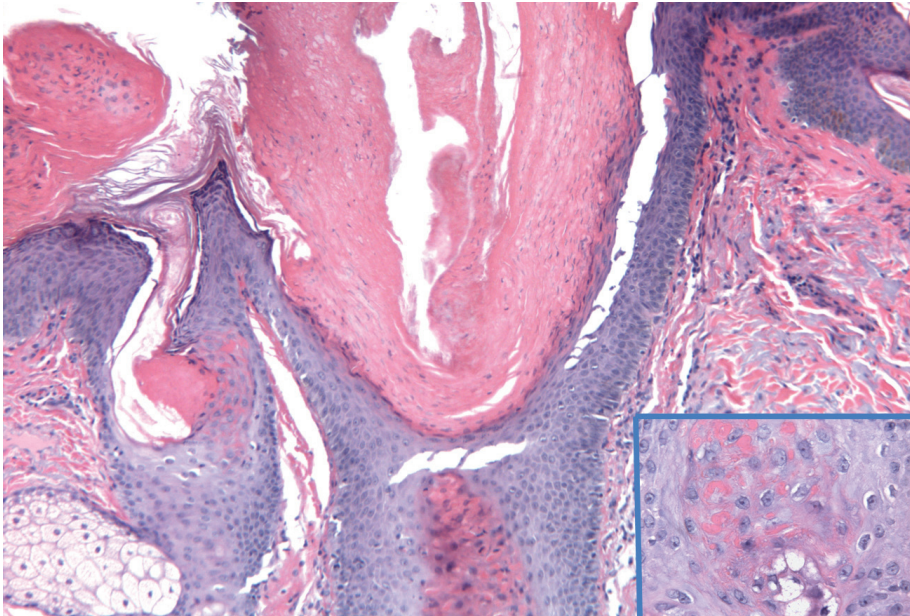


Figure 3. Histopathology of a spicule A widened hair follicle with a protruding spicule with dyskeratosis of follicular epithelium and excess inner root sheath-like cells surrounded by multiple eosinophilic bodies. Hematoxylin and eosin 40X. The interfollicular epidermis displays acanthosis (inset hematoxylin and eosin 100X, magnification of eosinophilic bodies in follicular epithelium of the infundibulum).

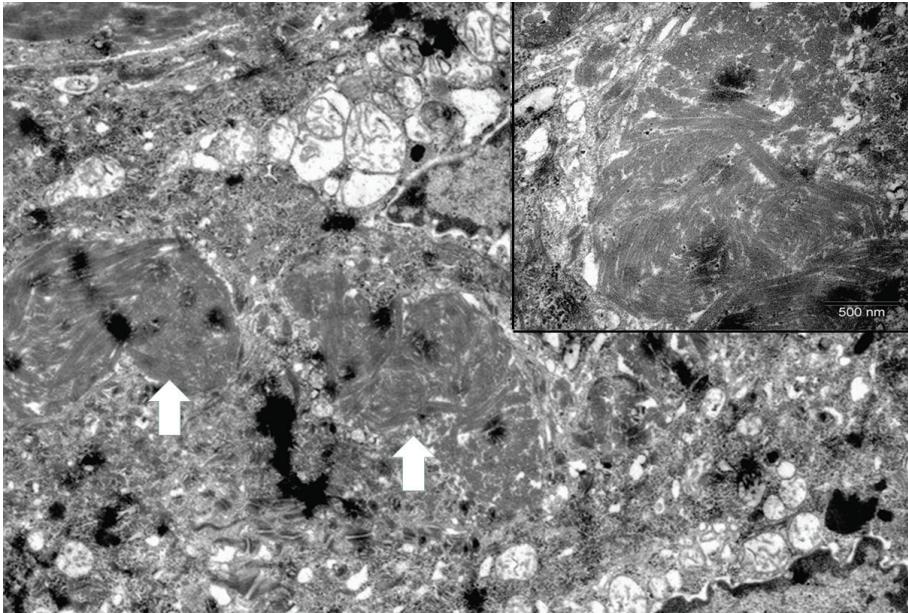


Figure 4. Electron microscopy of an affected follicle Inside a follicular keratinocyte fibrillar structures arranged in a paracrystalline configuration marked by arrows. See bars for magnification: 2 μ m for main and 500nm for inset (inset shows arrowed structures further magnified).

Having ruled out TSV as the causative agent we turned to our random amplification virus discovery approach to detect viral sequences. In this protocol, we enriched for virus particles and used a combined random-amplification and 454 deep-sequencing approach to look for both RNA and DNA viruses [81]. We used a spicule from a hair follicle to search for viral presence. The protocol for the discovery of RNA viruses and DNA viruses yielded 42.641 and 152.082 reads, respectively. After quality trimming, reads were assembled into contigs. Contigs and reads that were not assembled into contigs (so-called singletons), were submitted to the Basic Local Alignment Search Tool (BLASTn) for nucleotides. This algorithm finds regions of local similarity between the submitted sequence and known sequences in a nucleotide database, and makes calculations on statistical significance of possible matches. Sequences were classified into viruses, bacteria, and/or eukaryotes based on the taxonomic origin of the best-hit sequence using MEGAN software. The contigs from both the RNA and DNA virus discovery protocols and the singletons derived from the RNA virus discovery protocol did not result in any viral hits in the nucleotide BLAST. However, the singletons resulting from the DNA virus discovery protocol, showed 10 hits that were assigned to the polyomavirus family. All hits had the highest score to Merkel cell polyomavirus (MCV). A translated protein BLAST search based on the nucleotide sequences did not result in any additional viral hits. Using different BLAST parameters and MCV genome as a reference, another 4 reads were found to be assigned to MCV. The 14 sequence reads ranged in size from 109 to 462 nucleotides. The sequence identity of these hits to MCV ranged from 95 to 100%. Together, the 14 reads covered 1943 nucleotides of the viral genome. However, the reads that aligned to the genome were not equally distributed ($p > 0.05$). Instead, the first approximately 1800 nts of the genome were not covered by any reads. A PCR directed against the MCV Large-T gene [82] confirmed the presence of MCPyV and revealed a low average copy number of 0.1/cell. With a possible viral causative agent the patient was commenced on cidofovir 1% gel once daily for his face only. This therapy had shown success in cases of TS [83]. Within a week this lead to hemorrhagic crusts on his face and within 3 weeks it cleared his face almost completely of spicules while no effect was seen on untreated skin. Two weeks after the commencement of cidofovir treatment the patient was started on a combination of melpahalan, prednisone and bortezomib for the treatment of the patient's multiple myeloma.

DISCUSSION

Polyomaviruses are ubiquitous viruses capable of infecting humans and many animal species. Sequences from human polyomaviruses have been isolated from stool, respiratory secretions, blood, central nervous fluid, urine and skin [84]. Because of new technologies their involvement in various human diseases is easier to establish. Our findings suggest a role for a polyomavirus but not TSV in the pathophysiology of FSMM.

A recent article reported absence of TSV in a case of follicular spicules in a multiple myeloma confirming our findings [85]. Satta and colleagues suggested a role for paraprotein deposition in FSMM after finding the same electrophoresis pattern in a patient's spicules and serum paraprotein [86], although we found no immunoglobulin depositions in our patient's skin.

Although MCV cannot be appointed as the etiological agent of FSMM, several aspects do suggest a role for this virus in this dermatosis. First, the patient suffers from MM, a disease that predisposes to opportunistic infections. Second, the clinical picture resembled TS, caused by a related polyomavirus TSV. Third, the treatment with cidofovir, a viral DNA-polymerase inhibitor, cleared the spicules. This effect also seen in TS and HPV-infected skin, although polyomaviruses are not known to encode any polymerase. Although the patient was started on chemotherapy one week before the full clearing of spicules the confinement of the effect to the treated area is suggestive of some cidofovir effect. In addition, the patient had a notable response to cidofovir within one week, which was prior to the initiation of chemotherapy. And finally, MCV was discovered inside the affected tissue. Regarding the latter and the fact that MCV spread from healthy skin cannot be ruled out because asymptomatic carriage is highly prevalent in older individuals [87], further study into the observed association between MCV and FSMM as well as why cidofovir would have an effect on the lesions is warranted.

We have found evidence implicating a possible viral etiology, instead of a paraneoplastic origin for FSMM. The differential diagnosis of trichodysplasia spinulosa (TS) was rejected due to lack of presence of trichodysplasia spinulosa-associated polyomavirus (TSV) by PCR. However further research is necessary to confirm our findings.

ACKNOWLEDGEMENTS

We are indebted to Theo Bestebroer and Siamaque Kazem for excellent technical assistance, and Derek Smith and David Burke for essential discussions.

Financial Interests:

Design and conduct of the study: EMPERIE and ANTIGONE; collection, management, analysis, and interpretation of the data: EMPERIE and ANTIGONE; preparation, review, or approval of the manuscript: EMPERIE and ANTIGONE; and decision to submit the manuscript for publication: none.

Identification, Characterization, and Natural Selection of Mutations Driving Airborne Transmission of A/H5N1 virus

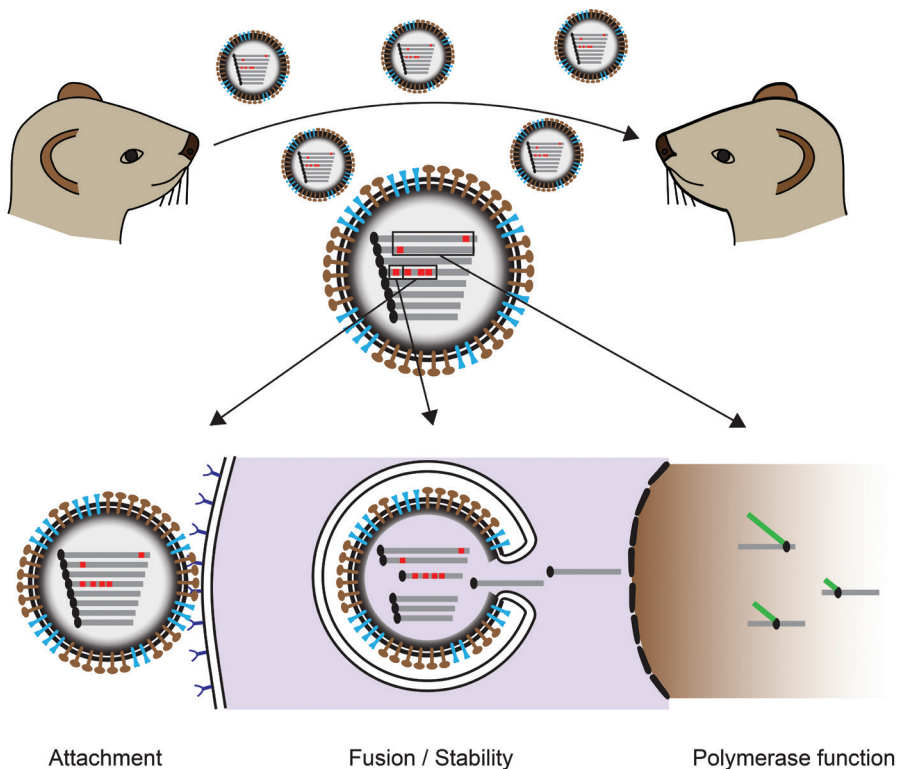
Sander van Boheemen*, Martin Linster*, Miranda de Graaf, Eefje J.A. Schrauwen, Pascal Lexmond, Benjamin Mänz, Theo M. Bestebroer, Jan Baumann, Debby van Riel, Guus F. Rimmelzwaan, Albert D.M.E. Osterhaus, Mikhail Matrosovich, Ron A.M. Fouchier, Sander Herfst.

* These authors contributed equally to this work

Cell. 2014 Apr 10;157(2):329-39.

SUMMARY

Recently, A/H5N1 influenza viruses were shown to acquire airborne transmissibility between ferrets upon targeted mutagenesis and virus passage. The critical genetic changes in airborne A/Indonesia/5/05 were not yet identified. Here, five substitutions proved to be sufficient to determine this airborne transmission phenotype. Substitutions in PB1 and PB2 collectively caused enhanced transcription and virus replication. One substitution increased HA thermostability and lowered the pH of membrane fusion. Two substitutions independently changed HA binding preference from $\alpha 2,3$ -linked to $\alpha 2,6$ -linked sialic acid receptors. The loss of a glycosylation site in HA enhanced overall binding to receptors. The acquired substitutions emerged early during ferret passage as minor variants and became dominant rapidly. Identification of substitutions that are essential for airborne transmission of avian influenza viruses between ferrets and their associated phenotypes advances our fundamental understanding of virus transmission and will increase the value of future surveillance programs and public health risk assessments.



INTRODUCTION

Since the first detection in the late 1990s [88], highly pathogenic avian influenza (HPAI) A/H5N1 viruses continue to circulate in poultry in Asia and the Middle East. Hundreds of millions of domestic birds have died as a result of infection and during culling activities to control the spread of the A/H5N1 virus. Occasional cross-species transmission events have been reported for several species of wild birds, pigs, felids, dogs, and mustelids. To date, 650 laboratory-confirmed cases of A/H5N1 virus infection in humans have been reported to the World Health Organization from 16 countries, of which ~60% had fatal outcome. Sustained human-to-human transmission has not yet been described.

However, the enzootic nature of A/H5N1 virus, its broad host range, the large number of infected hosts, and the observed accumulation of mammalian adaptive substitutions in the virus could potentially increase the risk of a future A/H5N1 virus pandemic.

For 15 years, one of the key questions for pandemic preparedness has been whether the A/H5N1 virus might acquire the ability to transmit via aerosols or respiratory droplets ("airborne transmission") among humans, a trait necessary for the virus to become pandemic. It was recently shown that a fully avian A/H5N1 virus can become airborne transmissible between ferrets [89]. Three other groups demonstrated that reassortant viruses between A/H5N1 and 2009 pandemic A/H1N1 viruses that contain the H5 hemagglutinin (HA) were also transmitted between ferrets or guinea pigs via the airborne route [90-92].

Herfst *et al.* (2012) introduced the well-known glutamic acid to lysine substitution at position 627 (E627K) of the basic polymerase 2 protein (PB2) that is associated with increased replication in mammalian cells at relatively low temperatures [93-95] and was shown to be important for airborne transmission of 1918 A/H1N1 and 1999 A/H3N2 viruses between ferrets and guinea pigs [96, 97]. In addition, two substitutions were introduced in the receptor binding site (RBS) of HA that are known to switch receptor specificity from "avian" α 2,3-linked sialic acids (α 2,3-SA) to "human" α 2,6-SA [98], glutamine to leucine at position 222, and glycine to serine at position 224 (Q222L, G224S in H5 HA numbering). These three substitutions or other polymerase and RBS mutations with similar phenotypes were found in all pandemic influenza viruses of the last century and were therefore postulated to represent minimal requirements for adaptation of animal influenza viruses to humans to yield pandemic strains [99, 100]. This "triple mutant" virus was passaged ten times in the upper respiratory tract of ferrets to yield mutant A/H5N1 viruses that were able to transmit via the airborne route between ferrets. In addition to the three substitutions introduced by reverse genetics, two substitutions in HA (H103Y and T156A) were consistently found in all transmitted A/H5N1 viruses. However, all airborne-transmitted viruses had accumulated additional substitutions. The transmissible virus with the lowest number of amino acid substitutions compared to the A/H5N1 wild-type virus had a total of nine substitutions.

Here, we describe the identification of a minimal set of substitutions required for airborne transmission of influenza virus A/Indonesia/5/05 between ferrets and provide a detailed characterization of the phenotypic changes caused by each of these substitutions. We show that the substitutions acquired upon ferret passage of the triple mutant A/Indonesia/5/05 virus emerged rapidly, which is suggestive for strong natural selection. The identification of previously unrecognized substitutions and phenotypic traits responsible for influenza virus transmission is key to increasing our fundamental understanding of airborne spread of influenza virus and may ultimately increase prognostic capabilities and diagnostic value of surveillance studies necessary for pandemic preparedness.

RESULTS

Airborne Transmission of A/H5N1 Virus between Ferrets Is Determined by a Minimum of Five Amino Acid Substitutions

To define a minimal number of substitutions in A/Indonesia/5/05 that confer airborne transmission between ferrets, transmission experiments were performed as described previously [101]. First, a recombinant virus was produced based on the consensus sequence of a previously identified virus that was airborne transmissible and that contained the lowest number of substitutions ($n = 9$) compared to the wild-type virus (PB2-E627K, PB1-H99Y, and PB1-I368V; HA-H103Y, HA-T156A, HA-Q222L, and HA-G224S; and NP-R99K and NP-S345N; [89]). This virus was transmitted to two out of two recipient ferrets, thus reproducing with a recombinant clonal virus our earlier results with a virus isolate (Figure 1A, V1). Next, we omitted either the two substitutions in nucleoprotein (NP) (V2) or the two substitutions in PB1 (V3). Whereas the recombinant virus missing two substitutions in NP was transmitted to two out of two animals, the virus missing two substitutions in PB1 was not transmitted (Figure 1A, V2 and V3). To investigate this further, the two substitutions in PB1 (H99Y and I368V) were tested individually. The virus harboring I368V in addition to the set of five substitutions consistently found in the airborne transmitted viruses (PB2-E627K, HA-H103Y, HA-T156A, HA-Q222L, HA-G224S) was not transmitted to recipient ferrets (Figure 1A, V4), whereas the addition of PB1-H99Y yielded a recombinant virus that was detected in three out of four exposed ferrets (Figure 1A, V5).

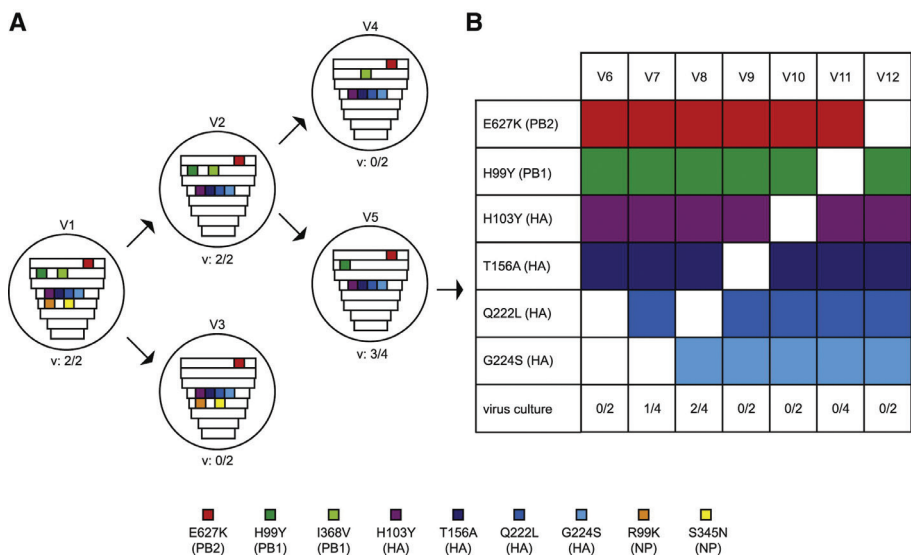
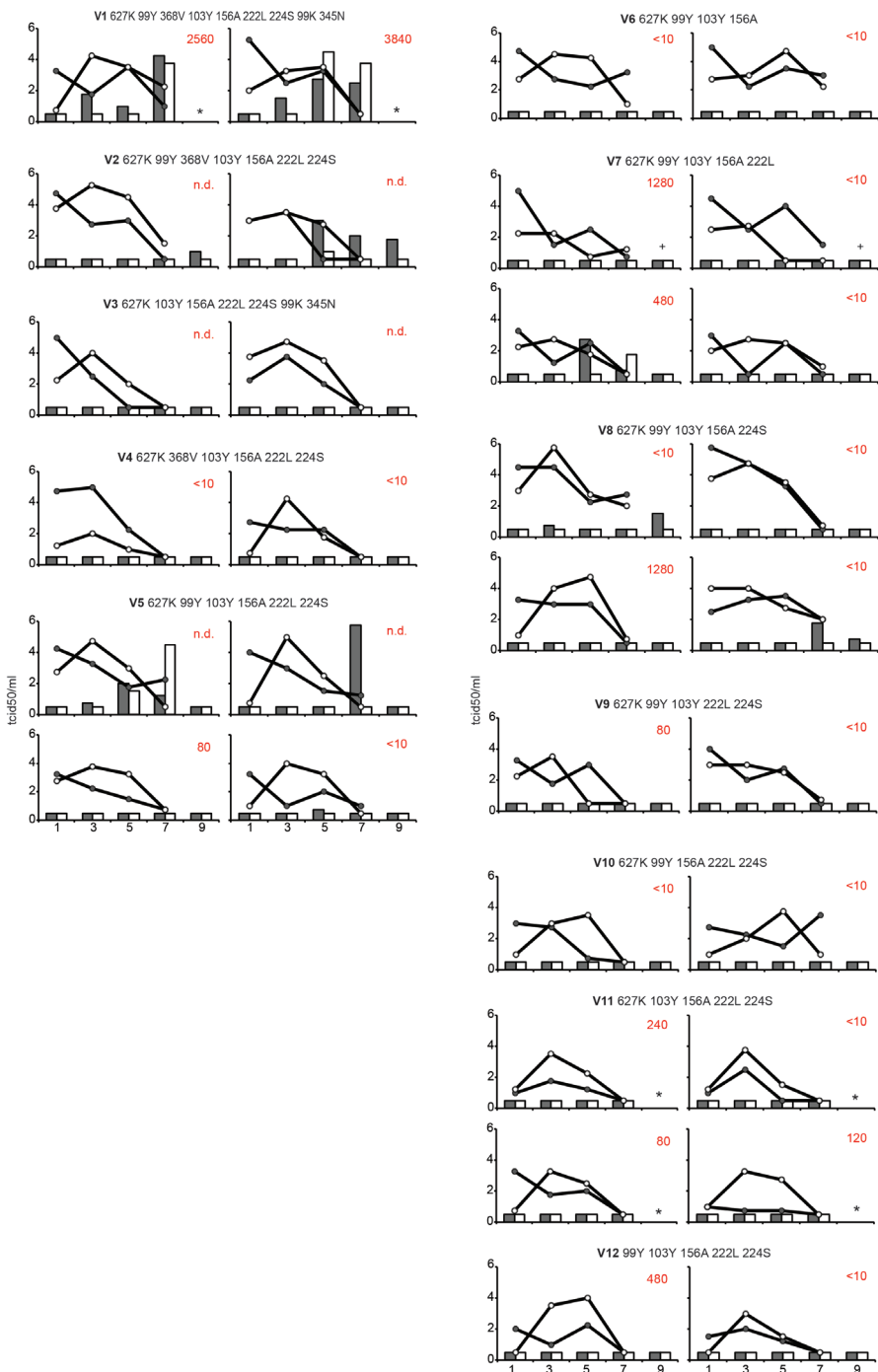


Figure 1. Summary of Results to Determine a Minimal Set of Substitutions Required for Airborne A/H5N1 Virus Transmission between Ferrets. (A) Starting with virus V1 that represents the airborne-transmitted virus with the lowest number of amino acid substitutions (nine) as compared to wild-type A/H5N1 [89], the requirement of substitutions in the PB1 and NP segments was investigated. Recombinant viruses V1-V5 are shown with eight gene segments, in which colored squares represent the presence or absence of indicated substitutions. The proportion of animals positive by virus isolation is indicated with "v." (B) All substitutions of virus V5 were omitted individually, and transmission was again tested in ferrets for viruses V6-V12. Virus-shedding patterns in donor and recipient ferrets for V1-V12 are provided in Figure S1.

Starting with virus V5 that had six substitutions compared to wild-type A/Indonesia/5/05 (HA-Q222L, HA-G224S, HA-H103Y, HA-T156A, PB2-E627K, and PB1-H99Y), all individual substitutions were omitted one by one (V7-V12). Viruses lacking either the receptor-binding substitution HA-G224S or HA-Q222L were detected in one and two out of four naive ferrets, respectively, upon exposure to inoculated ferrets (Figure 1B, V7 and V8). In contrast, when both HA-Q222L and HA-G224S were omitted, the virus was not transmitted between ferrets (Figure 1B, V6). Viruses lacking HA-T156A, HA-H103Y, PB1-H99Y, or PB2-E627K were not detected in recipient ferrets upon exposure either (Figure 1B, V9-V12). From this set of experiments, we conclude that PB2-E627K, PB1-H99Y, HA-H103Y, HA-T156A, and either HA-Q222L or HA-G224S in HA represent minimal sets of substitutions required for airborne transmission of A/Indonesia/5/05 between ferrets. Individual virus titers of nose and throat samples obtained from all donor-recipient pairs are shown in Figure S1.



Rapid Emergence of Substitutions Required for Airborne Transmission of A/H5N1 in Ferrets

We next determined at which passage substitutions H99Y in PB1 and H103Y and T156A in HA emerged during the ten repeated passages of the triple mutant A/Indonesia/5/05. To this end, primers were designed to amplify virus genome fragments covering these amino acid positions, and RT-PCR was performed on RNA isolated from nasal turbinates, lungs, and nasal swabs or washes collected from the ferrets after each passage. Amplicons were sequenced using the Roche 454 GS Junior platform.

PB1-99Y was already detected as a minor variant in the nasal turbinates in passage 1 and increased during subsequent passages to become the dominant variant in all tissues from passage 7 onward (Figure 2). HA-103Y was first detected in the nasal swabs of passage 2, was detected again in passages 5 and 6, and became the major variant from passage 7 onward. Of note, nasal washes instead of nasal turbinates were used to inoculate the ferrets from passage 6 onward, which may have contributed to rapid increase in the proportion of mutants containing HA-103Y after passage 7 (and perhaps also PB1-99Y). HA-156A was detected as a minor variant in passage 1 to rapidly become the dominant variant from passage 3 onward. Thus, the three substitutions that emerged during ferret passage and that contributed to aerosol transmissibility of A/H5N1 virus arose as early as after one or two passages and became dominant by passage 7.

Figure S1. Comparison of Airborne Transmission of A/Indonesia/5/05 Mutants in the Ferret Transmission Model, Related to Figure 1. Data for individual transmission experiments are shown in each panel, with virus shedding in inoculated and airborne recipient animals shown as lines and bars, respectively. Grey circles and bars represent shedding from the throat, white circles and bars represent shedding from the nose. The indicated virus numbers (V1 – V12) correspond to those used in Figure 1, an asterisk indicates that no samples for that day were collected. Plus signs denote the death of two animals on day 11 post exposure. These deaths appeared unrelated to the experiment since immunohistological analyses did not detect influenza virus positive cells in tissue samples collected from these animals. Red numbers on the top right of each graph represent the HA inhibiting antibody titers against A/Indonesia/5/05 HA H103Y, T156A, Q222L, G224S of sera collected from the airborne recipient animals at the end of the experiment.

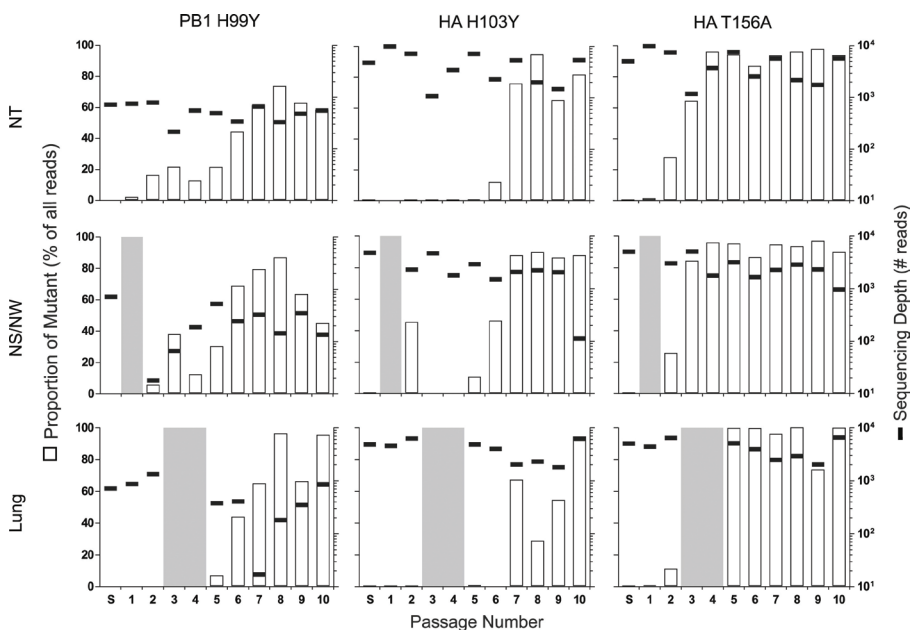


Figure 2. Single-Nucleotide Variations that Emerged upon Repeated Passaging of Influenza Virus A/H5N1_{HA} Q222L, G224S PB2 E627K in Ferrets, as Detected by Deep Sequencing. Passage number is indicated on the x axis. Left and right y axes indicated proportion of the mutant among all reads (white bars) and the sequencing depth in number of reads (black lines), respectively. The gray areas indicate that no virus sequences were amplified from these samples. S, virus stock used to inoculate the first ferret (P1); NT, nasal turbinate; NS, nasal swab; NW, nasal wash.

Nucleotide Variations in the Full Genomes of A/H5N1_{wild-type} and A/H5N1_{HA} Q222L, G224S PB2 E627K during Ferret Passaging and Transmission

We next compared the full genome sequences of viruses described in Herfst *et al.* (2012) present in passage 4, 8, and 10 nasal turbinates of ferrets inoculated with A/H5N1_{wild-type} or A/H5N1_{HA} Q222L, G224S PB2 E627K (Table S1, available at <http://www.cell.com/>). These samples were selected because they contained high copy numbers of viral RNA as determined by real-time PCR. Substitutions PB2-T23P, PA-T363P, PA-P211T, PA-L498F, NP-G462E, and NA-E239G emerged upon passage with both A/H5N1_{wild-type} and A/H5N1_{HA} Q222L, G224S PB2 E627K but did not become dominant variants. In contrast, T156A became a major variant upon passage of both viruses. Thus, whereas substitutions H99Y in PB1 and H103Y in HA emerged only upon passage of A/H5N1_{HA} Q222L, G224S PB2 E627K, HA-T156A emergence was independent of PB2-E627K, HA-Q222L, and HA-G224S, which were introduced by reverse genetics. The latter three substitutions remained dominant (>96.2%) throughout all passages tested and were not detected throughout passage of A/H5N1_{wild-type} (Table S1).

To study the effect of airborne transmission on the viral intrahost nucleotide sequence variation, we compared the full virus genome sequences present in nasal wash samples of ferrets inoculated with A/H5N1_{HA Q222L, G224S PB2 E627K} after ten passages in ferrets to nose swab samples of ferrets after two consecutive airborne transmission events (ferrets F5 and F6, Table S2, available at <http://www.cell.com/>). The number of nucleotide variations present in specimens from F5 (n = 12) and F6 (n = 16) was substantially less as compared to those in the passage 10 nasal wash sample (n = 28). Moreover, substitution PB1-99Y increased in frequency from 44.9% in passage 10 nasal wash to 100% and 73.3% in F5 and F6 after the two consecutive airborne transmission events. Similarly, HA-103Y increased from 87.6% to 100% and 100%, and HA-156A increased from 89.9% to 99% and 98.8%. These collective data are indicative of a strong selection bottleneck occurring on intrahost nucleotide sequence variation.

HA Substitutions Q222L, G224S, and T156A Affect Receptor Binding

To study the impact of the HA substitutions associated with airborne transmission on receptor preference, the attachment patterns of A/Puerto Rico/8/1934 (PR8) viruses harboring wild-type or mutant A/H5N1 HA proteins were first characterized using formalin-fixed paraffin-embedded tissue sections of ferret and human nasal turbinates, known to express “human-like” α 2,6-SA receptors abundantly. PR8 virus with the wild-type H5 HA did not attach to nasal turbinate sections, whereas the same virus with a control human H3 HA showed abundant attachment, as expected (Figure 3A). Introduction of Q222L and G224S in wild-type H5 HA resulted in abundant virus attachment to the nasal turbinate sections, comparable to the H3 HA control, as shown previously [99]. Introduction of H103Y and T156A in either the wild-type H5 HA or HA_{Q222L, G224S} did not result in obvious changes in these patterns of virus attachment. The attachment patterns of HAs to the ferret and human nasal turbinate tissue sections were indistinguishable.

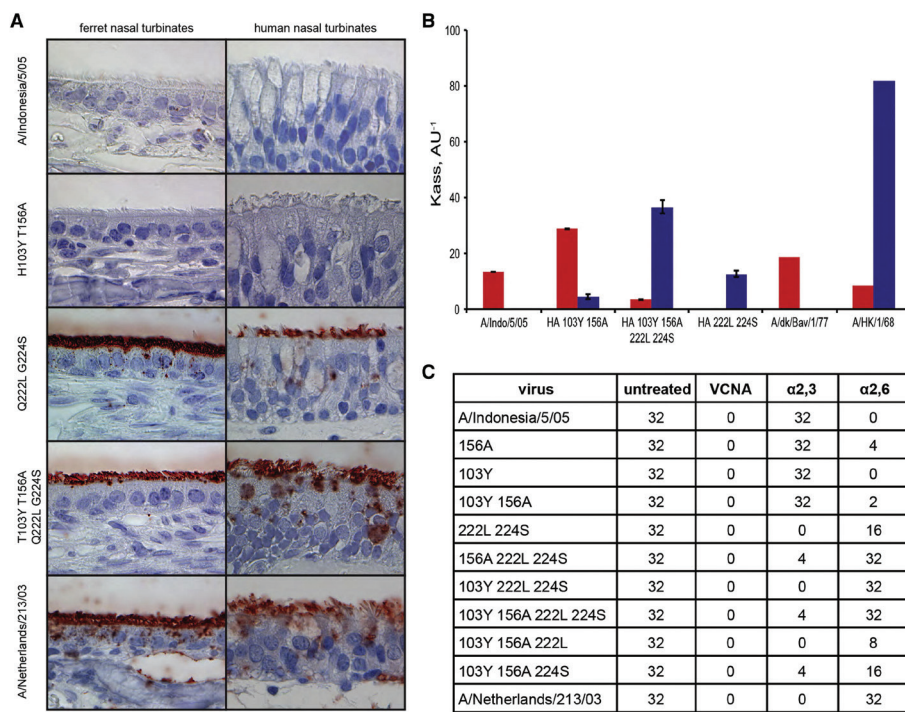


Figure 3. Receptor Binding Properties of Wild-Type and Mutant A/H5N1 HA Proteins. (A) Attachment patterns of viruses expressing wild-type or mutant H5 HA to tissue sections of ferret and human nasal turbinates. Red color represents binding of influenza viruses. Images were chosen to reflect representative attachment patterns. (B) Direct binding of viruses expressing wild-type or mutant H5 HA to fetuin containing either α 2,3-SA (red bars) or α 2,6-SA (blue bars). A/dk/Bav/1/77 and A/HK/1/68 represent avian and human prototype strains A/duck/Bavaria/1/1977 (A/H1N1) and A/HongKong/1/1968 (A/H3N2), respectively. Error bars represent the SD of the mean values ($n = 2$). See also Figure S2. (C) Agglutination of TRBCs by viruses expressing wild-type or mutant H5 HA. TRBCs were left untreated, stripped from SA using *Vibrio cholerae* neuraminidase (VCNA), or modified to contain either α 2,3-SA or α 2,6-SA. Numbers show the HA titers determined with the indicated TRBCs using various mutant viruses. A/Netherlands/213/03 served as a typical human virus with α 2,6 SA preference.

We used two solid-phase enzyme-linked receptor-binding assays to determine HA binding of virus immobilized on a 96-well plate to receptor analogs in solution: a direct binding assay and a binding inhibition assay [102]. PR8 viruses that harbor wild-type or mutant H5 HAs were tested in direct binding assays with receptor analogs 3'-fetuin and 6'-fetuin (Figure 3B). Introduction of the Q222L and G224S in HA_{wild-type} resulted in a switch in receptor binding specificity from α 2,3-SA to α 2,6-SA as expected, with no residual α 2,3-SA binding. Introduction of H103Y and T156A in HA_{Q222L, G224S} resulted in increased binding to α 2,6-SA receptors but also low binding to α 2,3-SA. Introduction of H103Y and T156A in HA_{wild-type} resulted in increased binding to α 2,3-SA and low binding to α 2,6-SA. Thus, in both the context of HA_{wild-type} and HA_{Q222L, G224S},

substitutions H103Y and T156A resulted in increased binding avidity and limited dual receptor specificity. A direct binding assay performed with 3'-sialyl-N-acetylglucosamine- and 6'-sialyl-N-acetylglucosamine-containing synthetic sialylglycopolymers rather than 3'-fetuin and 6'-fetuin yielded similar results (Figure S2A). In a fetuin-binding inhibition assay, we compared binding of the viruses to a panel of sialylglycopolymers containing several different sialyloligosaccharide moieties (Figure S2B). In this assay, the wild-type and mutant H5 viruses in general bound less avidly as compared to the control avian H1N1 virus A/Duck/Bavaria/1/1977 and human H3N2 virus A/HongKong/1/1968. HA_{wild-type} and HA_{H103Y, T156A} bound stronger to various α 2,3-SA analogs than to α 2,6-SA, whereas HA_{Q222L, G224S} and HA_{Q222L, G224S, H103Y, T156A} showed the opposite binding preference (Figure S2B), thus yielding similar results as the direct binding assays.

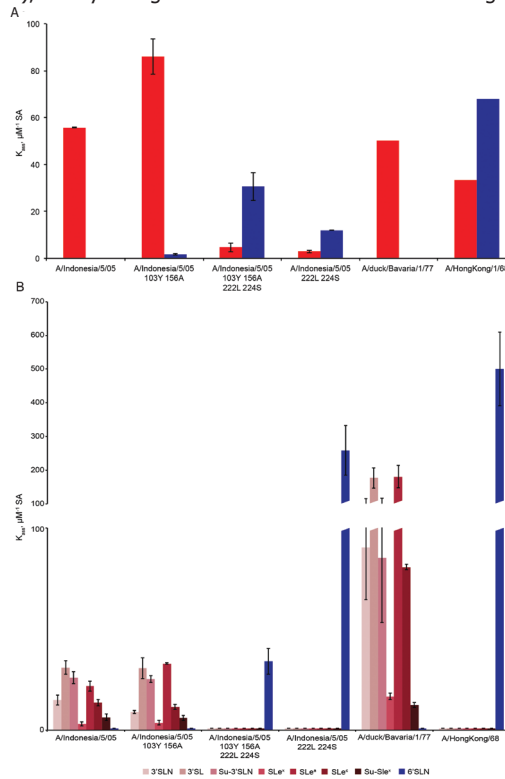


Figure S2. Receptor-Binding Properties of Wild-Type and Mutant A/H5N1 Viruses in Direct and Indirect Binding Assays, Related to Figure 3. (A) Binding of wild-type and mutant A/Indonesia/5/05, A/duck/Bavaria/1/1977 and A/HongKong/1/1968 to sialylglycopolymers containing α 2,3-SA (red bars) or α 2,6-SA (blue bars). This analysis is similar to that shown in Figure 3B using fetuin. Error bars represent the standard deviation from the mean values of two replicates. (B) Fetuin binding-inhibition assay indicating binding patterns of the mutant and wild-type A/Indonesia/5/05 viruses, A/duck/Bavaria/1/1977 (H1N1) and A/HongKong/1/1968 (H3N2) to the specified receptor analogs. Red color in different shades represents oligosaccharides containing α 2,3-SA, blue bars indicate an α 2,6-SA containing oligosaccharide. Data represent mean values and standard deviations of K_{dss} determined in the same experiment using two to four different concentrations of the inhibitor. 3'SLN: Neu5Aca2,3Gal β 1-4GlcNAc β ; Su-3'SLN Neu5Aca2,3Gal β 1-4(6-O-HSO₃)GlcNAc β ; SL^a: Neu5Aca2,3Gal β 1-4(Fuca1-3)GlcNAc β ; SL^b: Neu5Aca2,3Gal β 1-3(Fuca1-4)GlcNAc β ; SL^c: Neu5Aca2,3Gal β 1-3GlcNAc β ; Su-SL^a: Neu5Aca2,3Gal β 1-4(Fuca1-3)(6-O-HSO₃)GlcNAc β ; and 6'SLN: Neu5Aca2,6Gal β 1-4GlcNAc β .

In a third approach, we assessed the binding of PR8 viruses expressing wild-type or mutant H5 HA in a HA assay using normal turkey red blood cells (TRBC) that contain both α 2,3-SA and α 2,6-SA or modified TRBC that contain either α 2,3-SA or α 2,6-SA alone or no SA (Figure 3C). As shown previously, introduction of Q222L and G224S in H5 HA resulted in a switch in receptor binding preference from α 2,3-SA to α 2,6-SA [99]. Additional introduction of the H103Y and T156A substitutions required for airborne transmission increased the HA titers to both α 2,3-SA and α 2,6-SA containing TRBC. Each of the individual receptor-binding site substitutions Q222L and G224S in the context of changes H103Y and T156A displayed binding to α 2,6-SA containing TRBC, which is in agreement with the fact that single RBS substitutions were sufficient for airborne transmission (Figure 1B).

Introduction of H103Y in HA_{wild-type}, HA_{T156A}, HA_{Q222L, G224S}, and HA_{T156A, Q222L, G224S} did not result in consistent changes in HA titers. In contrast, introduction of T156A in HA_{wild-type}, HA_{H103Y}, HA_{Q222L, G224S}, and HA_{H103Y, Q222L, G224S} resulted in dual receptor specificity as indicated by a consistent ≥ 2 -fold increase in HA titers, irrespective of the HA used.

Collectively, we conclude from these studies that H103Y had no discernible effect on receptor binding preference, whereas T156A increased overall virus binding to both α 2,3-SA and α 2,6-SA, thus resulting in dual receptor specificity.

HA Substitution H103Y Affects Acid and Temperature Stability

Upon virus attachment to SA receptors on the cell surface and internalization into endosomes, a low-pH-triggered conformational change of HA mediates fusion of the viral and endosomal membranes to release the virus genome in the cytoplasm [103]. We measured the pH threshold required for fusion of wild-type and mutant HAs. Vero cells were transfected with HA-expression plasmids and exposed to trypsin to cleave and activate the HA, followed by acidification of the cell culture at a pH range of 5.2 to 6.0. Visual inspection of the cell cultures for the presence of syncytia (multinucleated cells) was used to determine the pH threshold required for fusion (Figure 4A). Fusion of HA_{wild-type} was triggered at pH ≤ 5.6 , similar to the threshold pH of the control H5 HA of A/Hongkong/156/97 (pH ≤ 5.8). HA of the control human H3N2 virus A/Netherlands/213/03 required a lower pH (≤ 5.2) for fusion to occur. These data are in agreement with the observation that avian influenza virus HAs generally trigger fusion at a higher pH than human virus HAs [104]. The three substitutions that affect receptor binding (T156A, Q222L, G224S; see above) did not result in a reduction of the threshold pH for fusion as compared to HA_{wild-type}. In contrast, upon introduction of H103Y in HA_{wild-type} and HA_{T156A, Q222L, G224S}, fusion was triggered only at pH 5.2 and lower. Moreover, HA_{H103Y, T156A, Q222L} and HA_{H103Y, T156A, G224S} also triggered fusion at relatively low pH (pH ≤ 5.2 and pH ≤ 5.4 , respectively).

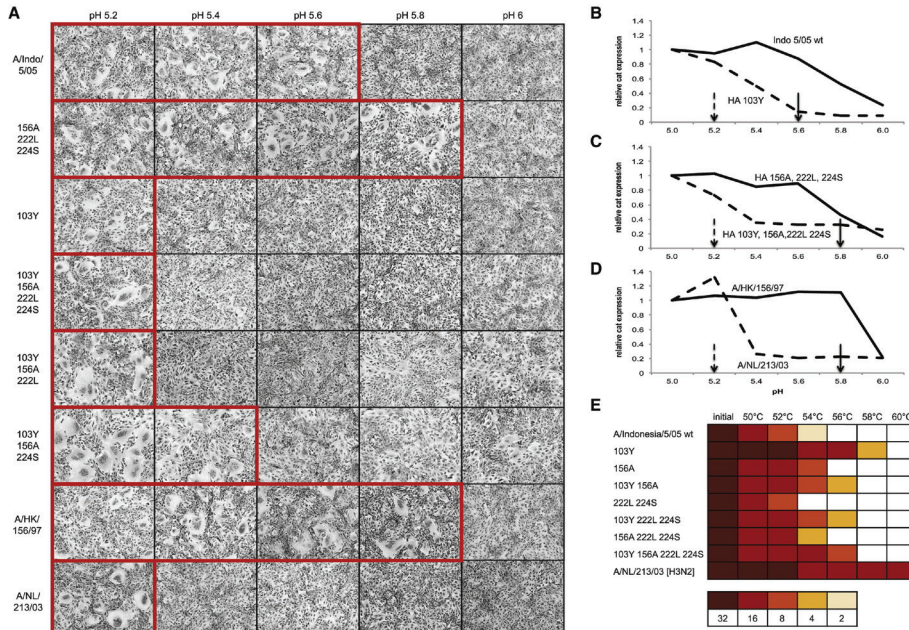


Figure 4. Analysis of pH Threshold for Fusion and Thermostability of Wild-Type and Mutant A/H5N1 HA Proteins. (A) Syncytium formation in MDCK cells upon expression of wild-type or mutant HA proteins after exposure to different pH. The red line marks the range of pH values at which fusion was detected microscopically. HA of A/HongKong/156/97 (H5N1) and A/Netherlands/213/03 (H3N2) were included as typical avian and human control viruses. (B–D) Quantification of fusion as measured by the expression of a CAT reporter gene in a cell content mixing assay for (B) influenza virus A/Indonesia/5/05 HA_{wild-type} (solid line) and HA_{H103Y} (dotted line); (C) HA_{T156A, Q222L, G224S} (solid line) and HA_{H103Y, T156A, Q222L, G224S} (dotted line); and (D) A/HongKong/156/97 (A/H5N1, solid line) and A/Netherlands/213/03 (A/H3N2, dotted line). Arrows indicate the pH threshold value at which syncytia were detected visually in (A). (E) HA protein stability as measured by the ability of viruses to agglutinate TRBCs after incubation at indicated temperatures for 30 min. Colors indicate the HA titers upon treatment at various temperatures for 30 min as shown in the legend.

As a second readout for fusion, we used a “cell content mixing assay,” in which two populations of Vero cells were transfected with HA and either a chloramphenicol-acetyl-transferase construct under control of the HIV-1 promoter (LTR-CAT) or an HIV-1 transactivator construct (pTat). Upon mixing of the two cell populations and after acidification of the cell culture at a pH range of 5.2 to 6.0, fusion was quantified by measuring CAT expression that is dependent on the expression of Tat. Also in this assay, introduction of H103Y in HA_{wild-type} and HA_{T156A, Q222L, G224S} resulted in a lower threshold pH for fusion, as indicated by the dotted lines in Figures 4B and 4C, respectively. The difference in pH required for fusion between HA with and without H103Y was similar to that

observed for the reference A/H3N2 and A/H5N1 HAs (Figure 4D).

The switch of influenza virus HA from a metastable nonfusogenic to a stable fusogenic conformation can also be triggered at neutral pH when the HA is exposed to increasing temperature. This conformational change of HA is biochemically indistinguishable from the change triggered by low pH [105] and results in a loss in the ability to bind receptor. To further investigate HA stability, PR8 viruses harboring wild-type or mutant HAs were incubated at increasing temperatures, after which the ability of the viruses to agglutinate TRBCs was quantified. The PR8 virus with the H3 HA of A/Netherlands/213/03 retained HA activity even upon treatment for 30 min at 60°C. In contrast, PR8 with H5 HA of A/Indonesia/5/05 lost HA activity upon treatment at 56°C for 30 min. Irrespective of the presence or absence of substitutions affecting receptor binding (HA_{wild-type}, HA_{T156A}, HA_{Q222L, G224S}, and HA_{T156A, Q222L, G224S} were tested), H103Y resulted in increased temperature stability as measured in the HA assay (Figure 4E).

Collectively, these data indicate that H103Y has a stabilizing effect on the HA of A/Indonesia/5/05 -with respect to both low pH and high temperature treatment- irrespective of the presence or absence of substitutions that affect receptor binding.

PB2-E627K and PB1-H99Y Affect Polymerase Activity

The influenza virus polymerase complex -consisting of the polymerase proteins PA, PB1, and PB2- transcribes the negative sense viral RNA ((-)vRNA) in mRNA and positive sense copy RNA ((+)cRNA), the latter of which is used as the template to yield newly synthesized (-)vRNA. To read out polymerase complex function, we used a (-)vRNA reporter construct consisting of the firefly luciferase open reading frame flanked by the noncoding regions of segment 8 of influenza A virus [106]. Upon cotransfection of the reporter with expression plasmids encoding PB1, PB2, PA, and NP, the (-)vRNA reporter is transcribed and the firefly luciferase protein is expressed. A plasmid that constitutively expresses *Renilla* luciferase was cotransfected as an internal control to standardize transfection efficiency and sample processing. Substitution E627K in PB2 resulted in a 12-fold increase in firefly luciferase expression as compared to the wild-type PB2. Substitution H99Y in PB1 resulted in a 3-fold increase. The polymerase complex with both PB2-E627K and PB1-H99Y yielded a 25-fold increase in firefly luciferase expression as compared to the polymerase complex of the wild-type virus (Figure 5A).

To study the levels of transcription of (-)vRNA, (+)cRNA, and mRNA during the course of a viral infection, we performed primer extension assays using total cellular RNA isolated upon virus inoculation of MDCK cells. Three hours after inoculation with A/H5N1_{wild-type}, all three RNA species were detected (Figure 5B). Introduction of E627K in PB2 resulted in elevated levels of the viral RNAs, predominantly increasing the amount of (-)vRNA and (+)cRNA, while marginally changing mRNA. Substitution H99Y in PB1 resulted in slightly reduced levels of all detected

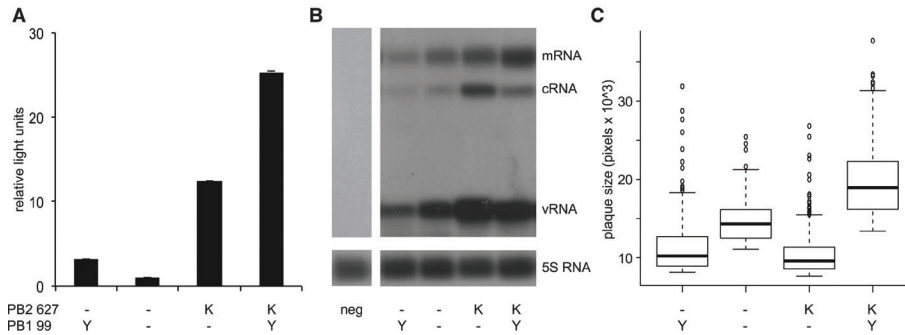


Figure 5. Effect of H99Y in PB1 and E627K in PB2 on Polymerase Activity and Virus Replication. (A) Minigenome reporter assay. Plasmids encoding PB2, PB1, PA, and NP were cotransfected with a vRNA reporter encoding firefly luciferase. Luminescence of firefly luciferase was standardized using a plasmid constitutively expressing *Renilla* luciferase. Results are calculated as relative light units (firefly luciferase/*Renilla* luciferase) and plotted as fold increase over wild-type. Error bars indicate the SD from the average of two independent experiments. (B) Primer extension assay. MDCK cells were inoculated at an MOI of 1 with wild-type or mutant A/Indonesia/5/05 viruses. At 3 hr postinoculation, cells were lysed, and viral RNA levels were determined by primer extension analysis using primers specific for mRNA, cRNA, or vRNA of PB1. Determination of the 5S RNA levels served as an internal loading control. (C) Plaque assay. MDCK cells were inoculated with A/Indonesia/5/05 viruses containing the indicated substitutions. After 48 hr, plaque formation was visualized by influenza NP-specific staining. Digital images were analyzed using ImageQuant TL software to determine plaque size. The surface of individual plaques in pixels is plotted using boxplots indicating the median value and quartiles. Differences between individual groups were significant ($p < 0.01$) in Student's *t* test. In all panels, the mutations in PB1 and PB2 are indicated, with dashes representing the wild-type sequences.

RNA species as compared with A/H5N1_{wild-type}. Upon inoculation with a virus containing both PB2-E627K and PB1-H99Y, the ratio between (–)vRNA, (+)cRNA, and mRNA was similar as observed for A/H5N1_{wild-type} but with higher overall levels of all three RNA species.

As a third test of polymerase function, we studied in vitro replication of the recombinant A/H5N1 viruses with and without polymerase substitutions by measuring plaque sizes upon inoculation of MDCK cells. At 48 hr after inoculation, cells were fixed and stained with an anti-NP antibody, and the number of pixels representing ~100 to 500 plaques on digital images was quantified. As compared to A/H5N1_{wild-type}, A/H5N1_{PB2 E627K} and A/H5N1_{PB1 H99Y} displayed a slight reduction in plaque size. However, combination of these two substitutions (A/H5N1_{PB2 E627K, PB1 H99Y}) yielded plaques that were significantly larger than those observed for A/H5N1_{wild-type} (Figure 5C).

From these assays, we conclude that PB2-E627K and PB1-H99Y collectively resulted in increased levels of (–)vRNA, (+)cRNA, and mRNA transcription and increased virus replication in MDCK

cells.

DISCUSSION

Here, we show that, of the nine substitutions observed in an airborne transmissible A/H5N1 virus [89], two alternative sets of five mutations are identified (E627K in PB2; H99Y in PB1; H103Y, T156A, and either Q222L or G224S in HA), either of which is sufficient to confer ferret transmissibility on A/Indonesia/5/05. Keeping in mind that the design of this ferret transmission model is qualitative rather than quantitative and that such studies need to take the principles of replacement, reduction, and refinement in animal experiments into account [107, 108], it should be noted that the present study is limited by the number of animals that were used and was purposely designed to define a minimal set of substitutions rather than the definitive minimal set of substitutions required for airborne transmission in ferrets. Starting with the recombinant A/H5N1 virus harboring nine substitutions, omission of two substitutions in NP still yielded virus transmission to five out of six airborne-exposed ferrets (V2 and V5). Subsequently, a virus lacking PB1-H99Y (V4) was not detected by virus isolation and serology in two out of two airborne-exposed ferrets, whereas a virus lacking PB1-I386V was detected by virus isolation in three out of four airborne-exposed ferrets, thus providing evidence that H99Y and not I386V was required for airborne transmission. In the subsequent experiment, all individual substitutions from the remaining set of six (PB2-E627K, PB1-H99Y, HA-H103Y, HA-T156A, HA-Q222L, and HA-G224S; Figure 1B) were eliminated one by one. Viruses lacking either HA-Q222L or HA-G224S still resulted in airborne transmission in two and one of four ferrets tested, indicating that a single receptor-binding site substitution was sufficient for transmission. In contrast, in transmission experiments using viruses from which each of the other single substitutions were eliminated, virus was not detected in exposed ferrets. However, some of these exposed animals seroconverted despite a lack of virus detection. Apparently, although transmission may occur for some viruses with five substitutions as measured by serology or single time points of virus detection (Figure S1), the viruses were insufficiently replication competent to cause robust infection and seroconversion consistently in the exposed ferrets. As a consequence, refining the minimal set of substitutions required for airborne transmission would require substantial numbers of ferret pairs, given the statistical considerations for this type of experiment [107, 109].

The substitutions required for airborne transmission in ferrets were determined in *in vitro* assays to either affect HA binding to receptors, HA stability, or activity of the polymerase complex. Two substitutions introduced by reverse genetics (Q222L and G224S) are known to change the receptor binding preference of the H5 HA from avian-like $\alpha 2,3$ -SA to human-like $\alpha 2,6$ -SA [99]. This change in receptor binding preference, either through single or double substitutions, was required for airborne transmissibility, which is in agreement with loss-of-function transmission studies using 1918 A/H1N1 and 1957 A/H2N2 viruses [110, 111]. Although HAs with single Q222L

or G224S substitutions were less efficient in binding $\alpha 2,6$ -SA-containing TRBC as compared to the double mutant, viruses with the single RBS substitutions were still transmissible. HA substitution T156A increased virus binding to both $\alpha 2,6$ -SA and $\alpha 2,3$ -SA in quantitative binding assays. Imai *et al.* (2012) showed that substitution N154D in HA also affected transmission of a reassortant H5 virus. T156A and N154D result in the loss of the same glycosylation site in HA, suggesting that loss of this glycosylation site rather than the specific amino acid substitutions were important for the change in phenotype. Loss of glycosylation in H5 HA in combination with substitutions Q222L and G224S was previously shown to enhance virus replication in ferrets [112]. The airborne-transmissible H5 virus of Imai *et al.* (2012) also contained Q222L in HA but had N220K as a second substitution in the RBS. In a third study, Q222L/G224S, along with Q192R in the context of A/Vietnam/1203/04 HA, also resulted in slightly increased transmission in ferrets [90]. These studies thus indicate that changes in receptor specificity critically contribute to airborne transmission of H5 viruses in ferrets. Affinity measurements using A/Indonesia/5/05, A/Vietnam/1203/04 and A/Vietnam/1194/04 HAs with substitutions associated with airborne transmission revealed a binding preference for human receptors (Figure 3) [113-116]. Structural studies of HA further showed that human and avian receptor analogs were bound in the RBS in the same “folded-back” conformation as seen for HA from H1, H2, and H3 pandemic viruses, which is distinct from typical avian virus HAs, including H5 [114-116]. In these studies, the affinity of the mutant H5 HA was relatively low as compared to HA of human H2 and H3 viruses [115-118]. Although it was speculated that an N-linked glycan at the tip of HA might sterically hinder SA binding [115], direct evidence from structural studies on HA of the airborne transmissible viruses is still lacking.

Arguably, one of the more intriguing findings of the Imai *et al.* (2012) and Herfst *et al.* (2012) studies is the requirement of HA mutations that affect stability in terms of temperature and pH. H103Y in A/Indonesia/5/05 HA resulted in increased temperature stability and requirement of lower pH treatment to trigger membrane fusion, similar as described for T315I by Imai *et al.* (2012). H103Y was recently shown to increase the thermostability of HA [113], and temperature-dependent circular dichroism spectroscopic experiments revealed hydrogen bond formation between 103Y and 413N in adjacent monomers that stabilized the trimeric protein [116]. In contrast, T315I stabilized the position of the fusion peptide within the HA monomer [115], indicating that multiple mechanisms can lead to increased HA stability. Galloway *et al.* (2013) suggested that an altered pH of fusion may be associated with virus adaptation to new hosts. Furthermore, substitutions that decrease the pH of fusion increased virus replication in the upper respiratory tract of ferrets and in mice [119-121]. Although the contribution of H103Y and T315I to increase airborne transmission between ferrets may be related to the pH of fusion or thermostability, these properties may merely be a surrogate for another -as yet unknown- phenotype, such as stability of HA in aerosols, resistance to drought, stability in mucus, or altered pH in the host environment.

The requirement of increased polymerase activity to yield an airborne transmissible H5 virus was also expected [100]. PB2 E627K has been identified as a major determinant of host adaptation of pandemic influenza viruses [93-97]. Here, E627K also resulted in increased polymerase activity. More importantly however, we identified a substitution acting in concert with E627K to increase polymerase activity. Like PB2 E627K, PB1 H99Y alone resulted in increased polymerase activity in minigenome assays but decreased virus replication. When combined, these two substitutions had a synergistic effect in minigenome assays and increased virus replication. In primer extension assays, PB2 E627K predominantly caused an increase in vRNA and cRNA, changing the ratio between RNA replication and mRNA transcription. Addition of PB1 H99Y lowered vRNA and cRNA while further increasing mRNA levels, yielding a similar ratio as observed for A/H5N1_{wild-type} but at overall increased levels. Here, we thus describe an “adaptive” substitution in PB1 that potentially improved the levels of, and balance among, vRNA, mRNA, and cRNA in concert with PB2-E627K in A/Indonesia/5/05. Beyond HA, only these two amino acid substitutions were required to generate airborne-transmissible A/H5N1 virus. Such conclusion cannot be obtained from the A/H5 virus transmission experiments that use reassortant viruses. Zhang *et al.* (2013b) showed that the PA or NS1 genes of a pH1N1 virus contributed to airborne transmission of reassortant A/H5N1 viruses in guinea pigs. At present, it is unclear how the guinea pig and ferret models compare with respect to A/H5 virus transmission studies.

Although some substitutions leading to an airborne phenotype have been observed in nature, the required combination of substitutions has not yet been detected [89, 122, 123]. Keeping in mind that the ferret model may not be predictive for airborne A/H5N1 virus emergence in humans, it is of interest to note that, upon acquisition of PB2 E627K and the receptor binding changes Q222L/G224S, the additional substitutions associated with the airborne phenotype of A/H5N1 emerged within only one or two passages in ferrets, became dominant after six passages, and appeared to be selected for during transmission events. Other sets of mutations than those identified here that similarly result in altered receptor specificity and stability of HA and polymerase function in mammalian cells may also be sufficient to increase airborne transmission. The specific mutations to increase transmission may be dependent on virus strain or subtype [124]. Given that some of the required substitutions (PB2 E627K and HA T156A) are commonly found in field isolates and that functionally equivalent substitutions exist for most of the identified substitutions, emergence of transmissible A/H5N1 influenza A viruses in nature cannot be excluded [123]. In this light, the A/H7N9 virus outbreak in China is a cause of concern, as some A/H7N9 field strains contain E627K in PB2 and Q222L in HA, and this H7 lineage lacks a glycosylation site at the tip of HA [125]. These viruses were shown to have increased preference for α 2,6-SA and decreased binding to α 2,3-SA [126-128]. Although airborne A/H7N9 virus transmission in ferrets was shown to be limited [127, 129-131], it cannot be excluded that fully avian viruses adapt upon repeated passage in mammals to gain transmissibility. In analogy to H5, it can be speculated that the A/H7N9 virus

needs to acquire increased binding preference for $\alpha 2,6$ -SA over $\alpha 2,3$ -SA, increased HA stability, and increased polymerase activity. For both A/H5N1 and A/H7N9, appropriate surveillance for emergence of mutations that affect HA receptor binding, HA stability, polymerase activity, and transmission is thus warranted. It has been argued that different sets of substitutions may lead to similar virus phenotypes, and hence, sequence-based virus surveillance may be misleading. However, such surveillance may be improved by including phenotyping assays -using relatively simple methods as described here- in the future. Such surveillance may be further improved by deep sequencing, in addition to sequencing consensus virus genomes.

Although it is clear that studies in ferrets may not be predictive for influenza virus outbreaks in humans, the ferret transmission model is one of the best models for influenza available today [132]. Experiments like the ones presented here are crucial to increase our basic understanding of airborne virus transmission, as such knowledge is currently very limited. The results of this study do not imply that an H5 influenza pandemic is imminent but warrant an intensified and broadened approach to detect emerging influenza viruses early and take immediate action once viruses naturally gain functions that might enable them to become a pandemic threat.

EXPERIMENTAL PROCEDURES

Biocontainment

All experiments involving highly pathogenic A/H5N1 viruses were conducted at enhanced animal biosafety level 3 (ABSL3+). The ABSL3+ facility of Erasmus MC consists of a negative pressurized (-30 Pa) laboratory in which all in vivo and in vitro experimental work is carried out in class 3 isolators or class 3 biosafety cabinets, which are also negative pressurized (<-200 Pa). Although the laboratory is considered "clean" because all experiments are conducted in closed class 3 cabinets and isolators, special personal protective equipment, including laboratory suits, gloves and FFP3 facemasks is used. Air released from the class 3 units is filtered by High Efficiency Particulate Air (HEPA) filters and then leaves the facility via a second set of HEPA filters. Only authorized personnel that received the appropriate training can access the ABSL3+ facility. All personnel working in the facility is vaccinated against seasonal and A/H5N1 influenza viruses. For animal handling in the facilities, personnel always work in pairs. The facility is secured by procedures recognized as appropriate by the institutional biosafety officers and facility management at Erasmus MC and Dutch and United States government inspectors. Antiviral drugs (oseltamivir and zanamivir) and

personnel isolation facilities are directly available to further mitigate risks upon incidents.

Publication Considerations

Prior to submission of the manuscript, all authors read and commented on all aspects of the study (design, realization, scientific value, ethical considerations, integrity of data, presentation of experimental outcomes, biosafety and biosecurity aspects etc.). We concluded that research involving airborne transmissible A/H5N1 influenza viruses must be considered DUR (dual-use research), but that the present study does not qualify as DURC (dual-use research of concern), since no increase in virulence or transmissibility compared to previously published studies is described. Furthermore, we also cannot foresee any direct misapplication of the presented work resulting in a threat to public health. In contrast, the gain of knowledge and future applications of this work will be beneficial for science and society.

The manuscript was sent for clearance to the Biosafety Office of Erasmus MC, representing the institutional Biosafety Advisory Board. Subsequently, we requested (under formal protest) an export license issued by the Dutch government for communication of the manuscript to NIAID (National Institutes of Allergy and Infectious Diseases) and the publisher. Next, members of the Biosecurity Committee and the Ethical Advisory Board of the EU FP7 program “ANTIGONE” endorsed publication of the manuscript and delivered comments that were carefully incorporated in the manuscript. NIAID as another funding agency requested an appendix 5 procedure to account for possible biosafety and biosecurity risks and to develop a communication plan. The final version of the manuscript was sent to the funders and all individuals involved in the evaluation process, after which the manuscript was submitted for publication.

Cells

Madin-Darby Canine kidney (MDCK) cells were cultured in Eagle’s minimal essential medium (EMEM, Lonza, Breda, the Netherlands) supplemented with 10% fetal calf serum (FCS), 100 IU/ml penicillin, 100 µg/ml streptomycin, 2 mM glutamine, 1.5 mg/ml sodiumbicarbonate (Lonza), 10 mM HEPES (Lonza), and non-essential amino acids (MP Biomedicals Europe, Illkirch, France). 293T cells were cultured in Dulbecco modified Eagle’s medium (DMEM, Lonza) supplemented with 10% FCS, 100 IU/ml penicillin, 100 mg/ml streptomycin, 2 mM glutamine, 1 mM sodium pyruvate, and non-essential amino acids. Subclone 118 of Vero-WHO cells (Vero-118) was cultured in Iscove’s modified Dulbecco’s medium (IMDM; BioWhittaker) supplemented with 10% fetal calf serum (FCS), 100 IU/ml penicillin, 100 µg streptomycin and 2 mM glutamine [133].

Viruses

Influenza virus A/Indonesia/5/05 (A/H5N1) was isolated from a human case of HPAI virus infection and passaged once in embryonated chicken eggs followed by one passage in MDCK cells. All eight gene segments were amplified by reverse transcription polymerase chain reaction and

cloned in a modified version of the bidirectional reverse genetics plasmid pHW2000 [99, 134]. Substitutions of interest (H103Y, T156A, Q222L and G224S in HA, E627K in PB2, R99K and S345N in NP and H99Y and I368V in PB1) were introduced by reverse genetics using the QuikChange multi-site-directed mutagenesis kit (Stratagene, Leusden, Netherlands) according to the instructions of the manufacturer. Recombinant viruses were produced upon transfection of 293T cells and virus stocks were propagated and titrated in MDCK cells. For binding assays, and stability assays, reassortant viruses consisting of seven gene segments of influenza virus A/PR/8/34 and the HA segment of A/H5N1 were produced using a previously described reverse genetics system for influenza virus A/PR/8/34 [134]. For binding studies, viruses were inactivated by treatment with β -propiolactone (BPL, Ferak Berlin) [135], concentrated by pelleting through a 20% sucrose cushion and stored in 0.02% sodium azide (Sigma-Aldrich).

Virus Titration in MDCK Cells

Virus titers were determined by end-point titration in MDCK cells in 96-well plates (Greiner Bio-One). MDCK cells were inoculated with tenfold serial dilutions of virus stocks, nose swabs, or throat swabs. Cells were washed with PBS one hour after inoculation and cultured in 200 μ l of infection media, consisting of EMEM supplemented with 100 U/ml penicillin, 100 μ g/ml streptomycin, 2 mM glutamine, 1.5 mg/ml sodiumbicarbonate, 10 mM HEPES, non-essential amino acids, and 20 μ g/ml trypsin (Lonza). Three days after inoculation, supernatants of infected cell cultures were tested for agglutinating activity using TRBCs as an indicator of virus replication in the cells. Infectious virus titers were calculated from 4 replicates (nose swabs, and throat swabs) or 10 replicates (virus stocks) by the method of Spearman-Kärber.

Airborne Transmission of A/H5N1 Virus in the Ferret Model

An independent animal experimentation ethical review committee (Dutch Stichting Dier Experimenten Commissie Consult) approved all animal studies. 64 female ferrets between 1 and 2 years of age were obtained from an accredited ferret breeder. All animals were tested for the presence of antibodies against A/H5, A/H3 and A/H1 prototype influenza A viruses and Aleutian Disease Virus, were microchipped and received hormonal treatment to prevent estrus.

Aerosol or respiratory droplet transmission experiments were performed as described previously [101]. In short, 2 or 4 seronegative female adult ferrets (*Mustela putorius furo*) were inoculated intranasally with 10^6 TCID₅₀ of virus by applying 250 μ l of virus suspension to each nostril. Each ferret was then placed in a transmission cage. One day after inoculation, one naive ferret was placed opposite to each inoculated ferret. Each transmission pair was housed in a separate transmission cage designed to prevent direct contact between the inoculated and naive ferrets but allowing airflow from the inoculated to the naive ferret. Nose and throat swabs were collected on 1, 3, 5, and 7 days postinoculation (dpi) for inoculated ferrets and on 1, 3, 5, 7, and 9 days post exposure (dpe) for the naive ferrets. Virus titers in swabs were determined by end-point titration

in MDCK cells.

Clinical manifestations observed in animals upon inoculation of a virus suspension or infection via aerosols or respiratory droplets included ruffled fur, loss of appetite, and lethargy. Two animals that died in the course of the experiments were examined for the presence of influenza virus in respiratory organs and were found to be free of detectable levels of viral protein. We concluded that these ferrets died due to reasons unrelated to the effect of virus infection. No animals required withdrawal from the study on animal welfare grounds. All animals were humanely killed at the end of the in-vivo phase of the study.

Statistical Considerations for Ferret Transmission Experiments

Given that airborne transmission in ferrets has never been observed for A/Indonesia/5/05 or any other avian A/H5N1 virus (Table S3, available at <http://www.cell.com/>), we define every single event of airborne transmission of mutant A/Indonesia/5/05 virus as detected by the presence of virus in naive ferrets as “airborne transmission”. The use of small group sizes could result in an underestimation of airborne transmission, e.g., when 0 of 2 animals show airborne transmission, it could be possible that upon testing of another 2 or 4 animals, airborne transmission would be detected. Thus, while “not transmissible” does not mean “will never transmit,” “transmissible” is a clearly defined virus phenotype.

In support of this and other statistical considerations, we performed an extensive analysis of influenza A virus airborne transmission studies in ferrets, extending beyond the analysis by Belser *et al.* (2013b) and discussions therein. A search for published literature via <http://www.ncbi.nlm.nih.gov/pubmed/> using search terms “influenza,” “transmission,” and “ferret” was performed on 27-12-2013. Studies that did not report airborne transmission between ferrets based on virus replication data or that tested swine viruses exclusively were discarded from the set. From the total of 48 full-text articles that were selected, we report the wild-type viruses that were tested, the group size of the experiment and the number of airborne transmission events that were reported (Table S3, available at <http://www.cell.com/>). From this analysis, it was clear that wild-type A/H5N1 viruses have never been shown to transmit via the airborne route between ferrets (0/28 ferret pairs in 7 published studies) and that the same holds true for various A/Indonesia/5/05 strains (0/8 pairs). Moreover, other avian influenza viruses (with the exception of 2013 A/H7N9 viruses) were shown to be not transmitted between 73 ferret pairs tested in 10 studies, with the exception of 1 study reporting transmission of a single avian A/H1N1 virus. In sharp contrast, human influenza viruses were transmitted in 132/149 ferret pairs in 32 published studies. Despite the small group sizes in most published studies, human influenza viruses were transmitted in 32/32 studies, while avian A/H5N1 viruses were transmitted in 0/7 studies. The 2013 A/H7N9 virus (which has several mammalian adaptation markers) was transmissible in 19 of 44 ferret pairs in 6 published studies, clearly intermediate in transmissibility between avian and human

influenza viruses. Despite small group sizes, also the results with A/H7N9 virus were comparable (transmission in 1/3, 4/8, 9/18, 3/9, 1/3, 1/3 ferret pairs in the individual studies). Thus, although the statistical power of transmission studies as performed here is insufficient for a quantitative analysis of transmission efficiency due to small group size, sufficiently robust data are generated to draw qualitative conclusions. We can therefore interpret any event of airborne transmission of mutant A/H5N1 virus as an indicator of increased transmission over wild-type A/H5N1 (never found to be transmitted), while we cannot draw strong conclusions about mutant A/H5N1 viruses that did not transmit between small numbers of ferret pairs.

Serology

The presence of antibodies elicited against the tested viruses was confirmed by HA inhibition (HI) assay using standard procedures in accordance with the WHO, and A/Indonesia/5/05_{HA} H103Y, T156A, Q222L, G224S as a test antigen. Briefly, ferret antisera were prepared upon intranasal inoculation and collecting blood 14 days later. Antisera were pre-treated overnight with receptor destroying enzyme *Vibrio cholerae* neuraminidase (VCNA) at 37°C, and incubated at 56°C for 1h the next day. Twofold serial dilutions of the antisera, starting at a 1:20 dilution, were mixed with 25 µl of a virus stock containing 4 hemagglutinating units and were incubated at 37°C for 30 min. Subsequently, 25 µl 1% TRBCs was added and the mixture was incubated at 4°C for 1h. HI was read and was expressed as the reciprocal value of the highest dilution of the serum that completely inhibited agglutination of virus and erythrocytes.

Analysis of Virus Nucleotide Sequence Variation Using a 454 Sequencing Platform

Viral RNA was extracted from ferret tracheal swabs, nasal swabs or nasal washes upon passaging with influenza viruses A/H5N1_{wild-type} and A/H5N1_{HA Q222L, G224S PB2 E627K}, using the High Pure RNA Isolation Kit (Roche). A SuperScript III One-Step reverse transcription kit (Invitrogen) was used to synthesize cDNA from extracted RNA. The RT mixture contained 46 µl of RNA extract, 4 µl (2 pmol/µL) primer AGCRAAGCAGG, 1 µl (40 U/µl) Ribonuclease Inhibitor (Promega), and 4 µl (10 mM each) deoxynucleoside triphosphates (Roche) in a 55 µl volume. After a 5 min incubation at 65°C for optimal primer hybridization to template, 16 µl (10x) First-Strand buffer, 14 µl (0.1 M) DTT, 1 µl (40 U/µl) Ribonuclease Inhibitor (Promega) and 4 µl (200 U/µL) SuperScript III Reverse Transcriptase was added to the mixture in a 80 µl volume. The RT mixture was sequentially incubated at 25°C for 5 min and 50°C for 1 hr to obtain cDNA. cDNA was subjected to polymerase chain reaction (PCR), using 32 primer sets (Table S4, available at <http://www.cell.com/>) that cover the full viral genome [136].

The PCR mixtures contained 10 pmol of each forward and reverse primer, 3 µl of cDNA, 1 µl (10 mM each) deoxynucleoside triphosphate, 5 µl 10 × PCR Gold buffer, 5 µl (25 mM) MgCl₂, and 1

μl (2.5 U/μL) AmpliTaq Gold DNA Polymerase (Applied Biosystems, Bleiswijk, The Netherlands). Water was then added to achieve a final volume of 50 μL. The PCR mixture was incubated at 95°C for 6 min, followed by 40 cycles at 95°C for 30 s, 45°C for 30 s, 72°C for 1 min, and a final extension at 72°C for 5 min. PCR fragments were visualized by blue light after electrophoresis on a 1% agarose gel containing 1 × SYBR® Safe DNA Gel Stain (Life Technologies, Bleiswijk, The Netherlands) in 1 × Tris-borate buffer (pH 8.0). SmartLadder (Eurogentec) was used to estimate amplicon size. Fragments ranged from 426 to 627 nucleotides in length.

Fragments with the correct size, were extracted from the gel using the MinElute Gel Extraction Kit (QIAGEN, Venlo, The Netherlands) according to the manufacturers protocol, and subsequently DNA concentrations were measured using a NanoDrop 1000 Spectrophotometer (Thermo Scientific).

Fragments for each virus were pooled in equal concentrations and libraries were created for each virus according to the manufacturer's protocol without DNA fragmentation (GS FLX Titanium Rapid Library Preparation, Roche). The emPCR (Amplification Method Lib-L) and GS Junior sequencing runs were performed according to instructions of the manufacturer (Roche). Sequence reads from the GS Junior sequencing data were sorted by barcode and aligned to reference sequence A/Indonesia/5/05 using CLC Genomics software 4.6.1. The sequence reads were trimmed at 30 nucleotides from the 3' and 5' ends to remove all primer sequences and the 3' end of the reads were trimmed to improve quality, using a Phred score of 20. The threshold for detection was manually set at 1%.

Virus Histochemistry

Viruses were purified on a sucrose gradient and labeled with fluorescein isothiocyanate (FITC, Sigma-Aldrich) as described previously [137]. Briefly, MDCK cells were inoculated and virus-containing supernatant was concentrated using a sucrose cushion and purified on sucrose gradients before inactivation by dialysis against 0.1% formalin and labeling with an equal volume of 0.1 mg/ml FITC.

All selected tissues were free of histological evidence for infection. Virus histochemistry was performed as described previously [137]. Briefly, formalin-fixed paraffin-embedded tissues were deparaffinized with xylene and rehydrated with graded alcohol. 50-100 hemagglutinating units of FITC-labeled influenza viruses were incubated with the respective tissue sections overnight at 4°C. The FITC label was detected with a peroxidase-labeled rabbit anti-FITC antibody (Dako, Heverlee, Belgium) and the signal was amplified using a Tyramide Signal Amplification System (Perkin Elmer, Groningen, The Netherlands) according to the instructions of the manufacturer. Peroxidase was revealed with 3-amino-9-ethyl-cabazole (Sigma-Aldrich) and tissues were counterstained with hematoxylin and embedded in glycerol-gelatin (Merck, Darmstadt, Germany). Attachment of

influenza virus to tissues was visible as granular to diffuse red staining of the epithelium.

Direct Receptor Binding Assay

For direct receptor binding assays and fetuin binding inhibition assay, virus working dilutions were prepared as described previously [102]. In addition, equal absorption of the wild-type and mutant A/H5N1 viruses in the wells of microplates was confirmed using a direct enzyme-linked immunosorbent assay with PR/8 neuraminidase-specific antibodies (BEI resources, Manassas, USA). Biotinylated synthetic poly-N-(2-hydroxyethyl)acrylamide-based sialylglycopolymers (SGPs) containing 20 mol% of Neu5Aca2,3Gal β 1-4GlcNAc β (3'SLN) or Neu5Aca2,6Gal β 1-4GlcNAc β (6'SLN) (Lectinity Holding, Inc., Moscow, Russia) were used as receptor analogs. Also re-sialylated fetuin preparations, containing either α 2,3-linked SAs (3'Fetuin) or α 2,6-linked SAs (6'Fetuin) labeled with horseradish peroxidase (HRP) were tested for their binding to the viruses. Ninety-six well plates were coated overnight at 4°C with bovine fetuin (5 μ g/ml) in PBS, washed with distilled water and dried at room temperature. Fetuin-coated plates were incubated overnight at 4°C with BPL-inactivated viruses in PBS, washed 3 times with PBS and blocked with 0.1% solution of desialylated bovine serum albumin (BSA-NA, Sigma-Aldrich) in PBS. After washing 3 times with ice-cold washing buffer (PBS with 0.05% Tween-80), the replicate virus-coated wells were incubated with two-fold dilutions of the receptor analogs in incubation buffer (1 μ M oseltamivir, 0.02% tween-80, 0.1% BSA) for 1 hr at 4°C. To quantify binding of HRP-labeled fetuin, plates were washed 3 times, after which substrate (tetramethylbenzidine) was added to detect HRP activity. To detect binding of biotinylated SGPs, plates were washed 3 times with ice-cold washing buffer, followed by incubation with streptavidin-HRP. After subsequent washing, tetramethylbenzidine was added to quantify binding. The absorbencies at 450 nm were determined, transferred to a PC and processed with Microsoft Excel software. The data were converted to Scatchard plots (A450/C versus A450), where C is the concentration of the receptor analog in solution and A450 is the absorbency in the corresponding well. Concentrations were expressed in arbitrary units (AU) for 3' and 6' fetuin and in μ M of sialic acid for SGPs. The apparent association constants (K_{ass}) of virus complexes with analogs were determined from the slopes of the Scatchard plots.

Fetuin Inhibition Assay

In the fetuin binding-inhibition assay, the association constant (K_{ass}) for the virus complex with receptor analog is calculated based on the level of its competition with fetuin-HRP [102]. As receptor analogs, several SGPs were used (Lectinity Holding, Inc., Moscow, Russia), which contained the following sialyloligosaccharide ligands: Neu5Aca2,3Gal β 1-4GlcNAc β (3'SLN), Neu5Aca2,6Gal β 1-4GlcNAc β (6'SLN), Neu5Aca2,3Gal β 1-4(6-O-HSO₃)GlcNAc β (Su-3'SLN), Neu5Aca2,3Gal β 1-4(Fuca1-3)GlcNAc β (SLe^x), Neu5Aca2,3Gal β 1-4(Fuca1-3)(6-O-HSO₃)GlcNAc β (Su-SLe^x), Neu5Aca2,3Gal β 1-3GlcNAc β (SLe^c), Neu5Aca2,3Gal β 1-3(6-O-HSO₃)GlcNAc β (Su-SLe^c) and Neu5Aca2,3Gal β 1-3(Fuca1-4)GlcNAc β (SLe^a). Viruses were first absorbed in fetuin-coated plates as described above. Subsequently, two-fold dilutions of SGPs in HRP-labeled fetuin-HRP-

containing buffer were added to the wells, followed by incubation for 1 hr at 4°C. After washing, tetramethylbenzidine was added and peroxidase activity was measured. Association constants were calculated from the binding inhibition data as described previously [102].

Modified Red Blood Cell Assay

Modified TRBCs were prepared as described [138] with slight modifications. Briefly, all SAs were removed from the surface of TRBC by incubation of 62.5 µl of 20% TRBC in PBS with 50 mU of VCNA (Roche, Almere, The Netherlands) in 8 mM calcium chloride at 37°C for one hour. Complete removal of SAs was confirmed by loss of HA of treated TRBC using control viruses. Resialylation was done using 0.25 mU α 2,3-(N)-sialyltransferase (Calbiochem, California, USA) or 12 mU α 2,6-(N)-sialyltransferase (Calbiochem, California, USA) and 1.5 mM CMP-SA (Sigma-Aldrich, Zwijndrecht, the Netherlands) at 37°C in 75 µl for 2 hr to produce α 2,3-TRBC and α 2,6-TRBC respectively. After washing, the TRBCs were resuspended in PBS containing 1% BSA to a final concentration of 0.5%. Resialylation of either α 2,3 or α 2,6 was confirmed by HA using viruses with known receptor specificity. Viruses were tested in standard HA assay using native and resialylated TRBCs. In brief, twofold dilutions of virus were made in PBS. An equal volume of 0.5% TRBCs was added and incubated at 4°C for one hour before reading the HA titer.

Fusion Assay

Fusion was tested as previously described [139] in a cell content mixing (CM) assay in which two 10 cm dishes containing Vero-118 cells were each transfected with 5 µg of pCAGGS-HA and 1 µg of pTS27 plasmid, a constitutive β -galactosidase (β -Gal) expression vector, using Xtremegene transfection reagent (Roche). The cells in one dish were cotransfected with 4 µg of pLTR-CAT (containing the chloramphenicol acetyltransferase [CAT] gene under the control of the human immunodeficiency virus type 1 [HIV-1] long terminal repeat [LTR]), and the cells in the other dish were cotransfected with 4 µg of pTat (expressing the HIV-1 transactivator of transcription Tat). One day after transfection, both cell populations were harvested using trypsin-EDTA, pooled, and plated in a six-well plate format. The next morning, cells were exposed to PBS at different pH for 10 min. Cell lysates were harvested 24 hr after the pH pulse, and CAT and β -Gal expression were quantified by enzyme-linked immunosorbent assays (Roche). HA-mediated fusion of cellular membranes resulted in polykaryon formation and Tat-mediated transactivation of the HIV-1 LTR resulted in induction of CAT expression. Alternatively to CAT quantification, cells were fixed using 70% ice-cold acetone, washed and stained using a 20% Giemsa mixture for microscopy (Merck Millipore, Darmstadt, Germany).

HA Stability Assay

Viruses were incubated for 30 min at different temperatures before performing an HA assay using TRBCs. Two-fold dilutions of virus in PBS containing 0.25% red blood cells were prepared in a

U-shaped 96 well plate and were incubated for one hour at 4°C and agglutination was recorded.

Minigenome Assay

A model vRNA, consisting of the firefly luciferase open reading frame flanked by the noncoding regions (NCRs) of segment 8 of influenza A virus, under the control of a T7 RNA polymerase promoter was used for minigenome assays [106]. The reporter plasmid (0.5 µg) was transfected into 293T cells in 6-well plates, along with 0.5 µg of each of the pHW2000 plasmids encoding PB2, PB1, PA, and NP; 1 µg of pAR3132 expressing T7 RNA polymerase [140]; and 0.02 µg of the *Renilla* luciferase expression plasmid pRL (Promega, Leiden, Netherlands) as an internal control. 24 hr after transfection, luminescence was measured using a Dual-Glo Luciferase Assay System (Promega) according to the instructions of the manufacturer in a TECAN Infinite F200 machine (Tecan Benelux bv, Giessen, Netherlands). Relative light units (RLU) were calculated as the ratio of firefly and *Renilla* luciferase luminescence.

Plaque Assay

The assay was performed as described [141]. In brief, MDCK cells (10⁶ per well) were seeded in a 6 well plate to reach 90% confluency the next day. One hour after inoculation with 10³ TCID₅₀/ml, the inoculum was replaced with 1:1 mixture of 2.4% Avicel (FMC biopolymers, Brussels, Belgium) with 2xEMEM infection medium. After 48 hr, cells were washed 3 times with PBS and 4% formalin solution was added. After two hours of incubation at room temperature (RT) cells were washed twice and then permeabilized using 0.1% Triton X-100 (Sigma Aldrich, Zwijndrecht, the Netherlands). Fixed and permeabilized cells were washed 3 times with PBS/Tween 0.05%, and then incubated for 1 hr at RT with mouse-anti-NP monoclonal antibody (1 mg/ml HB65, ATCC) diluted 1:1000 in PBS with 2% milk powder. Following three washes with PBS/Tween 0.05%, cells were incubated for 1 hr at RT with goat-anti-mouse-HRP (Invitrogen) 1:10.000 diluted in PBS/milk powder 2%. True blue reagent (KPL Inc., Maryland, USA) was added to the cells and incubated for 5 min or until full development of staining of the plaques. Cells were washed twice with H₂O and allowed to air dry. Digital images were taken and were analyzed using ImageQuant TL software (GE Healthcare Life Sciences). Plaque size was plotted as pixels in the original image.

Primer Extension Assay

For quantification of viral transcript levels in virus-infected MDCK cells, primer extension assays were performed as described previously [142] with slight modifications. Cells were seeded in 6-well plates and inoculated with a multiplicity of infection (MOI) of 1 in infection media. At 3 hr postinoculation, cells were collected in 1 ml of Trizol and RNA was purified according to the manufacturer's protocol (Invitrogen). 8,5 µl of RNA was incubated for 3 min at 95°C with 1,5 µl primer master mix containing 32P labeled specific primers for the PB1 segment (mRNA and cRNA: TCTGTTGACTGTGTCCATGG, vRNA: TCGAGTCTGGAAGGATTAAG) and cellular ribosomal RNA (5S rRNA: TCCCAGGCGGTCTCCCATCC). The mixture was cooled on ice and transferred to

50pC. After 5 min, 10 µl of 2x SuperScript III transcription mix (Invitrogen) was added and incubated for 1 hr. The reaction was stopped by addition of 10 µl of 90% formamide and heating to 95°C. Transcription products were separated on 6% polyacrylamide gels containing 7 M urea in trisborate-EDTA buffer and detected using autoradiography films.

Author Contributions

M.L. and S.v.B. are PhD students who were jointly responsible for the critical first series of experiments, the analysis of data, and drafting the manuscript under the supervision of S.H. In particular, M.L. identified a minimal set of mutations required for airborne transmission, whereas S.v.B. was responsible for sequence analyses and bioinformatics. Together, with S.H., they initiated the follow-up phenotypic analyses in collaboration with the other authors.

ACKNOWLEDGMENTS

We thank Dennis de Meulder for excellent technical assistance, Nicolai Bovin (Institute of Bio-organic Chemistry, Moscow, Russia) for providing sialylglycopolymers, and Malik Peiris (University of Hong Kong) for providing A/Indonesia/5/2005 with permission from I. Kandun of the Indonesian government. This work was supported by NIH/NIAID contract HHSN266200700010C, EU FP7 grants EMPIRE (223498), ANTIGONE (278976), FLUPIG (258084), and PREDEMICS (278433). MdG was funded in part by a Marie Curie fellowship (PIEF-GA-2009-237505). All experiments involving A/H5N1 virus were performed before January 2012 in agreement with a moratorium on A/H5N1 gain-of-function research and restarted after February 2013 but without NIH/NIAID funding. Special arrangements are in place with the NIH and the contractor at Icahn School of Medicine at Mount Sinai, New York, for sharing the viruses (and plasmids) in the present paper. A.D.M.E.O. and G.F.R. are CSO and part-time employees of ViroClinics Biosciences B.V. A.D.M.E.O. has advisory affiliations on behalf of ViroClinics Biosciences B.V. with GlaxoSmithKline, Novartis, and Roche. A.D.M.E.O. and R.A.M.F. are holders of certificates of shares in ViroClinics Biosciences B.V. To avoid any possible conflict of interests, Erasmus MC policy dictates that the shares as such are held by the Stichting Administratiekantoor Erasmus Personeelsparticipaties. The board of this foundation is appointed by the Board of Governors of the Erasmus MC and exercises all voting rights with regard to these shares.

Genomic Characterization of a Newly Discovered Coronavirus Associated with Acute Respiratory Distress Syndrome in Humans

Sander van Boheemen, Miranda de Graaf, Chris Lauber, Theo M. Bestebroer, V. Stalin Raj, Ali Moh Zaki, Albert D. M. E. Osterhaus, Bart L. Haagmans, Alexander E. Gorbalenya, Eric J. Snijder, and Ron A. M. Fouchier.

MBio 2012; 3(6):e00473-12.

ABSTRACT

A novel human coronavirus (HCoV-EMC/2012) was isolated from a man with acute pneumonia and renal failure in June 2012. This report describes the complete genome sequence, genome organization, and expression strategy of HCoV-EMC/2012 and its relation with known coronaviruses. The genome contains 30,119 nucleotides and contains at least 10 predicted open reading frames, 9 of which are predicted to be expressed from a nested set of seven subgenomic mRNAs. Phylogenetic analysis of the replicase gene of coronaviruses with completely sequenced genomes showed that HCoV-EMC/2012 is most closely related to *Tylonycteris* bat coronavirus HKU4 (BtCoV-HKU4) and *Pipistrellus* bat coronavirus HKU5 (BtCoV-HKU5), which prototype two species in lineage C of the genus *Betacoronavirus*. In accordance with the guidelines of the International Committee on Taxonomy of Viruses, and in view of the 75% and 77% amino acid sequence identity in 7 conserved replicase domains with BtCoV-HKU4 and BtCoV-HKU5, respectively, we propose that HCoV-EMC/2012 prototypes a novel species in the genus *Betacoronavirus*. HCoV-EMC/2012 may be most closely related to a coronavirus detected in *Pipistrellus pipistrellus* in The Netherlands, but because only a short sequence from the most conserved part of the RNA-dependent RNA polymerase-encoding region of the genome was reported for this bat virus, its genetic distance from HCoV-EMC remains uncertain. HCoV-EMC/2012 is the sixth coronavirus known to infect humans and the first human virus within *Betacoronavirus* lineage C.

IMPORTANCE

Coronaviruses are capable of infecting humans and many animal species. Most infections caused by human coronaviruses are relatively mild. However, the outbreak of severe acute respiratory syndrome (SARS) caused by SARS-CoV in 2002 to 2003 and the fatal infection of a human by HCoV-EMC/2012 in 2012 show that coronaviruses are able to cause severe, sometimes fatal disease in humans. We have determined the complete genome of HCoV-EMC/2012 using an unbiased virus discovery approach involving next-generation sequencing techniques, which enabled subsequent state-of-the-art bioinformatics, phylogenetics, and taxonomic analyses. By establishing its complete genome sequence, HCoV-EMC/2012 was characterized as a new genotype which is closely related to bat coronaviruses that are distant from SARS-CoV. We expect that this information will be vital to rapid advancement of both clinical and vital research on this emerging pathogen.

INTRODUCTION

Coronaviruses (CoVs) infect and cause disease in a wide variety of species, including bats, birds, cats, dogs, pigs, mice, horses, whales, and humans [56, 143]. Recent studies suggest that bats act as a natural reservoir for coronaviruses [59, 144-148]. Coronaviruses may cause respiratory, enteric, hepatic, or neurological diseases with highly variable severity in their hosts. Until 2003, only two coronaviruses were known to infect humans. Human coronaviruses (HCoVs) HCoV-229E and HCoV-OC43 were identified in the 1960s as the causative agents of -generally mild-respiratory illnesses [66, 67]. In 2002 to 2003, a previously unknown coronavirus -severe acute respiratory syndrome coronavirus (SARS-CoV)- caused a widespread outbreak of respiratory disease in humans, resulting in approximately 800 deaths and affecting around 30 countries [43, 58, 65, 149]. As a consequence of the renewed interest in coronaviruses after the SARS outbreak, two additional human coronaviruses were discovered after 2003: HCoV-NL63 in 2004 [68, 69] and HCoV-HKU1 in 2005 [70]. A recent analysis of a large collection of human nasopharyngeal specimens using a *Coronaviridae*-wide primer set suggested that HCoV-229E, -OC43, -NL63, and -HKU1 are the only coronaviruses circulating in the human population [150].

Coronaviruses are enveloped single-stranded positive-sense RNA viruses with genomes of 25 to 32 kb, and the group includes the largest known genomes among the RNA viruses [56, 151]. The coronaviruses form a subfamily (*Coronavirinae*) within the family *Coronaviridae* of the order *Nidovirales*. The International Committee on Taxonomy of Viruses (ICTV) has recognized four genera within the *Coronavirinae* subfamily: *Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus*, which were previously referred to as coronavirus groups 1, 2, and 3, and *Deltacoronavirus* [152]. Coronaviruses are assigned to a genus on the basis of rooted phylogeny and calculation of pairwise evolutionary distances for seven highly conserved domains in the replicase polyprotein [56, 153] (C. Lauber and A. E. Gorbalenya, unpublished data). HCoV-229E and HCoV-NL63 are viruses belonging to the genus *Alphacoronavirus* [56]. Four monophyletic lineages (A through D) with no formal taxonomic standing, some of them encompassing multiple virus species, are commonly recognized within the genus *Betacoronavirus*. Lineage A includes HCoV-OC43 and HCoV-HKU1 and lineage B SARS-CoV, all of which belong to different species. Lineages C and D include viruses detected only in bats, such as *Rousettus* bat coronavirus HKU9 (BtCoV-HKU9) (lineage D), *Tylonycteris* bat coronavirus HKU4 (BtCoV-HKU4), and *Pipistrellus* bat coronavirus HKU5 (BtCoV-HKU5) (both lineage C) [56]. The genetic diversity of coronaviruses is likely facilitated by a high frequency of RNA recombination and the ability of their unusually large RNA genomes to both gain and lose domains [56, 154, 155]. These factors are believed to have promoted the emergence of viruses with novel traits that are able to adapt to new hosts and ecological niches, sometimes causing zoonotic events.

For the present study, we report and analyze the complete genome sequence of the recently identified HCoV-EMC/2012, which was isolated from the sputum of a 60-year-old man who died in a hospital in Jeddah, Saudi Arabia, after developing acute respiratory distress syndrome (ARDS) and multiple organ dysfunction syndrome (MODS) in June 2012 [156]. This virus appears to be closely related to the HCoV detected in a second patient who was transported from a hospital in Qatar to a hospital in London, 3 months after hospitalization of the first patient [157]. These two cases of human infection with very similar or identical coronaviruses alarmed health authorities globally, as it was a reminder of the potential threat of coronaviruses to human health that was first highlighted by the SARS outbreak of 2003 [157]. The sequence analysis of a small reverse transcription-PCR (RT-PCR) fragment that was first amplified from the HCoV-EMC/2012 genome revealed the highest similarity to two betacoronaviruses circulating in bats, BtCoV-HKU4 and -HKU5 [156]. Here we present the complete genome sequence of the newly isolated HCoV-EMC/2012, accompanied by a detailed annotation of its genome organization and expression strategy. Furthermore, comparative genomic analysis and state-of-the-art classification and phylogenetic analyses were applied to determine the position of the novel agent with respect to previously characterized coronaviruses. We conclude that the HCoV-EMC/2012 genome organization and expression indeed most closely resemble those of BtCoV-HKU4 and -HKU5. However, based on our analysis and in line with the ICTV guidelines for the demarcation of coronavirus species, HCoV-EMC/2012 clearly qualifies to be recognized as the prototype of a novel species, which would thus constitute the first human coronavirus in lineage C of the genus *Betacoronavirus*.

RESULTS

Sequencing of the HCoV-EMC/2012 genome

Using a combination of approaches, including deep sequencing, cycle sequencing on a more traditional capillary sequencer, and determination of the genomic termini by rapid amplification of cDNA ends (RACE), the complete genome sequence of HCoV-EMC/2012 was determined from material that had been subjected to passage in cell culture 6 times. The data from the Roche 454 GS Junior deep-sequencing run yielded a total of 90,808 sequence reads, of which 87,256 were specific for HCoV-EMC/2012. Genome coverage ranged from 1 to 5,697 reads at single nucleotide positions, with an average of 1,006 reads in the deep-sequencing run. Based on the contigs assembled from these initial data, primers approximately 800 nucleotides (nt) apart were designed to amplify PCR fragments with 100-nt overlaps covering the entire virus genome (Table S1, available at <http://mbio.asm.org/>). These amplicons were sequenced using Sanger sequencing, and a total of 104 sequence runs were assembled -along with the original 90,808 deep-sequencing reads- into a single contiguous sequence of 30,119 nt, including the first 12 nt of the 3' poly(A) tail. Although 454 sequencing resulted in a higher single-read error rate than Sanger

sequencing, the high coverage in the first data set largely corrected for these errors. Occasionally, the correct number of bases in homopolymer stretches was difficult to determine, which is a typical problem in 454 sequencing. Nevertheless, there was excellent agreement between the deep-sequencing data and the confirmatory Sanger sequencing. The final consensus sequence was submitted to GenBank (see below). This sequence contains only two ambiguous positions, nt 11623 and 27162. The variation at position 11623 translates into a Val or Gly uncertainty at amino acid (aa) 3782 of pp1a/pp1ab. Position 27162 was either a G or an A, with the A creating a premature stop codon for translation of open reading frame 5 (ORF5) (see Discussion). The verification of our consensus sequence awaits the availability of a second HCoV-EMC/2012 virus isolate or original specimen. The overall content of G and C residues in the HCoV-EMC/2012 genome was 41%, which is similar to values reported for other coronaviruses (37% to 42%) [58].

Genome organization and expression strategy

Coronavirus genomes are polycistronic positive-stranded RNAs (Fig. 1A), of which the 5'-proximal three-fourths are occupied by the large replicase open reading frames ORF1a and ORF1b. These are translated from the genomic mRNA to produce polyproteins pp1a and, following -1 ribosomal frameshifting, pp1ab, which are subsequently cleaved into 15 or 16 nonstructural proteins (nsps) [151, 155, 158]. The region downstream of ORF1b is characterized by containing a variable number of smaller genes, always including those encoding the spike (S), envelope (E), membrane (M), and nucleocapsid (N) structural proteins. These genes are translated from subgenomic (sg) mRNAs that form a 5'- and 3'-coterminal nested set with the viral genome. Subgenomic mRNAs are composed of a common 5' leader sequence that is identical to the genomic 5' region and a variable part of the 3' quarter of the genome, with different sg mRNAs making different ORFs available for translation. The complement of the leader and "body" segments of the sg mRNAs are assumed to be joined during discontinuous negative-strand RNA synthesis. This step produces the templates for sg mRNA synthesis and is directed by a base-pairing interaction between conserved transcription-regulatory sequences (TRSs) [159-161]. Such TRSs are found at the 3' end of the leader sequence (leader TRS) and at different positions upstream of genes in the genomic 3'-proximal domain (body TRSs). The synthesis of subgenome-length negative-stranded RNAs is directed by the complement of a body TRS at the 3' end of a nascent minus-strand base pairing with the leader TRS, with the extent of sequence complementarity being an important determinant of the level at which a given sg mRNA is produced.

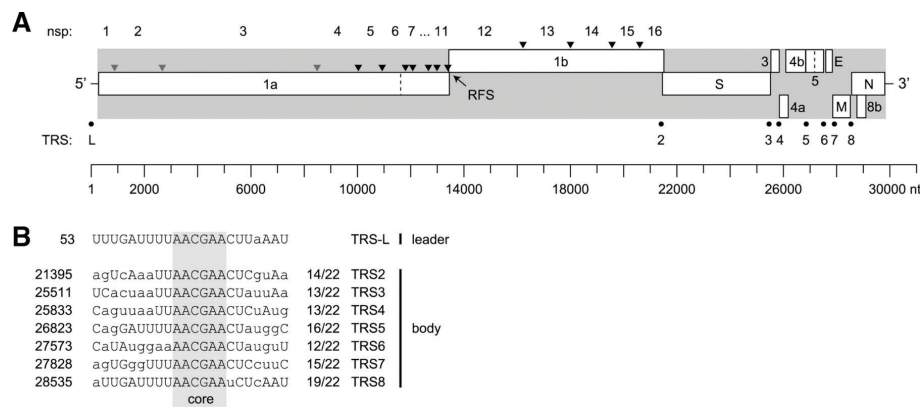


Figure 1. Genome organization and expression of HCoV-EMC/2012. (A) The coding part of the genome and terminal untranslated regions are depicted, respectively, by a gray background and horizontal lines. Rectangles indicate ORFs and their locations in three reading frames. The dashed lines in ORF1a and ORF5 indicate base ambiguities observed during sequencing. Triangles represent sites in the replicase polyproteins pp1a and pp1ab that are predicted to be cleaved by papain-like proteinases (gray) or the 3C-like cysteine proteinase (black). Cleavage products are numbered nsp1 to nsp16, according to the convention established for other coronaviruses [155]. The -1 ribosomal frameshift site (RFS) in the ORF1a/ORF1b overlap region is indicated. The location of the leader TRS (transcription-regulatory sequences) (L) and seven body TRSs (numbered) are highlighted by black dots. All coordinates correspond to the scale shown at the bottom. (B) Sequence comparison of leader TRS region and seven body TRSs. The fully conserved TRS core sequence AACGAA is highlighted. Nucleotides in the body TRSs are written in uppercase letters if the complementary nucleotide can base pair with the corresponding residue in the leader TRS region (including G-U base pairs). TRS starting coordinates in the HCoV-EMC/2012 genome are shown at the left; for the body TRSs, the numbers of (potential) base pairs with the leader TRS region are shown at the right.

Inspection of the genome sequence of HCoV-EMC/2012 revealed the two large, partially overlapping replicase open reading frames ORF1a and ORF1b, as well as (at least) nine downstream ORFs (Fig. 1A). The ORF1a sequence encodes the two protease domains conserved in all other coronaviruses, a papain-like protease (PL2pro) in nsp3 and a 3C-like protease (3CLpro; also known as the “main protease”) in nsp5. Sequence comparison with other coronaviruses allowed us to predict the putative pp1a/pp1ab cleavage sites and annotate the resulting nsp1 through -16 (Table 1). According to sequence conservation analyses performed with other coronaviruses, open reading frames ORF2, -6, -7, and -8a are predicted to encode the four canonical structural proteins of coronaviruses, the envelope proteins S, E, and M and the N protein, respectively (Fig. 1A). A leader TRS and seven putative body TRSs could be readily identified, with the sequence 5' AACGAA 3' forming the conserved TRS core and potential TRS duplexes during leader-body joining ranging from 14 to 19 matches over a 22-nt window that includes the core of the leader TRS (Fig. 1B). From this analysis, it can be predicted that seven subgenomic mRNAs carrying a 67-nt common leader sequence would be produced in HCoV-EMC/2012-infected cells, with sizes ranging from ~4.7 kb for mRNA2 to ~1.7 kb for mRNA8. Experimental studies are needed to confirm the correct identification of the TRSs in the genomes of HCoV-EMC/2012 and related lineage C betacoronaviruses.

TABLE 1 Cleavage products of the replicase polyproteins of HCoV-EMC/2012

| Cleavage product | Position in polyprotein pp1a/pp1ab ^a | Protein size (no. of amino acids) | Putative functional domain(s) ^b |
|------------------|---|-----------------------------------|--|
| nsp1 | 1Met-Gly193 | 193 | |
| nsp2 | 194Asp-Gly853 | 660 | |
| nsp3 | 854Ala-Gly2740 | 1887 | ADRP, PL2pro, TM1 |
| nsp4 | 2741Ala-Gln3247 | 507 | TM-2 |
| nsp5 | 3248Ser-Gln3553 | 306 | 3CLpro |
| nsp6 | 3554Ser-Gln3845 | 292 | TM-3 |
| nsp7 | 3846Ser-Gln3928 | 83 | |
| nsp8 | 3929Ala-Gln4127 | 199 | Putative primase |
| nsp9 | 4128Asn-Gln4237 | 110 | |
| nsp10 | 4238Ala-Gln4377 | 140 | |
| nsp11 | 4378Ser-Leu4391 | 14 | |
| nsp12 | 4378Ser-Gln5310 | 933 | RdRp |
| nsp13 | 5311Ala-Gln5908 | 598 | ZD, HEL1 |
| nsp14 | 5909Ser-Gln6432 | 524 | ExoN, NMT |
| nsp15 | 6433Gly-Gln6775 | 343 | NendoU |
| nsp16 | 6776Ala-Arg7078 | 303 | OMT |

^a Amino acids of the replicase proteins pp1a and pp1ab were numbered with the assumption that a –1 ribosomal frameshift occurs to express ORF1b, as in other coronaviruses (see text); the use of the slippery sequence UUUAAC is predicted to result in a peptide bond between Asn4385 and Arg4386 in pp1ab.

^b The major transmembrane domains and a selection of the most conserved domains with enzymatic activities that have been characterized functionally and/or structurally in coronaviruses are listed. Abbreviations: PL2pro, papain-like proteinase 2; ADRP, ADP-ribose 1"-phosphatase; TM, transmembrane domain; 3CLpro, 3C-like cysteine proteinase; RdRp, RNA-dependent RNA polymerase; ZD, putative zinc-binding domain; HEL1, superfamily 1 helicase; ExoN, 3'-to-5' exonuclease; NMT, N7-methyltransferase; NendoU, nidoviral endoribonuclease specific for U; OMT, S-adenosylmethionine-dependent ribose 2'-O-methyltransferase.

Furthermore, mRNA4 and -8 are predicted to be functionally bicistronic, with ribosomal leaky scanning being the likely translation initiation mechanism for both ORF4b and ORF8b. The ORF4b AUG codon is not preceded by a separate body TRS, and the 241-nt sequence separating the 5' ends of ORF4a and ORF4b is entirely devoid of AUG codons. The AUG codon of the current ORF8b, an internal ORF that is overlapped by the N protein gene (ORF8a) and is present in all betacoronaviruses, is the third AUG codon on mRNA8, but sequence analysis and comparison with the BtCov-HKU4 and -HKU5 sequences [59] suggests that the 5' end of ORF8b may have become truncated relatively recently (see Discussion).

Twenty-two additional putative ORFs of 150 to 432 nt in length were detected throughout the genome of HCoV-EMC/2012, overlapping the major ORFs. In contrast to the ORFs shown in Fig. 1A, these 22 additional ORFs are not positioned (immediately) downstream of a body TRS, and hence it is unlikely that they are expressed. The synthesis of the replicase pp1ab polyprotein of HCoV-EMC/2012 involves –1 programmed ribosomal frameshifting, with nt 13427 to 13433 predicted to form the conserved "slippery sequence" (5' UUUAAC 3') in the ORF1a/ORF1b overlap region that is typical for coronaviruses [162]. The frameshift region is followed by a predicted

RNA hairpin, formed by nucleotides at positions 13439 to 13450 base pairing with those at 13462 to 13473, with potential RNA pseudoknot formation occurring by base pairing of the loop of the hairpin (nt 13452 to 13460) with a downstream complementary sequence (nt 13506 to 13514). As is common in coronavirus genomes, nontranslated sequences are found only at the genomic termini, with the 5' and 3' untranslated regions (278 and 300 nt, respectively) having sizes similar to those found in other family members. The only other apparently untranslated region in the genome that is larger than 50 nucleotides concerns the intergenic region between ORF5 and -6 (nt 27515 to 27589). This region appears to be conserved between HCoV-EMC/2012, BtCoV-HKU4, and BtCoV-HKU5, with sequence identities ranging from 63% to 84%, but we have no explanation for this observation thus far.

Phylogenetic relations and taxonomic position of HCoV-EMC/2012

Phylogenetic trees were inferred using nucleotide sequences for ORF1ab (Fig. 2A) and a 332-nt fragment from ORF1b (Fig. 2B) encoding the most conserved part of the RNA-dependent RNA polymerase (RdRp) domain, which is commonly targeted in virus discovery studies. The first tree was produced for a representative set of coronaviruses for which complete genome sequences are available. In the second tree, we also included coronaviruses for which only partial genome sequences are known, particularly that of P.pipi/VM314/2008/NLD [163] which produced the best match with HCoV-EMC/2012. In both trees, HCoV-EMC/2012 clearly groups within lineage C of the genus *Betacoronavirus*, relatively close to BtCoV-HKU4 and BtCoV-HKU5. However, based on the 332-nt fragment from ORF1b, HCoV-EMC/2012 is more closely related to bat-derived isolate VM314/2008 (GenBank accession number GQ259977), which was isolated from *Pipistrellus* bats in The Netherlands 4 years ago. Phylogenetic trees were also constructed based on amino acid sequences, using coronavirus-wide conserved domains of replicative proteins in pp1ab (Fig. 3A) as well as using conserved parts of structural proteins (Fig. 3B). In both trees, HCoV-EMC/2012 clusters with betacoronaviruses, supporting its classification as a member of the genus *Betacoronavirus*.

ICTV assigns newly identified members of the family *Coronaviridae* to a subfamily and genus on the basis of rooted phylogeny and calculation of pairwise evolutionary distances for seven replicase domains [56, 153]. To establish whether HCoV-EMC/2012 indeed prototypes a new species, amino acid sequence alignments were generated for each of these domains and concatenated, after which the sequence identity of HCoV-EMC/2012 with closely related strains was calculated. For this purpose, the full genomes of 9 strains, derived from 3 species, belonging to *Betacoronavirus* lineage C were available (Table 2). Amino acid sequence identity between conserved replicase domains of HCoV-EMC/2012 and those of other lineage C viruses ranged from 57% (ADP-ribose 1"-phosphatase [ADRP]) to 94% (helicase [Hel]). Overall amino acid sequence identities to BtCoV-HKU4 and BtCoV-HKU5 strains across the conserved domains were around 75% and 76.7%, respectively. These percentages are well below the threshold of 90% amino

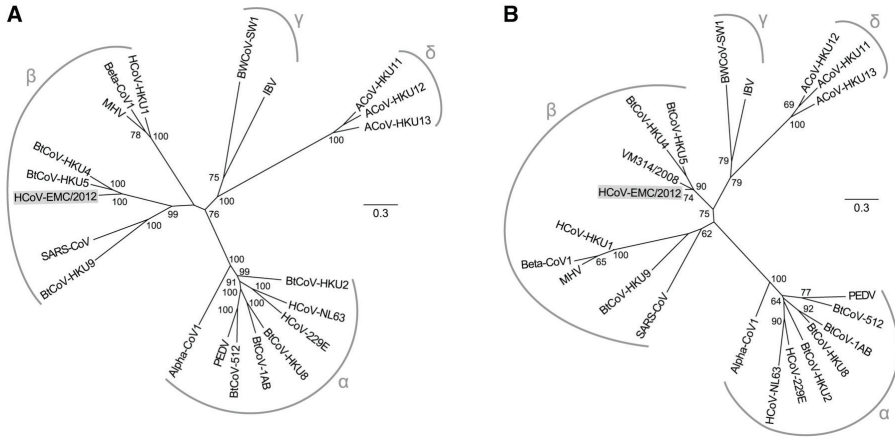


Figure 2. Phylogenetic trees for HCoV-EMC/2012 and selected other coronaviruses. Unrooted maximum likelihood phylogenies inferred from the nucleotide sequences of full-length ORF1ab (A) or a 332-nt fragment from the RdRp-encoding domain of ORF1b (B) are shown. HCoV-EMC/2012 and 20 viruses representing the recognized species diversity of coronaviruses were included, with bat-derived isolate VM314/2008 also included in the analysis presented in panel B [163]. The viruses and corresponding species used are *Alphacoronavirus 1* (Alpha-CoV1), *Human coronavirus 229E* (HCoV-229E), *Human coronavirus NL63* (HCoV-NL63), *Miniopterus bat coronavirus 1* (BtCoV-1AB), *Miniopterus bat coronavirus HKU8* (BtCoV-HKU8), *Porcine epidemic diarrhea virus* (PED), *Rhinolophus bat coronavirus HKU2* (BtCoV-HKU2), *Scotophilus bat coronavirus 512* (BtCoV-512), *Betacoronavirus 1* (Beta-CoV1), *Human coronavirus HKU1* (HCoV-HKU1), *Murine coronavirus* (MHV), *Tylonycteris bat coronavirus HKU4* (BtCoV-HKU4), *Pipistrellus bat coronavirus HKU5* (BtCoV-HKU5), *Rousettus bat coronavirus HKU9* (BtCoV-HKU9), *Severe acute respiratory syndrome-related coronavirus* (SARS-CoV), *Avian coronavirus* (IBV), *Beluga whale coronavirus SW1* (BWCoV-SW1), *Bulbul coronavirus HKU11* (ACoV-HKU11), *Thrush coronavirus HKU12* (ACoV-HKU12), and *Munia coronavirus HKU13* (ACoV-HKU13). Bootstrap values above 50 are shown. Arcs and symbols indicate the four coronavirus genera. The scale bar represents the number of nucleotide substitutions per site.

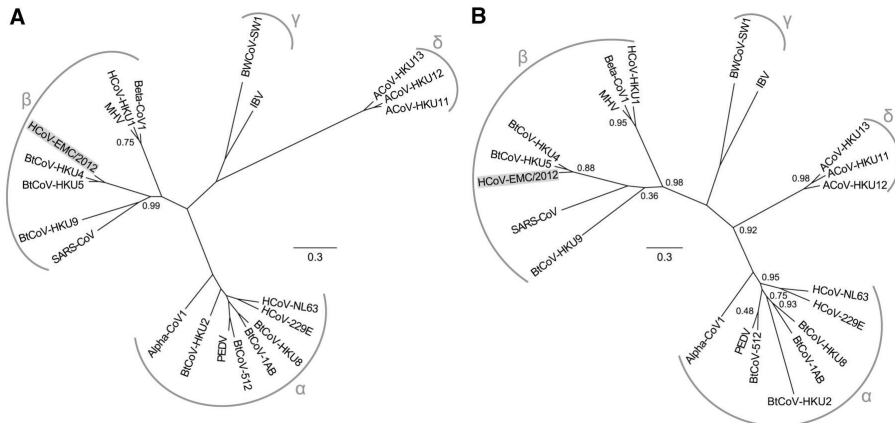


Figure 3. Phylogenetic trees for HCoV-EMC/2012 and selected other coronaviruses. Unrooted maximum likelihood phylogenies based on coronavirus-wide conserved protein domains in replicase pp1ab (A) or on the conserved parts of structural proteins S2, E, M, and N (B) for HCoV-EMC/2012 and 20 viruses representing the recognized species diversity of coronaviruses are shown (see Fig. 2 legend for names and abbreviations). Branch support values are based on the Shimodaira-Hasegawa-like procedure and are in the range of zero to one; only nonoptimal values smaller than one are shown. Arcs and symbols indicate the four coronavirus genera. The scale bars represent average numbers of substitutions per amino acid position.

acid sequence identity that is used for coronavirus species identification by the ICTV. The distance between HCoV-EMC/2012 and members of these two species is as large as that observed upon interspecies comparison of other species pairs, for example, *Murine coronavirus* versus *Human coronavirus HKU1* or *Porcine epidemic diarrhea virus* versus *Scotophilus bat coronavirus 512* (Fig. 3). Consequently, we propose that HCoV-EMC/2012 prototypes a novel species of lineage C of the genus *Betacoronavirus*.

TABLE 2 Percent amino acid sequence identity between conserved domains of the replicase polyprotein of HCoV-EMC/2012 and established betacoronaviruses^a

| Virus strain | % amino acid sequence identity with conserved domain of the indicated HCoV-EMC/2012 replicase polyprotein ^b | | | | | | | |
|----------------|--|--------|------|------|------|--------|------|-------------|
| | ADRP | 3CLpro | RdRp | Hel | ExoN | NendoU | O-MT | All domains |
| BtCoV-HKU4.1 | 57.4 | 81.0 | 90.0 | 92.1 | 85.4 | 72.6 | 83.4 | 75.1 |
| BtCoV-HKU4.2 | 57.5 | 81.0 | 90.0 | 92.1 | 85.4 | 72.6 | 83.4 | 75.1 |
| BtCoV-HKU4.3 | 57.4 | 81.0 | 90.0 | 92.1 | 85.4 | 72.6 | 83.4 | 75.1 |
| BtCoV-HKU4.4 | 57.5 | 81.0 | 89.9 | 92.1 | 85.4 | 72.6 | 83.4 | 74.9 |
| BtCoV/133/2005 | 57.6 | 80.7 | 89.9 | 91.6 | 86.4 | 72.0 | 83.4 | 74.9 |
| BtCoV-HKU5.1 | 57.6 | 82.6 | 92.1 | 93.8 | 91.7 | 79.7 | 85.3 | 76.7 |
| BtCoV-HKU5.2 | 57.6 | 82.0 | 92.2 | 93.8 | 91.7 | 80.0 | 85.3 | 76.7 |
| BtCoV-HKU5.3 | 57.2 | 82.0 | 92.2 | 93.8 | 91.7 | 80.0 | 85.3 | 76.7 |
| BtCoV-HKU5.5 | 57.3 | 82.0 | 92.2 | 93.8 | 91.7 | 80.0 | 85.3 | 76.7 |

^a Accession numbers used are as follows: for BtCoV-HKU4 strains, EF065505, EF065506, EF065507, and EF065508; for BtCoV/133/2005, DQ648794; and for BtCoV-HKU5 strains, EF065509, EF065510, EF065511, and EF065512.

^b For abbreviations, see Table 1.

Genome similarities between coronavirus HCoV-EMC/2012 and coronaviruses BtCoV-HKU4 and BtCoV-HKU5.

BtCoV-HKU4 and BtCoV-HKU5 [59] are the closest relatives of HCoV-EMC/2012 for which full-length genome sequences are available (see above). Accordingly, comparison of the genomes of these three viruses revealed important similarities, including the organization of the “accessory protein genes”, ORF3a through ORF5, residing between the S protein gene and those encoding the E, M, and N proteins. Upon annotating this region of the BtCoV-HKU4 and BtCoV-HKU5 genomes, Woo *et al.* [59] identified the body TRSs for sg mRNA3, mRNA4, and mRNA5 but unfortunately did not follow standard coronavirus nomenclature, naming the downstream open reading frames ORF3a through ORF3d (encoding ns3a through ns3d) rather than ORF3, ORF4a, ORF4b, and ORF5 (Fig. 1A). The similarities of all ORFs and proteins of HCoV-EMC/2012, BtCoV-HKU4, and BtCoV-HKU5 were calculated, and percentages of sequence identity are summarized in Table 3. The lowest percentages of sequence identity to BtCoV-HKU4 and BtCoV-HKU5 were observed for ORF3 at the nucleotide level (46.4% and 46.0%, respectively) and for ORF4b at the amino acid level (23.5% and 25.9%, respectively). The highest percentages of sequence identity to BtCoV-HKU4 and BtCoV-HKU5 were observed for the E ORF at the nucleotide level (74.6%

TABLE 3 Percent identity between open reading frames of coronavirus HCoV-EMC/2012 and coronaviruses BtCoV-HKU4 and BtCoV-HKU5 at the nucleotide and amino acid levels

| Annotation in HCoV-EMC/2012 | Annotation in BtCoV-HKU4 and BtCoV-HKU5 ^a | % identity to BtCoV-HKU4 ^b | | % identity to BtCoV-HKU5 ^b | |
|-----------------------------|--|---------------------------------------|------|---------------------------------------|------|
| | | nt | aa | nt | aa |
| ORF1ab | ORF1ab | 70.6 | 72.2 | 70.7 | 73.8 |
| S | S | 66.3 | 66.1 | 63.8 | 63.5 |
| ORF3 | NS3a | 46.4 | 34.9 | 46.0 | 31.4 |
| ORF4a | NS3b | 51.5 | 37.5 | 47.8 | 38.0 |
| ORF4b | NS3c | 35.1 | 23.5 | 45.2 | 25.9 |
| ORF5 | NS3d | 56.6 | 46.9 | 58.1 | 54.2 |
| E | E | 74.6 | 69.5 | 75.1 | 68.2 |
| M | M | 72.8 | 82.6 | 73.0 | 82.2 |
| N | N | 67.2 | 71.8 | 66.7 | 67.8 |
| ORF8b | Undescribed | 45.3 | 32.1 | 48.0 | 33.8 |

^a Annotations used for HCoV-EMC/2012 differ from those used for BtCoV-HKU4 and BtCoV-HKU5 (10).

^b Accession numbers used for BtCoV-HKU4 and BtCoV-HKU5 were EF065505 and EF065509.

and 75.1%, respectively) and for the M ORF at the amino acid level (82.6% and 82.2%, respectively). These data further supported the characterization of HCoV-EMC/2012 as a close relative of BtCoV-HKU4 and BtCoV-HKU5.

DISCUSSION

Coronaviruses have been known for quite some time as viruses that cause a variety of diseases in humans and animals [164, 165]. The discovery of a coronavirus as the causative agent of SARS revived the interest in coronaviruses and resulted in a rapid increase of the number of identified coronaviruses, as well as of the number of full coronavirus genome sequences. Until this study, lineage C of the genus *Betacoronavirus* (formerly known as subgroup 2c) included virus isolates from bats. Here, we determined and analyzed the complete genome sequence of a previously unknown lineage C *Betacoronavirus* that was isolated from the sputum of a 60-year-old male suffering from acute pneumonia and renal failure in the Kingdom of Saudi Arabia whose death was probably a consequence of this infection [156].

The sequencing of the full HCoV-EMC/2012 genome was greatly facilitated by the advent of high-throughput techniques. Using an optimized random amplification deep-sequencing approach, approximately 90% of the virus genome was covered with high accuracy in a single run. Using the data from this first run, primers could be designed to perform conventional Sanger sequencing for confirmation. This combination of techniques allowed the determination of the

complete virus genome within a few days, without a requirement for prior knowledge of the virus genome under investigation. The error rate in 454 deep sequencing was generally higher than in Sanger sequencing, but the high coverage across the HCoV-EMC/2012 virus genome (up to 5,697 reads per nucleotide position) corrected for most of the incorrect base callings. The sequence obtained using the 454 platform aligned almost perfectly with that obtained by Sanger sequencing, with the exception of two nucleotide positions. The deep-sequencing data revealed variation at position 11623 (U or G), with G occurring in 44% of the reads, suggesting that ORF1a-encoded residue 3782 can be either valine (codon GUC) or glycine (codon GGC). The valine codon was the more abundant codon at this position in HCoV-EMC/2012, and valine is also present in most other betacoronaviruses. At position 27162, both G and A were detected in different runs, with an A in 45% of the reads. This G-to-A substitution introduces a premature stop codon (UGG to UAG) in ORF5. The virus stock that we sequenced was derived from passage of the virus from a sputum specimen six times in Vero cell culture. Hence, the observed sequence variants may reflect either natural heterogeneity or emerging genetic changes that occurred during virus passage in cell culture. Additional HCoV-EMC/2012 virus isolates or patient materials are currently not available to verify these genome sequence ambiguities at positions 11623 and 27162.

Adaptation to cell culture leading to a loss of functionality of genes, and in particular in relation to the so-called “accessory protein genes”, has previously been described for a variety of coronaviruses, including SARS-CoV [143, 166]. These genes, like ORF3 through ORF5 of HCoV-EMC/2012, are dispersed between the structural protein genes [167] and in some cases may even overlap such a gene, as in the case of the ORF overlapping the N protein gene in betacoronaviruses (Fig. 1A) [155, 168]. The origin of most accessory protein genes remains unclear, although for some, acquisition by recombination with cellular or heterologous viral sequences seems plausible [169, 170]. Accessory gene functions have been probed by reverse genetics (knockout mutants) for a variety of coronaviruses, including SARS coronavirus [171], which established that they are not essential for replication in cell culture systems. In animal models, on the other hand, profound effects on pathogenesis after the inactivation (or transfer to a heterologous coronavirus) of accessory protein genes have been previously described [172-174]. In some cases, accessory gene products have been implicated in immune evasion, e.g., by interfering with cellular innate immune signaling [175].

The apparent absence of selection pressure on coronavirus accessory protein genes during cell culture passage may explain the relatively high frequency with which loss of functionality appears to occur. The detection of an internal termination codon in part of the HCoV-EMC/2012 ORF5 sequences (45% of the reads) may constitute another example of such an event, which would lead to the truncation of the ORF5 protein after 107 amino acids. This would resemble a 29-nt deletion that occurred in the SARS-CoV genome, which resulted in the truncation of ORF8 [166,

176], and a 45-nt in-frame deletion in ORF7b of the same virus that emerged upon cell culture passage [155].

Our analysis identified a potential ORF underlying the N protein gene (ORF8a), which is a common feature in betacoronaviruses. This ORF was not previously described for BtCoV-HKU4 and BtCoV-HKU5 [59] but is conserved in the genome sequences of both viruses (Fig. S1J, available at <http://mbio.asm.org/>). Remarkably, in HCoV-EMC/2012, both the 5' and 3' parts of the ORF appear to have been truncated. In BtCoV-HKU4 and BtCoV-HKU5, the ORF8b AUG codon would be the second AUG on sg mRNA8, making leaky ribosomal scanning a likely mechanism for translation initiation. In HCoV-EMC/2012, however, this AUG codon (positions 28606 to 28608) seems to have been mutated to AUA. Conservation of the sequence immediately downstream of this position, which is now formally upstream of ORF8b in HCoV-EMC/2012, was observed with BtCoV-HKU4 and BtCoV-HKU5, suggesting that the putative loss of this AUG codon may also have been a relatively recent event. In the 3' part of ORF8b, sequence alignment of HCoV-EMC/2012 with BtCoV-HKU4 and BtCoV-HKU5 suggests that the former acquired a premature termination codon at positions 29099 to 29101 (UAA). Although we cannot at present assess the timing of these events in HCoV-EMC/2012 evolution, due to the lack of alternative samples for this species, the presumed loss of ORF8b functionality may also be a consequence of virus passage in cell culture.

To classify newly identified coronaviruses as the prototype of a novel virus species, it is required that the amino acid sequence identity in the conserved replicase domains in all intervirus pairwise comparisons is below the 90% threshold [56]. Here, we propose HCoV-EMC/2012 to represent a novel species of the *Betacoronavirus* genus, since the amino acid sequence identities between HCoV-EMC/2012 and its closest relatives BtCoV-HKU4 and BtCoV-HKU5 in the seven conserved domains of ORF1ab were 75% and 77%, respectively. These viruses were originally detected in Asia in lesser bamboo bats (*Tylonycteris pachypus*) and Japanese house bats (*Pipistrellus abramus*), respectively [59]. This proposed classification will remain provisional until approved by ICTV.

The ICTV guidelines for coronavirus species demarcation require the availability of a (nearly) complete genome sequence prior to virus classification. However, there is considerable correlation between the results based on full-genome sequence analysis and those determined using the most conserved part of the ORF1b-encoded RdRp domain, which is commonly used in screening for new coronaviruses. In 2010, this partial sequence was reported for a *Betacoronavirus* (VM314/2008) that was isolated 2 years earlier from a *Pipistrellus pipistrellus* bat in The Netherlands. This virus was provisionally classified a *Betacoronavirus* based on a 332-nt fragment from the RdRp-encoding domain of ORF1b [163], which shares 88% nucleotide sequence

identity with HCoV-EMC/2012, the highest identity with any coronavirus sequence available in the public domain. Although this high similarity is not sufficient to resolve the taxonomic relation between HCoV-EMC/2012 and isolate VM314/2008, it suggests that they may both belong to the same coronavirus species. Establishing the genome sequence of VM314/2008, or closely related viruses, is urgently required to verify this hypothesis. Based on the genetic relation between HCoV-EMC/2012 and bat coronaviruses, it is tempting to speculate that HCoV-EMC/2012 emerged from bats- either directly or via an intermediate animal host, possibly *Pipistrellus* bats. This bat species is known to be present in the Kingdom of Saudi Arabia and neighboring countries.

Although most infections of human coronaviruses are relatively mild, the infection by HCoV-EMC/2012 with fatal outcome, and a similar severe case of an infection with a closely related coronavirus in London [157], is a reminder that certain coronaviruses may cause severe and sometimes fatal infections in humans. It is important to develop an animal model that can be used to fulfill Koch's postulates for the novel virus, by demonstrating that the isolated virus can indeed cause the observed disease. The availability of the HCoV-EMC/2012 genome sequence will facilitate the development of a variety of diagnostic assays that can be used to study the prevalence and clinical impact of HCoV-EMC/2012 infections in humans. The first generation of assays for this purpose has recently been described [77]. We anticipate that the availability of this full-length virus genome sequence will be valuable for the development of additional applied and fundamental research.

MATERIALS AND METHODS

Virus propagation

Patient material had been subjected to passage in Vero cells four times in the Dr. Soliman Fakeeh Hospital, Jeddah, Saudi Arabia. Subsequently, in the Erasmus Medical Center, Rotterdam, The Netherlands, LLC-MK2 cells were inoculated with HCoV-EMC/2012 in minimal essential medium (MEM-Eagle) with Earle's salts (BioWhittaker, Verviers, Belgium), supplemented with 2% serum, 100 U/ml penicillin, 100 mg/ml streptomycin, and 2 mM glutamine. Vero cells were inoculated with virus in Dulbecco's modified Eagle medium (BioWhittaker) supplemented with 1% serum, 100 U/ml penicillin, 100 mg/ml streptomycin, and 2 mM glutamine. After inoculation, the cultures were incubated at 37°C in a CO₂ incubator and checked daily for cytopathic changes. Three days after inoculation, supernatant from Vero cells was collected and used for virus genome characterization.

Arbitrarily primed PCR and virus genome sequencing

To characterize the viral genome, we used a random amplification deep-sequencing approach as the first step. Virus-containing supernatant was centrifuged for 10 min at 3,000 rpm to remove cellular debris. This supernatant was then filtered through a 0.45- μ m-pore-size centrifugal filter unit (Millipore, Amsterdam, The Netherlands) to minimize bacterial contamination. Omnicleave endonuclease (Epicenter Biotechnologies, Madison, WI) was used to remove any free DNA and RNA, according to the manufacturer's protocol. Subsequently, viral RNA was extracted from the purified, infected cell culture supernatant using a High Pure RNA isolation kit (Roche Diagnostics, Almere, The Netherlands). To remove contaminating mammalian rRNA, a Ribo-Zero RZH110424 rRNA removal kit (Epicenter Biotechnologies, Madison, WI) was used according to the manufacturer's protocol. RNA was reverse transcribed using circular permuted primers [64] that were extended with random hexamer sequences, namely, CCCACCACCAGAGAGAAAN(6), ACCAGAGAGAAACCCACCN(6), GAGAAACCCACCACCAGAN(6), GGAGGCAAGCGAACGCAAN(6), AAGCGAACGCAAGGAGGCN(6), and ACGCAAGGAGGCAAGCGAN(6). Per reaction, reverse transcription mixtures contained 6 μ l RNA, 1 μ l primer (20 pmol), 0.5 μ l (20 U) RNase inhibitor (Promega, Leiden, The Netherlands), 1 μ l (10 mM each) deoxynucleoside triphosphates (Roche), and 5 μ l water. After a 5 min incubation at 65°C for optimal primer hybridization to the template, 4 μ l (10 \times) first-strand buffer, 1 μ l (200 U/ μ l) SuperScript III reverse transcriptase (Invitrogen, Bleiswijk, The Netherlands), 1 μ l (0.1 M) dithiothreitol (DTT), and 0.5 μ l (20 U) RNase inhibitor (Promega) were added to the mixture in a 20- μ l volume. To obtain cDNA, the reverse transcription mixture was sequentially incubated at 25°C for 5 min and at 42°C for 1 h. After 3 min at 95°C and 2 min on ice, 1 μ l Klenow DNA polymerase (5 U) (New England BioLabs Inc., Ipswich, MA) was added and the mixture was sequentially incubated at 25°C for 5 min, 37°C for 1 h, and 75°C for 20 min to obtain double-stranded cDNA. The cDNA was purified using a MinElute PCR purification kit (Qiagen, Venlo, The Netherlands) according to the instructions of the manufacturer. To amplify the purified cDNA, a PCR with the individual circular permuted primers without the random hexamer was performed. The PCR mixture contained 2 μ l primer (40 pmol), 2 μ l purified cDNA, 1.25 μ l (10 mM each) deoxynucleoside triphosphate (Roche), 5 μ l (10 \times) PfuUltra II Rxn buffer, and 1 μ l (2.5 U) PfuUltra II DNA polymerase (Stratagene, Amsterdam, The Netherlands). Water was added to reach a final volume of 50 μ l. The PCR mixture was incubated at 95°C for 2 min and then for 40 cycles of 95°C for 20 s, 56°C for 1 min, and 72°C for 2 min, followed by a final extension at 72°C for 10 min. Fragments were purified using a MinElute PCR purification kit (Qiagen) according to the instructions of the manufacturer.

Amplified fragments were sequenced using a 454/Roche GS Junior sequencing platform. A fragment library was created according to the manufacturer's protocol without DNA fragmentation (GS FLX Titanium rapid library preparation; Roche), selecting for fragments larger than 100 bp. The emulsion-based PCR (emPCR) (amplification method Lib-L) and GS Junior sequencing run

were performed according to the instructions of the manufacturer (Roche). The sequence reads were trimmed at 30 nt from the 3' and 5' ends to remove all primer sequences. Sequence reads were assembled into contigs using CLC Genomics 5.5.1 software (CLC Bio, Aarhus, Denmark). Using this deep-sequencing approach, approximately 90% of the virus genome sequence was obtained.

As a second step, specific primers were designed to amplify overlapping fragments of approximately 800 bp by RT-PCR. These PCR products were purified from agarose gels and sequenced using a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Nieuwerkerk a/d IJssel, The Netherlands) and a 3130xl genetic analyzer (Applied Biosystems), according to the instructions of the manufacturers. The genomic 5'- and 3'-terminal sequences were determined using a FirstChoice RLM-RACE kit (Ambion, Bleiswijk, The Netherlands).

Virus classification

Newly identified members of the family *Coronaviridae* are generally assigned by the ICTV to a subfamily and genus on the basis of rooted phylogeny and calculation of pairwise evolutionary distances for seven replicase polyprotein domains [56]: the ADP-ribose 1"-phosphatase (ADRP) in nsp3, the coronavirus 3C-like (3 CL) protease (3CLpro, or "main protease") in nsp5, the RNA-dependent RNA polymerase (RdRp) in nsp12, the helicase (Hel) in nsp13, the exoribonuclease (ExoN) in nsp14, the nidoviral endoribonuclease specific for U (NendoU) in nsp15, and the ribose-2'-O-methyltransferase (O-MT) in nsp16. Amino acid sequence alignments were generated for each of these domains using ClustalW within the BioEdit (version 7.0.5.3) [47] program and concatenated, after which the sequence identity of HCoV-EMC/2012 with closely related strains was calculated. For this purpose, the full genomes of 9 strains, derived from 3 species, belonging to *Betacoronavirus* lineage C were available.

To support virus classification, protein-based phylogenetic trees were generated. Multiple amino acid alignments, including sequences of HCoV-EMC/2012 and one representative of each of the 20 recognized species of the subfamily *Coronavirinae*, were produced for the following proteins, using the Viralis platform [177] followed by manual correction: ADRP, the N-terminal part of PLP2, TM1, Y domain, nsp4 to nsp16, and the C-terminal part of the spike (S) protein (S2), envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein. From each protein alignment, the most informative blocks [178] were extracted using the BAGG program [179], and only these strongly conserved alignment regions were used for further analyses. Two concatenated alignments were used. The first included replicase pp1ab protein regions (4,110 aa positions, gap content of 0.9%), and the second included regions in the C-terminal domain of the S protein (S2) and the E, M, and N proteins (1,127 aa positions, gap content of 3.9%). ProtTest version 3.2 [180] was used to select the best-fitting model of protein evolution. For both datasets, the LG

model with rate heterogeneity (4 categories) ranked top among 112 models tested, with a relative weight of 0.98 under the Bayesian information criterion (BIC) and 0.74 under the corrected Akaike information criterion (AICc) for the first data set and 0.97/0.75 (BIC/AICc) for the second data set. Hence, this model was applied for maximum likelihood phylogeny reconstruction using PhyML version 3.0 [181].

Phylogenetic reconstruction

Nucleotide sequences were aligned using the ClustalW software running within the BioEdit (version 7.0.5.3) [47] program and MAFFT version 6 [182]. Maximum likelihood phylogenetic trees with 100 bootstrap replicates were estimated under the general time-reversible model (GTR) + I + Γ 4 and the transversion model (TVM) + I + Γ 4 (determined by ModelTest [54]), using PhyML 3.0 software [181]. For both the 332-nt ORF1ab alignment that included isolate VM314/2008 and the alignment of the complete ORF1ab, the GTR + I + Γ 4 model ranked top among 65 models tested, with relative weights of 0.8185 and 1.000 under AIC, respectively.

Nucleotide sequence accession number

The final HCoV-EMC/2012 consensus sequence was submitted to GenBank under accession number JX869059.

ACKNOWLEDGMENTS

We thank Igor Sidorov, Dmitry Samborskiy, and Alexander Kravchenko (partially supported through the MoBiLe program) for help and administration of the Viralis.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7) under EMPIRE grant agreement no. 223498 and SILVER grant agreement no. 260644.

Excessive Production and extreme editing of Human Metapneumovirus Defective Interfering RNA is Associated with Type I Interferon Induction

Bernadette G. van den Hoogen, Sander van Boheemen, Jonneke de Rijck,
Stefan van Nieuwkoop, Derek J. Smith, Brigitta Laksono, Alexander Gultyaev,
Albert D.M.E. Osterhaus and Ron A.M. Fouchier.

J Gen Virol. 2014 Aug;95(Pt 8):1625-33

SUMMARY

Type I interferon (IFN) production is one of the hallmarks of host innate immune responses upon virus infection. While most respiratory viruses carry IFN-antagonists, reports on human Metapneumovirus (HMPV) have been conflicting. Using deep sequencing we demonstrate that HMPV particles accumulate excessive amounts of defective interfering RNA (DIs) rapidly upon in-vitro passage, which are associated with IFN induction. Importantly, the DIs were edited extensively; up to 70% of the original A and T residues had mutated to G or C respectively. Such high editing rates of viral RNA have not been reported before. Bioinformatics and PCR assays indicated that Adenosine Deaminase acting on RNA (ADAR) is the most likely editing enzyme. HMPV thus has an unusually high propensity to generate DIs, which are edited at an unprecedented high frequency. The conflicting published data on HMPV IFN induction and antagonism are likely explained by DIs in virus stocks. The interaction of HMPV DIs with the RNA editing machinery and IFN responses warrants further investigation.

INTRODUCTION

Human metapneumovirus (HMPV), a paramyxovirus in the genus *Metapneumovirus* within the subfamily *Pneumovirinae* [183], is responsible for 5-15% of hospitalizations of children suffering from acute respiratory tract infections [183, 184]. Phylogenetic analysis has identified two main genetic lineages, with isolates NL/1/00 and NL/1/99 as prototypes of lineages A and B, respectively [185].

As a first line of defence, the innate immune system protects cells against viral infections through production of type-I interferons (IFN). For successful infection and replication, viruses use a variety of strategies to evade the host innate immune system (reviewed in [186]). Respiratory Syncytial Virus (RSV), the mammalian pneumovirus most closely related to HMPV, uses two non-structural proteins (NS1 and NS2) as IFN antagonists [187]. Homologues of NS1 and NS2 are absent in the genome of HMPV [188], indicating that HMPV must have a different strategy to evade the innate immune system. It has been suggested that HMPV is a strong activator of the RIG-I and/or MAVS signalling pathways leading to IFN production [189-191]. In contrast, the interaction of the attachment protein (G) of HMPV with RIG-I resulted in inhibition of RIG-I activation [192]. However, using siRNA silencing of G, an antagonistic role of G was not confirmed [193]. Goutagny *et al.* [194] reported that only NL/1/00 induced IFN production, in contrast to NL/1/99. Using the same virus strains, we could not reproduce IFN production upon NL/1/00 infection. These discrepancies between different studies lead us to hypothesize that defective interfering RNAs (DIs) present in virus stocks may be responsible for the reported activation of the IFN pathway. DIs consist of partially deleted viral genomes, which spontaneously arise when virus stocks are generated by passaging at high multiplicity of infection in mammalian cells, due to errors made by the replicase complex [195]. The presence of DIs in virus stocks has already been reported in the 60's and their generation by diverse DNA and RNA viruses is widely documented [196-198]. DIs with a "copyback" or "snapback" structure are strong IFN inducers, which correlated with their ability to self-anneal and form dsRNA [197, 198]. "Copyback" types are generated during generation of the (-) strand by the replicase which leaves its (+) stranded template and resumes synthesis at the beginning of the "daughter" genome (-) chain it is still carrying. The complementary ends of the two chains are responsible for the circular "panhandle" structures that these genomes form as naked RNA [199]. "Snapback" DIs are generated in a similar way as the "copyback" DIs, but the replicase does not only copy-back the 3' end, but when it crosses to the nascent strand it copies back the complete strand, resulting in a covalently linked plus and minus strand, forming a duplex [200].

Here we demonstrate for the first time that HMPV accumulates DIs with a snapback structure rapidly upon in-vitro passage causing activation of the IFN pathway upon infection. Strikingly, the genomes of each of the DIs displayed extensive Adenosine (A) to Guanidine (G) or Thymidine (T) to Cytosine (C) hypermutation. In addition, our data indicate that the RNA-editing enzyme Adenosine Deaminase Acting on RNA (ADAR) is the most likely enzyme responsible for editing of HMPV DIs.

RESULTS

IFN production upon HMPV infection.

The ability of HMPV to induce IFN production was tested upon inoculation of A549 cells with two different virus stocks of recombinant HMPV strain NL/1/00. One stock was generated with a maximum of two passages in Vero-118 cells at low multiplicity of infection (m.o.i.) 0.01 (hereafter named P2_{low}), and a second stock was generated by five passages in Vero-118 cells at a high m.o.i. of 3 (hereafter named P5_{high}). A549 cells were inoculated with these stocks as well as Sendai virus (SeV), measles virus Edmonston strain (MeV-Edm), and parainfluenza type 5 (PIV-5) as controls. PIV-5 does not activate the IFN pathway [201], while SeV (Cantell strain) and MeV-Edm are known to be strong inducers [198, 202, 203]. 24-Hours after inoculation at m.o.i. 3, supernatants were tested for IFN content compared to mock-inoculated cell-supernatants. The supernatant of P2_{low} inoculated cells displayed similar low amounts of IFN as the supernatants of cells inoculated with PIV-5. In contrast, P5_{high} inoculated cells produced nearly as much IFN as SeV or MeV-Edm inoculated cells (Fig.1). The lack of IFN induction by P2_{low} was not due to lower infection efficiency as FACS analysis 24 hrs after inoculation revealed infection of more than 90% of the cells for all viruses. HMPV NL/1/00 thus only activated the IFN pathway, when virus stocks were prepared at high m.o.i. The ability of SeV, vesicular stomatitis virus (VSV), MeV-Edm, and PIV-5 to strongly activate the IFN pathway has been associated with the presence of DIs in these virus stocks [204-207]. In addition, studies with VSV revealed that the IFN-inducing capacity of DIs was not affected by ultra violet (UV) treatment [208]. UV treatment of P5_{high}, P2_{low} and SeV virus stocks resulted in reduced infection of A549 cells with P5_{high} and SeV, but did not affect IFN induction by these viruses. This suggests that the IFN induction by P5_{high} may be attributed to the presence of DIs (Fig. S1).

Deep sequencing demonstrates presence of DIs.

To investigate the possible presence of DIs, RNA from virus stocks of P2_{low} and P5_{high} was subjected to arbitrarily primed 454 sequencing. Mapping of the sequence reads against the full length HMPV genome resulted in 28.121 and 8.200 sequence reads for P2_{low} and P5_{high}, respectively. The genomes of P2_{low} and P5_{high} were covered with an average of 588 and 168 reads per nucleotide, respectively. No coverage was obtained for the last 50 and 4 nucleotides for P2_{low} and P5_{high}, respectively. In addition, the end of the genomes were covered with less than 10 reads from nt 13.164 for P2_{low} and from 13.325 for P5_{high}. In Fig. 2(a), depicting the ratio between the depths of sequence obtained for P2_{low} and for P5_{high} per nucleotide position, demonstrates that, at the end of the genome, P5_{high} displayed a higher depth of sequence than P2_{low}. From nt 11.726 onwards P5_{high} displayed up to 31 times more depth of sequence than P2_{low}. In fact, for P5_{high} 16,3% of the 8.200 reads mapped to the region beyond position 11kb of the genome, compared to only 2,2% of 28.121 reads obtained for P2_{low}. Examination of the P5_{high} reads at this part of

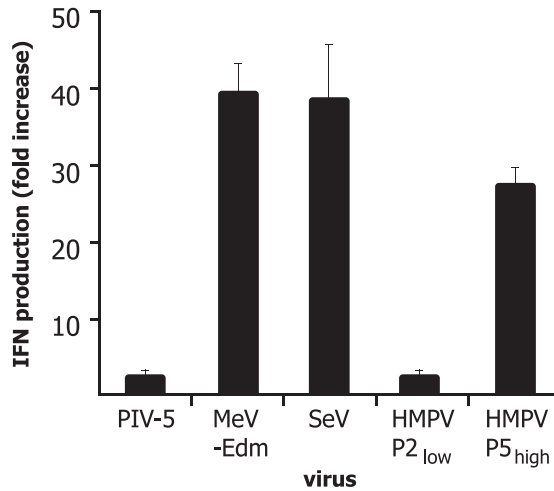


Figure 1. Induction of type I interferon in paramyxovirus infected A549 cells.

Production of IFN in A549 cells induced by PIV-5, MeV-Edm, SeV, HMPV P2_{low} and P5_{high} at 24 hours post inoculation. IFN production was measured using the ISRE-Firefly-Luciferase bioassay. Values indicate fold induction compared to mock infected cells. The experiment was conducted in duplo, a representative of three independent experiments is depicted.

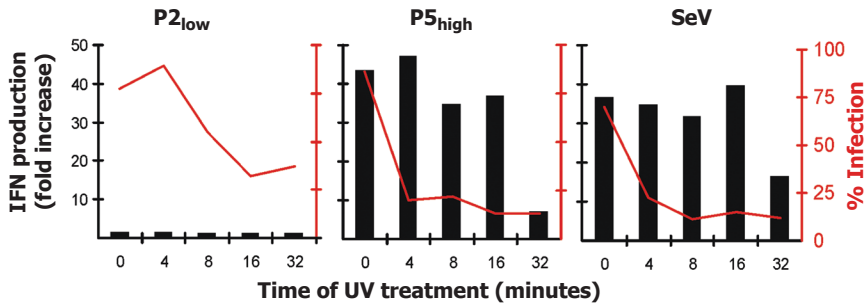


Figure S1. UV treatment blocks virus replication but not IFN induction.

To test whether HMPV P5_{high} harbored IFN inducing DIs, SeV, P2_{low} and P5_{high} viruses were treated with UV light for up to 32 minutes, after which A549 cells were inoculated at m.o.i. 3. UV treatment for 4 to 16 minutes resulted in reduced infection of A549 cells with P5_{high} and SeV, but did not affect IFN induction by these viruses. Virus stocks (equivalent to yield m.o.i. 3) were treated with UV light for 0 – 32 minutes after which A549 cells were inoculated. At 24 hours post inoculation cells were analyzed for infection efficiency by FACS analysis (red lines) and IFN production was measured using the ISRE-Luc bioassay. Values indicate fold induction compared to mock infected cells. A representative of 3 independent experiments is depicted.

the genome revealed the presence of reads that only aligned for approximately 50% to the reference sequence. Subsequent BLAST searches conducted with all sequence reads mapping for at least 20% to the reference sequence revealed the presence of reads that contained two copies of (almost) the same genome region: one part of the read mapping to the genomic (-) sequence and the other part mapping to the anti-genomic (+) sequence of the viral genome, with the two parts

being complementary. In total 47 double-stranded RNA structures were identified in P5_{high} and none in P2_{low}. All 47 reads aligned to the region upstream of position 11.200 of the viral genome, the same region where P5_{high} displayed deeper coverage than P2_{low}. Alignment of these reads revealed 12 unique structures that differed in sequence and position, as shown in Fig. 2(b). As assignment for the orientation of the two strands could not be done based on the deep sequencing results, the assignment was based on the results obtained in the BLAST search. The complementarity of the two strands indicated that P5_{high} contained DIs with a “snap-back” structure. The 12 DIs had variable lengths of the single stranded parts at the cross-over point (i.e. the position where the polymerase switches from the template to the daughter strand and where the plus and minus strands of the DI are connected). No specific sequence motif was detected at the position of cross-over points that may have served as a signal for the polymerase to switch strands.

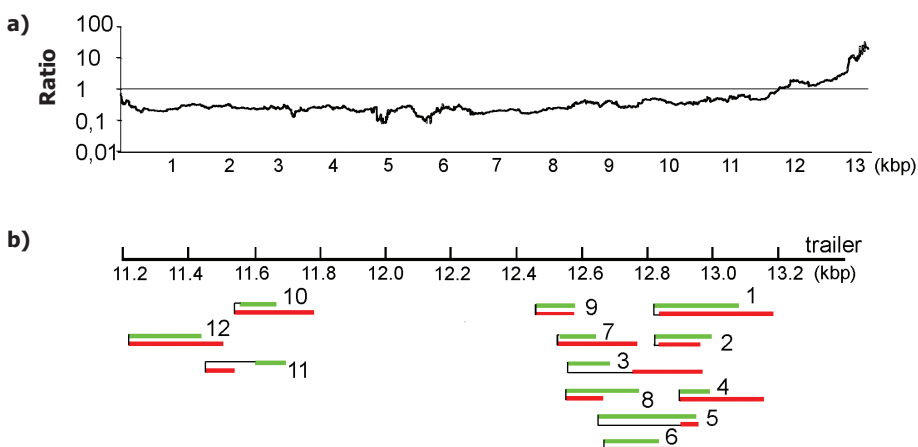


Figure 2. Deep sequencing of P2_{low} and P5_{high} demonstrated the presence of DIs in P5_{high}. a) Ratio between the coverage of depth per position between P2_{low} and P5_{high}. The x-axis corresponds to the 13.365 positions in the HMPV genome. b) The sequences obtained for 12 proposed defective interfering RNAs were mapped against the HMPV genome. Green bars indicate the genomic strand, the red bars the anti-genomic strand and the black lines indicate that the two bars were connected at that point.

Detection and quantification of DIs by Northern Blot assays.

To replicate, DIs need to contain polymerase initiation sites at the 5'end (genomic/minus strand) and the reverse complement of this trailer at the 3'end (antigenomic/plus strand). The deep-sequencing and bioinformatics approach only identified a read as a DI if the read overlapped with a cross-over point. All DI-reads were too short to reach from the cross-over point to the trailer sequences or the reverse complement thereof. To test whether the double stranded structures detected with deep-sequencing were indeed DIs with a “snap-back” structure, we performed electrophoresis of viral RNA and Northern blotting using a probe complementary to the trailer. This Northern blot assay demonstrated the presence of similar amounts of full length viral

genome in P5_{high} and P2_{low} (Fig. 3, top band) with an additional band, approx. size 7 kb, detected in both virus stocks. The identity of the 7 kb fragment is unknown. In addition to these large RNA molecules, smaller RNAs ranging in size from 0.5 to 4 kb were detected, but only in P5_{high}. As the RNA was analyzed on denaturing gels, fragments of this size may represent DIs, double-stranded in nature, of 0.25 to 2 kb in length. Quantification using the ChemiDoc Imaging system indicated that 80% of the total RNA, in the size range 0.5-4 kb for P5_{high}, was probably of DI origin.

In theory, the detection of double-stranded RNA structures could have been a result of the process of deep sequencing, as this method involves ligation of linkers to PCR products. RT-PCR assays directly on the virus stocks conducted with primers designed based on the consensus sequence obtained for DI#1 (Fig. S2, available at <http://vir.sgmjournals.org/>), confirmed the presence of DIs in the original virus stocks. Besides DI#1, these RT-PCR assays identified additional DIs that were not identified during analysis of the 454 sequencing reads (Fig. S3, available at <http://vir.sgmjournals.org/>).

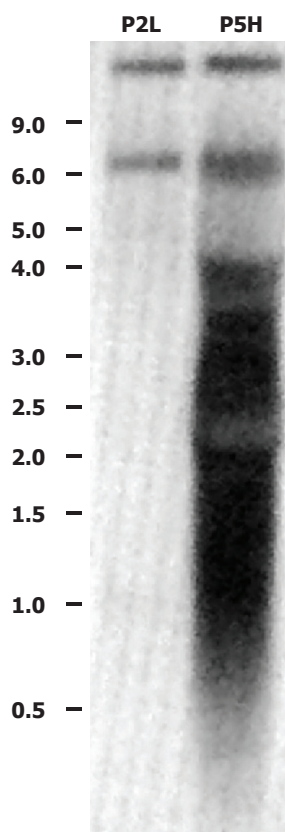


Figure 3. Northern blot analysis with a trailer probe showing the presence of small viral RNA. Upon RNA electrophoresis and hybridization with a probe complementary to the 5' end of the genome (trailer), the blot was exposed to a phosphorimager screen and analysed with a Storm 850 PhosphorImager and ImageQuant 1.2 software. Size markers (kb) are indicated on the left.

HMPV DI RNAs are hypermutated.

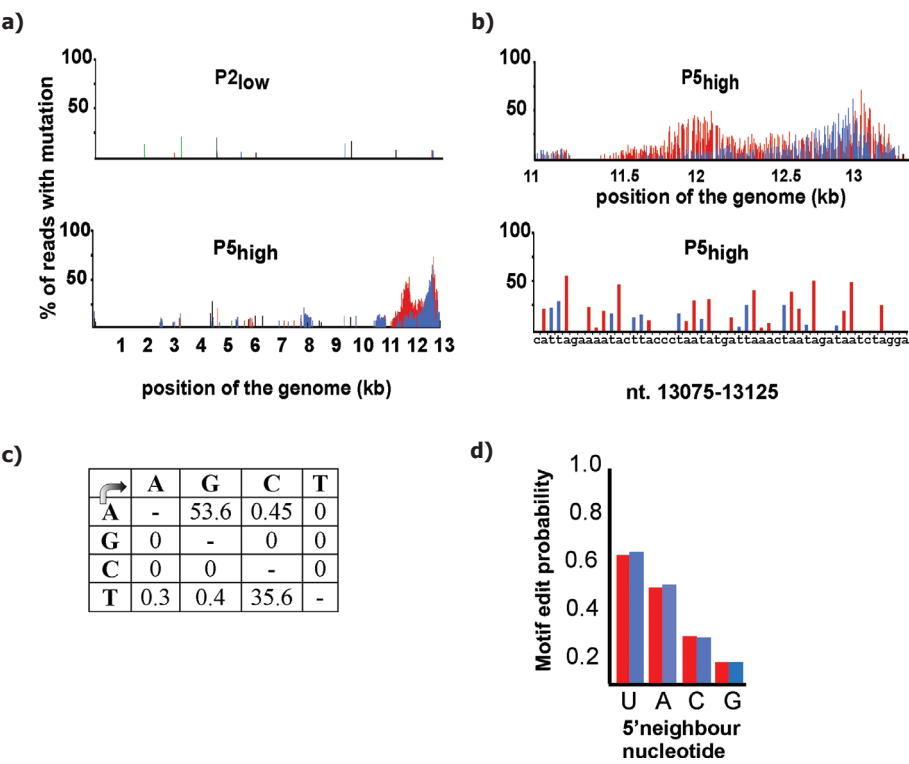
Analysis of the sequence variation of the deep-sequencing reads revealed extensive A to G and T to C mutations at the end of the genome of P5_{high} compared to the reference sequence, which was not observed in P2_{low} (Fig. 4a). The location of hypermutations matched with the region where P5_{high} displayed more sequence depth than P2_{low} and where DIs were detected. In the region downstream of nt 11.000, 53.6% of the A residues in the reference sequence displayed mutation to G and 35.7% of the T residues displayed mutation to C (in at least 2% of the reads mapping to that position) (Fig. 4b and c). In the region beyond position 12.000, 75% of the reads displayed hypermutation and the mutation percentages in this region were even higher: 87.3% A to G and 82.8% T to C mutation. Individual reads varied in the position and number of mutations, ranging between 5-77% of the A residues mutated to G or 5.7- 57.1% of the T residues mutated to C. We hypothesized that this kind of hypermutation was likely caused by the host RNA-editing enzyme Adenosine Deaminase Acting on RNA (ADAR). ADAR binds to dsRNA and deaminates A to inosine which is recognised as a G during virus replication [209]. Three human ADAR genes have been identified and two of these have editing activity, ADAR1 and 2.

Role of ADAR in editing of the DIs.

We analyzed whether the editing patterns in HMPV DIs were consistent with those described for ADAR editing [210]. Eggington *et al.*, refining earlier work [211], have determined that ADARs preferentially target adenosines with certain 5' and a variable 3' neighbor. They found that ADAR editing of an A to I, was more likely when the 5' neighbor was an U, followed by an A, C and G (U>A>C>G). The 3' preference was less specific, with a different preference for each of the four ADARs they considered and equal preference for some nucleotides. We analyzed all reads (n=954) mapping downstream of nt 11.000 for the editing probability depending on the 3' and 5' neighbor nucleotide. The average edit probability for each of the 16 combinations of 5' and 3' neighbors of an edited A or T was calculated. No pattern was identified for the 3' neighboring nucleotide. In contrast, the 5' neighbor preference was in accordance with the probabilities as described for ADAR [210], and were the same for A to G and T to C mutation (Fig. 4a).

In human cells, two isoforms of the ADAR1 protein are generated through the use of alternative promoters and alternative splicing. The long form (ADAR1-p150) is IFN inducible and localizes mostly in the cytoplasm, whereas a short form (ADAR1-p110) is constitutively expressed and localizes, like ADAR2, predominantly in the nucleus [212]. Vero cells are derived from African Green Monkeys, which might have different ADAR genes compared to human cells. RT-PCR assays were performed with primer sets designed based on the human ADAR1 sequences in combination with the available African Green Monkey ADAR1 sequences detected mRNA of both ADARs in Vero-118 cells (Fig. S4a, available at <http://vir.sgmjournals.org/>). Subsequent sequencing of the amplified fragments revealed that the Vero-ADAR1-p150 and Vero-ADAR-p110 mRNA displayed 95% nucleotide sequence identity with the human ADAR1-p150 and -p110 mRNAs and 99% with

those of macaques (Fig. S5, available at <http://vir.sgmjournals.org/>). In addition, Western blotting using an antibody against human ADAR1 revealed expression of ADAR1-p110 and ADAR1-p150 in Vero-118 cells, and upon IFN treatment a slight increase in expression of ADAR1-p150 was observed (Fig. S4b). Given the cytoplasmic replication of HMPV, the editing motifs and the expression of ADAR mRNA, we suggest that ADAR1 is the most likely enzyme responsible for editing of HMPV DIs.



DISCUSSION

As HMPV causes an acute respiratory infection in mammals it must -like other respiratory viruses- employ a mechanism to counteract the IFN response of the innate immune system. Several studies have reported on interactions between HMPV and the innate immune system, although with conflicting conclusions. We did not observe induction of IFN production for HMPV NL/1/00 and NL/1/99, upon passage at low m.o.i.. Only after a few passages at high m.o.i., HMPV- induced IFN production was observed. We showed that this induction was primarily caused by the presence of DIs with a “snapback” structure in the virus stocks, providing a plausible explanation for the discrepancies in the data reported so far.

In theory, the detection of double-stranded RNA structures could have been a result of the process of deep sequencing, as this method involves ligation of linkers to PCR products. However, no double-stranded structures were observed in P2_{low}, which was analysed using the same protocol. In addition, both the RT-PCR and the Northern blots were performed using RNA isolated directly from the virus stock. Of note, the deep-sequencing analyses, Northern blots, and RT-PCR assays performed using several independent P5_{high} stocks consistently revealed the presence of DIs and similar patterns of hypermutation. Our results indicated that different P5_{high} stocks contained various DIs.

Northern blot analyses revealed that DIs ranged in size between 0.5 and 2 kb and represented more than 80% of the RNA. Based on the Northern blot assays, the virus stock contained a larger number of DIs than detected with deep-sequencing. As smaller RNAs transfer more efficiently compared to larger RNAs with Northern blotting, 80% might be an overestimation. However, the deep-sequencing may be less efficient in obtaining reads mapping to the 5' end of the genome, which might result in an underestimation of the number of DI-reads. In addition, many more reads in the deep-sequencing analysis might have been of DI origin, but were not identified as such as they lacked the cross-over point.

Analysis of mutations in the deep-sequencing reads obtained for P5_{high} and P2_{low} revealed the presence of A to G and T to C mutation only in P5_{high} and only in the region where DIs were identified. In addition, all reads that were identified as DI-read and the DI-PCR products displayed hypermutation. Although we cannot exclude hypermutation of the viral genome, the location of observed hypermutation (in the polymerase gene) would, in case the viral genome is hypermutated, likely result in decreased viral replication. We observed similar replication of P2_{low} and P5_{high}, arguing against hypermutation of the viral genome itself (Fig. S6, available at <http://vir.sgmjournals.org/>).

So far, two mammalian RNA editing enzymes are known: apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC) and ADAR [213]. APOBEC deaminates cytosine to uridine on single-stranded DNA or RNA. APOBEC induced specific C to U editing has been described for retroviruses and DNA viruses but not for RNA viruses [214]. Commercial gene expression assays

did not detect expression of APOBEC mRNA in our Vero-118 cells (data not shown), while we did detect ADAR mRNAs. ADAR binds to dsRNA and deaminates adenosine to inosine that during replication is recognised as a guanine [209]. Three ADAR genes have been identified of which two have editing activity, ADAR1 and 2. The IFN inducible ADAR1-p150 resides in the cytoplasm, while the constitutively expressed ADAR1-p110 and ADAR2 reside in the nucleus [212]. For viruses that replicate in the cytoplasm of the infected host cell, such as HMPV, ADAR1-p150 is the most likely protein responsible for A-I editing [215]. We did not observe a clear up-regulation of ADAR1-p150 mRNA in IFN treated Vero-118 cells, and only a low level of up-regulation of protein in IFN treated A549 and Vero cells. In accordance, in studies detecting expression of endogenous ADAR1-p150, only low expression levels were observed [216]. Vero-118 cells are IFN-deficient [217], which would argue against a role for ADAR1-p150. But despite the absence of IFN-genes in Vero cells, it is well known that expression of interferon stimulated genes (e.g. ISG56) can be up-regulated in an IFN independent manner [218]. We did observe up-regulation of ISG56 mRNA upon SeV or P5_{high} inoculation of Vero-118 cells (Fig. S7, available at <http://vir.sgmjournals.org/>). Tripathi *et al.* (Poster J6-3044, Keystone Symposia; Innate Immunity to Viral Infections) found indications for IFN independent induction of ADAR expression through concerted action of IRF3 and IRF7, although the induction levels were low. However, the promoter that initiates the ADAR1-p150 transcript has a basal constitutive activity, which may be sufficient for editing activity [219].

ADAR induced hypermutation of MeV has been studied in detail, but it is unclear why this hypermutation was not observed upon inoculation of Vero cells with MeV [220, 221]. Although we have strong indications for the role of ADAR in HMPV infected Vero cells, studies in ADAR-deficient cells and RNA-binding-studies should elucidate the actual role of ADAR in editing HMPV (DI) RNA. The restricted host range of HMPV does not allow studies in ADAR deficient MEFs or Hela-cells, and the low expression of ADAR will be a challenge for RNA-binding studies.

DIs have been demonstrated to play a role in the pathogenesis of some virus infections, such as in acute dengue virus and influenza virus infections [222-224]. Given the relative ease by which HMPV accumulates DIs and the role demonstrated for DIs, the role that DIs may play in the pathogenesis of HMPV is clearly of interest.

In addition, DIs present in virus stocks have been shown to enable persistent infections both in vivo and in vitro ([225-227]. This is thought to be a result of activation of the innate immune system by DIs [197]. For HMPV, prolonged shedding of virus after experimental infection of mice has been reported [228-230]. It would be interesting to investigate whether the virus stocks used in these persistently infected animals contained DIs, as described for RSV ([231, 232].

Extensive A-to-I editing has been described in RNAs from MeV isolated from patients with persistent central nervous system infection [221, 233]. Subsequent studies demonstrated that ADAR1 acts as a proviral host factor in the context of measles virus infection [234]. In contrast, ADAR has an anti-viral function in other virus infections [215, 235]. Future research has to elucidate the role of ADAR in HMPV pathogenesis.

In conclusion, upon high m.o.i. passaging, HMPV accumulated DIs rapidly, and these DIs were responsible for a robust induction of IFN production. Our results should redirect research aiming to elucidate how HMPV subverts the innate immune system. Based on the high rate, and the patterns, of hypermutation, the preference of the 5' neighboring nucleotide, and the expression of ADAR in Vero-118 cells, we suggest that HMPV DIs are edited by ADAR. Based on the fact that HMPV (DIs) replicate in the cytoplasm of the infected cells, we speculate that HMPV (DI) genomes are hypermutated by ADAR1-p150.

METHODS

Cells and viruses

293-T cells and Vero-subclone 118 cells were grown as described [236]. A549 cells were grown in HAM's F-12 medium (Invitrogen) supplemented with 10% fetal calf serum, High Clone (HC-FCS, Greiner Bio-One) and PSG: 100 IU/ml penicillin, 100 µg/ml streptomycin and 2 mM glutamine. The generation of recombinant HMPV NL/1/00 has been described previously [236]. After virus rescue, supernatants were used to inoculate Vero-118 cells in infection medium: IMDM (Life Technologies) supplemented with PSG and 3.75 µg/ml trypsin (BioWhittaker). Upon reaching 80% cytopathic effect (CPE), cells and virus containing supernatants were harvested. One stock was generated with a maximum of two passages in Vero-118 cells at low multiplicity of infection (m.o.i.) 0.01, and a second stock was generated by five passages in Vero-118 cells at a high m.o.i. of 3. High titer virus stocks of MeV-Edm [237] and PIV-5 (strain W3) were generated with m.o.i 0.03 in Vero-118 cells in IMDM supplemented with 3% HI-FCS and PSG. After 4 days, cells and virus containing supernatants were harvested. For MeV-Edm, PIV-5 and HMPV, harvested cells and virus cultures were centrifuged 5 minutes at 1500 rpm and the cell-free supernatants were subsequently purified on sucrose gradients (60-30%) and aliquots were stored at -80°C. SeV strain Cantell was grown in 10-day-old embryonated chicken eggs at 37 °C for 48 hours. The titer of the virus stocks were determined by end-point titration in Vero-118 cells and expressed in tissue culture infectious dose 50 (TCID₅₀)/ml. Inoculation of A549 cells was done in HAM's F-12 supplemented with 2% HC-FCS and PSG.

Bioassays for IFN

IFN contents in supernatants were measured using the ISRE-Firefly-Luciferase reporter construct as described [238].

cDNA synthesis and PCR

RNA isolation and cDNA synthesis was performed as described previously [183]. Amplification of DI genomes was performed with primers BF568: '5-CTCTGCATTCCCTAGATTATC-3' (based on wt sequence), BF593: 5'-CCGCCGCCACAGAAGAATGGC-3' and BF595: 3'-GTTGACCATTTG GACTCTATGGCC-5', based on sequences obtained for DI#1 (Figure 4A). Primers for ADAR-1 p150: FWD (based on human ADAR-1 p150 [239]: 5'- GCGCAATGAATCCGCGG-3' and REV (based on AGM ADAR-1 p150 (Genbank Acc. Nr. EF190455): 5'-CCCCTGCCTTTCCATGTCAATTAGC-3'. Primers for ADAR-1 p110: FWD (start of ADAR-1 p150 in exon 1A): 5'-GAGAAGGCTACGTGGTGG-3' and REV (end of Exon 2; Genbank Acc. Nr. NM00111.4): 5'-AAAAAACTCAAGAGGATCTTCCAAGGC-3'. Thermocycling was performed with the following conditions: 94°C for 1 min, 40°C for 2 min, 72°C for 3 min (40 cycles). Gene expression assays were purchased from Applied Biosystems: Hs01020780-m1 for ADAR-1 p150, Hs01017596-m1 for ADAR-1 p110 and Hs00210562-m1 for ADAR-2. Real time PCR assays were performed with the 7000 Sequence Detection System (Applied Biosystems), using the Taqman Universal Mastermix according to the manufacturer's instructions.

Arbitrarily primed PCR and virus genome sequencing

Performed as described previously [81]. GS Junior sequence reads were trimmed 25 nt at both the 5' and 3' ends of the reads to remove primer sequences and aligned to the HMPV genome (Genbank Acc. 340 Nr. AF371337), using CLC Genomics software 4.6.1 (CLCbio, Aarhus, Denmark). Using this software, SNP analysis was conducted with a minimum of 2% variation frequency and a minimal 342 coverage of 5 reads per position.

Northern blot analysis

RNA was isolated from purified virus stock (10^7 TCID₅₀ per sample) using TRIzol (Invitrogen) according to the manufacturers' instructions. After electrophoresis of the RNA on a 1% agarose gel, RNA was transferred to a positively charged nylon membrane (Ambion, BrightStar-Plus) and incubated with the probe. The probe, complementary to the 5' end of genome, was generated with the Strip-EZ RNA system (Ambion) using a PCR fragment based on the full length cDNA construct of NL/1/00 [236] with PCR primers: FWD 5'-TAATACGACTCACTATAGGGGAAAATGATAAAATG-3' (T7 promoter underlined) and 5'-GGTTTTTTTGCCGT-3' (reverse complement to end of trailer). The blot was analyzed with the NorthernMax-Gly system (Ambion), exposed to a phosphorimager screen and analyzed with a Storm 850 PhosphorImager and ChemiDoc MP Imaging System and Image Lab 4.0.1 software (Molecular Dynamics Inc.).

ACKNOWLEDGEMENTS

We thank Adolfo García-Sastre, John Hiscott, Luis Martinez, Estanislao Nistal Villán, Jenish Patel, Frank van Kuppenveld and Eric Snijder for helpful discussions.

This research was supported in part by VENI grant # 916.66.092 of the Netherlands Organization for Scientific Research (NWO), the EU-FP7-Health project SILVER (Grant #260644), the EU-FP7-Health project EMPIRE (grant # 223498), the Virgo consortium, funded by the Dutch government project number FES0908, and by the Netherlands Genomics Initiative (NGI) project number 050-060-452.

Conflict of Interest

A.D.M.E Osterhaus is shareholder and C.E.O of ViroClinics Bioscience B.V. Other authors declare that they have no conflict of interest.

Summarizing Discussion

Global emergence of viral pathogens has caused, and will continue to cause, a major impact on human and animal health and welfare. The economic and social impact of both localized outbreaks and pandemics are immense. Rapid identification and characterization of the causative agent during an infectious disease outbreak is crucial to facilitate the implementation of measurements that may limit further spread and overall impact. Improvements in sequencing technologies have reduced the time needed to identify emerging pathogens, including several viruses. With current sequencing technologies being upgraded and new sequencing technologies arising, the field of virus discovery will continue to progress substantially.

VIRUS DIAGNOSTICS

Improved classical virus diagnostics

Current (classical) diagnostics remain vital for the detection of viruses. Although the initial detection of previously unknown viruses will be based on random amplification and NGS with increasing frequency, the use of more classical diagnostic techniques continue to be crucial for the rapid diagnostics of large numbers of samples. The use of RT-PCR in this case is a good example. Upon discovery of the MERS-CoV and publication of its complete genome [81, 156], this genome sequence was immediately used to design specific RT-PCR assays to rapidly test for the presence of MERS-CoV in large numbers of other samples from humans and animals [77]. The use of these assays greatly reduced the time needed to detect the virus as compared to the initial detection using random amplification. Also, for high-throughput use, RT-PCR assays have become sufficiently cheap to become widely applied.

The RT-PCR assay detecting all known paramyxoviruses described in this thesis has been successfully used to test relatively large numbers of both human and animal samples for the detection of paramyxoviruses [240, 241]. With the addition of this family-wide PCR-based assay to detect paramyxoviruses, family-wide assays are now available for emerging viruses belonging to most important virus families, facilitating the rapid discovery of unknown pathogens during virus outbreaks in the future.

It is important to note that not just the rapid identification of pathogens has improved recently, but also the detection of virus-specific antibodies, to investigate previous exposure of human or animal populations to newly emerging pathogens, which is important for risk assessments and epidemiological studies. Protein micro-arrays to detect virus-specific antibodies have for example been used to investigate the circulation of MERS-CoV in animal populations in several parts of the world [242]. The advantage of serological assays for this purpose, is that serology provides information on past virus circulation, whereas direct virus detection only provides information on current virus infections. The development of luciferase immunoprecipitation systems (LIPS) [243] provides a similar improvement over classical enzyme-linked immunosorbent assays (ELISA).

LIPS is based on the fusion of protein antigens to a light-emitting enzyme reporter. The antigen fusions are immunoprecipitated with serum samples and protein A/G beads, and the amount of light production that is measured is related to antibody titers.

Although methods for both virus detection and serology will continue to be improved with new technologies, classical techniques such as virus isolation in cell cultures can never be fully replaced. Fulfilling Koch's postulates is a requirement to causally link a newly identified virus to the disease observed. Live virus preparations are needed to test Koch's postulates, and hence the use of cell cultures will remain indispensable. In addition, when the infectious virus load needs to be determined in specimens, only the use of cell cultures allows the efficient quantification of live viruses, because genome sequences or genome copy numbers do not provide information about virus infectivity. Furthermore, obtaining live viruses will remain crucial for basic scientific research and the development of model systems to evaluate antiviral drugs, vaccines, and other intervention options.

Virus diagnostics using NGS

The choice of NGS platform used for sequencing depends on the specific aims of the investigator. When it comes to virus discovery, long sequence reads are beneficial as they generally result in longer contigs, which provide more statistical power to find related sequences in nucleotide databases. Errors introduced during amplification or sequencing are less important for initial virus identification. However, when one is interested in detecting single-nucleotide variations (SNVs) e.g. for virus characterization, the quality of sequencing and sequencing depth become more important parameters.

So far, the most used platforms in sequencing new virus genomes are Sanger, Illumina and Roche 454 sequencing. Illumina sequencing has a high sequence yield and the cost per base is relatively low. Disadvantages of this platform are characteristic systematic base calling biases [17], differences in sequencing quality of the sequencing plate in reads from different tiles [18], a higher sequencing error rate at the 3' ends of the reads as compared to the 5' ends [244], and increased single-base errors in association with GGC motifs [245]. 454 sequencing produces longer reads, and produces them relatively fast. Disadvantages are the high error rate in regions containing homopolymer stretches [19, 20] and the fact that up to 15% of the reads produced are a consequence of artificial amplification [21]. A comparison between Sanger, Illumina and Roche 454 sequencing is presented in table 1. This thesis focused on the Roche 454 platform, primarily because of the long read lengths. Unfortunately, the Roche 454 platform is expected to be discontinued shortly, and hence new investments will need to be made to optimize virus identification and characterization protocols using alternative NGS technologies. However, several of the steps described below will be useful using other NGS platforms.

Table 1. An overview of the comparison of costs and performances of Sanger 3130xl, Roche 454 and Illumina MiSeq DNA sequencers.

| | Purchase Cost | Reagent Cost/run | Run time | Reads/ run | Bases/ readrun | Yield MB/ Errors | Primary rate | Error |
|----------------|---------------|------------------|----------|---------------------|----------------|------------------|--------------|--------|
| Sanger 3730xl | \$376,000 | \$144 | 2 hrs. | 96 | 650 | 0,06 | Substitution | 0,1-1% |
| Roche 454 | \$108,000 | \$1,100 | 11 hrs. | 1 x 10 ⁵ | 400 | 50 | Indel | 1% |
| Illumina MiSeq | \$125,000 | \$1,040 | 26 hrs. | 4 x 10 ⁶ | 150+150 | 1,200 | Substitution | 0,1% |

Based on Glenn *et al.*, 2011 [290], and updated for 2013.

VIRUS DISCOVERY PROTOCOL

We used NGS technology for the rapid identification and characterization of emerging viruses that involves a number of steps from the acquisition of the sample until the identification and/or characterization of a virus. These steps include the purification of nucleic acids from the sample, the construction of a nucleic acid library, the amplification of this library, the sequencing step itself, and the use of computational algorithms to process the large datasets in order to determine the presence of a virus and identify and characterize it. Figure 1 describes the steps required from sample acquisition until virus identification and the time required for completing each step. Two steps were further optimized in the context of this PhD program using the Roche 454 platform and are further discussed below; sample preparation and data analysis. The other steps -library construction, emulsion PCR, and the 454 sequencing itself- were performed using generic protocols.

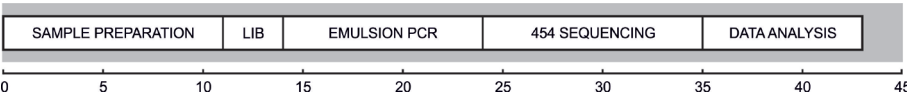


Figure 1. Overview of the virus discovery workflow from arrival of a sample to pathogen discovery using the Roche 454 platform. Time scale is shown in hours. LIB: library preparation.

Sample preparation

Before a sample from e.g. humans, animals, or the environment can be sequenced to detect possible virus presence, a procedure to remove as many non-viral nucleic acids as possible is required. Current procedures rely on the physical properties of virus particles in which the nucleic acids are protected from nuclease treatment, and that allows separation of viruses from cellular debris and microorganisms. The procedure starts off with low-speed centrifugation of the sample -that consists of either liquid material (e.g. respiratory specimens) or lysed cells from tissues-, pelleting most of the bacterial and mammalian cells. The supernatant is passed through a filter

to further remove microorganisms and large debris from the sample. Next, the filtrate is treated with RNases and DNases to remove free RNA and DNA from microorganisms or the host that were present in the filtrate. Once this treatment is completed, the remaining nucleic acids (DNA and/or RNA) that were protected from nucleases in the previous step are isolated from the sample. The virus discovery workflow can now be split to detect either RNA viruses or DNA viruses, or both. To detect DNA viruses, the nucleic acids can be amplified by PCR directly (see below). For the RNA virus detection procedure, a bacterial ribosomal RNA (rRNA) removal step is applied. The bacterial ribosomal proteins protect the rRNA in a similar way as viral genomes are protected by viral structural proteins, and as a consequence bacterial rRNA can represent a substantial proportion of RNA in sample collections. After this rRNA removal step, the remaining RNA is reverse transcribed using random primers. These primers have a random combination of six nucleotides, which will anneal to arbitrary regions of RNA and DNA sequences. The random hexamer primers were extended with specific sequences of 18 nucleotides [64]. After this reverse transcription step, a second strand DNA synthesis step is performed. The double-stranded DNA (either obtained from the reverse transcription reaction or directly from the DNA in the sample) is amplified randomly by PCR using primers specific for the 18 nucleotide sequences. After PCR, the amplified DNA is purified to remove primers and salts, and its concentration is measured.

Library preparation and emulsion PCR

The next step is the construction of a 454 library, where adaptors are ligated to the DNA fragments for use in subsequent purification, quantitation, amplification and sequencing. After library construction, the DNA fragments are bound to beads, which are incorporated in lipid droplets of an oil-in-water emulsion. PCR will create millions of clonal copies of each library fragment on each bead. After this step, the beads are purified from the emulsion and sequencing can commence. All of these steps are generic within the 454 platform and did not require optimization for virus discovery and virus characterization.

Bioinformatics

While it is relatively easy to generate millions of bases of sequence, it is a challenge to analyze that data efficiently. Key processes in the analysis of NGS data are quality scoring, sequence assembly, and annotation.

Several approaches can be used to analyze reads once a run is completed. If a reference virus genome is available, the sequences can be mapped directly to this reference using a mapper such as the Burrows-Wheeler Alignment tool (BWA) [246]. This alignment can provide information about substitutions, insertions, deletions, variations and gene loss. However, if one wants to look at novel genes or sequences, this method is not applicable. In this case, or where a suitable reference genome is not available altogether, sequencing reads can be assembled into contigs using a de novo assembler [247]. De novo assembly algorithms find overlapping information between

reads, leading to the generation of contiguous sequences or contigs. However, because virus genomes may be highly diverse, sequencing reads or contigs may not show much homology to known viruses and therefore it may still be hard to identify new virus sequences. Because mapping to a reference genome takes less computational power as compared to de novo assembly, an obvious primary step is to align all reads from a sequencing run to bacterial or host genomes (i.e. human genome for samples originating from humans). One way to do this is to use the short oligonucleotide alignment program (SOAP) [248]. Alternatively, commercial software packages such as the CLC genomics software package integrate multiple algorithms for NGS data analyses. Viral metagenomics is a common method to identify and characterize all virus genomes within a certain sample. The metagenomics field had grown tremendously over the last 10 years. While there were 13 publications on metagenomics in 2003, the number of publications had increased to well over 1,500 by 2013. An interesting aspect of viral metagenomics is the fact that a large number of sequence reads do not have significant similarity to any known sequences available from the nucleotide or protein databases. A typical human viral metagenomics sample can contain between 60% and 99% of unknown sequences [249-252]. This is caused by variation of the sample type, the length of the sequence reads, the homology search method used, the similarity threshold (i.e. E-value cutoff), and the database used. Although many researchers usually ignore these reads of unknown origin, they could represent novel virus species or even novel viral families or orders [21, 24, 252]. To date, suitable bioinformatic methods to characterize these unknown sequences are not available. Investments should be made to better characterize the “dark matter” of NGS metagenomics, perhaps by combining sequences from multiple centers for analysis. Larger comparable datasets may produce longer contigs for analysis. Comparison of datasets from different hosts or sample types combined with searches for residual homology with known sequences will be a challenging but interesting task.

EMERGING VIRUSES

MERS-CoV

On June 13, 2012, a 60-year-old man from Saudi Arabia was admitted to a private hospital in Jeddah, Saudi Arabia. He had a 7-day history of fever, cough, expectoration, and shortness of breath. 11 days after admission, on June 24, 2012, the patient died of progressive respiratory and renal failure. From respiratory samples of the patient, a previously unidentified coronavirus was isolated (named HCoV-EMC) which was determined to be the most likely cause of the disease. After this patient from Jeddah was reported [156], a person from Qatar who became unwell during a visit to Saudi Arabia with severe respiratory disease and renal failure was transferred to a hospital in London, where he was diagnosed with an infection caused by the same virus [157]. There was no direct link between the two infected patients or a potential common source of

infection. The viruses associated with this outbreak are now collectively known as Middle East Respiratory Syndrome (MERS)-CoV [253].

Upon determination of the first full-genome of the MERS-CoV and immediate submission of the sequence to GenBank after completion, other research teams were able to design primersets to amplify and sequence additional MERS-CoV isolates for analysis [254]. Moreover, the full-genome sequence facilitated the design of real-time reverse-transcription polymerase chain reaction assays targeting diverse regions of the MERS-CoV genome, one upstream of the E gene (upE) and one within the open reading frame (ORF) 1b, as well as synthetic nucleic acid sequences to be used as positive controls in diagnostic units [77]. These primers, nucleic acids, and assays are instrumental to trace new cases of infection, virus transmission, to identify animal reservoir host(s), and trace the epidemiology.

As a consequence of the availability of sensitive and specific diagnostic tests for MERS-CoV infections, the first known MERS-CoV infections were detected retrospectively in Jordan. An undiagnosed outbreak of Severe Acute Respiratory Infection (SARI) occurred in the Intensive Care Unit (ICU) of a hospital in Zarga, Jordan, in April 2012. The outbreak involved 11 individuals including nurses, a doctor and a relative of one of the nurses. This outbreak led to one death of a nurse with underlying medical conditions. Two specimens of the cases in this cluster were tested later, and were found to be positive for the MERS-CoV. Both cases had fatal outcomes; the nurse and the original patient that was admitted to the ICU with SARI [255].

It became clear rapidly that the MERS-CoV infections were not the result of a single continuous human-to-human transmission chain [256]. Nevertheless, the MERS-CoV has been detected in patients from countries in the Middle East, Europe, North Africa, and North America, all with direct connections to the Arabian peninsula [156, 157, 257]. From April 2012 onwards, a total number of 699 laboratory-confirmed cases have been reported to the WHO, including 204 fatalities [258]. The outbreak is particularly concerning to health officials during the Hajj pilgrimage in Mecca, Saudi Arabia [259]. Although most infections of human coronaviruses are mild, the emergence of MERS-CoV is a reminder that coronaviruses may sometimes cause severe and fatal infections in humans.

Although coronaviruses of bats (HKU4, HKU5) were initially described to be close relatives of MERS-CoV, it was clear from the beginning that other (intermediate) animal hosts might play a role in the emergence of MERS-CoV. Recently, MERS-CoV neutralizing antibodies in dromedary camels provided clues for the presence of this virus in these animals in Oman, Egypt, and the Canary Islands [260, 261]. Subsequently, dromedary camels from a farm in Qatar proved to be positive for fragments of MERS-CoV and virus neutralizing antibodies, and the viral sequences from these dromedary camels showed high sequence identity to sequences from two human MERS-CoV cases linked to this farm [262]. Although the direct transmission of virus from dromedary camels to humans could not be proven with certainty, this is clearly the most likely route of transmission from animals to humans.

The sequence identity between the first sequenced MERS-CoV genome and that of other published full-genomes is relatively low; the identity between the genome sequence of MERS-CoV EMC and other human isolates deposited in the GenBank database is approximately 99,4%. The full-genome sequence of a MERS-CoV isolated from the nasal cavity of an 8-month-old dromedary camel in Doha, Qatar, shared 98,8% sequence identity. Figure 2 shows a phylogenetic analysis of full-genomes of MERS-CoV, compared with the first strain, HCoV-EMC. The fact that the MERS-CoV from a camel was similar to those isolated from humans who were infected more than one year earlier in the same region, suggested that this virus is maintained within camel populations. Indeed, the most recent data indicate that multiple genetic variants of MERS-CoV are circulating in dromedary camels, and that close relatives of several of these variants also affected humans [263]. This supports the hypothesis that MERS-CoV can be transmitted from dromedary camels to humans.



Figure 2. Phylogenetic analysis of complete genomes of Middle East respiratory syndrome coronaviruses (MERS-CoVs). Adapted from V. Stalin Raj *et al.*, 2014, in press.

Currently, much basic scientific research is ongoing. A crucial study recently showed that in contrast to the SARS-CoV -that used the angiotensin converting enzyme 2 (ACE2) as a receptor for binding to human cells [264]- the MERS-CoV uses dipeptidyl peptidase 4 (DPP4), also known as CD26, as a functional receptor [265]. The evolutionarily conserved DPP4 protein of *Pipistrellus* *Pipistrellus* bats was also shown to be used by MERS-CoV to infect cells. Given that DPP4 is highly variable among various animal species, identification of the functional receptor may help to identify appropriate animal models to study intervention options (e.g. drugs and vaccines), which currently are not yet available.

Polyomavirus

The use of random amplification combined with NGS was also used to identify a possible viral infection of a 70-year-old male, who had coarse outgrowths on his face which had been there for several months. Even though viral sequences from Merkel cell polyomavirus (MCV) were found, this virus could not be appointed as the etiological agent. This uncertainty stems from the fact that MCV was also detected on the surface of the normal-appearing skin of healthy individuals using a high throughput metagenomic sequencing approach [266]. Uncertainty about causation is a common problem when using metagenomic approaches, because among the viruses found in a sample, many can be apathogenic. However, due to specific circumstances such as immunosuppression, such viruses can suddenly cause disease. Given the very low frequency of detection of the symptoms seen in the 70-year-old man, it was still important to report the -inconclusive- results, in the hope that other cases elsewhere in the world would benefit from the reported data on diagnosis and potential treatment, and continued investigations on potential MCV-associated disease in immunocompromised individuals.

Other virus discovery programs

Numerous other virus discovery approaches combined with NGS have also been successful. The VIDISCA method is a virus discovery tool to identify RNA or DNA viruses. The method uses restriction enzyme recognition sites and subsequent PCR amplification, and was recently combined with NGS to identify new viruses in canine livers [267] as well as other virus discoveries.

Additionally, a mix of 96 hexamer primers has been designed to avoid targeting of rRNA but still ensures the capability of amplification of all known viruses by RT-PCR [268]. However, the exclusion of hexamers that do not align with any known virus, the possibility of detecting new virus families decreases substantially, hence such approaches may be less preferable than truly generic (random) approaches.

The method for virus discovery described in this thesis, i.e. removal of ribosomal RNA, cDNA synthesis with random primers extended with a known sequence, PCR amplification, and subsequent NGS, is now widely used by many research groups to identify previously unknown viruses from humans and animals from a wide variety of samples [269-272].

VIRUS CHARACTERIZATION

RNA viruses evolve rapidly, complicating the management of chronic infections and the control of emerging infectious agents [273]. This rapid evolution is driven by high mutation rates, rapid replication kinetics, and large population sizes, leading to many single nucleotide polymorphisms (SNPs). Combined with the complex evolutionary dynamics, this genetic diversity presents a challenge to conventional population genetics [274]. The quasispecies theory builds on classical population genetics, but explores the consequences of error-prone replication and large population sizes for genomic evolution [275, 276]. A viral quasispecies is defined as a cloud of diverse variants that are genetically linked through mutation, interact cooperatively on a functional level, and collectively contribute to the characteristics of the population [277]. NGS techniques (coined “deep sequencing”) provide unique opportunities to study viral quasispecies for emerging viruses. Influenza virus is an RNA virus that causes local and global emerging and reemerging disease and is exemplary for rapidly evolving viruses. Influenza A/H1N1, A/H3N2, and influenza B viruses currently circulate in humans. Worldwide, these annual epidemics cause around three to five million cases of severe illness, resulting in about 250.000 to 500.000 deaths [278]. At the same time, multiple virus subtypes (A/H5N1, A/H7N9, A/H10N8, A/H6N1, A/H3N2sw) continue to cause zoonotic infections. SNP detection has become an increasingly interesting tool for a number of applications in influenza surveillance. Molecular assays have been developed to detect SNPs in the influenza virus that lead to oseltamivir or zanamivir resistance using classical RT-PCR techniques [279, 280]. Deep sequencing now provides the opportunity to monitor and quantify the emergence of multiple resistance mutations simultaneously. Secondly, SNPs conferring single amino acid substitutions in hemagglutinin that lead to antigenic change were recently mapped for human influenza A/H3N2 virus from 1968 to 2003 [281] and for clade 2.1 A/H5N1 virus that continues to cause outbreaks in poultry and human cases of infection in Indonesia (Koel *et al.*, 2014, MBio in press). Since the hemagglutinin surface glycoprotein is the main component of influenza vaccines, SNPs that have been characterized to lead to escape of influenza viruses from vaccine-induced immunity might be detected at an early stage using SNP detection to predict future antigenic drift variants, allowing early updates of the influenza vaccine. For several zoonotic influenza viruses, mutations that may be associated with altered host range, replicative ability, virulence, and/or host responses to infection have been identified [282]. Critical adaptive mutations were also identified in the recently emerging A/H7N9 viruses [283, 284]. Using deep sequencing of clinical specimens from humans and animals, the detection of emergence of such “adaptive” mutations may function as early warning signals in surveillance studies. Two independent studies recently showed that avian A/H5N1 influenza viruses can become transmissible between ferrets via respiratory droplets. The A/Indonesia/5/2005 avian A/H5N1 influenza virus required five amino acid substitutions, and the A/Vietnam/1203/2004 A/H5N1 influenza virus required four substitutions and reassortment to acquire airborne transmissibility [89, 91]. The identified mutations and the

methods developed to detect the SNPs in clinical specimens by NGS [285, 286] could further improve the (predictive) value of surveillance studies. However, it has also become clear from mathematical models that were developed to determine the probability of emergence of airborne A/H5N1 virus during natural infections [123], that NGS techniques and surveillance intensity need to be improved and additional laboratory work is needed to get the most value out of such enhanced surveillance activities. The methods we described for of detecting acquisition of mutations driving mammalian adaptation and airborne transmission of A/H5N1 can be implemented in the high-throughput screening of field isolates for mutations leading to vaccine escape, anti-viral resistance and airborne transmission, as well as whole genome sequencing.

FUTURE OUTLOOK

Historically, the discovery of viruses as causative agents of disease occurred many years after a disease was recognized. For example, poliomyelitis has been known for centuries, but only much later a poliovirus was identified as the causative agent. Now and in the future, metagenomics studies will generate a surplus of genetic information on previously unknown virus sequences. While some of these unknown viral sequences could be associated with human disease, many will not, and some may only become associated with disease at a later stage. On one hand, NGS techniques will lead to increased preparedness, because sequences from potentially pathogenic (animal) viruses can be identified quickly or might even already be available from public sequence databases long before they cause disease in humans. On the other hand, proving which particular virus out of many is the causative agent of a specific disease may become increasingly difficult, especially if representative viable viruses are not available.

Since the launch of commercially available NGS in 2000, the number of published papers using this technology has increased tremendously. After the scientific community discovered the techniques for basic science, slowly but steadily NGS is being adapted to routine diagnostics. One example is the diagnosis of cancer by detecting a cryptic fusion oncogene by massive parallel sequencing [287]. NGS can also serve diverse applications in clinical virology. An example is the monitoring of a patient with an influenza virus infection to detect minor variations in neuraminidase that confer resistance against neuraminidase inhibitors. Detection of such resistance mutations as minor variants may change the treatment of this patient. As another example, in A/H5N1 virus surveillance studies using clinical samples from humans upon zoonotic infection, mutations that provide early warning signals for pandemic potential (e.g. associated with human-to-human transmission), may trigger aggressive intervention options such as quarantining patients and stamping out animal host populations, intervention options that may be too costly, complicated, and controversial to undertake routinely. Given the increasing availability of NGS technology and

decreasing costs of its use, even more routine testing of viruses in clinical samples may become feasible. When random amplification is used, there is no bias towards certain viral species or families, thus reducing the need for species-specific assays. In addition to a qualitative diagnosis (i.e. what virus is present), NGS techniques may provide quantitative indications (i.e. relative virus load). Given the fact that only a single PCR may be needed and that multiplexing allows for testing of multiple samples/patients per run, NGS techniques could potentially become a cost-effective approach in virus diagnostics in the near future.

For further integration of NGS in clinical diagnostics, numerous problems need to be solved. The first problem is the turn-around time between sample collection and final diagnosis. Ideally, results should be obtained within a single day, but most laboratory protocols are still very time-consuming, several sequencing machines have very long processing times, and hurdles also still exist in bioinformatics. For some applications, the current error frequencies associated with amplification and sequencing combined is around 0,1%, which may be too high if rare alleles are to be detected [123]. However, the use of circular sequencing (CirSeq), where the error introduced by amplification and sequencing is reduced to 10^{-12} because tandem repeats are generated from the same fragment at the reverse transcriptase step [288], is an excellent example of how innovative thinking can solve problems.

On the post-sequencing side, the fast growing size of data from NGS platforms requires increased analysis space (e.g. through cloud computing) and improved bioinformatics tools. An example of the latter is the SLIM iterative BLAST processing program [289], that removes abundant sequences and facilitates identification of rare virus sequences amidst large numbers of non-viral sequences. However, input from bioinformaticians and programmers is needed to improve sequence quality scoring, assembly, and annotation. Of note, the needs for virology applications as mentioned above may be quite distinct from the needs of other fields, e.g. related to required sequence depths and accuracy, to assembly of small variable genomes, or to distantly related species by BLAST.

Even though many technical parameters still need to be optimized, the use of NGS in both research and diagnostics will continue to develop and, in time, replace several of the current sequencing and diagnostic techniques. Continued investment in NGS technology will enable the screening of large numbers of samples, with increased value to human and animal health, and potentially at relative low cost.

1. Cleaveland, S., M.K. Laurenson, and L.H. Taylor, Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence. *Philos Trans R Soc Lond B Biol Sci*, 2001. 356(1411): p. 991-9.
2. Jones, K.E., et al., Global trends in emerging infectious diseases. *Nature*, 2008. 451(7181): p. 990-3.
3. Taylor, L.H., S.M. Latham, and M.E. Woolhouse, Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci*, 2001. 356(1411): p. 983-9.
4. Woolhouse, M.E. and S. Gowtage-Sequeria, Host range and emerging and reemerging pathogens. *Emerg Infect Dis*, 2005. 11(12): p. 1842-7.
5. Field, H.E., J.S. Mackenzie, and P. Daszak, Henipaviruses: emerging paramyxoviruses associated with fruit bats. *Curr Top Microbiol Immunol*, 2007. 315: p. 133-59.
6. Wang, L.F. and B.T. Eaton, Bats, civets and the emergence of SARS. *Curr Top Microbiol Immunol*, 2007. 315: p. 325-44.
7. Parrish, C.R., et al., Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiol Mol Biol Rev*, 2008. 72(3): p. 457-70.
8. Kuiken, T., et al., Public health. Pathogen surveillance in animals. *Science*, 2005. 309(5741): p. 1680-1.
9. Organization, W.H. WHO annual World Health Report 2007. 2007; Available from: <http://www.who.int/whr/2007/en/>.
10. EMPERIE (European Management Platform for Emerging and Re-emerging Infectious disease Entities). Available from: <http://emperie.eu/>.
11. Baltimore, D., Expression of animal virus genomes. *Bacteriol Rev*, 1971. 35(3): p. 235-41.
12. Virus Taxonomy: the International Committee on Taxonomy of Viruses. 2013; Available from: <http://www.ictvonline.org/>.
13. Liu, C., Rapid diagnosis of human influenza infection from nasal smears by means of fluorescein-labeled antibody. *Proc Soc Exp Biol Med*, 1956. 92(4): p. 883-7.
14. Saiki, R.K., et al., Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*, 1985. 230(4732): p. 1350-4.
15. Sanger, F., S. Nicklen, and A.R. Coulson, DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*, 1977. 74(12): p. 5463-7.
16. Maxam, A.M. and W. Gilbert, A new method for sequencing DNA. 1977. *Biotechnology*, 1992. 24: p. 99-103.
17. Erlich, Y., et al., Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nat Methods*, 2008. 5(8): p. 679-82.
18. Dolan, P.C. and D.R. Denver, TileQC: a system for tile-based quality control of Solexa data. *BMC Bioinformatics*, 2008. 9: p. 250.
19. Margulies, M., et al., Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 2005. 437(7057): p. 376-80.
20. Quince, C., et al., Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat Methods*, 2009. 6(9): p. 639-41.
21. Gomez-Alvarez, V., T.K. Teal, and T.M. Schmidt, Systematic artifacts in metagenomes from complex microbial communities. *ISME J*, 2009. 3(11): p. 1314-7.
22. Thurber, R.V., et al., Laboratory procedures to generate viral metagenomes. *Nat Protoc*, 2009. 4(4): p. 470-83.
23. Cox-Foster, D.L., et al., A metagenomic survey of microbes in honey bee colony collapse disorder. *Science*, 2007. 318(5848): p. 283-7.
24. Wooley, J.C. and Y. Ye, Metagenomics: Facts and Artifacts, and Computational Challenges. *J Comput Sci Technol*, 2009. 25(1): p. 71-81.
25. Schmidt, T.M., E.F. DeLong, and N.R. Pace, Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *J Bacteriol*, 1991. 173(14): p. 4371-8.
26. Victoria, J.G., et al., Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol*, 2009. 83(9): p. 4642-51.

27. Breitbart, M., et al., Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A*, 2002. 99(22): p. 14250-5.
28. Altschul, S.F., et al., Basic local alignment search tool. *J Mol Biol*, 1990. 215(3): p. 403-10.
29. Koonin, E.V., T.G. Senkevich, and V.V. Dolja, The ancient Virus World and evolution of cells. *Biol Direct*, 2006. 1: p. 29.
30. Iyer, L.M., et al., Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res*, 2006. 117(1): p. 156-84.
31. van Nimwegen, E., J.P. Crutchfield, and M. Huynen, Neutral evolution of mutational robustness. *Proc Natl Acad Sci U S A*, 1999. 96(17): p. 9716-20.
32. Wilke, C.O., et al., Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, 2001. 412(6844): p. 331-3.
33. Varghese, V., et al., Minority variants associated with transmitted and acquired HIV-1 nonnucleoside reverse transcriptase inhibitor resistance: implications for the use of second-generation nonnucleoside reverse transcriptase inhibitors. *Journal of acquired immune deficiency syndromes*, 2009. 52(3): p. 309-15.
34. Wang, C., et al., Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome research*, 2007. 17(8): p. 1195-201.
35. Lok, A.S., et al., Antiviral drug-resistant HBV: standardization of nomenclature and assays and recommendations for management. *Hepatology*, 2007. 46(1): p. 254-65.
36. Lamb, R.A. and G.D. Parks, Paramyxoviridae: The viruses and their replication, in *Fields Virology*, D.M. Knipe and P.M. Howley, Editors. 2007, Wolters Kluwer: Lippincott Williams & Wilkins: Philadelphia. p. 1449-1496.
37. Tregoning, J.S. and J. Schwarze, Respiratory viral infections in infants: causes, clinical symptoms, virology, and immunology. *Clin Microbiol Rev*, 2010. 23(1): p. 74-98.
38. Chua, K.B., et al., Nipah virus: a recently emergent deadly paramyxovirus. *Science*, 2000. 288(5470): p. 1432-5.
39. Chua, K.B., et al., Fatal encephalitis due to Nipah virus among pig-farmers in Malaysia. *Lancet*, 1999. 354(9186): p. 1257-9.
40. Murray, K., et al., A morbillivirus that caused fatal disease in horses and humans. *Science*, 1995. 268(5207): p. 94-7.
41. Kitchen, A., L.A. Shackelton, and E.C. Holmes, Family level phylogenies reveal modes of macroevolution in RNA viruses. *Proc Natl Acad Sci U S A*, 2011. 108(1): p. 238-43.
42. Scaramozzino, N., et al., Comparison of flavivirus universal primer pairs and development of a rapid, highly sensitive heminested reverse transcription-PCR assay for detection of flaviviruses targeted to a conserved region of the NS5 gene sequences. *J Clin Microbiol*, 2001. 39(5): p. 1922-7.
43. Drosten, C., et al., Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med*, 2003. 348(20): p. 1967-76.
44. Ksiazek, T.G., et al., A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med*, 2003. 348(20): p. 1953-66.
45. Kidd, A.H., et al., Rapid subgenus identification of human adenovirus isolates by a general PCR. *J Clin Microbiol*, 1996. 34(3): p. 622-7.
46. Tong, S., et al., Sensitive and broadly reactive reverse transcription-PCR assays to detect novel paramyxoviruses. *J Clin Microbiol*, 2008. 46(8): p. 2652-8.
47. Hall, T.A., BioEdit: A user friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, 1999. 41: p. 95-98.
48. Schneider, T.D. and R.M. Stephens, Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*, 1990. 18(20): p. 6097-100.
49. Malur, A.G., et al., Analysis of the mutations in the active site of the RNA-dependent RNA polymerase of human parainfluenza virus type 3 (HPIV3). *Gene Expr*, 2002. 10(3): p. 93-100.
50. Munster, V.J., et al., Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS Pathog*, 2007. 3(5): p. e61.

51. Poch, O., et al., Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. *J Gen Virol*, 1990. 71 (Pt 5): p. 1153-62.
52. Svenda, M., et al., Analysis of the large (L) protein gene of the porcine rubulavirus LPMV: identification of possible functional domains. *Virus Res*, 1997. 48(1): p. 57-70.
53. Swofford, D.L., PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0 Beta. 2002, Sinauer Associated, Inc.: Sunderland, MA.
54. Posada, D. and K.A. Crandall, MODELTEST: testing the model of DNA substitution. *Bioinformatics*, 1998. 14(9): p. 817-8.
55. Pierce, J.R., *An Introduction to Information Theory: Symbols, Signals and Noise*. 2nd ed. 1980, New York, N.Y.: Dover Publications, Inc.
56. de Groot, R.J., et al., Family Coronaviridae, in *Virus taxonomy: classification and nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Viruses.*, A.M.Q. King, et al., Editors. 2012, Elsevier Academic Press: San Diego. p. 806-828.
57. Peiris, J.S., et al., Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet*, 2003. 361(9366): p. 1319-25.
58. Rota, P.A., et al., Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science*, 2003. 300(5624): p. 1394-9.
59. Woo, P.C., et al., Comparative analysis of twelve genomes of three novel group 2c and group 2d corona viruses reveals unique group and subgroup features. *J Virol*, 2007. 81(4): p. 1574-85.
60. van Boheemen, S., et al., A family-wide RT-PCR assay for detection of paramyxoviruses and application to a large-scale surveillance study. *PLoS One*, 2012. 7(4): p. e34961.
61. Vijgen, L., et al., A pancoronavirus RT-PCR assay for detection of all known coronaviruses. *Methods Mol Biol*, 2008. 454: p. 3-12.
62. Verstrepen, W.A., et al., Rapid detection of enterovirus RNA in cerebrospinal fluid specimens with a novel single-tube real-time reverse transcription-PCR assay. *J Clin Microbiol*, 2001. 39(11): p. 4093-6.
63. Jothikumar, N., et al., Quantitative real-time PCR assays for detection of human adenoviruses and identification of serotypes 40 and 41. *Appl Environ Microbiol*, 2005. 71(6): p. 3131-6.
64. Welsh, J. and M. McClelland, Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Res*, 1990. 18(24): p. 7213-8.
65. Peiris, J.S., Y. Guan, and K.Y. Yuen, Severe acute respiratory syndrome. *Nat Med*, 2004. 10(12 Suppl): p. S88-97.
66. Hamre, D. and J.J. Procknow, A new virus isolated from the human respiratory tract. *Proc Soc Exp Biol Med*, 1966. 121(1): p. 190-3.
67. McIntosh, K., et al., Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc Natl Acad Sci U S A*, 1967. 57(4): p. 933-40.
68. Fouchier, R.A., et al., A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci U S A*, 2004. 101(16): p. 6212-6.
69. van der Hoek, L., et al., Identification of a new human coronavirus. *Nat Med*, 2004. 10(4): p. 368-73.
70. Woo, P.C., et al., Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol*, 2005. 79(2): p. 884-95.
71. Woo, P.C., et al., Molecular diversity of coronaviruses in bats. *Virology*, 2006. 351(1): p. 180-7.
72. Davies, A., et al., Extracorporeal Membrane Oxygenation for 2009 Influenza A(H1N1) Acute Respiratory Distress Syndrome. *JAMA*, 2009. 302(17): p. 1888-95.
73. Fraaij, P.L. and T. Heikkinen, Seasonal influenza: the burden of disease in children. *Vaccine*, 2011. 29(43): p. 7524-8.
74. Kaw, G.J., et al., Chest radiographic findings of a case of severe acute respiratory syndrome (SARS) in Singapore. *Singapore Med J*, 2003. 44(4): p. 201-4.
75. Peiris, J.S., et al., Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet*, 2003. 361(9371): p. 1767-72.
76. Wong, R.S., et al., Haematological manifestations in patients with severe acute respiratory syndrome: retrospective analysis. *BMJ*, 2003. 326(7403): p. 1358-62.

77. Corman, V.M., et al., Detection of a novel human coronavirus by real-time reverse-transcription polymerase chain reaction. *Euro Surveill*, 2012. 17(39).
78. Organization, W.H. Novel coronavirus infection - update - revised interim case definition. 2012; Available from: http://www.who.int/csr/don/2012_09_29/en/index.html.
79. Requena, L., et al., Follicular spicules of the nose: a peculiar cutaneous manifestation of multiple myeloma with cryoglobulinemia. *J Am Acad Dermatol*, 1995. 32(5 Pt 2): p. 834-9.
80. van der Meijden, E., et al., Discovery of a new human polyomavirus associated with trichodysplasia spinulosa in an immunocompromized patient. *PLoS Pathog*, 2010. 6(7): p. e1001024.
81. van Boheemen, S., et al., Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio*, 2012. 3(6).
82. Goh, S., et al., Merkel cell polyomavirus in respiratory tract secretions. *Emerg Infect Dis*, 2009. 15(3): p. 489-91.
83. Wanat, K.A., et al., Viral-associated trichodysplasia: characterization of a novel polyomavirus infection with therapeutic insights. *Arch Dermatol*, 2012. 148(2): p. 219-23.
84. Feltkamp, M.C., et al., From Stockholm to Malawi: recent developments in studying human polyomaviruses. *J Gen Virol*, 2013. 94(Pt 3): p. 482-96.
85. Linke, M., et al., Follicular Erythematous Papules with Keratotic Spicules: A Quiz. *Acta Derm Venereol*, 2014. 94(4): p. 492-496.
86. Satta, R., et al., Follicular spicules and multiple ulcers: cutaneous manifestations of multiple myeloma. *J Am Acad Dermatol*, 2003. 49(4): p. 736-40.
87. Viscidi, R.P., et al., Age-specific seroprevalence of Merkel cell polyomavirus, BK virus, and JC virus. *Clin Vaccine Immunol*, 2011. 18(10): p. 1737-43.
88. Claas, E.C., et al., Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet*, 1998. 351(9101): p. 472-7.
89. Herfst, S., et al., Airborne transmission of influenza A/H5N1 virus between ferrets. *Science*, 2012. 336(6088): p. 1534-41.
90. Chen, L.M., et al., In vitro evolution of H5N1 avian influenza virus toward human-type receptor specificity. *Virology*, 2012. 422(1): p. 105-13.
91. Imai, M., et al., Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature*, 2012. 486(7403): p. 420-8.
92. Zhang, Y., et al., H5N1 hybrid viruses bearing 2009/H1N1 virus genes transmit in guinea pigs by respiratory droplet. *Science*, 2013. 340(6139): p. 1459-63.
93. Aggarwal, S., et al., Biochemical impact of the host adaptation-associated PB2 E627K mutation on the temperature-dependent RNA synthesis kinetics of influenza A virus polymerase complex. *The Journal of biological chemistry*, 2011. 286(40): p. 34504-13.
94. Subbarao, E.K., W. London, and B.R. Murphy, A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *Journal of virology*, 1993. 67(4): p. 1761-4.
95. Taubenberger, J.K., et al., Characterization of the 1918 influenza virus polymerase genes. *Nature*, 2005. 437(7060): p. 889-93.
96. Steel, J., et al., Transmission of influenza virus in a mammalian host is increased by PB2 amino acids 627K or 627E/701N. *PLoS Pathog*, 2009. 5(1): p. e1000252.
97. Van Hoeven, N., et al., Human HA and polymerase subunit PB2 proteins confer transmission of an avian influenza virus through the air. *Proc Natl Acad Sci U S A*, 2009. 106(9): p. 3366-71.
98. Matrosovich, M., et al., Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J Virol*, 2000. 74(18): p. 8502-12.
99. Chutinimitkul, S., et al., In vitro assessment of attachment pattern and replication efficiency of H5N1 influenza A viruses with altered receptor specificity. *Journal of virology*, 2010. 84(13): p. 6825-33.
100. Sorrell, E.M., et al., Predicting 'airborne' influenza viruses: (trans-) mission impossible? *Current opinion in virology*, 2011. 1(6): p. 635-42.
101. Munster, V.J., et al., Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. *Science*, 2009. 325(5939): p. 481-3.

102. Matrosovich, M.N. and A.S. Gambaryan, Solid-phase assays of receptor-binding specificity. *Methods in molecular biology*, 2012. 865: p. 71-94.
103. Shaw, M.L. and P. Palese, Fields Virology, in Fields Virology, D.M. Knipe and P.M. Howley, Editors. 2013, Wolters Kluwer. p. 1151-1185.
104. Galloway, S.E., et al., Influenza HA subtypes demonstrate divergent phenotypes for cleavage activation and pH of fusion: implications for host range and adaptation. *PLoS pathogens*, 2013. 9(2): p. e1003151.
105. Carr, C.M., C. Chaudhry, and P.S. Kim, Influenza hemagglutinin is spring-loaded by a metastable native conformation. *Proceedings of the National Academy of Sciences of the United States of America*, 1997. 94(26): p. 14306-13.
106. de Wit, E., et al., Molecular determinants of adaptation of highly pathogenic avian influenza H7N7 viruses to efficient replication in the human host. *Journal of virology*, 2010. 84(3): p. 1597-606.
107. Belser, J.A., et al., Considerations regarding appropriate sample size for conducting ferret transmission experiments. *Future microbiology*, 2013. 8: p. 961-5.
108. Russell, W.M.S. and R.L. Burch, *The Principles of Humane Experimental Technique*. 1959, UK: Methuen & Co. Ltd.
109. Nishiura, H., H.L. Yen, and B.J. Cowling, Sample size considerations for one-to-one animal transmission studies of the influenza A viruses. *PLoS one*, 2013. 8(1): p. e55358.
110. Pappas, C., et al., Receptor specificity and transmission of H2N2 subtype viruses isolated from the pandemic of 1957. *PLoS one*, 2010. 5(6): p. e11158.
111. Tumpey, T.M., et al., A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science*, 2007. 315(5812): p. 655-9.
112. Wu, B., et al., New evidence suggests Southern China as a common source of multiple clusters of highly pathogenic H5N1 avian influenza virus. *The Journal of infectious diseases*, 2010. 202(3): p. 452-8.
113. de Vries, R.P., et al., Hemagglutinin receptor specificity and structural analyses of respiratory droplet-transmissible H5N1 viruses. *J Virol*, 2014. 88(1): p. 768-73.
114. Lu, X., et al., Structure and receptor-binding properties of an airborne transmissible avian influenza A virus hemagglutinin H5 (VN1203mut). *Protein & cell*, 2013. 4(7): p. 502-11.
115. Xiong, X., et al., Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature*, 2013. 497(7449): p. 392-6.
116. Zhang, W., et al., An airborne transmissible avian influenza H5 hemagglutinin seen at the atomic level. *Science*, 2013. 340(6139): p. 1463-7.
117. Liu, J., et al., Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. *Proceedings of the National Academy of Sciences of the United States of America*, 2009. 106(40): p. 17175-80.
118. Skehel, J.J. and D.C. Wiley, Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual review of biochemistry*, 2000. 69: p. 531-69.
119. Krenn, B.M., et al., Single HA2 mutation increases the infectivity and immunogenicity of a live attenuated H5N1 intranasal influenza vaccine candidate lacking NS1. *PLoS One*, 2011. 6(4): p. e18577.
120. Shelton, H., et al., Mutations in haemagglutinin that affect receptor binding and pH stability increase replication of a PR8 influenza virus with H5 HA in the upper respiratory tract of ferrets and may contribute to transmissibility. *The Journal of general virology*, 2013. 94(Pt 6): p. 1220-9.
121. Zaraket, H., et al., Increased Acid Stability of the Hemagglutinin Protein Enhances H5N1 Influenza Virus Growth in the Upper Respiratory Tract but Is Insufficient for Transmission in Ferrets. *Journal of virology*, 2013. 87(17): p. 9911-22.
122. Neumann, G., et al., Egyptian H5N1 influenza viruses-cause for concern? *PLoS pathogens*, 2012. 8(11): p. e1002932.
123. Russell, C.A., et al., The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science*, 2012. 336(6088): p. 1541-7.
124. Tharakaraman, K., et al., Structural determinants for naturally evolving H5N1 hemagglutinin to switch its receptor specificity. *Cell*, 2013. 153(7): p. 1475-85.
125. Liu, D., et al., Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses. *Lancet*, 2013. 381(9881): p. 1926-32.

126. Kageyama, T., et al., Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin*, 2013. 18(15): p. 20453.
127. Richard, M., et al., Limited airborne transmission of H7N9 influenza A virus between ferrets. *Nature*, 2013.
128. van Riel, D., et al., Novel avian-origin influenza A (H7N9) virus attaches to epithelium in both upper and lower respiratory tract of humans. *Am J Pathol*, 2013. 183(4): p. 1137-43.
129. Lam, T.T., et al., The genesis and source of the H7N9 influenza viruses causing human infections in China. *Nature*, 2013.
130. Shi, Y., et al., Structures and Receptor Binding of Hemagglutinins from Human-Infecting H7N9 Influenza Viruses. *Science*, 2013.
131. Zhu, H., et al., Infectivity, transmission, and pathology of human-isolated H7N9 influenza virus in ferrets and pigs. *Science*, 2013. 341(6142): p. 183-6.
132. Maher, J.A. and J. DeStefano, The ferret: an animal model to study influenza virus. *Lab Anim (NY)*, 2004. 33(9): p. 50-3.
133. Kuiken, T., et al., Experimental human metapneumovirus infection of cynomolgus macaques (*Macaca fascicularis*) results in virus replication in ciliated epithelial cells and pneumocytes with associated lesions throughout the respiratory tract. *Am J Pathol*, 2004. 164(6): p. 1893-900.
134. de Wit, E., et al., Efficient generation and growth of influenza virus A/PR/8/34 from eight cDNA fragments. *Virus research*, 2004. 103(1-2): p. 155-61.
135. Matrosovich, M., et al., The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *Journal of virology*, 1999. 73(2): p. 1146-55.
136. Salzberg, S.L., et al., Genome analysis linking recent European and African influenza (H5N1) viruses. *Emerging infectious diseases*, 2007. 13(5): p. 713-8.
137. van Riel, D., et al., Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *The American journal of pathology*, 2007. 171(4): p. 1215-23.
138. Nobusawa, E., et al., Change in receptor-binding specificity of recent human influenza A viruses (H3N2): a single amino acid change in hemagglutinin altered its recognition of sialyloligosaccharides. *Virology*, 2000. 278(2): p. 587-96.
139. Herfst, S., et al., Low-pH-induced membrane fusion mediated by human metapneumovirus F protein is a rare, strain-dependent phenomenon. *J Virol*, 2008. 82(17): p. 8891-5.
140. Dunn, J.J., et al., Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. *Gene*, 1988. 68(2): p. 259-66.
141. Matrosovich, M., et al., New low-viscosity overlay medium for viral plaque assays. *Viro J*, 2006. 3: p. 63.
142. Turell, L., et al., The role and assembly mechanism of nucleoprotein in influenza A virus ribonucleoprotein complexes. *Nat Commun*, 2013. 4: p. 1591.
143. Perlman, S. and J. Netland, Coronaviruses post-SARS: update on replication and pathogenesis. *Nat Rev Microbiol*, 2009. 7(6): p. 439-50.
144. Gloza-Rausch, F., et al., Detection and prevalence patterns of group I coronaviruses in bats, northern Germany. *Emerg Infect Dis*, 2008. 14(4): p. 626-31.
145. Lau, S.K., et al., Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A*, 2005. 102(39): p. 14040-5.
146. Li, W., et al., Bats are natural reservoirs of SARS-like coronaviruses. *Science*, 2005. 310(5748): p. 676-9.
147. Pfefferle, S., et al., Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. *Emerg Infect Dis*, 2009. 15(9): p. 1377-84.
148. Vijaykrishna, D., et al., Evolutionary insights into the ecology of coronaviruses. *J Virol*, 2007. 81(8): p. 4012-20.
149. Marra, M.A., et al., The Genome sequence of the SARS-associated coronavirus. *Science*, 2003. 300(5624): p. 1399-404.
150. Zlateva, K.T., et al., No novel coronaviruses identified in a large collection of human nasopharyngeal specimens using family-wide CODEHOP-based primers. *Arch Virol*, 2013. 158(1): p. 251-5.

151. Gorbalenya, A.E., et al., Nidovirales: evolving the largest RNA virus genome. *Virus Res*, 2006. 117(1): p. 17-37.
152. Adams, M.J. and E.B. Carstens, Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2012). *Arch Virol*, 2012. 157(7): p. 1411-22.
153. Lauber, C. and A.E. Gorbalenya, Partitioning the genetic diversity of a virus family: approach and evaluation through a case study of picornaviruses. *J Virol*, 2012. 86(7): p. 3890-904.
154. Masters, P.S., The molecular biology of coronaviruses. *Adv Virus Res*, 2006. 66: p. 193-292.
155. Snijder, E.J., et al., Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol*, 2003. 331(5): p. 991-1004.
156. Zaki, A.M., et al., Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*, 2012. 367(19): p. 1814-20.
157. Bermingham, A., et al., Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012. *Euro Surveill*, 2012. 17(40): p. 20290.
158. Ziebuhr, J., E.J. Snijder, and A.E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J Gen Virol*, 2000. 81(Pt 4): p. 853-79.
159. Pasternak, A.O., W.J. Spaan, and E.J. Snijder, Nidovirus transcription: how to make sense...? *J Gen Virol*, 2006. 87(Pt 6): p. 1403-21.
160. Sawicki, S.G., D.L. Sawicki, and S.G. Siddell, A contemporary view of coronavirus transcription. *J Virol*, 2007. 81(1): p. 20-9.
161. Sola, I., et al., RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol*, 2011. 8(2): p. 237-48.
162. Firth, A.E. and I. Brierley, Non-canonical translation in RNA viruses. *J Gen Virol*, 2012. 93(Pt 7): p. 1385-409.
163. Reusken, C.B., et al., Circulation of group 2 coronaviruses in a bat species common to urban areas in Western Europe. *Vector Borne Zoonotic Dis*, 2010. 10(8): p. 785-91.
164. Holmes, K.V. and M.C. Lai, Coronaviridae: the viruses and their replication, in *Field virology*, B.N. Fields, D.M. Knipe, and P.M. Howley, Editors. 1996, Lippincott-Raven: Philadelphia, PA. p. 1075-1093.
165. McIntosh, K., Coronaviruses, in *Fields virology*, B.N. Fields, D.M. Knipe, and P.M. Howley, Editors. 1996, Lippincott-Raven: Philadelphia, PA. p. 1095-1103.
166. Keng, C.T., et al., SARS coronavirus 8b reduces viral replication by down-regulating E via an ubiquitin-independent proteasome pathway. *Microbes Infect*, 2011. 13(2): p. 179-88.
167. Narayanan, K., C. Huang, and S. Makino, SARS coronavirus accessory proteins. *Virus Res*, 2008. 133(1): p. 113-21.
168. Senanayake, S.D. and D.A. Brian, Bovine coronavirus I protein synthesis follows ribosomal scanning on the bicistronic N mRNA. *Virus Res*, 1997. 48(1): p. 101-5.
169. Mazumder, R., et al., Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily. *Nucleic Acids Res*, 2002. 30(23): p. 5229-43.
170. Snijder, E.J., et al., Comparison of the genome organization of toro- and coronaviruses: evidence for two nonhomologous RNA recombination events during Berne virus evolution. *Virology*, 1991. 180(1): p. 448-52.
171. Yount, B., et al., Severe acute respiratory syndrome coronavirus group-specific open reading frames encode nonessential functions for replication in cell cultures and mice. *J Virol*, 2005. 79(23): p. 14909-22.
172. Cruz, J.L., et al., Coronavirus gene 7 counteracts host defenses and modulates virus virulence. *PLoS Pathog*, 2011. 7(6): p. e1002090.
173. de Haan, C.A., et al., The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. *Virology*, 2002. 296(1): p. 177-89.
174. Pewe, L., et al., A severe acute respiratory syndrome-associated coronavirus-specific protein enhances virulence of an attenuated murine coronavirus. *J Virol*, 2005. 79(17): p. 11335-42.
175. Frieman, M., et al., Severe acute respiratory syndrome coronavirus ORF6 antagonizes STAT1 function by sequestering nuclear import factors on the rough endoplasmic reticulum/Golgi membrane. *J Virol*, 2007. 81(18): p. 9812-24.

176. Oostra, M., C.A. de Haan, and P.J. Rottier, The 29-nucleotide deletion present in human but not in animal severe acute respiratory syndrome coronaviruses disrupts the functional expression of open reading frame 8. *J Virol*, 2007. 81(24): p. 13876-88.
177. Gorbalenya, A.E., et al., Practical application of bioinformatics by the multidisciplinary VIZIER consortium. *Antiviral Res*, 2010. 87(2): p. 95-110.
178. Castresana, J., Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol*, 2000. 17(4): p. 540-52.
179. Antonov, I.V., A.M. Leontovich, and A.E. Gorbalenya. BAGG - Blocks Accepting Gaps Generator, version 1.0. 2008; Available from: <http://www.genebee.msu.su/~antonov/bagg/cgi/bagg.cgi>.
180. Darriba, D., et al., ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, 2011. 27(8): p. 1164-5.
181. Guindon, S., et al., New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*, 2010. 59(3): p. 307-21.
182. Katoh, K., et al., MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*, 2002. 30(14): p. 3059-66.
183. van den Hoogen, B.G., et al., A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med*, 2001. 7(6): p. 719-24.
184. Williams, J.V., et al., Human metapneumovirus and lower respiratory tract disease in otherwise healthy infants and children. *N Engl J Med*, 2004. 350(5): p. 443-50.
185. van den Hoogen, B.G., et al., Antigenic and genetic variability of human metapneumoviruses. *Emerg Infect Dis*, 2004. 10(4): p. 658-66.
186. Randall, R.E. and S. Goodbourn, Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *J Gen Virol*, 2008. 89(Pt 1): p. 1-47.
187. Spann, K.M., K.C. Tran, and P.L. Collins, Effects of nonstructural proteins NS1 and NS2 of human respiratory syncytial virus on interferon regulatory factor 3, NF-kappaB, and proinflammatory cytokines. *J Virol*, 2005. 79(9): p. 5353-62.
188. van den Hoogen, B.G., et al., Analysis of the genomic sequence of a human metapneumovirus. *Virology*, 2002. 295(1): p. 119-32.
189. Banos-Lara Mdel, R., A. Ghosh, and A. Guerrero-Plata, Critical role of MDA5 in the interferon response induced by human metapneumovirus infection in dendritic cells and in vivo. *J Virol*, 2013. 87(2): p. 1242-51.
190. Bao, X., et al., Airway epithelial cell response to human metapneumovirus infection. *Virology*, 2007. 368(1): p. 91-101.
191. Liao, S., et al., Role of retinoic acid inducible gene-I in human metapneumovirus-induced cellular signalling. *J Gen Virol*, 2008. 89(Pt 8): p. 1978-86.
192. Bao, X., et al., Human metapneumovirus glycoprotein G inhibits innate immune responses. *PLoS Pathog*, 2008. 4(5): p. e1000077.
193. Preston, F.M., et al., siRNA against the G gene of human metapneumovirus. *Virol J*, 2012. 9: p. 105.
194. Goutagny, N., et al., Cell type-specific recognition of human metapneumoviruses (HMPVs) by retinoic acid-inducible gene I (RIG-I) and TLR7 and viral interference of RIG-I ligand recognition by HMPV-B1 phosphoprotein. *J Immunol*, 2010. 184(3): p. 1168-79.
195. Huang, A.S., Defective interfering viruses. *Annu Rev Microbiol*, 1973. 27: p. 101-17.
196. Rima, B.K., W.B. Davidson, and S.J. Martin, The role of defective interfering particles in persistent infection of Vero cells by measles virus. *J Gen Virol*, 1977. 35(1): p. 89-97.
197. Sekellick, M.J. and P.I. Marcus, Persistent infection. I Interferon-inducing defective-interfering particles as mediators of cell sparing: possible role in persistent infection by vesicular stomatitis virus. *Virology*, 1978. 85(1): p. 175-86.
198. Strahle, L., D. Garcin, and D. Kolakofsky, Sendai virus defective-interfering genomes and the activation of interferon-beta. *Virology*, 2006. 351(1): p. 101-11.
199. Kolakofsky, D., Isolation and characterization of Sendai virus DI-RNAs. *Cell*, 1976. 8(4): p. 547-55.
200. Lazzarini, R.A., J.D. Keene, and M. Schubert, The origins of defective interfering particles of the negative-strand RNA viruses. *Cell*, 1981. 26(2 Pt 2): p. 145-54.

201. Childs, K., R. Randall, and S. Goodbourn, Paramyxovirus V proteins interact with the RNA Helicase LGP2 to inhibit RIG-I-dependent interferon induction. *J Virol*, 2012. 86(7): p. 3411-21.
202. Takaki, H., et al., Strain-to-strain difference of V protein of measles virus affects MDA5-mediated IFN-beta-inducing potential. *Mol Immunol*, 2011. 48(4): p. 497-504.
203. Shingai, M., et al., Differential type I IFN-inducing abilities of wild-type versus vaccine strains of measles virus. *J Immunol*, 2007. 179(9): p. 6123-33.
204. Baum, A., R. Sachidanandam, and A. Garcia-Sastre, Preference of RIG-I for short viral RNA molecules in infected cells revealed by next-generation sequencing. *Proc Natl Acad Sci U S A*, 2010. 107(37): p. 16303-8.
205. Killip, M.J., et al., Deep sequencing analysis of defective genomes of parainfluenza virus 5 and their role in interferon induction. *J Virol*, 2013. 87(9): p. 4798-807.
206. Marcus, P.I. and M.J. Sekellick, Defective interfering particles with covalently linked [+/-]RNA induce interferon. *Nature*, 1977. 266(5605): p. 815-9.
207. Sekellick, M.J. and P.I. Marcus, Interferon induction by viruses. VIII. Vesicular stomatitis virus: [+/-]DI-011 particles induce interferon in the absence of standard virions. *Virology*, 1982. 117(1): p. 280-5.
208. Marcus, P.I. and C. Gaccione, Interferon induction by viruses. XIX. Vesicular stomatitis virus--New Jersey: high multiplicity passages generate interferon-inducing, defective-interfering particles. *Virology*, 1989. 171(2): p. 630-3.
209. Bass, B.L., RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem*, 2002. 71: p. 817-46.
210. Eggington, J.M., T. Greene, and B.L. Bass, Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun*, 2011. 2: p. 319.
211. Lehmann, K.A. and B.L. Bass, Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry*, 2000. 39(42): p. 12875-84.
212. George, C.X., et al., Adenosine deaminases acting on RNA, RNA editing, and interferon action. *J Interferon Cytokine Res*, 2011. 31(1): p. 99-117.
213. Hamilton, C.E., F.N. Papavasiliou, and B.R. Rosenberg, Diverse functions for DNA and RNA editing in the immune system. *RNA Biol*, 2010. 7(2): p. 220-8.
214. Bransteitter, R., C. Prochnow, and X.S. Chen, The current structural and functional understanding of APOBEC deaminases. *Cell Mol Life Sci*, 2009. 66(19): p. 3137-47.
215. Samuel, C.E., Adenosine deaminases acting on RNA (ADARs) are both antiviral and proviral. *Virology*, 2011. 411(2): p. 180-93.
216. Cavarec, L., et al., In vitro screening for drug-induced depression and/or suicidal adverse effects: a new toxicogenomic assay based on CE-SSCP analysis of HTR2C mRNA editing in SH-SY5Y cells. *Neurotox Res*, 2013. 23(1): p. 49-62.
217. Desmyter, J., J.L. Melnick, and W.E. Rawls, Defectiveness of interferon production and of rubella virus interference in a line of African green monkey kidney cells (Vero). *J Virol*, 1968. 2(10): p. 955-61.
218. Gelinas, J.F., et al., Enhancement of replication of RNA viruses by ADAR1 via RNA editing and inhibition of RNA-activated protein kinase. *J Virol*, 2011. 85(17): p. 8460-6.
219. Patterson, J.B. and C.E. Samuel, Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol Cell Biol*, 1995. 15(10): p. 5376-88.
220. Suspene, R., et al., Double-stranded RNA adenosine deaminase ADAR-1-induced hypermutated genomes among inactivated seasonal influenza and live attenuated measles virus vaccines. *J Virol*, 2011. 85(5): p. 2458-62.
221. Wong, T.C., et al., Generalized and localized biased hypermutation affecting the matrix gene of a measles virus strain that causes subacute sclerosing panencephalitis. *J Virol*, 1989. 63(12): p. 5464-8.
222. Li, D., et al., Defective interfering viral particles in acute dengue infections. *PLoS One*, 2011. 6(4): p. e19447.
223. Saira, K., et al., Sequence Analysis of In Vivo Defective Interfering-Like RNA of Influenza A H1N1 Pandemic Virus. *J Virol*, 2013. 87(14): p. 8064-74.
224. de Chasse, B., et al., The Interactomes of Influenza Virus NS1 and NS2 Proteins Identify New Host Factors and Provide Insights for ADAR1 Playing a Supportive Role in Virus Replication. *PLoS Pathog*, 2013. 9(7): p. e1003440.

225. Cave, D.R., F.M. Hendrickson, and A.S. Huang, Defective interfering virus particles modulate virulence. *J Virol*, 1985. 55(2): p. 366-73.
226. Roux, L., A.E. Simon, and J.J. Holland, Effects of defective interfering viruses on virus replication and pathogenesis in vitro and in vivo. *Adv Virus Res*, 1991. 40: p. 181-211.
227. Weiss, B., R. Levis, and S. Schlesinger, Evolution of virus and defective-interfering RNAs in BHK cells persistently infected with Sindbis virus. *J Virol*, 1983. 48(3): p. 676-84.
228. Alvarez, R., et al., Human metapneumovirus persists in BALB/c mice despite the presence of neutralizing antibodies. *J Virol*, 2004. 78(24): p. 14003-11.
229. Hamelin, M.E., et al., Human metapneumovirus infection induces long-term pulmonary inflammation as associated with airway obstruction and hyperresponsiveness in mice. *J Infect Dis*, 2006. 193(12): p. 1634-42.
230. Liu, Y., et al., Human metapneumovirus establishes persistent infection in the lungs of mice and is reactivated by glucocorticoid treatment. *J Virol*, 2009. 83(13): p. 6837-48.
231. Estripeaut, D., et al., Respiratory syncytial virus persistence in the lungs correlates with airway hyperactivity in the mouse model. *J Infect Dis*, 2008. 198(10): p. 1435-43.
232. Mejias, A., et al., Respiratory syncytial virus persistence: evidence in the mouse model. *Pediatr Infect Dis J*, 2008. 27(10 Suppl): p. S60-2.
233. Cattaneo, R., et al., Biased hypermutation and other genetic changes in defective measles viruses in human brain infections. *Cell*, 1988. 55(2): p. 255-65.
234. Toth, A.M., et al., RNA-specific adenosine deaminase ADAR1 suppresses measles virus-induced apoptosis and activation of protein kinase PKR. *J Biol Chem*, 2009. 284(43): p. 29350-6.
235. Taylor, D.R., et al., New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1. *J Virol*, 2005. 79(10): p. 6291-8.
236. Herfst, S., et al., Recovery of human metapneumovirus genetic lineages a and B from cloned cDNA. *J Virol*, 2004. 78(15): p. 8264-70.
237. de Vries, R.D., et al., In vivo tropism of attenuated and pathogenic measles virus expressing green fluorescent protein in macaques. *J Virol*, 2010. 84(9): p. 4714-24.
238. Patel, J.R., et al., ATPase-driven oligomerization of RIG-I on RNA allows optimal activation of type-I interferon. *EMBO Rep*, 2013. 14(9): p. 780-7.
239. George, C.X. and C.E. Samuel, Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible. *Proc Natl Acad Sci U S A*, 1999. 96(8): p. 4621-6.
240. Cha, S.Y., et al., Respiratory disease due to current egg drop syndrome virus in Pekin ducks. *Vet Microbiol*, 2013. 165(3-4): p. 305-11.
241. Lednicky, J.A., et al., Isolation and genetic characterization of human coronavirus NL63 in primary human renal proximal tubular epithelial cells obtained from a commercial supplier, and confirmation of its replication in two different types of human primary kidney cells. *Virol J*, 2013. 10: p. 213.
242. Reusken, C., et al., Specific serology for emerging human coronaviruses by protein microarray. *Euro Surveill*, 2013. 18(14): p. 20441.
243. Burbelo, P.D., et al., Antibody profiling by Luciferase Immunoprecipitation Systems (LIPS). *J Vis Exp*, 2009(32).
244. Schroder, J., et al., Reference-free validation of short read data. *PLoS One*, 2010. 5(9): p. e12681.
245. Nakamura, K., et al., Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res*, 2011. 39(13): p. e90.
246. Li, H. and R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 2009. 25(14): p. 1754-60.
247. Miller, J.R., S. Koren, and G. Sutton, Assembly algorithms for next-generation sequencing data. *Genomics*, 2010. 95(6): p. 315-27.
248. Li, R., et al., SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, 2009. 25(15): p. 1966-7.
249. Breitbart, M., et al., Metagenomic analyses of an uncultured viral community from human feces. *J Bacteriol*, 2003. 185(20): p. 6220-3.

250. Zhang, T., et al., RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS Biol*, 2006. 4(1): p. e3.
251. Allander, T., et al., Identification of a third human polyomavirus. *J Virol*, 2007. 81(8): p. 4130-6.
252. Finkbeiner, S.R., et al., Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog*, 2008. 4(2): p. e1000011.
253. de Groot, R.J., et al., Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol*, 2013. 87(14): p. 7790-2.
254. Cotten, M., et al., Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. *Emerg Infect Dis*, 2013. 19(5): p. 736-42B.
255. Pollack, M.P., et al., Latest outbreak news from ProMED-mail: novel coronavirus -- Middle East. *Int J Infect Dis*, 2013. 17(2): p. e143-4.
256. Cotten, M., et al., Transmission and evolution of the Middle East respiratory syndrome coronavirus in Saudi Arabia: a descriptive genomic study. *Lancet*, 2013. 382(9909): p. 1993-2002.
257. Gulland, A., Novel coronavirus spreads to Tunisia. *BMJ*, 2013. 346: p. f3372.
258. Organization, W.H. MERS-CoV summary updates. 2014; Available from: http://www.who.int/csr/disease/coronavirus_infections/archive_updates/en/index.html.
259. Al-Tawfiq, J.A., A. Zumla, and Z.A. Memish, Respiratory tract infections during the annual Hajj: potential risks and mitigation strategies. *Curr Opin Pulm Med*, 2013. 19(3): p. 192-7.
260. Perera, R.A., et al., Seroepidemiology for MERS coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in Egypt, June 2013. *Euro Surveill*, 2013. 18(36): p. pii=20574.
261. Reusken, C.B., et al., Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *Lancet Infect Dis*, 2013. 13(10): p. 859-66.
262. Haagmans, B.L., et al., Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. *Lancet Infect Dis*, 2014. 14(2): p. 140-5.
263. Alagaili, A.N., et al., Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *MBio*, 2014. 5(2): p. e00884-14.
264. Li, W., et al., Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature*, 2003. 426(6965): p. 450-4.
265. Raj, V.S., et al., Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature*, 2013. 495(7440): p. 251-4.
266. Foulongne, V., et al., Human skin microbiota: high diversity of DNA viruses identified on the human skin by high throughput sequencing. *PLoS One*, 2012. 7(6): p. e38499.
267. van der Heijden, M., et al., Sequence-independent VIDISCA-454 technique to discover new viruses in canine livers. *J Virol Methods*, 2012. 185(1): p. 152-5.
268. Endoh, D., et al., Species-independent detection of RNA virus by representational difference analysis using non-ribosomal hexanucleotides for reverse transcription. *Nucleic Acids Res*, 2005. 33(6): p. e65.
269. Mishra, N., et al., Identification of a Novel Polyomavirus in a Pancreatic Transplant Recipient With Retinal Blindness and Vasculitic Myopathy. *J Infect Dis*, 2014.
270. Tokarz, R., et al., Genome characterization of Long Island tick rhabdovirus, a new virus identified in *Amblyomma americanum* ticks. *Virol J*, 2014. 11: p. 26.
271. Ng, T.F., et al., Divergent picobirnaviruses in human feces. *Genome Announc*, 2014. 2(3).
272. Phan, T.G., et al., Cyclovirus in nasopharyngeal aspirates of Chilean children with respiratory infections. *J Gen Virol*, 2014. 95(Pt 4): p. 922-7.
273. Duffy, S., L.A. Shackelton, and E.C. Holmes, Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet*, 2008. 9(4): p. 267-76.
274. Holland, J., et al., Rapid evolution of RNA genomes. *Science*, 1982. 215(4540): p. 1577-85.
275. Biebricher, C.K. and M. Eigen, What is a quasispecies? *Curr Top Microbiol Immunol*, 2006. 299: p. 1-31.
276. Eigen, M., Viral quasispecies. *Sci Am*, 1993. 269(1): p. 42-9.

277. Lauring, A.S. and R. Andino, Quasispecies theory and the behavior of RNA viruses. *PLoS pathogens*, 2010. 6(7): p. e1001005.
278. Organization, W.H. Influenza (Seasonal). 2009; Available from: <http://www.who.int/mediacentre/factsheets/fs211/en/>.
279. van der Vries, E., et al., Molecular assays for quantitative and qualitative detection of influenza virus and oseltamivir resistance mutations. *J Mol Diagn*, 2013. 15(3): p. 347-54.
280. Wang, W., et al., PCR for detection of oseltamivir resistance mutation in influenza A(H7N9) virus. *Emerg Infect Dis*, 2014. 20(5): p. 847-9.
281. Koel, B.F., et al., Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science*, 2013. 342(6161): p. 976-9.
282. Prevention, C.f.D.C.a. H5N1 Genetic Changes Inventory. 2012; Available from: <http://www.cdc.gov/flu/avianflu/h5n1/inventory.htm>.
283. Neumann, G., C.A. Macken, and Y. Kawaoka, Identification of amino acid changes that may have been critical for the genesis of A(H7N9) influenza viruses. *J Virol*, 2014. 88(9): p. 4877-96.
284. Yamayoshi, S., et al., Virulence-affecting amino acid changes in the PA protein of H7N9 influenza A viruses. *J Virol*, 2014. 88(6): p. 3127-34.
285. Linster, M., et al., Identification, characterization, and natural selection of mutations driving airborne transmission of A/H5N1 virus. *Cell*, 2014. 157(2): p. 329-39.
286. Wilker, P.R., et al., Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. *Nat Commun*, 2013. 4: p. 2636.
287. Welch, J.S., et al., Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. *JAMA*, 2011. 305(15): p. 1577-84.
288. Acevedo, A., L. Brodsky, and R. Andino, Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*, 2013.
289. Cotten, M., et al., Full genome virus detection in fecal samples using sensitive nucleic Acid preparation, deep sequencing, and a novel iterative sequence classification algorithm. *PLoS One*, 2014. 9(4): p. e93269.
290. Glenn, T.C., Field guide to next-generation DNA sequencers. *Mol Ecol Resour*, 2011. 11(5): p. 759-69.

Nederlandse Samenvatting

De opkomst van virale pathogenen heeft wereldwijd grote gevolgen voor de gezondheid van mens en dier en voor de maatschappij en economie. Zo leidde de uitbraak van het SARS coronavirus in 2003 tot 8.096 besmette gevallen waarvan 774 met dodelijke afloop. Tegelijkertijd leidde het tot veel publieke onrust en enorme economische schade in de betrokken landen. Het is van cruciaal belang om tijdens een dergelijke uitbraak het pathogeen zo snel mogelijk te identificeren. De identificatie van het pathogeen kan ervoor zorgen dat specifieke maatregelen genomen kunnen worden om verdere verspreiding te minimaliseren. Ook kunnen dan specifieke tests worden ontwikkeld en behandelingsstrategieën worden ontwikkeld en uitgevoerd.

KLASSIEKE VIRUS DIAGNOSTIEK

De huidige diagnostiek van virussen is vooral gericht op de specifieke amplificatie van het erfelijk materiaal van virussen waarvan het genoom (het geheel van erfelijke informatie) bekend is en op het kweken van virussen uit klinische monsters. Door specifieke amplificatie van het erfelijk materiaal van virussen door middel van zogenaamde (RT-)PCR tests kunnen in een relatief korte tijd veel monsters getest worden op de aanwezigheid van virussen.

Hoofdstuk 2.1 beschrijft een nieuwe RT-PCR test voor de detectie van alle virussen binnen de paramyxovirus familie. Het genoom van paramyxovirussen bestaat uit RNA. Deze virussen zijn in staat om veel verschillende soorten organismen te infecteren, zoals mensen, paarden, honden, muizen, dolfinen en vleermuizen. De familie bevat een aantal virussen die vaak voorkomen en een grote impact hebben op de volksgezondheid, zoals het respiratoir syncytieel virus, het mazelenvirus, het bofvirus, het humaan metapneumovirus en de humane para-influenza virussen. Het Newcastle disease virus, dat vooral kippen en kalkoenen infecteert, kan grote schade aanrichten in de pluimvee-industrie. De nieuw ontwikkelde test richt zich op een zeer geconserveerd deel van het genoom van paramyxovirussen, waardoor alle virussen van deze familie in één snelle reactie gedetecteerd kunnen worden. Het deel van het virusgenoom dat geamplificeerd wordt, is voldoende variabel, waardoor de test toch een indicatie geeft van welk paramyxovirus gedetecteerd is. Hierdoor kunnen potentiële nieuwe leden van de familie snel gedetecteerd worden. De test is gevalideerd op een groot aantal monsters van mensen en vogels, waarbij in sommige gevallen virussen werden aangetoond die in de reguliere diagnostische tests gemist waren.

SEQUENCEN

Sequenzen is het bepalen van de volgorde van nucleotiden in genetisch materiaal. Deze nucleotiden zijn adenine, guanine, cytosine en thymine in DNA en adenine, guanine, cytosine en uracil in RNA. Sanger sequenzen, waarbij de volgorde van een consensus sequentie wordt bepaald, is in

1977 door Frederick Sanger beschreven en is lange tijd de meest gebruikte manier van sequencen gebleven. De ontwikkeling van “next-generation sequencers” heeft een grote invloed gehad op het sequencen. Door de komst van deze sequencers is de hoeveelheid genetisch materiaal dat in een tijdseenheid kan worden geanalyseerd enorm toegenomen, terwijl de kosten daarvan juist zijn afgenomen. In de volgende hoofdstukken van dit proefschrift is deze “next-generation sequencing” technologie ontwikkeld en toegepast voor de ontdekking en karakterisering van virussen.

ONTDEKKING VAN VIRUSSEN

Bij ziekte-uitbraken door onbekende oorzaak is het van belang om door middel van genetische karakterisering snel vast te stellen of een virus de veroorzaker is en welk virus het betreft. Vóór de opkomst van next-generation sequencing werd dit vaak gedaan aan de hand van willekeurige amplificatie, gevolgd door klonering, transformatie naar bacteriën en Sanger sequencen. In hoofdstuk 3 wordt de ontwikkeling van een methode beschreven die het proces van virusidentificatie in een klinisch monster versnelt. De methode combineert willekeurige amplificatie van erfelijk materiaal met next-generation sequencing, waarbij er op voorhand geen informatie nodig is over de genetische samenstelling van het virus in het monster. Deze methode is in hoofdstuk 3.1 toegepast om een virus te identificeren in een 60-jarige man uit Jeddah, Saudi-Arabië, die leidde aan acute longontsteking en nierfalen. Elf dagen na de opname in het ziekenhuis overleed de patiënt. In het serum en sputum van de patiënt konden met specifieke tests geen van de reeds bekende virussen worden geïdentificeerd. Een test voor coronavirussen, in combinatie met de nieuw ontwikkelde methode, liet zien dat de man geïnfecteerd was met een tot dan toe onbekend virus behorende tot de familie van de coronavirussen. Door middel van een analyse van 2.400 serummonsters van andere patiënten in hetzelfde ziekenhuis in Jeddah kon worden vastgesteld dat het nieuwe virus zich nog niet verder had verspreid. Het virus bleek genetisch gezien het meest overeen te komen met een virus dat eerder was aangetoond in een vleermuis: bCoV-HKU5. Het nieuwe virus werd humaan betacoronavirus 2c EMC/2012 genoemd. Nadat het virus in meer patiënten met een vergelijkbaar ziektebeeld werd gevonden, werd er een nieuwe naam gegeven aan de ziekte en het virus, namelijk Middle East Respiratory Syndrome coronavirus (MERS-CoV).

In hoofdstuk 3.2 is de methode voor virusontdekking toegepast op weefsel van uitstekende spicules op de huid van een 70-jarige man uit Nederland. Dit ziektebeeld kwam overeen met een ziekte veroorzaakt door het Trichodysplasia spinulosa-geassocieerde polyomavirus. Echter, een specifieke diagnostische test voor dit virus gaf een negatief resultaat. De spicules waren aanwezig op het gezicht, de borstkas en de armen van de patiënt. Bij de patiënt werd multipel myeloom geconstateerd, een kwaadaardige woekering van plasmacellen, waardoor de diagnose folliculaire spicules van multipel myeloom werd gesteld. De etiologie hiervan is onbekend. Met behulp van

de willekeurige virusontdekkingsmethode werd Merkelcelpolyomavirus gedetecteerd. Dit virus werd tot dusver nog niet bij een dergelijk klinisch beeld beschreven. Na toediening van de DNA-polymeraseremmer Cidofovir verdwenen de spicules binnen drie weken van het gezicht. Of het Merkelcelpolyomavirus daadwerkelijk de veroorzaker is van de spicules, is echter nog onduidelijk.

VIRUS KARAKTERISERING

Naast het ontdekken van virussen kan next-generation sequencen ook worden toegepast bij het karakteriseren van virussen. Dit laatste werd uitgewerkt in hoofdstuk 4.

Het influenza A/H5N1 is een virus dat normaliter vogels infecteert. Door nauw contact met pluimvee worden mensen soms ook geïnfecteerd. Wanneer een infectieziekte kan worden overgedragen van dieren op mensen spreken we van een zoonose. Ongeveer 60% van de gerapporteerde gevallen van A/H5N1 virusbesmetting in mensen had de dood tot gevolg. Gelukkig is het A/H5N1 virus tot dusver niet in staat gebleken efficiënt van mens op mens over te gaan, zodat het aantal besmettingen van mensen nog altijd beperkt is gebleven. In 2012 werd voor het eerst aangetoond dat het virus wel dusdanig kan evolueren dat het tussen zoogdieren (fretten) via de luchtwegen kan worden overgedragen. In hoofdstuk 4.1 is uitgezocht welke veranderingen dit A/H5N1 virus had ondergaan om tussen fretten overdraagbaar te worden. Vijf mutaties bleken hiervoor verantwoordelijk. Voor ieder van deze mutaties werd vastgesteld wat het effect was op de biologische eigenschappen van het virus. Drie mutaties bevonden zich in het oppervlakte-eiwit hemagglutinine en resulteerden in verhoogde stabiliteit van dit eiwit en betere binding van het virus aan zoogdiercellen. De andere twee mutaties bevonden zich in de polymerase eiwitten en veroorzaakten een betere vermeerdering van het virus. Next-generation sequencing werd niet alleen gebruikt om deze mutaties in kaart te brengen, maar ook om te onderzoeken hoe snel deze veranderingen zich voordeden na virusinfecties in fretten. De natuurlijk opkomende mutaties hadden effectief maar twee tot drie herhaalde infecties van fretten nodig om te ontstaan en vervolgens dominant te worden.

Hoofdstuk 4.2 beschrijft de complete karakterisering van het nieuwe coronavirus dat is beschreven in hoofdstuk 3.1 door middel van next-generation sequencing. Diagnostische tests konden ontwikkeld worden nadat het volledige genoom in kaart werd gebracht, waarmee het virus tijdens uitbraken sneller opgespoord kan worden. De totale lengte van het genoom is 30.119 nucleotiden. Bioinformatische analyses lieten zien dat het virus waarschijnlijk elf eiwitten tot expressie brengt. Aan de hand van vergelijkingen met bekende coronavirussen en in overeenstemming met de richtlijnen van het Internationaal Comité voor de Taxonomie van Virussen kon het virus als een nieuw soort binnen de virusfamilie worden voorgesteld.

In hoofdstuk 4.3 werden de ontwikkelde next-generation sequencing technieken gebruikt om een humaan metapneumovirus isolaat te karakteriseren dat een hoge interferon-productie liet zien in

weefselkweek, terwijl humane metapneumovirussen normaliter weinig of geen interferon induceren. Interferon speelt een belangrijke rol in het immuunsysteem van gewervelde dieren als eerste verdedigingsmechanisme tegen infecties. Het isolaat bleek alleen na herhaalde weefselkweek verhoogd interferon te induceren. De verhoogde interferonproductie kon worden verklaard door de aanwezigheid van defectief interfererend RNA in het virusisolaat. Defectief interfererend RNA ontstaat door foute genoomrelicatie, waarna kleine stukjes van het genoom in virusdeeltjes worden ingepakt. Het defectief interfererend RNA in het virusisolaat bleek een enorm aantal mutaties te hebben ondergaan; veel van de adenine en thymine residuen waren gemuteerd naar guanine en cytosine. Met behulp van bioinformatica en experimenteel onderzoek werd het eiwit Adenosine Deaminase Acting on RNA (ADAR) aangewezen als waarschijnlijke oorzaak voor de hypermutatie.

In de basale wetenschap hebben de toepassingen van next-generation sequenzen inmiddels hun weg gevonden. Het zal niet lang meer duren voordat next-generation sequenzen ook in de dagelijkse diagnostiek geïmplementeerd zal worden. Hoewel de technologie nog verder ontwikkeld moet worden, kan next-generation sequenzen tegen relatief lage kosten een belangrijke bijdrage leveren voor onderzoek en volksgezondheid.

Eindelijk is het zover! Vanaf het moment dat ik het dankwoord kan gaan schrijven, weet ik dat het meeste werk aan het proefschrift is gedaan. In eerste instantie dacht ik het te houden bij dat ene woord "bedankt", maar naarmate ik er verder over nadenk, zijn er toch een aantal mensen die ik persoonlijk wil bedanken. Een mooi moment dus om eens rustig achterover te gaan zitten en te bedenken welke mensen hebben bijgedragen aan dit proefschrift.

Ron, net als op een artikel zijn de eerste en de laatste mensen het belangrijkste. Zo ook nu. Na mijn aanstelling was je misschien niet helemaal overtuigd van mijn kunnen, maar ik hoop dat ik dat in de afgelopen vier jaar goed heb kunnen maken. Jij bent een familieman, een echte wetenschapper met het hart op de juiste plaats, met een helder geweten en met beslissingen in het belang van de mens. Dat klinkt misschien allemaal zwaar, maar zo zie ik het wel. MERS sequentie voltooid? Rondsturen over de hele wereld zodat iedereen er gelijk mee aan de slag kan. Tegenstrijdige (met de te willen) resultaten? Onderdeel van het artikel maken, zodat anderen niet ook tegen dat probleem aanlopen. Ik hoop dat je nog veel meer onderzoekers zal inspireren zoals je dat bij mij hebt gedaan. Bedankt voor alles!

Ab, bedankt dat ik op "Het lab van Ab" heb mogen werken. Het waren vier mooie jaren. Op een afdeling die als motto "work hard, party hard" heeft, is er nooit een saai moment. Veel hebben we niet gepraat, maar leerzaam was het zeker. Ik kan nu meer de actualiteiten nuanceren wanneer iemand in de politiek weer eens wat roept zonder dat hij/zij er goede argumenten voor heeft en er klaarblijkelijk zelf niets vanaf weet. Daar stond ik zo van te kijken, dat ik er een stelling van gemaakt heb.

Alle mensen op de afdeling Viroscience, ook zij die de afdeling kort geleden hebben verlaten. Teveel om individueel op te noemen. Van Robert to Byron, en van Wim tot Frank. Jullie hebben de afdeling gemaakt tot wat het is. Vele borrels zijn er gehouden, waarvan ik aan het merendeel met plezier terug denk (als ik ze me kan herinneren).

Saskia, bedankt voor de samenwerking op het gebied van de ontwikkeling van het virus discovery protocol. Ook bedankt dat je mijn introductie en discussie wilde doornemen. Uiteraard ook Marije, Claudia, Rogier en Debby.

De dames van het secretariaat: Simone, Sabine, Loubna, Carola, Anouk en Maria. Maria, bedankt dat je mij zoveel geholpen hebt met alle procedures voor de promotie.

Sasha (Gulyaev), mijn koffiemaatje. Vele dubbele espresso's hebben wij genuttigd in het Erasmus MC. Altijd mooie gesprekken en goede discussies gevoerd over de Russische geschiedenis en de actuele gebeurtenissen. Een hoop is mij duidelijke geworden in de internationale relaties, en vele

vooroordelen heb jij bij mij weten te slechten. Gelukkig kunnen we deze discussies voortzetten in Leiden! Спасибо! И в следующий раз обед вернулся в LUMC за мой счет.

De grote flu groep, het waren goede werkdiscussies, groepuitjes en diners bij de baas thuis. Sander, jij bent grotendeels verantwoordelijk voor de beroemde en beruchte gezelligheid op de afdeling. Mocht een borrel mislopen, dan was jij er altijd om er toch nog een topfeest van te maken. Ik zal jou onwijs gaan missen gast! Eefje, jij bent een van de liefste roodhoofden die ik ken. Nogmaals gefeliciteerd met je tweede kleintje. Extra bedankt voor alle tips voor mijn promotie! Miranda, veel succes met je volgende stappen en we gaan binnenkort een bakkie doen in het LUMC. Bernadette, het was een vruchtbare samenwerking met goede discussies. Josanne, fijn gekletst te hebben over de invulling van onze OIO. En natuurlijk alle andere toppers in de groep: Pascal, Benjamin, Bernike, Bjorn, Commando, Mathilde, Monique, Oahn, Ramona, Chinmoy en Bri. Special thanks voor mijn kamergenoten: Anna, Marine, Marina, Pascal, Jonneke, Kim en Stefan! Rogier en Joost, twee postdocs die onwijs hard gewerkt hebben en hun aandeel in de gezelligheid van de afdeling mogen meepikken. Joost, heel veel succes met je nieuwe baan en plezier met je twee kanjers!

All the people from Cambridge. Thanks! Especially Terry and Barbara. If we would've continued working together for a couple of years, we would've published a great paper on the dark matter. For now, we'll just have to do it with a case report. Terry, you're definitely one of a kind. I've enjoyed your company both in Cambridge and in Rotterdam.

Eric en Sasha (Gorbalenya), our collaborations leading to the mBio paper made it as great as it is now. Although the draft took significantly longer to finish, the quality of the paper increased proportional with the time it took to draft it. Ik blijf nog even in Leiden. Marjolein, jij bent het startpunt van mijn virologische ambities en stiekem van mijn wetenschappelijke ambities geweest. Jij hebt de interesse en het plezier in de wetenschap zo kunnen overbrengen dat ik uiteindelijk voor het onderzoek heb gekozen. Onwijs bedankt daarvoor! Je zorgde ook weer voor een link naar het lab van Adolfo, waar ik bij Gijs terecht kwam voor een vervolgstage. Gijs, wat heb je mij een toptijd in New York bezorgd man! Hetzelfde enthousiasme dat ik van Marjolein meekreeg, heb jij aangevuld met de jouwe. Ik had daar meer het idee onderzoek te doen met een vriend dan met een begeleider. Uiteindelijk nog een eerste officiële publicatie ook! Om dan toch weer terug te komen bij jou Eric. Ook al was jou aandeel tijdens mijn stage meer in het faciliteren van mij naar Marjolein, toen was je al iemand van wie ik wist waarbij een onderzoeker zoals ik, goed op z'n plek zou zitten. Nu weet ik dat zeker en wil ik je bedanken voor de kans die ik bij jou krijg, één die perfect is om mijn wetenschappelijke carrière verder te ontwikkelen.

Egon, vriendje Eeg. Veel heb je niet in de eerste vier jaar van mijn OIO voor het proefschrift betekend. Des te meer in de laatste weken. Dat de opmaak nog zoveel voeten in de aarde zou hebben, ben ik mij in de afgelopen drie weken gaan realiseren. Hele dagen heb je voor mij vrijgehouden om samen aan de opmaak van dit proefwerk te gaan zitten! En al die tijd dat je zowel thuis als op je werk veel belangrijkere dingen te doen had. Maar zoals je het zelf zegt; "dat is wat vriendjes voor elkaar doen" :).

Ook wil ik de rest van mijn vriendjes en vriendinnetjes bedanken, velen ken ik al van lang geleden. Jullie zijn topvrienden! Gezellig barbecuen in het Vroesepark, stappen, voetballen op de vereniging of op het veldje voor de deur, of gekke (vrijgezellen)weekenden weg door Europa. En volgend jaar Ibiza zijn wij er gewoon weer bij!

Somtin els, Theo. Wat zou mijn boekje zijn geweest zonder jou? In elk artikel in dit proefschrift heb jij een groot aandeel gehad. Jouw immens grote drive om dingen voor elkaar te krijgen hebben menig OIO en postdoc uit de brand geholpen. Jouw kennis en hands-on aanpak is een voorbeeld voor iedere researcher. Daarnaast kunnen wij het goed met elkaar vinden, aangezien wij veel gemeen hebben in de typische mannendingen: voetbal, bier en... iets anders ;).

Mijn paranimfen. Marco, grote broer, vriend. Het zal wel door de genen komen, maar veel hebben wij gemeen. Jij koos voor de HAVO na de MAVO, ik deed dat ook. Jij koos voor het VWO na de HAVO, ik deed dat ook. Jij koos voor een universitaire studie na het VWO, ik deed dat ook. Niet alleen school gerelateerd deden wij veel hetzelfde, maar nog meer daarbuiten. Broer, ik ben zó blij dat jij straks naast mij staat tijdens de plechtigheid, dan weet ik dat ik heel de wereld aan kan! Herr Linster, de afgelopen tijd hebben wij in hetzelfde schuitje gezeten. Wij hebben veel persoonlijke overeenkomsten, die het onderling makkelijk praten maakt. Aangezien je ook nog eens op mijn kamer zat, heeft dit tot vele mooie gesprekken geleid. Ik denk dat jij evenveel plezier haalde uit mijn opzettelijk slechte Duits, als dat ik jou daarmee wilde pesten. De inzichten die jij hebt met betrekking tot het onderzoek, zijn zeer goed en ik zie jou het daarom ook nog ver schoppen in de wereld van de wetenschap. Singapore is daar een goed begin van. Ik wil jou en jouw frau Linster, Marina, heel veel succes en (duik)plezier in Singapore wensen!

Mijn familie: papa, mama, Jeroen, Else, Zelda, Arvid, Marco, Marjolein, Sanna, Stan, Laurien, Pjotr, Valerio, Sterre, de oma's, Arie, Joke, Linda en Danny. Bedankt voor alle liefde en steun! En al mijn lieve tantes, ooms, neven en nichten.

Papa en mama, jullie zijn de basis van mijn hele ontwikkeling. De normen en waarden die wij als kinderen van jullie hebben meegekregen hebben geleid tot vier kinderen die zichzelf hebben ontwikkeld tot wat zij echt wilden zijn. Jullie gaven ons genoeg vrijheid om onze grenzen op te zoeken. Jullie staan altijd voor ons klaar. Ik dank jullie voor deze onvoorwaardelijke steun in mijn keuzes en hoop dat jullie met trots op de afgelopen jaren kunnen terugkijken. Ik hou van jullie.

Zoals aangegeven, eindig ik met de belangrijkste(n). Krista, lieve lieve schat. Mensen beginnen vaak zo'n zin met "zonder jou..". Maar in mijn geval is dat ook echt zo. Zonder jou zou ik nooit zo immens gelukkig zijn als dat ik nu ben. Zonder jou zou ik nooit de vader kunnen zijn van zo'n verschrikkelijk mooi ventje. Zonder jou had ik deze zinnen nooit kunnen schrijven. Jij en Bas zijn voor mij het belangrijkste in het leven. Als ik weer eens wat te zeuren had over het werk, hoorde jij mijn gebrom met geduld en liefde aan. Als ik weer eens vanaf mijn laptop de kamer in begon te schelden, was een "hummm hummm" van jou vanaf de bank genoeg voor mij om te weten dat er iemand om mijn gezeur gaf. Ook in de begintijden van mijn OIO, toen ik niet wist of ik goed bezig was, wist je mij altijd gerust te stellen en kon ik weer terugkeren naar mijn relax-stand. Ik hou zo verschrikkelijk veel van je lieverd, onvoorwaardelijk! Bas, veel heb je nog niet door van de rare dingen die papa allemaal aan het doen is, maar veel heb ik aan jou te danken. Ik was al zo gelukkig, maar jou komst heeft dat nog zoveel beter gemaakt. Een blik in jou prachtige blauwe ogen heeft hetzelfde effect als mama haar "hummm hummm": alles is gelijk weer goed. Je slaapt al, maar ik ga je nu toch nog heel stilletjes een kus geven.

About The Author

CURRICULUM VITAE

The author of this thesis was born on August 5th, 1984 in Enkhuizen, the Netherlands. In 2004 he finished pre-university secondary education at the Maerlant College in Brielle and started to study Biology at the Leiden University. During his Bachelor phase, he did a research rotation at the department of molecular microbiology on the identification of *Aspergillus niger* rmsA protein interactions, and wrote a small review on *situs inversus* for the department of Integrative Zoology. He continued to specialize into molecular and cellular biology during his Master phase. Here, he performed a research rotation at the department of medical microbiology at the Leiden University Medical Center unraveling the deubiquitinating and deISG15ylating activity of Equine Arteritis Virus non-structural protein 2, and a research rotation at the department of medical microbiology at the Mount Sinai School of Medicine, New York, USA, deciphering influenza A virus mediated regulation of the ISG15 conjugation system. In 2009 he started as a PhD student at the department of virology in the Erasmus Medical Center in Rotterdam under the supervision of prof. dr. Ron Fouchier and prof. dr. Ab Osterhaus. The focus of this project was the discovery and characterization of viruses using next-generation sequencing, and has resulted in the present thesis.

PHD PORTOFOLIO

| | |
|------------------------|---|
| Name: | Sander van Boheemen |
| Erasmus MC department: | Department of Viroscience |
| Research school: | Post-graduate Molecular Medicine |
| PhD period: | 2009-2014 |
| Promotors: | Prof. dr. R.A.M. Fouchier Prof. dr. A.D.M.E. Osterhaus |

Education

| | |
|-----------|---|
| 2009-2014 | PhD program, Erasmus Medical Center, Rotterdam, the Netherlands. PhD thesis: Virus Discovery and Characterization using Next-Generation Sequencing. |
| 2004-2009 | Master of Science, Leiden University, Leiden, the Netherlands. Study: Biology: Molecular and Cellular Biology. |
| 1996-2004 | Pre-university secondary education, Maerlant College, Brielle, the Netherlands. |

In-depth courses

| | |
|---|------|
| Course in Virology | 2010 |
| Basic Course on 'R' | 2010 |
| Photoshop and Illustrator CS5 Workshop | 2011 |
| Next Generation Sequencing Training: CLC bio II | 2011 |
| Next Generation Sequencing (NGS) Data analysis Course | 2012 |
| Metagenomics Approaches and Data Analysis | 2013 |

Poster presentations

| | |
|--|------|
| 15th Molecular Medicine Day, Rotterdam | 2011 |
| XV IUMS International Congress of Virology, Sapporo, Japan | 2011 |
| 17th Molecular Medicine Day, Rotterdam | 2013 |

Oral presentations

| | |
|--|------|
| 3rd International Influenza Meeting, Munster, Germany | 2012 |
| 6th Orthomyxovirus Conference, Montreal, Canada | 2012 |
| EMPERIE Annual Meeting, Rotterdam | 2012 |
| Dutch Annual Virology Symposium, Amsterdam | 2013 |
| EMPERIE Annual Meeting, Rotterdam | 2013 |
| Roche 454 Experience Event, Utrecht | 2013 |
| Roche APAC Genetics Day, Bangkok, Thailand | 2013 |
| 32nd American Society for Virology Annual Meeting, Penn State, USA | 2013 |

Attended

| | |
|--|------|
| Dynamics in 454 Sequencing, Leiden | 2010 |
| 14th Molecular Medicine Day, Rotterdam | 2010 |
| Dutch Annual Virology Symposium, Amsterdam | 2011 |
| EMPERIE Annual Meeting, Amsterdam | 2011 |
| Dutch Annual Virology Symposium, Amsterdam | 2012 |
| Copenhagenomics, Copenhagen, Denmark | 2012 |

LIST OF PUBLICATIONS

van den Hoogen BG, **van Boheemen S**, de Rijck J, van Nieuwkoop S, Smith DJ, Laksono B, Gultyaev A, Osterhaus AD, Fouchier RA. Excessive Production and extreme editing of Human Metapneumovirus Defective Interfering RNA is Associated with Type I Interferon Induction. *J Gen Virol*. 2014 Apr 23.

van Boheemen S, Linster M, de Graaf M, Schrauwen EJ, Lexmond P, Mänz B, Bestebroer TM, Baumann J, van Riel D, Rimmelzwaan GF, Osterhaus AD, Matrosovich M, Fouchier RA, Herfst S. Identification, Characterization, and Natural Selection of Mutations Driving Airborne Transmission of A/H5N1 Virus. *Cell*. 2014 Apr 10.

Martín-Valls GE, Simon-Grifé M, **van Boheemen S**, de Graaf M, Bestebroer TM, Busquets N, Martín M, Casal J, Fouchier RA, Mateu E. Phylogeny of Spanish swine influenza viruses isolated from respiratory disease outbreaks and evolution of swine influenza virus within an endemically infected farm. *Vet Microbiol*. 2014 Jun 4.

Lim SM, Koraka P, **van Boheemen S**, Roose JM, Jaarsma D, van de Vijver DA, Osterhaus AD, Martina BE. Characterization of the mouse neuroinvasiveness of selected European strains of West Nile virus. *PLoS One*. 2013 Sep 18.

Richard M, Schrauwen EJ, de Graaf M, Bestebroer TM, Spronken MI, **van Boheemen S**, de Meulder D, Lexmond P, Linster M, Herfst S, Smith DJ, van den Brand JM, Burke DF, Kuiken T, Rimmelzwaan GF, Osterhaus AD, Fouchier RA. Limited airborne transmission of H7N9 influenza A virus between ferrets. *Nature*. 2013 Sep 26.

Bodewes R, van de Bildt MW, Schapendonk CM, van Leeuwen M, **van Boheemen S**, de Jong AA, Osterhaus AD, Smits SL, Kuiken T. Identification and characterization of a novel adenovirus in the cloacal bursa of gulls. *Virology*. 2013 May 25.

van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, Osterhaus AD, Haagmans BL, Gorbalenya AE, Snijder EJ, Fouchier RA. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio*. 2012 Nov 20.

Zaki AM, **van Boheemen S**, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 2012 Nov 8.

Corman VM, Eckerle I, Bleicker T, Zaki A, Landt O, Eschbach-Bludau M, **van Boheemen S**, Gopal R, Ballhause M, Bestebroer TM, Muth D, Müller MA, Drexler JF, Zambon M, Osterhaus AD, Fouchier RM, Drosten C. Detection of a novel human coronavirus by real-time reverse-transcription polymerase chain reaction. *Euro Surveill.* 2012 Sep 27.

Russell CA, Fonville JM, Brown AE, Burke DF, Smith DL, James SL, Herfst S, **van Boheemen S**, Linster M, Schrauwen EJ, Katzelnick L, Mosterín A, Kuiken T, Maher E, Neumann G, Osterhaus AD, Kawaoka Y, Fouchier RA, Smith DJ. The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science.* 2012 Jun 22.

van Boheemen S, Bestebroer TM, Verhagen JH, Osterhaus AD, Pas SD, Herfst S, Fouchier RA. A family-wide RT-PCR assay for detection of paramyxoviruses and application to a large-scale surveillance study. *PLoS One.* 2012 Apr 4.

Oudshoorn D, **van Boheemen S**, Sánchez-Aparicio MT, Rajsbaum R, García-Sastre A, Versteeg GA. HERC6 is the main E3 ligase for global ISG15 conjugation in mouse cells. *PLoS One.* 2012 Jan 17.

van Kasteren PB, Beugeling C, Ninaber DK, Frias-Staheli N, **van Boheemen S**, García-Sastre A, Snijder EJ, Kikkert M. Arterivirus and nairovirus ovarian tumor domain-containing Deubiquitinases target activated RIG-I to control innate immune signaling. *J Virol.* 2012 Jan.

Versteeg GA, Hale BG, **van Boheemen S**, Wolff T, Lenschow DJ, García-Sastre A. Species-specific antagonism of host ISGylation by the influenza B virus NS1 protein. *J Virol.* 2010 May.