Human respiratory and enteric viruses: methods for diagnostics and discovery

Dissertation

Submitted in partial fulfilment of the requirements for the doctoral degree

- Dr. rer. Nat. -

Department of Biology,

Faculty of Mathematics, Informatics, and Natural Sciences

University of Hamburg, Germany

by

Luciano Kleber de Souza Luna

from São Paulo, Brazil

Hamburg, Germany

2008

Genehmigt vom Department Biologie
der Fakultät für Mathematik, Informatik und Naturwissenschaften
an der Universität Hamburg
auf Antrag von Professor Dr. B. FLEISCHER
Weiterer Gutachterin der Dissertation:
Frau Professor Dr. I. BRUCHHAUS
Tag der Disputation: 28. März 2008

Hamburg, den 10. März 2008

Professor Dr. Jörg Ganzhorn
Leiter des Departments Biologie

First reviewer: Prof. Dr. Bernhard Fleischer

Second reviewer: Prof. Dr. Iris Bruchhaus

Date of disputation: 28 March 2008.

The contents of this work were published in the following journals:

Luna, L. K.*, Baumgarte, S.*, Grywna, K., Panning, M., Drexler, J. F. & Drosten, C. (2008). Identification of a contemporary human parechovirus type 1 by VIDISCA and characterisation of its full genome. *Virol J* 5(1), 26.

Baumgarte, S.*, **de Souza Luna, L. K.***, Grywna, K., Panning, M., Drexler, J. F., Karsten, C., Huppertz, H. I. & Drosten, C. (2008). Prevalence, types, and RNA concentrations of human parechoviruses, including a sixth parechovirus type, in stool samples from patients with acute enteritis. *J Clin Microbiol* 46(1), 242-248.

Luna, L. K., Panning, M., Grywna, K., Pfefferle, S. & Drosten, C. (2007). Spectrum of viruses and atypical bacteria in intercontinental air travelers with symptoms of acute respiratory infection. *J Infect Dis* 195(5), 675-679.

de Souza Luna, L. K., Heiser, V., Regamey, N., Panning, M., Drexler, J. F., Mulangu, S., Poon, L., Baumgarte, S., Haijema, B. J., Kaiser, L. & Drosten, C. (2007). Generic detection of coronaviruses and differentiation at the prototype strain level by reverse transcription-PCR and nonfluorescent low-density microarray. *J Clin Microbiol* 45(3), 1049-1052.

Papa, A., Papadimitriou, E., **Luna, L. K.**, Al Masri, M., Souliou, E., Eboriadou, M., Antoniadis, A. & Drosten, C. (2007). Coronaviruses in children, Greece. *Emerg Infect Dis* 13(6), 947-949.

*These authors contributed equally

Acknowledgments

I am deeply indebted to Prof. Dr. Christian Drosten for giving me the opportunity to conduct my work at his laboratory and who was an excellent supervisor and scientific mentor, supporting me throughout the entire time of my doctorate.

I want to thank Prof. Dr. Bernhard Fleischer and Prof. Dr. Iris Bruchhaus for accepting to be the reviews of my dissertation.

I want to thank Dr. Marcus Panning for helping me during my first steps in the laboratory.

I thank all my colleagues Britta Liedigk, Felix Drexler, Klaus Grywna, Maike Prange, Nadine Petersen, and Susanne Pfefferle for all support and collaboration in the laboratory.

My thanks to Dr. Volker Heiser from Chipron GmbH, Berlin, Germany, to provide the LCD arrays for the universal coronavirus detection assay.

My thanks to Dr. Anna Papa from Aristotle University of Thessaloniki, Thessaloniki, Greece, Dr. Bert Jan Haijema from Institute of Virology, Faculty of Veterinary Medicine, University of Utrecht, The Netherlands, Dr. Laurent Kaiser from Central Laboratory of Virology, University Hospitals of Geneva, Switzerland, and Dr. Leo Poon from Department of Microbiology, University of Hong Kong, China, for providing me coronavirus prototypes and clinical samples.

My thanks to Dr. Sabue Mulangu from Institut National de Recherche Biomédicale, Kinshasa, Democratic Republic of Congo, and Dr. Sigrid Baumgarte from Institute of Hygiene and Environmental Health, Hamburg, Germany, for providing me with clinical samples.

My thanks to Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for providing me the fellowship to carry out my doctoral project at Bernhard Nocht Institute for Tropical Medicine in Hamburg, Germany.

My thanks to the Deutsche Akademische Austauschdienst (DAAD) for providing me with a German language course realised in the city of Göttingen at Goethe-Institut. This period was very important for my adaptation in Germany.

I would like to thank my parents Oscar and Vania for all the love, moral and educational support through all my life, and also to my brother Emerson and sister Judite. You are very special to me.

I would like to thank specially my wife Keyla for all her love, support and patience, and for having presented me with our sun Luan, who brought a new joy for our lives.

Table of contents

Human respiratory and enteric viruses: methods for diagnostics and discovery

Abstract

Viral respiratory and enteric infections are the most common illnesses in humans and a major public health problem worldwide due to their occurrence, ease of spread and considerable morbidity and mortality. They are caused by a large spectrum of different viruses but only a very small fraction of them has been determined. In the last years, novel viruses causing respiratory and enteric infections have been discovered, including new coronaviruses and human parechoviruses.

The objectives of this study were: (1) investigate the spectrum of respiratory viruses and atypical bacteria in samples collected in Germany from air travellers with symptoms of respiratory infection during SARS epidemic; (2) develop a universal reverse transcription, polymerase chain reaction (RT-PCR) for the genus coronavirus combined with species identification by a non-fluorescent, low-density oligonucleotide array; (3) implement and use the Virus-Discovery-cDNA-AFLP (VIDISCA) method to identify untypeable viruses growing in cell culture; (4) develop the first broad-range real-time RT-PCR assay for all human parechoviruses and investigate their prevalence in patients with gastroenteritis; (5) establish a new method for discovery of viruses with single-stranded, positive strand RNA genomes (GDD method), employing the unnatural nucleotides isoC and isoG in PCR primers and using chikungunya virus as a model organism.

In respiratory samples from 172 patients presenting symptoms of respiratory infection after air travel, a pathogen was identified in 67 travellers (43.2%). Influenza and parainfluenza viruses were most prevalent, at 14.2% and 15.5%, respectively. Prevalence of adenoviruses, human metapneumovirus, coronaviruses, and rhinoviruses ranged between 2.6% and 4.8%. Human bocavirus, respiratory syncytial virus, and *Legionella*, *Mycoplasma*, and *Chlamydophila* species were absent or appeared at frequencies of <1%. These were the first specific baseline data for the mentioned agents in the context of air travel.

The universal coronavirus RT-PCR presented high specificity and sensitivity, detecting a minimum of 16 viruses per reaction test. The clinical detection limit in patients with severe acute respiratory syndrome was about 100 viruses per sample. In 46 children suffering from human coronaviruses 229E, NL63, OC43, or HKU1, the sensitivity was as high as that of individual real-time RT-PCR assays. Coronavirus identification by oligonucleotide array was uniform in all samples and prototypes tested, and in full concordance with nucleotide sequencing.

The study of patients with gastroenteritis was conducted with samples collected during a period of one year and covered all age groups. The overall detection rate of human parechovirus infection was 1.6%. Positive samples occurred only in summer and autumn. All positive patients except one were <2 years of age, with a neutral gender ratio. In children <2 years of age, the detection rate was 11.6%. Phylogenetic analysis showed mainly contemporary human parechovirus type 1 strains. A novel human parechovirus type 6 could be identified and characterized.

The VIDISCA method amplified a 188 nucleotide fragment from an untypeable virus growing in cell culture. Sequence analysis identified an HPeV type 1. Whole genome sequence analysis of this isolate revealed it as a divergent strain in relation to the prototype 1 described in 1961, as well as a genomic recombination between human parechovirus types 1 and 3.

The GDD method was successfully established in the chikungunya model. Genome fragments from chikungunya virus could be obtained from a full-length *in vitro* transcript RNA, as well as from virus growing in cell culture. Genome fragments could be amplified, cloned and sequenced.

Humanpathogene respiratorische und enteritische Viren: Methoden zur Diagnostik und Identifizierung

Zusammenfassung

Virale Infektionen der Atemwege und des Magen-Darm-Traktes zählen zu den häufigsten menschlichen Krankheitsbildern. Sie stellen aufgrund von Prävalenz, Infektiosität und hoher Morbidität und Mortalität weltweit eines der gravierendsten Probleme des öffentlichen Gesundheitswesens dar. Bislang konnte nur ein geringer Teil des großen Spektrums viraler Erreger beschrieben werden. Neue molekularbiologische Methoden machten in den letzten Jahren die Entdeckung und Klassifizierung verschiedener bislang unbekannter Viren möglich, darunter auch Coronaviren und neue humanpathogene Parechoviren.

Die Ziele dieser Arbeit waren: (1) die Prävalenz respiratorischer Viren und atypischer Bakterien bei deutschen Flugreisenden mit respiratorischer Symptomatik während der SARS-Epidemie zu untersuchen; (2) einen universellen RT-PCR (*reverse transcription, polymerase chain reaction*) Test für Coronaviren zu entwickeln, verbunden mit einer Spezies-spezifischen Typisierung durch eine nicht-Fluoreszenz basierte *low density DNA oligonucleotide microarray*; (3) Die neu beschriebene VIDISCA (*virus-discovery-cDNA-AFLP*) Methode zur Entdeckung unbekannter Viren zu etablieren und bei nicht typisierbaren Isolaten aus Zellkulturen anzuwenden; (4) Den ersten real time RT-PCR Test für alle humanpathogenen Parechoviren (HPeV) zu entwickeln und deren Prävalenz bei Patienten mit dem klinischen Leitbild Diarrhö zu bestimmen; (5) Eine neuartige Methode zur Entdeckung von Viren mit einem (+)-RNA Einzelstrang Genom zu entwickeln (sog. *GDD Methode*), basierend auf der Inkorporation der nicht natürlich vorkommenden Nukleotide Iso-G und Iso-C in PCR Primern, unter Verwendung von Chikungunya Virus als Modellorganismus.

In klinischen Proben von 172 Flugreisenden mit respiratorischer Symptomatik konnte in 67 Fällen (43,2%) ein Erreger identifiziert werden. Hierbei wiesen Influenza- und Parainfluenzaviren mit 14,2% und 15,5% die höchste Prävalenz auf. Die Prävalenz von Adenoviren, humanem Metapneumovirus, Coronaviren und Rhinoviren schwankte zwischen 2,6% und 4,8%. Humanes Bocavirus, *Respiratory Syncytial Virus* sowie *Legionella*, *Mycoplasma* und *Chlamydophila* Spezies konnten nicht oder nur in Einzelfällen (<1%) identifiziert werden. Die Studie lieferte hiermit die ersten verfügbaren Daten zur Prävalenz dieser Erreger bei respiratorischen Erkrankungen nach Flugreisen.

Die universelle Coronavirus RT-PCR zeichnete sich durch hohe Sensitivität und Spezifität aus, die untere Nachweisgrenze lag bei 16 Viren pro Reaktion. Die klinische Nachweisgrenze bei *Schweren Akuten Respiratorischen Syndroms* (SARS)-Patienten lag bei rund 100 viralen Partikeln pro eingesandtem Probenmaterial. Die Sensitivität der universellen RT-PCR war identisch mit individuellen spezies-spezifischen real time RT-PCR Tests bei 46 Kindern, die mit humanen Coronaviren 229E, NL63, OC43 oder HKU1 infiziert waren. Die Identifizierung der jeweiligen Coronavirusspezies war in jedem der getesteten Isolate einheitlich und in voller Übereinstimmung mit den jeweiligen Coronavirus Prototypen und konnte auch durch Nukleotid-Sequenzierung bestätigt werden.

Die Prävalenzstudie zu akuten Gastroenteritiden wurde anhand von 499 Patientenproben aller Altersstufen durchgeführt, welche im Zeitraum eines Jahres gesammelt wurden. Die HPeV Grundprävalenz lag bei 1,6%. Positive Proben entstammten nur aus den Jahreszeiten Sommer und Herbst. Mit einer Ausnahme waren alle positiven Patienten unter 2 Jahren alt, bei gleichmäßiger Verteilung auf beide Geschlechter. Bei Kindern unter 2 Jahren konnte eine 11,6% Prävalenz ermittelt werden. Die Mehrzahl der Isolate konnte als gegenwärtig zirkulierendes HPeV Typ 1 identifiziert werden. Zusätzlich konnte ein neues sechstes HPeV identifiziert und beschrieben werden.

Die VIDISCA Methode lieferte ein 188 Nukleotid-Fragment aus regulär nicht typisierbarem Zellkulturmaterial, welches nach initialer Sequenzierung HPeV Typ1 zugeordnet werden konnte. Eine phylogenetische Analyse des Vollgenoms dieses Kulturvirus zeigte deutliche Abweichung vom 1961 erstbeschriebenen HPeV Typ1 Prototypen und insbesondere eine Rekombination zwischen HPeV Typ 1 und 3.

Die GDD Methode konnte erfolgreich auf das Chikungunya Virus Modell angewendet werden. Es war möglich, Chikungunya Fragmente sowohl aus Vollgenom *in vitro* RNA Transkripten, als auch aus Zellkulturmaterial zu amplifizieren, klonieren und zu sequenzieren.

Abbreviations

Viruses

AdV	Adenovirus
ALC	Avian leopard cat coronavirus
AMV	Avian myeloblastosis virus
BCoV	Bovine coronavirus
BtCoV	Bat coronavirus
CCoV	Canine coronavirus
CFBCoV	Chinese ferret badger coronavirus
CMV	Cytomegalovirus
CoV	Coronavirus
CRCoV	Canine respiratory coronavirus
DCoV	Duck coronavirus
ECoV	Equine coronavirus
EV	Enterovirus
FCoV	Feline coronavirus
FECV	Ferret enteric coronavirus
FIPV	Feline infectious peritonitis virus
GBV	Hepatitis G virus
GCoV	Goose coronavirus
GiCoV	Giraffe coronavirus
HBoV	Human boca virus
HCoV	Human coronavirus
HCV	Human hepatitis C virus
HEV	Porcine hemagglutinating encephalomyelitis
HEV	Human hepatitis E virus
HHV	Human herpes virus
hMPV	Human metapneumovirus
HPeV	Human parechovirus
HRV	Human retrovirus
HSV	Herpes simplex virus
IBV	Avian bronchitis virus
Inf	Influenza virus
KSHV	Kaposi’s sarcoma associated virus
MHV	Murine hepatitis virus

MuLV	Murine leukemia virus
PaCoV	Parrot coronavirus
PARV	Parvovirus
PuCoV	Pigeon coronavirus
PiCoV	Puffinosis coronavirus
PEDV	Porcine epidemic diarrhoea virus
PhCoV	Pheasant coronavirus
PIV	Parainfluenza virus
PRCoV	Porcine respiratory coronavirus
QCoV	Quail coronavirus
RCoV	Rat coronavirus
RDCoV	Raccoon dog coronavirus
RM-Bt-CoV	Rocky mountain bat coronavirusS
RSV	Respiratory syncytial virus
SARS-CoV	SARS coronavirus
SDAV	Rat sialodacryoadenitis coronavirus
SHCoV	Spotted yena coronavirus
SIV	Simian immunodeficiency virus
SNV	Sin nombre virus
TCoV	Turkey coronavirus
TGEV	Transmissible gastroenteritis virus
TTV	Torque teno virus
VZV	Varicella zoster virus

Other

°C	Degree Celsius
2’-O-MT	S-adenosylmethionine-dependent 2’-O-ribose mehytltransferase
ACE	Angiotensin-converting enzyme
AEGIS	Expanded genetic information system
AFLP	Amplified fragment length polymorphism
APN	Aminopeptidase N
bDNA	Branched DNA
BSA	Bovine serum albumin
CDD	Conserve domain database
cDNA	Complementary DNA

CEACAM	Carcinoembryonic antigen-related cell adhesion molecule
CFU	Colony-forming units
CI	Confidence intervals
CODEHOP	Consensus-degenerated hybrid oligonucleotide primer
CPE	Cytophatic effect
cre	*cis*-acting replication element
CSF	Cerebrospinal fluid
*Cvi*AII	Chlorella virus PBCV-1
DAF	Decay accelerating factor
DEPC	Diethyl pyrocarbonate
DMEM	Dulbecco's Modified Eagle's Medium
dNTP	Deoxyribonucleoside triphosphate
DTT	1,4-dithiothreitol
EHA	Enzyme hybridization assay
eIF	Eukaryotic initiation factor
ER	Endoplasmic reticulum
ERGIC	Endoplasmic reticulum-Golgi intermediate compartment
ExoN	3’ to 5’ exonuclease
FRET	Fluorescence resonance energy transfer
HE	Hemagglutinin esterase
HEL	Helicase gene
*Hin*P1I	*Haemophilus influenzae* P1
HRP	Horseradish peroxidase
IFU	Infection-forming units
IPTG	Isopropyl-β-D-thiogalactoside
IRES	Internal ribosome entry site
isoC	Isocytosine
isoG	Isoguanosine
kDa	Kilo Dalton
LB	Luria broth
LCD	Low-cost, low-density
LOD	Limit of detection
MEM	Lipid envelope membrane
*Mse*I	*Micrococcus* species
NendoU	Uridylate-specific endonuclease

NTP	Nucleoside triphosphate
NTPase	Nucleoside-triphosphatase
ORF	Open reading frame
$P[dN]_6$	Random hexamer primers
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PLP	Papain-like cysteine proteinase
pu	purine[
py	Pyrimidine
RACE	rapid amplification of cDNA ends
RDA	Representational difference analysis
RdRp	RNA-dependent RNA polymerase
RGD	Arginine- glycine- aspartic acid motif
RNP	RNA-nucleocapsid
RPM	Respiratory pathogen microarray
RT	Reverse transcriptase
RT- PCR	Reverse transcriptase polymerase chain reaction
SARS	Severe acute respiratory syndrome
SISPA	Sequence-independent, single-primer amplification
ssDNA	Single-stranded DNA
ssRNA	Single-stranded RNA
TAE	Tris-Acetate-EDTA
Taq	*Thermus aquaticus*
TBE	Tris-Borate-EDTA
TCID	Tissue culture infective dose
TdT	Terminal deoxynucleotidyl transferase enzyme
Tfl	*Thermus flavus*
TMB	3,3',5,5'-tetramethylbenzidine
TOPO	Topoisomerase
TRS	Transcription-regulatory sequence
UTR	Untranslated region
VIDISCA	Virus- Discovery- cDNA- AFLP
VPg	Virion protein, genome linked
WHO	World Health Organization

List of figures

List of tables

1. Introduction

1.1 Viral respiratory infections

Viral respiratory infections represent a major public health problem because of their worldwide occurrence, ease of spread in the community and considerable morbidity and mortality. Mortality from respiratory virus infections in healthy individuals in developed countries is rare; however, in less developed countries childhood mortality can be very high with an estimated 5 million children under 5 years dying annually worldwide from respiratory virus infections (WHO, 2007). Their worldwide occurrence results in people of all ages being susceptible to respiratory virus infections. However, on average, children are infected two to three times more frequently than adults are, with acute respiratory virus infections being the most common infections experienced by healthy children (Kesson, 2007).

Most viruses are transmitted by direct contact or droplets, although some are transmitted by aerosols. Viruses that primarily infect the respiratory tract include influenza viruses (Inf), adenoviruses (AdVs), parainfluenza viruses (PIVs), respiratory syncytial virus (RSV), coronaviruses (CoVs), human metapneumovirus (hMPV), rhinoviruses (HRVs) and enteroviruses (EVs). More recently, a new virus, bocavirus (HBoV), has been identified (Allander *et al.*, 2005). Inf, PIV, hMPV and RSV occur in epidemics while AdVs, CoVs and HRVs occur endemically. Other viruses that can cause respiratory diseases, more commonly in immunosuppressed people, include measles, varicella zoster virus (VZV), herpes simplex virus (HSV) and cytomegalovirus (CMV).

Respiratory virus infections often have a seasonal distribution especially in temperate climates and while the peak incidence varies year to year, there is often a predominant seasonal occurrence. RSV and Inf both have a peak incidence in winter and these peaks usually do not coincide but overlap. PIV 3 usually peaks in winter with PIV 1 and PIV 2 peaking in autumn and early winter. The *Picornaviridae* cause infections all year round with EVs more common in summer and autumn while HRVs are more common in winter and spring. AdVs tend to cause infections all over the year as do the *Herpesviridae*, except for varicella that is more prevalent in the late winter and early spring (Kesson, 2007).

1.1.1 The respiratory infection viruses

1.1.1.1 Influenza viruses (Inf)

Inf A virus was isolated in 1933, Inf B in 1940 and Inf C virus in 1951. Inf A, B and C viruses belong to the family *Orthomyxoviridae*. Influenza virions are enveloped particles containing a single-stranded, negative-sense segmented RNA genome that is surrounded by a

helical capsid, with Inf A and B containing eight segments of RNA and Inf C containing seven segments. Only Inf A and B are clinically important. Inf C infection occurs uncommonly and is usually associated with the mild upper respiratory tract illness; it can rarely cause bronchitis or pneumonia.

The enveloped virion has peplomers or spikes consisting of two glycoproteins, the hemagglutinin (H) which is involved in the attachment of the virus to cells and the initiation of infection, and neuraminidase (N) which facilitates release of newly formed virions from the cells. The two glycoproteins, H and N, exhibit substantial antigenic variation among Inf A viruses with 16 H subtypes and 9 N subtypes recognized. Inf B has only one type of H and N glycoproteins. The occurrence of annual influenza epidemics throughout the world is due to the high rate of antigenic variation from a stepwise mutation of the H and/or N genes and reflected in variations of the antigenic characteristics of these proteins by the H and N glycoproteins of Inf A and B viruses. This stepwise mutation of H and N results in antigenic drift and the recurrent annual influenza epidemics seen each winter. The virus can infect and produce disease among populations who would otherwise possess immunity from previous infection because their antibodies fail to recognize the new antigenic variations. This is the reason for the necessity for annual influenza vaccination with differing serotypes of Inf A and B viruses. In addition, the segmented nature of the influenza genome allows antigenic shift, which is the reassortment of genome segments from two different Inf A viruses with major changes in the H or N proteins or both. It is this major variation in genetic make-up, which gives the Inf A virus the potential for the development of global pandemics.

1.1.1.2 Parainfluenza viruses (PIVs)

PIV types 1, 2 and 3 occur worldwide and among persons from all age groups. PIV 4A and 4B are much less frequent. PIVs are belonged to the family *Paramyxoviridae* and genus Respirovirus. PIVs are enveloped particles containing a single-stranded, negative-sense, non-segmented RNA genome that is surrounded by a helical capsid.

PIV 1 occurs in epidemics usually during autumn in alternate years, PIV 2 occurs sporadically and PIV 3 tends to cause annual winter epidemics in temperate climates. PIV 1-3 are the main causes of croup in infants and young children under 5 years of age (Vainionpaa & Hyypia, 1994). PIV 3 can also cause viral pneumonia and bronchiolitis in infants and small children. PIV 4 occurs very infrequently and is usually associated with mild symptoms of upper respiratory tract illness (rhinorrhea, pharyngitis and cough). Primary infection with PIV provides

some measure of immunity but this immunity is not complete or long lasting and reinfections occur commonly but they are rarely as severe as the illnesses seen with primary infection.

1.1.1.3 Respiratory syncytial virus (RSV)

RSV was first isolated in 1956 from a laboratory chimpanzee with an illness resembling the common cold, and shortly after it was demonstrated to be a human pathogen. Epidemiological studies have shown that RSV represents the single most important cause of serious lower respiratory tract disease, especially bronchiolitis and pneumonia, in infants and children (Singh *et al.*, 2007). RSV can cause severe pneumonia and death in persons with underlying immune deficiency.

RSV belongs to the family *Paramyxoviridae* and the genus Pneumovirus. It is an enveloped virus with a single-stranded, negative-sense, non-segmented RNA genome with at least 10 viral proteins. Antigenic analysis of RSV has identified two subgroups, A and B, based on their reactivity to a panel of monoclonal antibodies with the B strain further differentiated into two variants, B1 and B2. Annual epidemics of RSV occur during winter in temperate climates and the two subgroups usually co-circulate in the same geographical area, often with a predominance of subgroup A.

1.1.1.4 Adenoviruses (AdVs)

AdVs were isolated from the primary cell cultures of adenoids from children in the early 1950s. AdVs are non-enveloped particles that contain linear double-stranded DNA (dsDNA) surrounded by an icosahedral capsid with fibre-like projections. AdVs infections occur worldwide and transmission varies from sporadic to epidemic. Since AdVs are very stable, they can be easily transmitted in the environment by fomites. AdVs are an important cause of upper and lower respiratory tract disease with types 1, 2, 3, 5 and 7 accounting for about 85% of all infections.

1.1.1.5 Rhinoviruses (HRVs)

HRVs are a group of more than 100 serotypes and cause more common cold (minor upper respiratory tract) illness than any other virus that infects the respiratory tract. They account for about one half of common colds occurring in children (Greenberg, 2003). In persons with underlying lung disease and immune disorders, HRVs can cause pneumonia.

HRVs belong to the family *Picornaviridae*, which are small, non-enveloped viruses with a single-stranded, positive-sense RNA genome. Unlike EVs, HRVs are inactivated when exposed to mild acid (pH <5), accounting for their failure to infect the gastrointestinal tract.

HRVs also caused otitis media in infants and children and they have been recovered from middle ear fluid in 10% of sub-acute or chronic cases that are negative for bacteria. Recent studies associate HRVs infection with exacerbations of asthma and with acute lower respiratory tract disease, especially in persons with chronic obstructive airway disease and cystic fibrosis, and those who are immunocompromised (Brownlee & Turner, 2008; Mallia & Johnston, 2006; Tan, 2005). HRVs prefer to grow at lower temperatures than many other respiratory viruses (33 to 34°C), and grow less well at the higher temperature of the lungs than in the nose; however, they can establish infection in the lung (Brownlee & Turner, 2008).

1.1.1.6 Human metapneumovirus (hMPV)

hMPV was first isolated from 28 respiratory specimens collected over a 20-year period in the Netherlands (van den Hoogen *et al.*, 2001). Electron microscopy revealed filamentous viral-like particles, suggesting that the agent was a virus. hMPV was found to be a member of the *Paramyxoviridae* family and the first human pathogen of the genus Metapneumovirus. Since its initial discovery, hMPV has been identified worldwide. In general, hMPV infection accounts for approximately 2–12% of paediatric lower respiratory illness (Alto, 2004).

1.1.1.7 Bocavirus (HBoV)

HBoV is a newly discovered parvovirus that was first identified in Sweden but which occurs globally (Allander *et al.*, 2005; Foulongne *et al.*, 2006; Longtin *et al.*, 2008; Redshaw *et al.*, 2007). This virus is closely related to bovine parvovirus and canine minute virus and is a member of the genus Bocavirus, subfamily *Parvovirinae*, family *Parvoviridae*. HBoV has been detected in young children with respiratory distress, many with pneumonia with interstitial infiltrates noted on chest X-ray. HBoV in children with gastroenteritis have also been reported (Albuquerque *et al.*, 2007). However, whether HBoV is a cause of respiratory disease and gastroenteritis is yet to be fully determined.

1.2 Coronaviruses (CoVs)

CoVs are large enveloped viruses with RNA genomes of non-segmented, single-stranded molecules of positive sense. They are currently classified as one of the two genera in the family *Coronaviridae*, together with the genus Torovirus. The family *Coronaviridae* is part of the Nidovirales order that also includes two other families, *Arteriviridae* and *Roniviridae* (Cavanagh, 1997; Gonzalez *et al.*, 2003).

CoVs infect many species of animals, including humans, and are important causes of diseases that include upper and lower respiratory infections, gastroenteritis, hepatic and neurological disorders (Weiss & Navas-Martin, 2005). In animals, CoVs are responsible for a number of economically important diseases. Avian infectious bronchitis virus (IBV), a CoV isolated from domestic fowls (*Gallus gallus*), is one of the foremost causes of economic loss within the poultry industry (Cavanagh, 2007). Porcine transmissible gastroenteritis virus (TGEV) causes devastating disease in newborn pigs, with mortality rates approaching 100% (Enjuanes *et al.*, 1995; Shibata *et al.*, 2000). Bovine CoV (BCoV) causes severe diarrhoea in newborn calves, winter dysentery in adult cattle, and respiratory infections in calves and feedlot cattle (Park *et al.*, 2006). Several other animal CoVs also cause enteritis, such as turkey CoV (TCoV), feline CoV (FCoV), canine CoV, and porcine epidemic diarrhoea CoV (PEDV). In addition to respiratory and enteric diseases, murine hepatitis virus (MHV) causes hepatitis in mice and rats, and some strains can cause encephalitis (Weiss and Navas-Martin, 2006). Porcine hemagglutinating encephalomyelitis virus (HEV) causes encephalitis in pigs (Sasseville *et al.*, 2002).

In humans, CoVs were known to cause the common cold, a mild infection without important pathotological consequences except in immunocompromised patients. Before 2003, only two human CoVs (HCoVs) were characterized. HCoV-229E and HCoV-OC43 were discovered in the 1960s and subsequently have been found to cause mainly mild upper respiratory tract diseases (Masters, 2006). In 2003, the World Health Organization (WHO) received reports from China of an atypical pneumonia outbreak in Guangdong Province (China) in November 2002. This illness was later denominated as severe acute respiratory syndrome (SARS). A novel virus was isolated causing SARS and later identified as a CoV (SARS-CoV), which infected the lower respiratory tract and caused fever, dry cough, dyspnoea, headache and hypoxemia (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Peiris *et al.*, 2003; Rota *et al.*, 2003). By March 2003, the disease had spread worldwide and by July there were over 8,000 infected individuals leading to more than 800 deaths in 32 countries. The epidemic was controlled in August 2003 (Satija & Lal, 2007). Additional sporadic cases occurred in the period between the winter of 2003 and early spring of 2004 (Gu & Korteweg, 2007).

The discovery of SARS-CoV has triggered systematic investigations for new CoVs. In 2004, a novel HCoV was isolated from a child with bronchiolitis in The Netherlands and termed HCoV-NL63 (van der Hoek *et al.*, 2004). In the same year, in an independent investigation in the Netherlands, HCoV-NL63 was also isolated from a child suffering pneumonia (Fouchier *et al.*, 2004). In 2005, other novel HCoV (HKU1) was described from an adult with chronic pulmonary disease in Hong Kong (Woo *et al.*, 2005). In addition, a variety of novel animal CoVs from mammals and birds has been described (Table 1).

1.2.1 Taxonomy

Based on genotypic and serological characteristics, CoVs were classified into three distinct groups (Brian & Baric, 2005; Lai & Cavanagh, 1997; Ziebuhr, 2004). In recent studies, however, new subgroups within groups 1 and 2 are proposed based on phylogenetic analyses of partial and full genome sequences of novel discovered CoVs (Guan *et al.*, 2003; Lau *et al.*, 2005; Poon *et al.*, 2005; Tang *et al.*, 2006; Vijaykrishna *et al.*, 2007; Woo *et al.*, 2006). Group 1 and 2 include CoVs from mammals and group 3 includes avian CoVs. Moreover, phylogenetic analyses of partial genome sequences of a novel CoV detected from leopard cats (*Prionailurus bengalensis*) and Chinese ferret badgers (*Melogale moschata*), in Southern China, showed that some structural proteins and replicase genes were most closely related to group 3 CoVs, whereas the spike protein was most closely related to group 1 CoVs (Dong *et al.*, 2007). Table 1 shows CoV species distribution within groups.

Table 1: Distribution of coronaviruses within groups.

Group	Virus*	Host	GenBank accession no.	Reference
1	BtCoV-61	Bat	AY864196	(Poon *et al.*, 2005)
	BtCoV-HKU6	Bat	DQ249224	(Woo *et al.*, 2006)
	CCoV	Dog	D13096	(Horsburgh *et al.*, 1992)
	CRCoV	Dog	DQ471938	(Escutenaire *et al.*, 2007)
	FECV	Ferret	DQ340560	(Wise *et al.*, 2006)
	RM-Bt-CoV 11	Bat	EF544563	(Dominguez *et al.*, 2007)
	RM-Bt-CoV 27	Bat	EF544564	(Dominguez *et al.*, 2007)
	RM-Bt-CoV 3	Bat	EF544567	(Dominguez *et al.*, 2007)
	RM-Bt-CoV 48	Bat	EF544565	(Dominguez *et al.*, 2007)
	RM-Bt-CoV 6	Bat	EF544568	(Dominguez *et al.*, 2007)
	RM-Bt-CoV 65	Bat	EF544566	(Dominguez *et al.*, 2007)
	SHCoV	Hyena	DQ317972	(East *et al.*, 2004)
1a	CFBCoV	Ferret	EF192160	(Vijaykrishna *et al.*, 2007)
	FCoV	Cat	DQ010921	(Dye & Siddell, 2005)
	FIPV	Cat	AY994055	Unpublished
	RDCoV	Raccoon dog	EF192159	(Vijaykrishna *et al.*, 2007)
	TGEV	Pig	AJ271965	(Almazan *et al.*, 2000)
1b	BtCoV-HKU2	Bat	EF203064	(Lau *et al.*, 2007)
	BtCoV-HKU7	Bat	DQ249226	(Woo *et al.*, 2006)
	BtCoV-HKU8	Bat	DQ249228	(Woo *et al.*, 2006)
	BtCoV/512/2005	Bat	DQ648858	(Tang *et al.*, 2006)
	BtCoV/A515/2005	Bat	DQ648822	(Tang *et al.*, 2006)
	BtCoV/A527/2005	Bat	DQ648823	(Tang *et al.*, 2006)
	BtCoV/A773/2005	Bat	DQ648835	(Tang *et al.*, 2006)
	BtCoV/A911/2005	Bat	DQ648850	(Tang *et al.*, 2006)
	BtCoV/A977/2005	Bat	DQ648855	(Tang *et al.*, 2006)
	HCoV-229E	Human	AF304460	(Thiel *et al.*, 2001)
	HCoV-NL63	Human	AY567487	(van der Hoek *et al.*, 2004)
	PEDV	Pig	AF353511	(Kocherhans *et al.*, 2001)
	PRCoV	Pig	Z24675	(Rasschaert *et al.*, 1990)
2	ECoV	Horse	EF446615	(Zhang *et al.*, 2007)
	PuCoV	Puffin	AJ005960	(Klausegger *et al.*, 1999)
	RCoV	Rat	AF088984	(Compton *et al.*, 1999)
	SDAV	Rat	AF207551	(Yoo *et al.*, 2000)
2a	BCoV	Cow	AF391541	(Chouljenko *et al.*, 2001)
	GiCoV-OH3	Giraffe	EF424623	(Hasoksuz *et al.*, 2007)
	HCoV-HKU1	Human	AY597011	(Woo *et al.*, 2005)
	HCoV-OC43	Human	AY391777	(Vijgen *et al.*, 2005)
	HEV	Pig	AF481863	(Sasseville *et al.*, 2002)
	MHV	Mouse	AF029248	(Leparc-Goffart *et al.*, 1997)
2b	Bat SARS-CoV Rf1/2004	Bat	DQ412042	(Li *et al.*, 2005)
	Bat SARS-CoV Rm1/2004	Bat	DQ412043	(Li *et al.*, 2005)
	Bat SARS-CoV Rp3/2004	Bat	DQ071615	(Li *et al.*, 2005)
	BtCoV/273/2005	Bat	DQ648856	(Tang *et al.*, 2006)
	BtCoV/279/2005	Bat	DQ648857	(Tang *et al.*, 2006)
	BtCoV-HKU3	Bat	DQ022305	(Lau *et al.*, 2005)
	Civet SARS-CoV 007	Civet	AY572034	(Wang *et al.*, 2005)
	Civet SARS-CoV SZ16	Civet	AY304488	(Guan *et al.*, 2003)
	Civet SARS-CoV SZ3	Civet	AY304486	(Guan *et al.*, 2003)
	SARS-CoV	Human	AY274119	(He *et al.*, 2004)
2c	BtCoV/133/2005	Bat	DQ648794	(Lau *et al.*, 2005)
	BtCoV-HKU4	Bat	EF065505	(Lau *et al.*, 2005)
	BtCoV-HKU5	Bat	EF065509	(Lau *et al.*, 2005)
2d	BtCoV-HKU9	Bat	EF065513	(Woo *et al.*, 2007)

(continued)

Group	Virus*	Host	GenBank accession no.	Reference
3	ALC	Leopard	EF584908	(Dong *et al.*, 2007)
	DCoV	Duck	AJ871024	(Jonassen *et al.*, 2005)
	GCoV	Goose	AJ871017	(Jonassen *et al.*, 2005)
	IBV	Chicken	M95169	(Boursnell *et al.*, 1987)
	PaCoV	Parrot	DQ233651	(Gough *et al.*, 2006)
	PiCoV	Pigeon	AJ871022	(Jonassen *et al.*, 2005)
	PhCoV	Pheasant	AJ619602	(Cavanagh *et al.*, 2002)
	QCoV	Quail	EF446156	(Circella *et al.*, 2007)
	TCoV	Turkey	AF124991	(Stephensen *et al.*, 1999)

*Abbreviations: ALC, Avian leopard cat coronavirus; BCoV, Bovine coronavirus; BtCoV, Bat coronavirus; CCoV, Canine coronavirus; CFBCoV, Chinese ferret badger coronavirus; CRCoV, Canine respiratory coronavirus; DCoV, Duck coronavirus; ECoV, Equine coronavirus; FCoV, Feline coronavirus; FECV, Ferret enteric coronavirus; FIPV, Feline infectious peritonitis virus; GCoV, Goose coronavirus; GiCoV, Giraffe coronavirus; HEV, Hemagglutinating encephalomyelitis virus; IBV, Avian infectious bronchitis virus; MHV, Mouse hepatitis virus; PaCoV, Parrot coronavirus; PiCoV, Pigeon coronavirus; PuCoV, Puffinosis coronavirus; PEDV, Porcine epidemic diarrhoea virus; PhCoV, Pheasant coronavirus; PRCoV, Porcine respiratory coronavirus; QCoV, Quail coronavirus; RCoV, Rat coronavirus; RDCoV, Raccoon dog coronavirus; RM-Bt-CoV, Rocky mountain bat coronavirus strain; SDAV, Rat sialodacryoadenitis virus; SHCoV, Spotted yena coronavirus; TCoV, Turkey coronavirus; TGEV, Transmissible gastroenteritis virus.

1.2.2 Structure and genomic organization of CoVs

CoVs are large spherical enveloped particles of approximately 80 to 120 nm in diameter. When examined by electron microscopy, the enveloped virion is seen to be surrounded by a crown-like structure, or from Latin *corona*, of studded projections, and therefore named as CoV (figure 1).

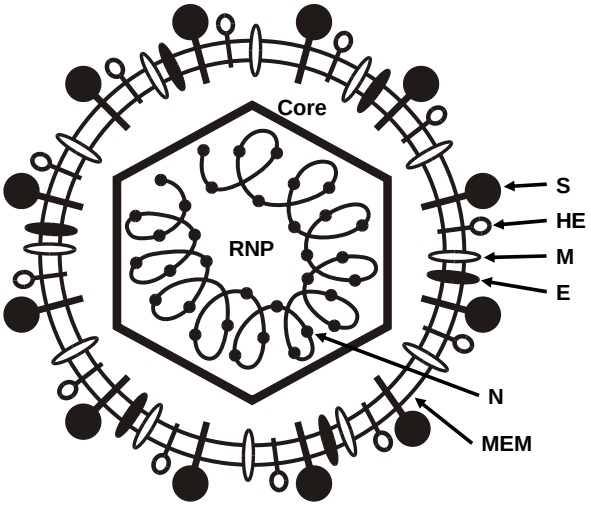

Figure 1: Diagram model of CoV structure. Abbreviations: S, spike protein; HE, hemagglutinin-esterase; M, membrane protein; E, small envelope protein; N, Nucleocapsid; MEM, lipid envelope membrane; RNP; RNA-nucleocapsid protein (Lai *et al.*, 2007).

The genome RNA is associated with the basic nucleocapsid (N) protein to form a long, flexible and coiled helical capsid found within the viral membrane. There is evidence that in TGEV and MHV, the nucleocapsid is enclosed within a spherical or possibly icosahedral core structure (Risco *et al.*, 1996). The virus core is enclosed by a lipoprotein envelope or membrane. The envelope of all CoVs contains at least three viral proteins. These are spike (S), a type I glycoprotein that forms the peplomers on the virion surface, giving the virus its *corona* or crown-like appearance in electron microscopy; the membrane protein (M), a protein that spans the membrane three times and has a short N-terminal ectodomain and a cytoplasmic tail; and small membrane protein (E), a highly hydrophobic protein. Some group II CoVs have an additional membrane protein, hemagglutinin-esterase (HE). There is an additional group II virion protein called Internal protein (I), as it is encoded within the nucleocapsid open reading frame (ORF). This is a nonessential protein of unknown function (Fischer *et al.*, 1997). In addition, the 3a protein encoded by SARS-CoV is an additional structural protein (Ito *et al.*, 2005).

CoVs have the largest RNA genomes of any known virus family, at lengths ranging from 27,164 nucleotides for some strains of BtCoV-HKU2 to 31,526 nucleotides for MHV strain JHM. The non-segmented, single-stranded, positive-sense RNA genome resemble most eukaryotic messenger RNAs (mRNAs) in containing a 5’ methyl-guanosine cap (Lai & Stohlman, 1981) and a 3’-terminal poly(A) tail (Lai & Stohlman, 1978; Wege *et al.*, 1978). Like other positive-sense RNA viruses, the genomic RNA of CoV is infectious when transfected in permissive host cells (Brian *et al.*, 1980; Schochetman *et al.*, 1977).

The 5’ untranslated region (UTR) of CoVs range from 209 to 528 nucleotides. Interestingly, the 5’ UTR contains a short ORF of 65 to 98 nucleotides, AUG-initiated, called *leader region*, which is involved in regulation of mRNA translation and is also present at the 5’ end of all subgenomic mRNAs (Morris & Geballe, 2000). The 3’ UTR of CoVs range from 288 to 506 nucleotides and possess an octameric sequence duplication beginning at base 73 to 80 upstream from the poly(A) tail. The sequences of both 5’ and 3’ UTRs are important for RNA replication and transcription. The remaining genome sequences contain multiple ORFs, and the order of genes is highly conserved among different CoVs.

The first gene (Gene 1), which comprises two thirds of the genome from the 5’ end, is about 20,000 to 22,000 nucleotides in length. It consists of two ORFs (1a and 1b) that are partially overlapping by a region of 40 to 60 nucleotides. The translation of ORF 1b depends on a ribosomal frame shift caused by a complex RNA structure pseudoknot located at the junction of ORF 1a and 1b (Bredenbeek *et al.*, 1990). This frame shift mechanism occurs at frequencies ranging from 20 to 30% (Bredenbeek *et al.*, 1990; Brian & Baric, 2005). These ORFs are

translated into a polyprotein, which is the precursor of proteins necessary for RNA replication. The polyproteins are processed by virus-coded proteinases at cleavage sites flanking the protein domains. ORF 1a encodes two papain-like cysteine proteinases (PLP1 an PLP2) and a chymotrypsin-like cysteine proteinase that resembles the picornavirus 3C proteinases ($3CL^{pro}$ or M^{pro}). The proteolytic cleavage of ORF1b produces a set o proteins, including the RNA-dependent RNA polymerase (RdRp), a multifunctional helicase (HEL), a putative 3' to 5' exonuclease (ExoN), an uridylate-specific endonuclease (NendoU), and an S-adenosylmethionine-dependent 2'-O-ribose mehytltransferase (2'-O-MT) (Ivanov *et al.*, 2004; Minskaia *et al.*, 2006; Snijder *et al.*, 2003). The genes encoding the ORF 1a and 1b are collectively called replicase. The structural proteins are encoded within the 3' one third of the genome. The invariant gene order in all members of the CoV family is 5'-replicase-S-E-M-N-3'. Figure 2 shows the genome organization of the prototype group 1, 2 and 3 CoVs. These genes are interspersed with several ORFs encoding non-structural proteins, unique for each group of CoVs. Most of these proteins are not essential for virus replication and their functions are scarcely described. In some cases, their deletion causes virus attenuation in natural hosts (de Haan *et al.*, 2002).

The genome or CoVs has multiple functions during infection. It acts initially as an mRNA that is translated into the huge replicase polyprotein. The replicase is the only translation product derived from the genome; all downstream ORFs are expressed from subgenomic RNAs. Next, the genome serves as the template for replication and transcription. Finally, the genome plays a role in assembly, as progeny genomes are incorporated into progeny virions.

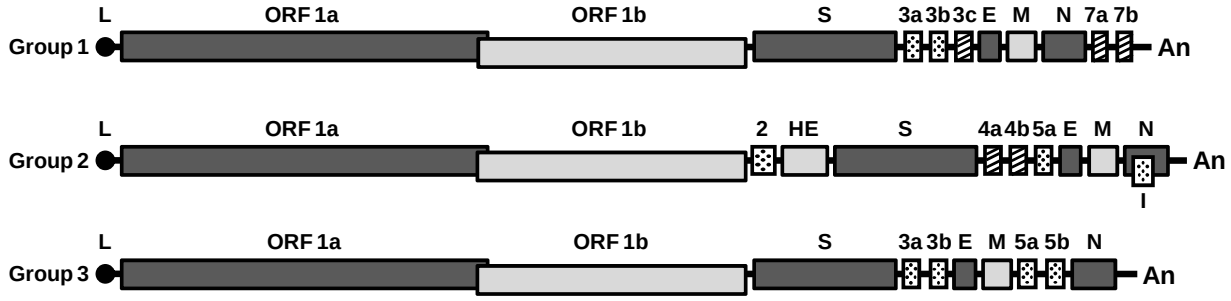

Figure 2: Genome organization of the prototype group 1, 2 and 3 CoVs. Indications on the top of boxes indicate the genes for the replicase, structural and non-structural proteins. Patterned boxes stand for non-structural protein genes that are common to all (dotted boxes) or only some (striped boxes) species in that group. L, leader sequence. (Gonzalez *et al.*, 2003).

1.2.3 Structural proteins

1.2.3.1 Hemagglutinin-esterase (HE)

The HE protein is present in some group 2 CoVs and group 3 TCoV as a disulfide-linked dimer of a 65 to 70 kDa protein that forms short spikes on the virions (Hogue *et al.*, 1989). This protein has functional similarities to both the receptor-biding virus hemagglutinin protein (HA), and to the influenza receptor-destroying enzyme, neuraminidase (NA), of Inf C virus (Vlasak *et al.*, 1988a; Vlasak *et al.*, 1988b). HE may play a role in biding and entry of CoVs into target cells, and could facilitate virus release from the cell surface (Cornelissen *et al.*, 1997). In a study with recombinant strains of MHV, it was found that the presence of HE dramatically enhanced neurovirulence, as measured by viral spread and lethality (Kazi *et al.*, 2005).

1.2.3.2 Spike (S)

The spike glycoprotein is the major surface transmembrane protein protruding from the viral envelope, which forms trimers with a stalk-and-club morphology, giving the virion a crown-like fringe of proteins (Delmas & Laude, 1990). S protein has a molecular mass of about 150 to 180 kDa. It is responsible for virus entry and cell-to-cell spread, and is the primary determinant of host range and tissue tropism. It is the primary target for immune responses of the host and neutralising antibody induction (Weiss & Navas-Martin, 2005).

The S protein is synthesized as a single polypeptide chain that in most group 2 and all group 3 CoVs is cleaved by a trypsin-like host protease into two polypeptides, the amino-terminal S1 domain and a carboxy-terminal S2 domain of roughly equal sizes. Even for uncleaved S proteins of the group 1 CoVs and SARS-CoV, the designations S1 and S2 are used for the amino-terminal and carboxy-terminal halves of the S protein, respectively. The S1 subunit, which forms the clubbed end of the spike, is the most divergent region of the molecule within and across the CoV groups. It recognizes specific cellular receptors and initiates attachment (Wang *et al.*, 1994). The S2 subunit, which is conserved among all CoVs and forms the stalk of the spike, has a short C-terminal tail, a hydrophobic transmembrane domain, and an exterior domain with two long alpha-helical regions. Adjacent alpha helices within the trimer interact with each other to form a structure called a coiled-coil, similar to the transmembrane portion of the envelope glycoproteins of a number of the RNA viruses (de Groot *et al.*, 1987).

1.2.3.3 Small envelope protein (E)

The small envelope protein is an 8.4 to 12 kDa protein that is present only in very limited numbers in the virion. Virus-like particles can be formed by expressing only the E and M proteins together in cells, suggesting a critical role for E in virion formation and budding. The E protein may serve to facilitate budding of the virion into the lumen of the endoplasmic reticulum and Golgi membranes during particle maturation (Liu *et al.*, 2007a).

1.2.3.4 Membrane protein (M)

The membrane glycoprotein is the most abundant structural protein in the CoVs, and gives the virion envelope its shape. It is one of the structural proteins, together with E protein, that are essential for the production of virus-like particles. M protein is a triple-spanning membrane protein containing three hydrophobic domains separated by short hydrophilic regions. Rigid constraints on its structure are likely essential for the functional requirements of M in virus assembly. The amino terminus of M is located on the exterior of the virion and is in contact with the spike protein. The carboxy terminus comprises the major part of the molecule and extends within the virion underneath the envelope, and is in contact with the viral nucleocapsid (Masters, 2006).

1.2.3.5 Nucleocapsid protein (N)

The N protein is a 43 to 50 kDa phosphoprotein component of the helical nucleocapsid and it binds to the genomic RNA resulting in a “beads-on-a-string” structure (Masters, 2006). Based on a comparison of sequences of multiple strains, it has been proposed that the N protein is divided into three conserved domains, including an RNA-biding domain in the middle that bonds to the leader sequence of viral RNA. The domains are separated by two highly variable spacer regions (Parker & Masters, 1990). The main role of N protein is to bind to the viral RNA, but also has been implicated in virus RNA synthesis. Unlike the helical nucleocapsids of non-segmented negative-strand RNA viruses, the N protein in ribonucleoprotein complexes provides only limited protection to the RNA genome against the action of ribonucleases (Macneughton & Davies, 1978).

1.2.4 Coronavirus replication cycle

The viral cycle of CoVs are illustrated in the figure 3.

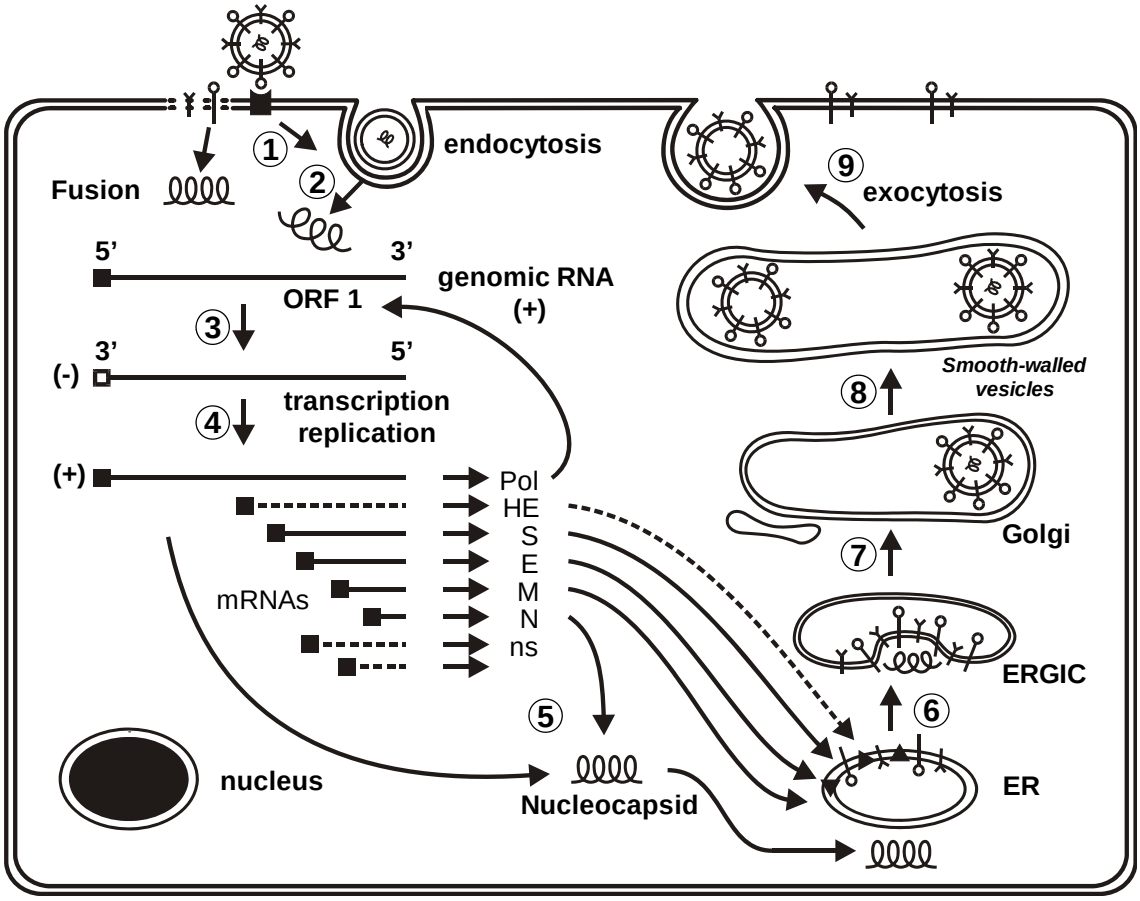

Figure 3: coronavirus replication cycle. After adsorption of viral particles (1), the positive-stranded RNA genome is uncoated (2) and replication starts in the cytoplasm. A nested set of subgenomic mRNA is produced by discontinuous transcription during minus-strand synthesis (3 to 4) resulting in structural and non-structural proteins. Positive-stranded genomic RNA and viral proteins assemble and bud at the ER Golgi and ERGIC (5 to 7), and are processed in the Golgi apparatus (8). Newly produced viral particles are incorporated in vesicles and released by exocytosis (9). (Lai *et al.*, 2007).

1.2.4.1 Virus entry

The viral replication cycle initiates with the biding of virions to the plasma membranes of target cells, by means of the interaction between S protein to a specific receptor glycoprotein on the cell surface.

The spike protein of group 1 CoVs bind to aminopeptidase N (APN), a family of zinc-biding metalloproteinases found in cellular plasma membranes. These receptors are widely distributed on the apical surfaces of respiratory and intestinal epithelium, as well as on myelocytic cells, on kidney tubular epithelium, on monocyte and granulocyte lineages, and at synaptic junctions. Despite the high degree of similarity among members of the APN family, the utilization of the receptors by CoVs is for the most part species-specific.

Group 2 CoVs utilizes a variety of cell receptors. MHV binds to carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1 or CD66a) (Tan *et al.*, 2002). CEACAM1 is a member of the immunoglobulin family and is widely distributed in all mammalian species.

SARS-CoV and HCoV-NL63 use angiotensin-converting enzyme 2 (ACE2) (Hamming *et al.*, 2007; Li *et al.*, 2003), another metalloproteinase, as a receptor. ACE2 is expressed in the lungs, heart, kidney, and small intestine as well as other tissues (Hamming *et al.*, 2004). SARS-CoV can also infect certain cell types that lack ACE2 and recent studies indicate that L-SIGN, a cell surface lectin, may also serve as a receptor for the SARS-CoV (Gramberg *et al.*, 2005).

CoVs that express HE can also bind to cells by interaction between HE and 9-O-acetylated sialic acid, found on a variety of cell membrane glycoproteins and glycolipids (Lai & Cavanagh, 1997). Sialic acid is a common target bound by surface proteins of many animal viruses. This esterase can cleave an acetyl group from 9-O-acetylated sialic acid, resulting in the release of bound virus from the cell surface (Vlasak *et al.*, 1988b). This may help in the release and dispersal of virions from some cell types.

Following virus attachment, the S protein mediates fusion of the virus envelope with either the plasma membrane or the endosomal membranes. While the external S1 subunit is responsible for attachment, the S2 subunit that forms the stalk facilitates fusion through a series of conformational changes that result in the insertion of S2 into the target cell membrane, bringing the cell membrane and the viral envelope into close contact (Zelus *et al.*, 2003). Some S proteins can induce cell-cell fusion at neutral or alkaline pH values, suggesting that this virions fuse directly with the plasma membrane (Lai & Cavanagh, 1997). When these S proteins are incorporated into the plasma membrane of an infected cell, cell fusion can occur, resulting in the formation of a large multinucleated giant cells, or syncytia, which are observed in some CoV infections. Other S proteins require low pH to induce fusion, suggesting that they enter via endosomes and fuse with the endosomal membrane once the internal pH has decreased (Gallagher *et al.*, 1991).

After virus-membrane fusion either at the plasma membrane or at the endosome, the viral nucleocapsid is released into the cytoplasm and the RNA is uncoated to become available for translation. This process is not well understood. Removal of phosporylation or ubiquitination of certain cellular factors may be involved (Mohandas & Dales, 1991; Yu & Lai, 2005).

1.2.4.2 Replication

All known CoV replication proteins can be found in membrane-associated complexes in the cytoplasm. These complexes are the sites of viral RNA synthesis, and therefore called replication complexes. Nucleocapsid protein is also found in abundance in replication complexes, suggesting that encapsidation of newly synthesized genome RNA for packing into virions also occurs at these sites. The proteins in replication complexes direct both genome replication and mRNA synthesis, or transcription. Replication of the genome involves the synthesis of a full-length negative-strand RNA that is present at a low concentration and serves as template for full-length positive-strand genomic RNA (Brian & Baric, 2005).

The expression of all CoV genes beyond ORF 1a/b occurs from a series of subgenomic mRNAs. These mRNAs have a unique structure. Each mRNA has a common leader sequence at its 5' end consisting of 65 to 98 nucleotides, which is identical to the sequence present at the 5' end of genome RNA (Sawicki *et al.,* 2007).

This untranslated leader sequence is joined to a subgenomic mRNA sequence that includes one or more of the ORFs. In addition, the initiation point of each mRNA corresponds to a stretch of consensus sequences, called intergenic sequences or transcription-regulatory sequences (TRSs). TRSs are moderately well conserved within each CoV group. The core consensus TRS for group 1 is 5'-AACUAAAC-3'; for group 2, 5'-AAUCUAAAC-3; except for SARS-CoV (5'-AAACGAAC-3'); and group 3, 5'-CUUAACAA-3' (Masters, 2006). All the subgenomic mRNAs extend to the 3' end of the genome RNA and have poly(A) tails. Because subgenomic mRNAs have different lengths and common sequences at their 3' end, they are referred to as a nested set of mRNA. The nested set of subgenomic mRNAs, with or without a leader sequence, is a defining feature of the order *Nidovirales* (Lai & Cavanagh, 1997).

The mechanism by which the group of positive- and negative-strand RNAs are synthesized involves a unique discontinuous transcription mechanism that is not completely understood. However, subgenomic mRNA synthesis is believed to be regulated by TRSs (Sawicki *et al.,* 2007). The current model is that discontinuous transcription occurs during the synthesis of subgenomic negative-strand RNAs, with the antileader sequences being added onto the 3' ends of negative-strand RNAs, which then serve as templates for synthesis of mRNAs (Sawicki *et al.,* 2007). Viral proteins are translated from individual mRNAs, generally from the 5' ORF only. In some cases there may be two ORFs carried on and translated from one mRNA. Expression levels of subgenomic mRNAs are relative to their length. Thus, subgenomic mRNA of S gene has the lowest expression level, whereas that of N gene has the highest one.

1.2.4.3 Virus assembly and release

After CoVs replication and transcription have taken place, assembly and budding of virions occurs into a different set of intracellular membranes. The N protein and newly synthesized genomic RNA assemble to form helical nucleocapsids. The membrane glycoprotein M is inserted in the endoplasmic reticulum (ER) and anchored in the Golgi apparatus (Klumperman *et al.*, 1994). The nucleocapsid probably first binds to M protein at the budding compartment, the endoplasmic reticulum-Golgi intermediate compartment (ERGIC), that lies between the ER and the Golgi apparatus (Stertz *et al.*, 2007). E protein is also transported to the ERGIC, where E and M proteins interact to trigger the budding of virions, enclosing the nucleocapsid. Spikeless virus-like particles can form in the absence of S protein, indicating that S protein is not essential for virus assembly but is required for the formation of infectious virus particles. The S and HE glycoproteins are glycosylated, trimerized, and transported through the Golgi apparatus. Along the transport pathway, S and HE proteins associate with M protein and are incorporated into the maturing virus particles (Opstelten *et al.*, 1995). Excess of S and HE proteins that are not incorporated into virions are transported to the plasma membrane, where they may participate in cell-cell fusion or hemadsorption, respectively.

After budding, virus particles undergo further morphological changes within the Golgi resulting in the appearance of mature virus particles with a compact, electron-dense internal core, as they reach the secretory vesicles. Virions accumulate in large, smooth-walled vesicles, which eventually fuse with the plasma through exocytosis to release virus into extracellular space (Sawicki *et al.*, 2007). The entire cycle of CoV replication occurs in the cytoplasm.

1.2.5 Coronavirus diversity and ecology

The spectrum of diseases observed in CoV infections reflects the capacity of these viruses to evolve and adapt to different environments. As a result of the unique mechanism of viral replication, CoVs have a high frequency of recombination (Lai, 1992; Lai & Cavanagh, 1997). For example, recombination frequencies of MHV have been calculated to be as high as 25% (Baric *et al.*, 1990). Such a high recombination rate, coupled with the infidelity of the polymerases of RNA viruses, may allow them to adapt to new hosts and ecological niches (Herrewegh *et al.*, 1998).

To date, RNA recombination has been demonstrated for MHV, TGEV, and IBV, both in tissue culture and experimental and natural animal infections (Banner *et al.*, 1990; Keck *et al.*, 1988; Kottier *et al.*, 1995). The high frequency of RNA recombination in CoVs is probably the result of the unique mechanism of CoV RNA synthesis, which involves discontinuous

transcription and polymerase jumping (Sawicki *et al.*, 2007). RNA recombination is an important mechanism in the natural evolution of CoVs. For example, new strains of IBV in poultry flocks are the result of natural recombination between different field and vaccine strains (Kusters *et al.*, 1989; Wang *et al.*, 1993). RNA recombination may also play a role in the evolution of different CoV species. Recombination may also explain the acquisition of the HE gene from an mRNA of Inf C virus by progenitor of the group 2 CoVs (Luytjes *et al.*, 1988). In addition, some strains of FCoV are likely to have arisen by recombination between FCoV and CCoV (Herrewegh *et al.*, 1998).

Since the emergence of SARS epidemic in 2002 - 2003, intensified scientific efforts have been focused on in the identification of the zoonotic origin of SARS-CoV and its transmission pathway to humans (Guan *et al.*, 2003; Song *et al.*, 2005; Tu *et al.*, 2004). Surveys of domestic and wild animals in southern China revealed that civet cats, raccoon dogs, and ferret badgers from wet markets were vectors for SARS-CoV outbreaks in 2002 and 2003 (Guan *et al.*, 2003; Song *et al.*, 2005; Tu *et al.*, 2004). However, phylogenetic relationships and low genetic similarities to human and civet SARS-CoVs indicate that the natural reservoir of these viruses has still not been determined. Furthermore, SARS-like CoVs have been identified from four bat species of *Rhinolophus* genus, in both northern and southern China (Lau *et al.*, 2005; Li *et al.*, 2005).

In a recent study Vijaykrishna *et al.* (2007), have estimated and compared the divergence times and population dynamics of all CoV groups, including new data from wild animals, to better understand the ecology and evolutionary pathways of CoVs. They concluded that:

(1) CoVs from bats have the most genetic diversity and are older than all CoVs recognized from any other animal species, while analysis of population dynamics revealed that CoV populations in bats have constant population size and that viruses from all other hosts show epidemic-like increases in population;

(2) bats are the natural hosts for all CoV lineages and that all CoVs recognized in other species were derived from viruses residing in bats. The ancestor of all of the established CoV lineages has not been identified, however, the data indicate that the hypothetical ancestor is likely from a bat;

(3) diverse coronaviruses are endemic in different bat species, with repeated introductions to other animals and occasional establishment in other species. In addition, divergence dates for SARS-CoV and SARS-like bat coronaviruses indicated that the SARS-CoV precursor may have circulated in an unidentified host before the 2003 epidemic;

(4) Analyses of HEL, S and N genes showed that all three CoV groups are monophyletic, with high statistical support;

(5) The most recent common ancestor dates for major CoV lineages, calculated for the HEL domain, was estimated to be 1586, approximately 500 years before the present, while the major divergence between group 1 and group 2 was estimated to occur in 1647;

(6) For group 1, there have been two interspecies transmissions from bats to other animals, including humans, and most interestingly between wild animals and domestic pets;

(7) no precursor from bats has been identified for group 2 viruses, with interspecies transmissions between domesticated animals, mice, and humans;

(8) group 3 coronaviruses hypothetically arose from a single introduction from an ancestral BtCoV. A possible transmission pathway from bats to poultry may have occurred via raptors, which are known to prey on bats.

1.2.6 Generic detection of CoVs

The CoV aetiology of SARS and the recent discoveries of the novel HCoVs NL63 and HKU1 (Fouchier *et al.*, 2004; van der Hoek *et al.*, 2004; Woo *et al.*, 2005) have triggered intensified efforts in virus identification and diagnostics. Generic reverse transcription (RT)-PCR assays with a very broad detection range are required, but few such assays are available. None of them has been previously validated in a diagnostic setting (Moes *et al.*, 2005; Stephensen *et al.*, 1999). The requirement for sequencing in order to achieve strain identification limits the applicability of generic PCR assays in general. Alternative techniques, such as mass spectrometry or complex fluorescent DNA microarrays, have been proposed (Sampath *et al.*, 2005), but these will often be too sophisticated for medical facilities.

We describe in this dissertation a simple and feasible approach to detecting the full spectrum of CoVs with diagnostic sensitivity, combining generic RT-PCR and low-cost, low-density (LCD) DNA microarrays, which can be read with the naked eye.

1.3 Viral respiratory infections after air travel

Acute respiratory infections are frequently experienced after air travel. Because most patients do not see a doctor, current knowledge about incidence and aetiology is imprecise and anecdotal (Leder & Newman, 2005). The very few systematic studies available suggested that up to 20% of passengers may develop respiratory infections within 1 week after air travel and that flight attendants show significantly higher incidence rates of respiratory infections than do control groups (Whelan *et al.*, 2003; Zitter *et al.*, 2002). Even though antibiotic therapy against common bacterial respiratory disease is often administered, laboratory data on the spectrum of causative agents are actually not available (File, 2003; Leder & Newman, 2005; Mangili & Gendreau, 2005).

There is a greater concern about the aircraft as a vector for spread of influenza strains than is in-flight transmission. The fact that influenza outbreaks worldwide have been affected by influenza strains imported by air travel is well established (Laurel *et al.*, 2001; Perz *et al.*, 2001; Sato *et al.*, 2000). However, only three studies of in-flight transmission of influenza have been reported (Klontz *et al.*, 1989; Marsden, 2003; Moser *et al.*, 1979).

The epidemic of SARS in 2003 involved a period of heightened awareness of respiratory infections after air travel. The global spread by air travellers and in-flight spread of SARS has been documented (Olsen *et al.*, 2003; WHO, 2003a; Wilder-Smith *et al.*, 2003). The disease is believed to usually be spread by large aerosolised droplets or by direct and indirect contact, but airborne or small droplet transmission better explains the distribution of SARS cases that has occurred on commercial airlines (Olsen *et al.*, 2003; WHO, 2003a). Evidence suggests that transmission of SARS during the Amoy Gardens outbreak in Hong Kong was a result of airborne spread via a viral plume (Yu *et al.*, 2004). A total of 40 flights have been investigated for carrying SARS-infected passengers (Olsen *et al.*, 2003; WHO, 2003a; Wilder-Smith *et al.*, 2003). Five of these flights have been associated with probable on-board transmission of SARS in 37 passengers (Olsen *et al.*, 2003; WHO, 2003a). No on-board transmissions have occurred since late March 2003, when the WHO issued specific guidelines for in-flight containment of SARS (WHO, 2003b; c).

1.3.1 Aetiology investigation of respiratory infections after air travel during SARS epidemic

During SARS epidemic, samples for laboratory testing were routinely obtained from patients with symptoms of respiratory infection after air travel, which provided a unique opportunity for studying their disease aetiologies. After initial characterization of the causative

agent of SARS, we acted as a diagnostic reference laboratory for the WHO (Drosten *et al.*, 2003). During the epidemic, we accepted samples for initial and confirmatory testing exclusively from those patients who fulfilled the WHO case definition of suspected or probable SARS. The definition was designed to be sensitive and thus to prevent any possible transmission, but its specificity was low. It covered most respiratory illnesses compatible with viral or atypical bacterial infection.

Respiratory agents are best diagnosed by direct assays and the most sensitive of them is PCR. Because of the rapid pace at which the technique evolves, we searched the literature for the most up-to-date PCR assays that cover the broadest possible range of genetic diversity for each respective agent. High sensitivities had to be clearly proved in studies. Where assays fulfilling these criteria were not available, we established sensitive real-time PCRs *de novo*. As described in this dissertation, these assays were used to determine a point prevalence of the full spectrum of respiratory viruses and atypical bacteria in SARS-compatible patients after air travel.

1.4 Viral gastroenteritis

Viral gastroenteritis is one of the most common illnesses in humans worldwide and it has a great impact on people, particularly in infants and the elderly. The mortality among children due to acute gastroenteritis is greater in developing than in developed countries (Villena *et al.*, 2003). Millions of deaths per year are associated with acute gastroenteritis, placing it in the top five causes of death worldwide, with most occurring in young children in non-industrialized countries. In industrialized countries, viral gastroenteritis is one of the most common illnesses in all age groups, and an important cause of morbidity (Chang *et al.*, 2003; Oh *et al.*, 2003).

Many different viruses have been recognized as important causes of gastroenteritis, including Noroviruses, Rotaviruses, Astroviruses, enteric AdVs, EVs, and more recently, human parechoviruses (HPeVs).

1.4.1 The gastroenteritis viruses

1.4.1.1 Noroviruses

Noroviruses are a genetically diverse group of non-enveloped particles with single-stranded, and non-segmented RNA genome. They belong to the family *Caliciviridae* and genus Norovirus. Noroviruses comprise four genogroups: genogroup I, II, and IV (GI, GII, and GIV) infect humans, and genotype III (GIII) only affect cattle. The groups are further classified according to amino acid sequences from the capsid gene, and the location where the virus was first described, for example GI/1 (Norwalk virus), GII/4 (Bristol virus).

Noroviruses are the most common cause of outbreaks of nonbacterial gastroenteritis and it is estimated that they are responsible for 68-80% of all outbreaks of gastroenteritis in industrialized countries. The emergence and detection of new strains often coincide with the increase in norovirus outbreaks (Lopman *et al.*, 2004). When these outbreaks occur, thousands of people can be infected, causing the closure of facilities and businesses (Khanna *et al.*, 2003; Nygard *et al.*, 2003). It is for this reason that noroviruses have since been described as being the most important cause of viral gastroenteritis worldwide.

1.4.1.2 Rotaviruses

Rotaviruses were first described as a human pathogen in 1973, and are now classified as a genus within the family *Reoviridae* (Bishop *et al.*, 1973). The rotavirus particles are non-enveloped, icosahedral structures. An inner and outer capsid gives a double layer, surrounding a core containing the viral genome. The double-stranded RNA consists of 11 segments. Rotaviruses are classified into seven serogroups (A-G) based upon the antigenic properties of

VP6, an inner capsid protein, of which groups A, B, and C are human pathogens. Within the groups, viruses are classified into serotypes based on differing outer capsid antigens. To date, 15 group A VP7 antigens (termed G types, G1-G15) and 20 VP4 antigens (termed P types, P1-P20) have been described (Anderson & Weber, 2004; Wilhelmi *et al.*, 2003). Rotaviruses are a common cause of infectious gastroenteritis, responsible for 600,000 to 875,000 deaths per year. The burden is most severe in the very young children and in developing countries. In children under 5 years, rotavirus is responsible for over 2 million hospitalizations and up to 600,000 deaths per year (Parashar *et al.*, 2003).

1.4.1.3 Astroviruses

The *Astroviridae* family is divided into two genera: Mamastrovirus, which encompasses human astroviruses and animal astroviruses; and Avastrovirus, the avian astroviruses. Astroviruses are non-enveloped particles, appearing as a five or six-pointed star by electron microscopy. The astrovirus genome is a single-stranded, non-segmented, positive sense RNA molecule.

Although not as important as other causes concerning disease severity, astroviruses probably cause more cases of gastroenteritis than noroviruses (Jakab *et al.*, 2003). Human astrovirus 1 remains the most prevalent serotype, although detection of others is increasing due to newer assays rather than the emergence of new types (Mendez-Toss *et al.*, 2004; Roman *et al.*, 2003). Mixed infection with rotavirus is often seen.

1.4.1.4 Enteric AdVs

At least 51 AdV serotypes (AdV 1-51) in six subgenera (A-F) have been described in humans (De Jong *et al.*, 1999). Although diarrhoea may be a feature of infection by other AdVs, for example AdV 3 and AdV 7, most AdV gastroenteritis is caused by the so-called enteric AdVs, AdV 40 and AdV 41, which are members of subgenus F. Up to 15% of diarrhoea is caused by AdVs (Cunliffe *et al.*, 2002; Simpson *et al.*, 2003).

1.5 Parechoviruses

Parechoviruses are small, non-enveloped, single-stranded RNA (ssRNA) viruses of positive sense. They are classified as one of the nine genera of the *Picornaviridae* family, which is a highly diverse family of important pathogens of humans and animals. Among the nine genera within the family, there are important pathogens such as the HRVs, EVs, hepatoviruses, and aphthoviruses.

The Parechovirus genus has been defined in the early 1990s (Hyypia *et al.*, 1992; Joki-Korpela & Hyypia, 2001). Within this genus, there are two species, Ljungan virus, isolated from bank voles (*Clethrionomys glareolus*) (Niklasson *et al.*, 1999) and Human parechovirus (HPeV), a frequent human pathogen. The HPeVs are classified based on serological characteristics and until 2004, only two HPeVs have been characterized. They were first classified as echovirus types 22 and 23, within the EV genus. However, they were reclassified as HPeVs 1 and 2 respectively, based on distinctive biological and molecular properties.

Improved molecular diagnostic methods and better clinical surveillance have recently led to the identification of novel parechoviruses. HPeV3 was identified in Japan in 2004 by molecular and serological methods (Ito *et al.*, 2004), and confirmed in North America and Europe short thereafter (Abed & Boivin, 2005; Benschop *et al.*, 2006b; Boivin *et al.*, 2005). HPeV4 was detected in the Netherlands by serological and molecular methods (Benschop *et al.*, 2006a), and confirmed in a phylogenetic study from stored isolated virus stocks, along with the characterization of HPeV5 by molecular genotyping (Al-Sunaidi *et al.*, 2007). HPeV6 was isolated in Japan in 2007 (Watanabe *et al.*, 2007) and recently confirmed in Europe (Baumgarte *et al.*, 2008; de Vries *et al.*, 2008).

For characterization of novel HPeVs, genetic analysis has been suggested for HPeV typing rather than sero-reactivity (Al-Sunaidi *et al.*, 2007; Benschop *et al.*, 2006b). Two arguments in favour of molecular typing are: first, once virus detection is now mainly being performed using PCR, molecular features become increasingly important in defining differences between viruses; and second, the potential of recombination among parechoviruses, which was recently detected for the first time (Al-Sunaidi *et al.*, 2007; Baumgarte *et al.*, 2008).

To date, a variable spectrum of symptoms caused by HPeVs have been described. The common symptoms are similar to that caused by some EVs, including mostly enteritis with diarrhoea, and respiratory disease (Baumgarte *et al.*, 2008; Benschop *et al.*, 2006b; Joki-Korpela & Hyypia, 1998; Stanway & Hyypia, 1999; Stanway *et al.*, 2000). Other symptoms also have been reported such as meningoencephalitis (Ehrnst & Eriksson, 1993; Figueroa *et al.*, 1989; Grist *et al.*, 1978; Koskiniemi *et al.*, 1989; Watanabe *et al.*, 2007), encephalomyelitis (Legay *et*

al., 2002), flaccid transient paralysis (Figueroa *et al.*, 1989; Ito *et al.*, 2004; Watanabe *et al.*, 2007), nosocomial infection (Ehrnst & Eriksson, 1996), neonatal sepsis (Boivin *et al.*, 2005), myocarditis (Ehrnst & Eriksson, 1993; Maller *et al.*, 1967; Russell & Bell, 1970), myositis (Grist *et al.*, 1978; Watanabe *et al.*, 2007), haemolytic uremic syndrome (O'Regan *et al.*, 1980), lymphadenopathy, hand-and-mouth disease, rash, fever of unknown origin, Influenza-like illness, and Reye's syndrome (Watanabe *et al.*, 2007).

In general, HPeV infection is most restricted to young children (Abed & Boivin, 2005; Baumgarte *et al.*, 2008; Benschop *et al.*, 2006b; Tauriainen *et al.*, 2007; Watanabe *et al.*, 2007). Infections with HPeV3 seem to peak at earlier age than HPeV1, may be associated with male gender, and seem to involve more sepsis-like illness and central nervous symptoms (Benschop *et al.*, 2006b). In studies conducted in different countries worldwide, almost the whole adult population had anti-HPeV antibodies (Joki-Korpela & Hyypia, 1998). A recent study on children in Finland showed that the median age of infection with HPeV1 was 18 months, with 20% being infected after the first year of life (Tauriainen *et al.*, 2007). In another study, 95% of newborn infants had maternal antibodies against HPeV1 (Joki-Korpela & Hyypia, 1998).

1.5.1 Structure and genome organization

HPeVs are small and naked icosahedral viruses with about 28 nm in diameter (Hyypia *et al.*, 1992; Ito *et al.*, 2004). The genome of HPeVs is a single positive-stranded RNA, which has an average length of 7,300 nucleotides. The viral RNA is infectious because it is translated on entry into the cell to produce all the viral proteins required for viral replication. The genome is packaged into an icosahedral capsid made up of multiple copies of each of the capsid proteins VP0, VP3 and VP1. The overall genomic organization is similar to that of other picornaviruses (figure 4), but there are differences in protein processing of capsid proteins. For example, during production of infectious picornavirus particles, the capsid protein VP0 is cleaved to produce the proteins VP4 and VP2, which is not observed in HPeVs (Hyypia *et al.*, 1992; Stanway *et al.*, 1994).

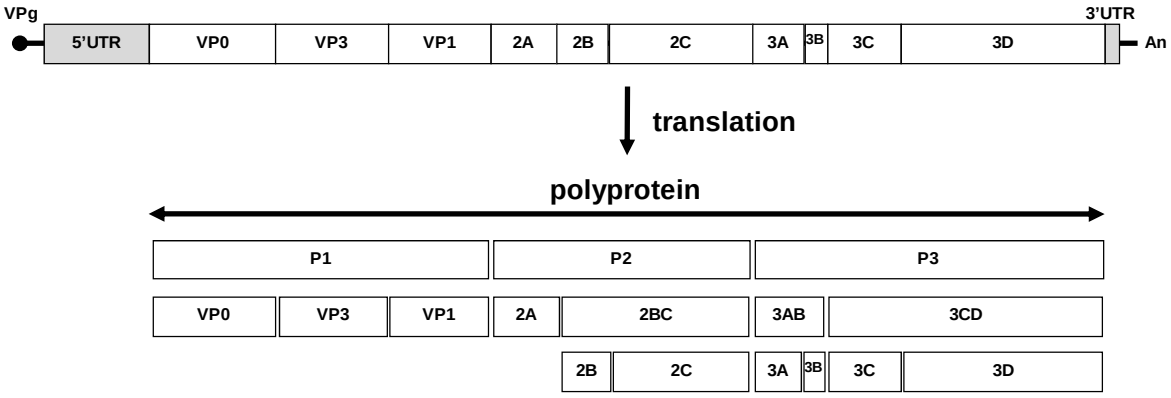

Figure 4: Genome organization of HPeVs and processing pattern of polyprotein.

The genome of HPeVs have a small virus-encoded protein of 20 amino acids, called VPg (virion protein, genome linked), covalently linked to the 5' end of the genome RNA. Four regions or domains could be found in HPeVs: a 5' UTR, followed by a single ORF, a 3' UTR and a poly(A) tract.

The 5' UTR of HPeVs are long (700 to 709 nucleotides) and highly structured by a complex folding pattern of secondary (stem-loops) and tertiary structure elements. This region of the genome contains sequences that control genome replication and translation. The 5' UTR contains the internal ribosome entry site (IRES) that directs translation of the mRNA by internal ribosome binding. All picornavirus IRES elements contain a polypyrimidine-rich tract 17 to 25 nucleotides upstream from a conserved AUG initiation codon, which is found in a Kozak context (ANNAUGG). The IRES stem-loop elements of HPeVs resemble those already described for cardioviruses and aphthoviruses, constituting a type II IRES (Duke *et al.*, 1992; Jang & Wimmer, 1990; Pilipenko *et al.*, 1990).

The 3' UTR of HPeVs are short (88 to 106 nucleotides) and also contains a secondary structure, that should be implicated in controlling viral RNA synthesis as already described for other picornaviruses (Jacobson *et al.*, 1993; Mirmomeni *et al.*, 1997). Analysis of the HPeV sequences have showed that much of the 3' UTR is made up of two highly conserved repeats of the sequence AUUAGAACACUAAUUUG arranged in tandem (Al-Sunaidi *et al.*, 2007). In addition, the 3' UTR of HPeVs is predicted to be composed of a single stem-loop (Abed & Boivin, 2005). The 3' UTR of HPeVs are followed by a poly(A) tract. Unlike most cellular mRNAs, the poly(A) tract of picornaviruses is part of the genome sequence and is not added by a separate poly(A) polymerase.

The genome of HPeVs are constituted of a large ORF that encodes a single polyprotein that is subsequently processed to form individual viral proteins. The polyprotein is cleavage during translation. Cleavage is carried out by virus-encoded proteinases to yield 10 cleavage

products. According to the picornavirus nomenclature, the polyprotein has been divided in three regions: P1, P2 and P3. The P1 region encodes the viral capsid proteins, whereas the P2 and P3 regions encode the non-structural proteins that are involved in protein processing and genome replication.

The viral polyprotein precursor is processed by intramolecular reactions carried out by proteinases that are themselves part of the polyprotein. These proteinases fold into enzymatically active structures while still part of the polyprotein. The polyprotein is first cleavage into the fragments P1, P2 and P3. P1 is subsequently cleavage into the capsid proteins VP0, VP3 and VP1. P2 are cleavage into the proteins 2A, 2B and 2C; and P3 into 3A, 3B, 3C and 3D.

The two major proteinases encoded by picornaviruses are the 2A ($2A^{pro}$) and 3C ($3C^{pro}$). However, the 2A protein of HPeVs does not have proteolytic activity (Schultheiss *et al.*, 1995). The primary cleavage at P1/P2, and P2/P3 junction regions and subsequent secondary cleavage are carried out by $3C^{pro}$ (Schultheiss *et al.*, 1995).

1.5.2 Non-structural proteins

Much of the function and structure of non-structural proteins of HPeVs are based on studies with other picornaviruses, most of them carried out with poliovirus. Thus, the description of non-structural proteins bellow are manly based in other picornaviruses.

1.5.2.1 2A protein

In EVs and HRVs, the 2A protein is a trypsin-like cysteine protease that is responsible for the *cis* cleavage at its own N-terminus between the VP1 capsid protein and 2A. In cardioviruses and aphthoviruses, 2A is involved in an unusual C-terminal proteolytic processing event between the 2A and 2B proteins (Donnelly *et al.*, 1997). The HPeV 2A protein lacks the motifs in the trypsin-like 2A of EV and HRVs associated with protein cleavage or with the unusual 2A/2B cleavage activity of cardioviruses and aphthoviruses. In addition, *in vitro* translation experiments have shown that HPeV1 2A protein has no autocatalytic proteolytic activity (Schultheiss *et al.*, 1995). During HPeV infection, the 2A protein is typically found diffusely in the cytoplasm and nucleus, with the bulk of the protein localized to the perinuclear area, where partial co-localization with viral RNA is detected (Krogerus *et al.*, 2007). This observation implies a possible association of 2A with the viral replication complex. It has been demonstrated that 2A protein is a ssRNA-binding protein with preferential binding to 3' UTR of HPeV1 RNA (Samuilova *et al.*, 2004). In addition, 2A protein has the ability to bind to the duplex containing 3' UTR(+) - 3' UTR(-). However, the affinity to the duplex is significantly

lower than to ssRNA. The localization of the protein in the infected cells and its ability to interact with RNA suggest that 2A participates in viral replication events (Krogerus *et al.*, 2007).

1.5.2.2 2B protein

The 2B protein is a small, hydrophobic membrane associated protein that is involved at early step of viral RNA synthesis. The C-terminus of 2B contains a hydrophobic region and a conserved putative amphipatic α-helix that appear to be crucial to the function of the protein and its association with membranes (van Kuppeveld *et al.*, 1996). Protein 2B has been called a viroporin, a protein that oligomerizes and inserts into membranes to create channels (Agirre *et al.*, 2002). The exact role of 2B in RNA synthesis, however, is not known. Synthesis of protein 2B leads to an inhibition of protein secretion from the Golgi apparatus (Doedens & Kirkegaard, 1995) and permeabilization of membranes (Agirre *et al.*, 2002; van kuppeveld *et al.*, 1997), which may play a role in release of the virion from cells. In transfected cells, 2B protein could be localized in the ER (Krogerus *et al.*, 2007). Protein 2B is also partly responsible for the proliferation of membranous vesicles in infected cells, which are the sites of viral RNA replication.

1.5.2.3 2C protein

Protein 2C is a highly conserved protein with biding regions to membrane, RNA, and nucleoside triphosphate (NTP) (Echeverri *et al.*, 1998; Rodriguez & Carrasco, 1993; 1995). The structure of 2C suggests that it contains three domains, with amphipatic α-helices at both the N and C-termini that mediate peripheral association with membranes (Kusov *et al.*, 1998; Teterina *et al.*, 1997), and a central region with NTP biding domains. Protein 2C shares amino acid homology with known RNA helicases, proteins encoded by most positive-stranded RNA viruses with large genomes. These enzymes are believed to be necessary to unwind double-stranded RNA structures that form during RNA replication. Protein 2C may have two functions during viral RNA synthesis: as a nucleoside-triphosphatase (NTPase), and directing replication complexes to cell membranes. Synthesis of 2C causes disassembly of the Golgi apparatus and ER, and formation of vesicular structures to those that constitute the replication complex. (Aldabe & Carrasco, 1995).

1.5.2.4 2BC protein

Much of protein 2BC, the precursor to 2B and 2C, remains uncleaved during infection, and its presence is critical to viral replication (Paul *et al.*, 1998). Synthesis of 2BC causes membrane permeabilization to a greater degree than 2B synthesis alone (Aldabe *et al.*, 1996),

and leads to the formation of vesicles that are more similar to those formed during viral infection than does protein 2C (Aldabe & Carrasco, 1995; Cho *et al.*, 1994). The C-terminus of 2B and the N-terminus of 2C may interact intramolecularly in 2BC, and protein cleavage may cause a conformational change that alters the properties of 2B and 2C individually (Barco & Carrasco, 1998). Larger 2BC-containing precursors are also required for certain steps of replication (Jurgens & Flanegan, 2003), although an intact 2C/3A junction is not strictly required (Paul *et al.*, 1998).

1.5.2.5 3AB protein

A strongly hydrophobic region in the C-terminus of 3A mediates the association of 3A and its precursor, 3AB, with membranes (Lama *et al.*, 1998). The solution structure of 3A demonstrates that it is a homodimer, with the dimer interface located in the central region of the protein (Strauss *et al.*, 2003). The requirement for 3A dimerization during infection had not been investigated. Protein 3B, also now as VPg, plays an important role in viral replication by acting as a protein primer for viral RNA synthesis. The *cis*-acting replication element (*cre*) serves as a template for the uridylylation of VPg. Protein 3AB is believed to anchor VPg in membranes for the priming of RNA synthesis.

1.5.2.6 3D protein

The 3D protein is the picornavirus RdRp, also known as $3D^{pol}$. The structure of this enzyme, like other polymerases such as DNA-dependent DNA polymerase, DNA-dependent RNA polymerase, and RNA-dependent DNA polymerase (reverse transcriptase), is characterized by analogy with a right hand, consisting of a palm, fingers, and thumb domains (Hansen *et al.*, 1997). The palm domain contains the active site of the enzyme. $3D^{pol}$ is a template and primer-dependent enzyme. The RNA polymerase $3D^{pol}$ is produced by cleavage of a precursor protein, the $3CD^{pro}$, which is highly active as a protease but has no polymerase activity. As observed in poliovirus, the $3D^{pol}$ is a cooperative ssRNA-binding protein and can unwind duplex RNA without the need for adenosine triphosphate hydrolysis (Cho *et al.*, 1993; Pata *et al.*, 1995).

1.5.3 Parechovirus replication cycle

1.5.3.1 Virus entry

The first step of virus infection is interaction with cellular receptors. A variety of cell surface molecules, such as members of the immunoglobulin superfamily, integrins, low density lipoprotein receptor family and decay accelerating factor (DAF) have been reported to be receptors for different picornaviruses (Evans & Almond, 1998). In addition to the primary

receptors, additional molecules may be involved in attachment, entry or uncoating. Sequence analysis of HPeVs indicated the presence of the arginine-glycine-aspartic acid (RGD) motif at the C-terminus of the VP1 capsid protein, except in HPeV3. RGD motif is known to participate frequently in cell-cell and cell-matrix interactions and to play a role in host cell recognition by several viruses trough interactions with cell surface integrins (Hynes, 1992; Ruoslahti & Pierschbacher, 1987). There is evidence that HPeV1 RGD motif interacts with αvβ1 and αvβ3 integrins receptors (Boonyakiat *et al.*, 2001; Joki-Korpela & Hyypia, 2001; Triantafilou *et al.*, 2000). The RGD motif is absent in HPeV3 strains indicating that HPeVs entry occur by a RGD-dependent and RGD-independent pathway (Abed & Boivin, 2005; Al-Sunaidi *et al.*, 2007; Ito *et al.*, 2004).

After attachment to the receptor, the virus penetrates into the host cell, and this process is followed by uncoating and release of the viral genome. HPeVs are believed to enter the host cell through the clathrin-dependent endocytic pathway (Joki-Korpela *et al.*, 2001). As observed for other picornaviruses, acidification within the endosomal vesicle causes a conformational change in the capsid, resulting in the dissociation of capsid subunits and release of the viral RNA. Exactly how the RNA subsequently passes from the vesicle into the cytoplasm is unknown.

1.5.3.2 Virus replication

Messenger RNA and genome synthesis of picornaviruses, like that of most RNA viruses, occurs in the cytoplasm of the cell. Once in the cytoplasm, the genome RNA must be translated by cellular ribosomes to produce viral proteins. All picornaviruses initiate protein synthesis by a cap-independent mechanism. In the case of HPeVs, host cell proteins bind to the type II IRES found in the 5' UTR and help to dock the 40S ribosomal subunit onto the RNA. In this model, the 40S subunit moves by translocation or scanning from its site of initial contact with the IRES to reach the AUG codon that initiates the ORF. The dynamic of this movement are not understood, but results in the assembly of 80S ribosomes at the initiator AUG codon and initiation of protein synthesis.

In picornaviruses, cap-independent translation also utilizes some of the same eukaryotic initiation factors that are involved in cap-dependent translation (Bedard & Semler, 2004; Belsham & Sonenberg, 1996). Eukaryotic initiation factors eIF-2α and eIF-4B are known to interact with picornavirus IRES elements, and eIF-2, eIF-4E, and eIF-4F all stimulate cap-independent translation. Binding of eIF-4F to IRES is specified by the central portion of its eIF-4G subunit, replacing the function of eIF-4E, which normally recognizes the cap and directs biding of this complex to capped mRNAs (Ali *et al.*, 2001).

Once viral proteins have been made, they proceed to replicate the viral RNA. Most of the proteins made from P2 and P3 regions are involved in picornavirus RNA replication (Bedard & Semler, 2004; Belsham & Sonenberg, 1996). Replication starts with full-length negative-strand RNA, using the initial positive-strand RNA as a template. This negative-strand RNA is used as a template for the synthesis of additional positive-strand RNAs via multi-stranded replicative intermediate structures. These multi-stranded structures result from highly efficient reinitiating of synthesis of positive strands on the template RNA, before the synthesis of previously initiated positive strands is completed. The number of positive strands synthesized in infected cells exceeds the number of negative strands 30 to 70 times (Novak & Kirkegaard, 1991).

Viral RNA is synthesized is association with membrane vesicles that are induced during virus infection. These vesicles are required for picornavirus RNA synthesis and may serve as a nucleation site for the formation of replication complexes (Cho *et al.*, 1994; Krogerus *et al.*, 2003). Association of viral RNA synthesis with membranous vesicles in the cytoplasm is a common characteristic found in a variety of positive-strand RNA viruses. These vesicles help to localize all the elements required for efficient viral RNA replication. Picornavirus proteins 2B, 2C, and 3A participate in the induction of these vesicles. Proteins 2B and 3A both mediate inhibition of cellular protein secretion, maybe by blocking membrane trafficking from the ER through the Golgi apparatus to the plasma membrane. This could contribute to the build-up of vesicles in infected cells. In addition to inducting vesicle formation, 2C is an RNA-binding protein implicated specifically in the initiation of negative-strand RNA synthesis. It may be involved in bringing genome RNAs in contact with the vesicles.

The proteins that actually direct viral RNA synthesis are derived from P3 region. Protein 3AB is anchored to the proliferated cytoplasmic vesicles via a hydrophobic region in the 3A portion. The protein VPg, which is covalently attached to the 5’ end of genome and antigenome viral RNAs, becomes covalently attached to uridine residues via a tyrosine side chain near its N-terminus. These protein-bound uridine residues hybridize to the poly(A) tract at the 3’ end of viral RNA, and serve as a primer to initiate RNA synthesis (Crawford & Baltimore, 1983; Paul *et al.*, 2000).

The $3D^{pol}$ adds nucleotides to the primer, copying the template RNA strand. Since the $3D^{pol}$ begins copying the positive-strand RNA at its 3’ end, it is possible that the 5’ and 3’ ends of the genome RNA are brought together at this point. Protein 2C is also believed to bind to the RNA and bring it to the vesicles where RNA synthesis takes place (Aldabe & Carrasco, 1995; Kusov *et al.*, 1998).

Other cellular proteins may also function in the replication complex. A *cre* sequence, putatively located in the VP0-encoding gene of HPeVs (Al-Sunaidi *et al.*, 2007), is required to initiate picornavirus RNA synthesis (Mason *et al.*, 2002). This element, along with IRES stem-loop I, may serve to distinguish viral RNA from the many molecules of polyadenylated cellular mRNA present in the cytoplasm (Andino *et al.*, 1993; Andino *et al.*, 1990).

1.5.3.3 Assembly of virus particles

Once sufficient genome RNAs are made, virions can be assembled. Following cleavage of P1, five protomers, each containing one molecule of VP0, VP3 and VP1, self-assemble to form a 14S pentamer structure (Palmenberg, 1982). There are two proposed pathways leading to the production of infectious picornavirus particles. One pathway have 12 of the pentamers coming together to form an 80S empty procapsid. A single molecule of newly synthesized viral RNA would then be threaded into the empty procapsid to form a provirion (Jacobson & Baltimore, 1968). This assembly model would seemingly require an opening in the empty capsid through which the RNA can enter. Examination of the high-resolution x-ray crystallographic structure of these particles does not provide evidence for such an opening (Basavappa *et al.*, 1994). The alternative assembly pathway proposes that the 14S pentamers first associate with viral genome RNA and subsequently combine with other 14S pentamers to form the provirion directly (Nugent & Kirkegaard, 1995). Whether the viral RNA is threaded into an empty capsid or is bound by 14S pentamers prior to particle assembly, only positive-strand RNAs with VPg protein linked to the 5' end are package (Novak & Kirkegaard, 1991). The final proteolytic cleavage of VP0 to produces VP4 and VP2 proteins to generate a fully infectious virion does not occur in HPeVs.

1.5.4 Diversity and recombination of picornaviruses

As with all other RNA viruses, picornaviruses have very high error rates, because of misincorporation during chain elongation and the lack of proofreading ability in RNA polymerases. With error frequencies as high as one misincorporation per 10^3 to 10^4 nucleotides, RNA virus populations exists as *quasispecies*, or mixtures of many different genome sequences (Domingo & Holland, 1997). It has been suggested that RNA viruses exist on the threshold of "error catastrophe", to maximize diversity and adaptability (Domingo & Holland, 1997). A moderate increase in error frequency would be expected to destroy the virus population. High RNA virus error rates are believed to be necessary to enable survival of the virus population under selective pressure. Consequently, viruses with less error-prone RNA polymerases should

be at a competitive disadvantage in complex environments such as an infected animal (Pfeiffer & Kirkegaard, 2005).

Recombination, the exchange of nucleotide sequences among different genome RNA molecules, was first discovered in cells infected with poliovirus, and was subsequently found to occur during infection with other positive and negative-stranded RNA viruses. The frequency of recombination, which is calculated by dividing the yield of recombinant virus by the sum of the yields of parental viruses, can be relatively high. In one study of poliovirus and foot-and-mouth disease virus, the recombination frequency was 0.9%, leading to the estimation that 10% to 20% of the viral genomes recombine in one growth cycle (Jarvis & Kirkegaard, 1992).

RNA recombination also occurs in natural infection. For example, intertypic recombinants among the three serotypes of Sabin poliovirus vaccine strain are readily isolated from the intestines of vaccines; some recombinants contain sequences from all three serotypes (Cammack *et al.*, 1988). The significance of these recombinants is unknown, but it has been suggested that such viruses are selected for their improved ability to replicate in the human alimentary tract compared with the parental viruses. Recombination in nature has also been demonstrated among nonpolio EVs (Simmonds & Welch, 2006) and recently for HPeVs (Al-Sunaidi *et al.*, 2007; Baumgarte *et al.*, 2008).

RNA recombination is believed to be coupled with the process of genome RNA replication: it occurs by template switching during negative strand synthesis, as first demonstrated in poliovirus-infected cells (Kirkegaard & Baltimore, 1986) and subsequently in cell-free extracts (Tang *et al.*, 1997). The RNA polymerase first copies the 3’ end of one parental positive strand, then switches templates and continues synthesis at the corresponding position on a second parental positive strand. Template switching in poliovirus-infected cells occurs predominantly during negative-strand synthesis because the concentration of positive-strand acceptors for template switching is 30 to 70 times higher than that of negative-strand acceptors. The cause of template switching is unknown, but it might be triggered by pausing of the polymerase during chain elongation.

1.5.5 Generic detection of HPeVs

As described above, improved molecular diagnostic methods and better clinical surveillance have recently led to the identification of novel HPeVs. In this dissertation, a diagnostic real-time reverse transcriptase PCR (RT-PCR) assay covering all known HPeVs was developed and used to test two large cohorts of patients with enteritis, as well as the genetic characterization of novel HPeV6 independently identified in this study, that was described very recently (Watanabe *et al.*, 2007). In addition, we describe a contemporary strain of an HPeV1 divergent from the historical prototype HPeV1 strain Harris.

1.6 Discovery of new viruses

Respiratory tract and enteric infections are responsible for considerable morbidity and mortality in humans and animals. These infections are caused by an enormous variety of microorganisms, with most of them belonging to a vast number of viral species of different virus families. In contrast, only a very small fraction of these viruses have been identified and characterized to date.

Since the early 1980s, viral genomes were recovered by molecular biological methods from clinical specimens and their sequences have subsequently been deduced. Often, this has been the first step in the characterization of previously unknown viruses, and has led to the expression of viral proteins for use in diagnostic assays, such as enzyme immunoassays (EIAs), and even to growth of the virus in cell culture from molecular clones.

The genomes of the viruses that have been discovered by molecular methods are diverse, and encompass almost the full range of RNA and DNA molecules, single or double stranded, segmented, linear or circular. The first methods were devised before the implementation of the polymerase chain reaction (PCR). They ranged from simple cloning methods to recombinant cDNA library constructions that have been used to characterize new viruses (Choo *et al.*, 1989; Cotmore & Tattersall, 1984).

With the advent of PCR, primers based on conserved regions of viral genomes, such as polymerase genes, were designed to amplify sequences from previously unknown viruses (Clewley *et al.*, 1998; Donehower *et al.*, 1990; Wichman & Van Den Bussche, 1992). The use of degenerated or consensus primers are generally only useful when searching for a specific sort of virus genome.

Several related PCR methods were developed in the 1990s that involved ligation of DNA adapter-linkers containing primer biding sites to DNA fragments, and sequence enrichment by amplification. Some of the related PCR methods successfully implemented to the discovery of new viruses are representational difference analysis (RDA), sequence-independent single primer amplification (SISPA) and virus-discovery-cDNA-AFLP method (VIDISCA). In addition to these methods, other molecular approaches have been proposed for the discovery of new viruses, including a diverse number of DNA microarray-based viral detection platforms and ultra-high-throughput sequencing. Table 2 shows some examples of viral genomes initially discovered or characterized by molecular methods.

A description of some viral discovery methods is summarized below.

Table 2: Some examples of viral genomes initially discovered or characterized by molecular methods.

Virus	Disease	Genome nucleic acid	Method used	Year	Reference
PARV B19	Fifth disease	ssDNA	Molecular cloning and DNA hybridization	1984	(Cotmore & Tattersall, 1984)
HCV	Non-A, non-B hepatitis	ssRNA	Transmission in primates, molecular cloning and immunoscreening	1989	(Choo *et al.*, 1989)
HEV	Non-A, non-B hepatitis	ssRNA	Transmission in primates, molecular cloning and sequence similarity	1990	(Reyes *et al.*, 1990; Tam *et al.*, 1991)
Retrovirus	Various	ssRNA and proviral dsDNA	Degenerate or consensus primer PCR and sequencing	1990	(Donehower *et al.*, 1990; Wichman & Van Den Bussche, 1992)
Rotavirus	Gastroenteritis	dsDNA segmented	SISPA	1992	(Lambden *et al.*, 1992)
Astrovirus	Gastroenteritis	ssRNA	SISPA and Immunescreening	1993	(Matsui *et al.*, 1993)
HHV-8, KSHV	Kaposi's sarcoma, B cell lymphomas	dsDNA	RDA	1994	(Chang *et al.*, 1994)
GBV-A GBV-B	Not know	ssRNA	Transmission in primates and RDA	1995	(Simons *et al.*, 1995)
GBV-C	Not know	ssRNA	SISPA, molecular cloning and sequencing	1996	(Linnen *et al.*, 1996)
TTV	Not know	circular ssDNA	RDA	1997	(Nishizawa *et al.*, 1997)
SIVdrl	Immunodeficiency	ssRNA and proviral dsDNA	Degenerate primer PCR and sequencing	1998	(Clewley *et al.*, 1998)
hMPV	Respiratory disease	ssRNA	Cell culture, random PCR and sequencing	2001	(van den Hoogen *et al.*, 2001)
Bovine parvovirus	Not know	ssDNA	SISPA, molecular cloning and sequencing	2001	(Allander *et al.*, 2001)
HCoV-NL63	Respiratory disease	ssRNA	Cell culture VIDISCA	2004	(van der Hoek *et al.*, 2004)
HCoV-HKU1	Respiratory disease	ssRNA	Consensus primer PCR and sequencing	2005	(Woo *et al.*, 2005)
HBoV	Respiratory disease? Gastroenteritis?	ssDNA	Random PCR, cloning and sequencing	2005	(Allander *et al.*, 2005)
PARV4, TT-like viruses	Not know	ssDNA	SISPA, molecular cloning and sequencing	2005	(Jones *et al.*, 2005)
KI polyomavirus	Not know	circular dsDNA	SISPA, molecular cloning and sequencing	2007	(Allander *et al.*, 2007)
WU polyomavirus	Not know	circular dsDNA	pan-viral microarray-based detection and sequencing	2007	(Gaynor *et al.*, 2007)
HPV102 and 106	Not know	Circular dsDNA	Consensus primer PCR, molecular cloning and sequencing	2007	(Chen *et al.*, 2007)
HPeV1 strain BNI-788St	Gastroenteritis	ssRNA	Cell culture VIDISCA	2008	(Baumgarte *et al.*, 2008)
HPeV5 isolate 2000-1108-4	late-onset sepsis neonatorum	ssRNA	Cell culture VIDISCA	2008	(de Vries *et al.*, 2008)
New Arenavirus	Fatal transplant-associated diseases	ssRNA	Ultra-high-throughput sequencing and cell culture	2008	(Palacios *et al.*, 2008)

Abbreviations: HBoV, human boca virus; HCV, human hepatitis C virus; HEV, human hepatitis E virus; HHV, human herpes virus; hMPV, human metapneumovirus, KSHV, Kaposi's sarcoma-associated herpesvirus; PARV, parvovirus; SIV, simian immunodeficiency virus; TTV, Torque teno virus; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; ssRNA, single-stranded RNA.

1.6.1 Virus discovery methods

1.6.1.1 Sequence-independent, single-primer amplification (SISPA)

SISPA was introduced by Reyes & Kim (1991), as a technique to identify viral nucleic acid of unknown sequence present at low concentration. Since then, many variations have been described and used to identify new viruses (Table 2). The SISPA procedure consists initially in pre-treat the sample by a combination of methods to aid purification of the viral nucleic acid. These methods include filtration of the sample to remove host cells and mitochondria, removal of extracellular DNA by DNase treatment and the isolation of viral particles by centrifugation. Finally, the remaining viral nucleic acid can be further purified by silica particle and guanidinium extraction methods. The sample may need to be divided at this stage if it is not known whether the viral nucleic acid is either RNA or DNA. If ssDNA is present, a second strand of DNA needs to be synthesized; if the sample contains RNA, cDNA needs to be generated. Once dsDNA has been produced, it can be digested by restriction enzymes. This enables the ligation of adapter-linkers with the relevant overhangs. Alternatively, adapter-linkers with blunt ends can be ligated onto the termini of undigested dsDNA. To amplify the unknown viral nucleic acid, a primer can be used that is complementary to the known sequence of the adapter-linker. A further selective round of amplification can be performed using a primer with an additional nucleotide at the 3' end. Consequently, in theory, only a quarter of the DNA fragments will be amplified. Finally, the sequence of the amplified viral nucleic acid can be characterised by downstream cloning and sequencing.

1.6.1.2 Virus-discovery-cDNA-AFLP (VIDISCA)

VIDISCA was introduced to successfully identify a previously unknown CoV, the HCoV-NL63 (van der Hoek *et al.*, 2004). This method is based on the same principles as SISPA but uses two primers rather than one in the PCR amplification step, as is done in the amplified fragment length polymorphism (AFLP) technique (Bachem *et al.*, 1996). In this method, the starting template is converted to dsDNA and then digested with two restriction enzymes. The two anchors or adapter-linkers are prepared by annealing together the top and the bottom strand oligonucleotides, and they are then ligated onto the digested dsDNA fragments. The first PCR amplification is done with the two primers, each of them complementary to an adapter-linker, and the second PCR is performed with the same primer sequences except that they have an additional 3'-base (N). It results in 16 different possible primer combinations. The second PCR may be done with an input of the first reaction, using touchdown conditions. The use of two

adapters and primers, and also a nested PCR step, makes VIDISCA more sensitive and specific than SISPA.

1.6.1.3 Representational difference analysis (RDA)

The RDA method combines subtractive hybridisation with gene amplification to detect differences between two similar clinical samples (Lisitsyn & Wigler, 1993). In virology, the most likely samples to be compared are pre- and post-infection samples. The DNA isolated from the pre-infection sample acts as the "driver" and is compared with the DNA isolated from the post-infection sample, the "tester". The two DNA samples are hybridised together to reduce common sequences, leaving mainly viral sequences for downstream analysis.

For RDA, the driver is used at a higher concentration that the tester, to drive the reaction. The DNA samples are digested with a restriction enzyme. A linker-adapter (to provide the primer sequence in further PCR) is added only to the tester DNA digest. The two DNA populations are mixed, heated and annealed to form three kinds of molecules: tester-tester sequences; hybrids of tester and driver; and driver-driver sequences. As there is an excess of driver DNA, the tester-tester molecules should be enriched for pathogen sequences because the tester non-pathogen sequences will hybridise to the corresponding driver sequences. The ends of the re-annealed DNA are filled in and then amplified by PCR with a primer specific for the linker-adapter sequence. The tester-tester molecules with the pathogen sequence should be preferentially and exponentially amplified. Nuclease digestion is used to remove unwanted ssDNA and further PCR is performed. More rounds of this procedure may be carried out by combining the resultant pathogen-enriched difference amplicons with an excess of driver DNA restriction enzyme fragments. RDA has been used to identify a HHV-8 (KSHV), GBV-A and GBV-B and a TTV (Table 2).

1.6.1.4 Random PCR

The method and its applicability were demonstrated by detection of a herpes simplex virus type 1 (HSV-1) in a mouthwash sample from a patient with chronic fatigue syndrome (Stang *et al.,* 2005). Random PCR, a technique similar to SISPA, can be used to amplify both DNA and RNA viral genomes. Samples are prepared as described for SISPA. A primer with a unique 5' end nucleotide universal sequence, containing restriction enzyme sites for subsequent cloning, followed by a degenerate hexa- or heptamer sequence at the 3' end is used in a PCR reaction to amplify viral DNA. Usually a lower annealing temperature or touchdown conditions are used. The degenerate part of the primer anneals to complementary DNA sequences that occur

stochastically throughout the viral genome. The primer is extended using T4 DNA polymerase. Then the generated double-stranded sequences are separated by denaturation and complementary sequences hybridise to form finite portions of dsDNA with the first primer present at each end. A primer representing only the unique sequence of the first primer is used for subsequent amplification of the fragments of the target.

1.6.1.5 Degenerate PCR

Degenerate PCR primers contain mixed combinations of nucleotides at defined residues of the primer. These mixed bases correspond to regions of variation that occur between related viruses within a virus family. Mixed base (degenerate) oligonucleotides are synthesized by the reaction of equimolar amounts of the appropriate nucleotide with a synthetic DNA chain. For example, to make the sequence 5'..G(C or T)A..3', an A residue is added, followed by equal amounts of C and T, followed by G. This produces a population of two distinct oligonucleotide primers (GCA and GTA). Degenerate PCR primers can be designed by comparing different virus genomes. Greater degeneracy can be obtained by designing primers based on conserved viral amino acid sequences following back-translation, rather than by simply comparing viral nucleic acids.

Designing a complete PCR protocol with degenerate primers involves some strategic choices. The overall aim is to achieve a balance between covering all possible variants (i.e. primers with high degeneracy) and making the number of different primers unsuitably large. At high levels of degeneracy, only a very small proportion of the pool of primers are able to prime DNA synthesis, whereas a large proportion of the remaining primers will still be able to anneal but, because of sequence mismatches, will be refractory to PCR extension. It is possible to reduce degeneracy further by using deoxyriboinosine as an alternative nucleotide in the PCR primer. Overall degeneracy is less important at the 5' end of the primer than at the discriminatory 3' end. Finally, the 3' nucleoside of the primer should not be degenerate.

Degenerate PCR have been used successfully to identify new human viruses, such as hepatitis G virus (GBV-C), sin nombre virus (SNV) and the human retrovirus 5 (HRV-5) (Griffiths *et al.,* 1997; Linnen *et al.*, 1996; Nichol *et al.*, 1993).

1.6.1.6 Consensus-degenerated hybrid oligonucleotide primer (CODEHOP) PCR

CODEHOP PCR is an elaborated variant of degenerate PCR (Rose *et al.*, 1998). This method was developed to identify and characterize distantly related gene sequences based on consensus-degenerated hybrid oligonucleotide primers (CODEHOPs). CODEHOPs are designed

from amino acid sequence motifs that are highly conserved within members of a gene family, and are used in PCR amplification to identify unknown related family members. Each CODEHOP consists of a pool of primers where each primer contains one of the possible coding sequences across a 3-4 amino acid motif at the 3' end (degenerate core). Each primer also contains a longer sequence derived from a consensus of the possible coding sequences 5' to the core motif (consensus clamp). Thus, each primer has a different 3' sequence coding for the amino acid motif and the same 5' consensus sequence. The 5' consensus clamp stabilizes hybridization of the 3' degenerate core with the target DNA template during the initial PCR amplification reaction.

Using CODEHOPs derived from conserved motifs within retroviral reverse transcriptases, a diverse family of retroviral elements in the human genome have been identified, as well as a novel endogenous pig retrovirus, and a new retrovirus in Talapoin monkeys (Osterhaus *et al.*, 1999; Rose *et al.*, 1998; Wilson *et al.*, 1998). In addition, three homologs of the Kaposi's sarcoma-associated herpesvirus have been discovered in macaques by CODEHOP method (Rose *et al.*, 1997; Schultz *et al.*, 2000).

1.6.1.7 Microarray-based detection

In the last years, many DNA microarray methods have been described to detect viral specific groups or to identify unknown viruses.

Wang *et al.* (2002), have designed a DNA microarray for viral discovery and sequence recovery, using available sequence data from more than 140 sequenced viral genomes. The DNA microarray consisted of 70-mer long oligonucleotides with the potential to simultaneously detect hundreds of viruses, including essentially all respiratory tract viruses. Afterwards, a second generation of the DNA microarray heve been elaborated (Wang *et al.*, 2003). They included new 70-mer oligonucleotides derived from every fully sequenced reference viral genome in GenBank until 15 August 2002. To maximise the probability of detecting unknown and unsequenced members of existing families by cross-hybridization to these array elements, the most highly conserved 70-mers from each virus were selected. On average, ten 70-mers were selected for each virus, totalling approximately 10,000 oligonucleotides from about 1,000 viruses. In addition, a random PCR amplification strategy was developed in order to amplify viral nucleic acid to be identified using the microarray (Wang *et al.*, 2002). This pan-viral array was used as part of the global effort to identify the novel SARS-CoV associated with SARS (Wang *et al.*, 2003).

Boriskin *et al.* (2004), have constructed a low-density, high-resolution diagnostic DNA microarray comprising 38 gene targets for 13 viral causes of meningitis and encephalitis. It has been used for the detection of multiplex PCR-amplified viruses in cerebrospinal fluid (CSF) and non-CSF specimens, and was tested using 41 clinical specimens. The clinical sensitivity, specificity and negative and positive predictive values of the assay were determined to be 93%, 100%, 100%, and 83%, respectively, when the results were compared to those of the single virus PCR. However, it was reported that the interpretation of a negative result is difficult because it is affected by assay sensitivity, low viral genome number and sample-specific inhibition.

Lin *et al.* (2004), have combined a fluorogenic real-time multiplex PCR for detection of respiratory disease-associated AdVs combined with an oligonucleotide microarray for the rapid identification of common serotypes. The microarray was composed of two long unique oligonucleotide probes, between 60 and 72 nucleotides, specific for three AdV genes of each serotype. Microarray-based sensitivity assessments revealed lower detection limits, between 1 and 100 genomic copies, for AdV of cell culture lysates, clinical nasal and throat swabs, and purified DNA from clinical samples.

Korimbocus *et al.* (2005), have reported the simultaneous detection of herpesviruses, EVs and flaviviruses using amplification by PCR and detection of amplified products using a high-density DNA microarray. This method was successfully used to identify herpesviruses, namely, herpes simplex virus type 1 (HSV-1), HSV-2, and CMV; all serotypes of human EVs; and five flaviviruses (West Nile virus, dengue viruses, and Langat virus)

Conejero-Goldberg *et al.* (2005), have developed an array-based pathogen chip for the detection of viral RNA or DNA relevant to pathologies of the central nervous system. Open reading frames (ORFs) with highly conserved and heterogenic sequence regions within viral families were used to design a total of 715 unique oligonucleotides (60-mers), representing approximately 100 pathogens. Viral genes representing different stages of pathogen infection were also included on the chip to potentially define the stage of the viral infection. The array was tested with six post-mortem brain samples from which human CMV and Herpes simplex virus (HSV) were detected.

Lin *et al.* (2006), have demonstrated an approach that combined the use of a custom-designed Affymetrix resequencing Respiratory Pathogen Microarray (RPM v.1) with methods for microbial nucleic acid enrichment, random nucleic acid amplification, and automated sequence similarity searching for broad-spectrum respiratory pathogen surveillance. This method was tested with clinical samples obtained from patients presenting AdV or influenza virus-

induced febrile respiratory illness, and was able to identify the pathogens correctly at species and strain levels, with supported statistical interpretation and clinically relevant sensitivity levels.

Liu *et al.* (2007b), have developed a novel PCR assay called microarray-in-a-tube system, which integrates multiple PCR processes and DNA microarrays for multiple virus detection. A 5X5 oligonucleotide microarray for detecting four viruses causing human acute respiratory tract infection (SARS-CoV, Inf A and B viruses, and enterovirus [EV]) with inner controls was arranged on the inner surface of a specially designed Eppendorf cap with a flat, optically transparent window. The whole detection processes are performed in the encapsulated system without opening the cap. This method consists in a one-step RT-PCR, followed by hybridization of amplified products on the microarray. The hybridization signals are obtained with an ordinary fluorescent microscope. The sensitivity of the system for virus detection was estimated in 10^2 genome-equivalent copies per µl.

Kistler *et al.* (2007), have developed a DNA microarray-based viral detection platform named Virochip, to characterize the viral diversity in respiratory tract infections in adults with and without asthma. The Virochip microarray contains oligonucleotides for the most conserved sequences of all known viruses of humans, animals, plants, and microbes. In addition, this method is potentially able to detect new members of known virus families by cross-hybridization. In this method, the samples are randomly amplified, labelled, and then hybridized to the Virochip microarray. In the same study, Virochip could detect viruses in a higher proportion of samples (65%) than did culture isolation (17%) while exhibiting high concordance (98%) with and comparable sensitivity (97%) and specificity (98%) to pathogen-specific polymerase chain reaction. The method was also able to detect 5 divergent isolates of HRV that formed a distinct genetic subgroup.

1.6.1.8 Ultra-high-throughput sequencing

The ultra-high-throughput sequencing method was introduced by Margulies *et al.* (2005). This method is applied to produce millions of raw bases per hour on a single instrument. This method uses an emulsion-based method to isolate and amplify DNA fragments *in vitro* and then performs pyrophosphate-based-sequencing in picolitre-sized wells. In this method, DNA is isolated, fragmented, ligated to adapters and separated into single strands. Fragments are bound to beads under conditions that favour one fragment per bead. The beads are captured in the droplets of a PCR-reaction-mixture-in-oil emulsion and PCR amplification occurs within each droplet, resulting in beads each carrying ten million copies of a unique DNA template. The emulsion is broken, the DNA strands are denatured, and beads carrying ssDNA clones are

deposited into wells of a fibre-optic slide. Smaller beads carrying immobilized enzymes required for pyrophosphate sequencing are deposited into each well. The enzymes convert the chemicals generated during nucleotide incorporation into a chemilluminescent signal that can be detected by a CCD camera. Finally, the generated sequences are assembled in single contigs with the use of a proprietary software. This methodology has been successfully used to the identify a new Arenavirus transmitted through solid-organ transplantation (Palacios *et al.*, 2008).

1.7 The unnatural nucleotides isocytosine and isoguanosine

The advance of organic chemistry in the 1980s has made possible the synthesis of molecules that expand the DNA chemistry, including additional base pairs. In natural DNA, two complementary strands are joined by a sequence of Watson-Crick base pairs. These obey two rules of complementarity: size (large purines pair with smaller pyrimidines) and hydrogen binding (hydrogen bond donors from one nucleobase pair with hydrogen bond acceptors from the other). The creation of an artificially expanded genetic information system (AEGIS) has lead to the synthesis of new base pairs that are not found anywhere in nature. The properties of the synthetic or unnatural nucleotides isocytosine (isoC) and isoguanosine (isoG) have been extensively reported (Johnson *et al.*, 2004b; Moser & Prudent, 2003; Roberts *et al.*, 1995; Switzer *et al.*, 1989; Switzer *et al.*, 1993). IsoC and isoG are isomeric forms of the nucleobases cytosine and guanosine, respectively. They differ from each other in the form that the amino and carboxyl group are transposed. Consequently, isoC and isoG have altered hydrogen-binding directionality in relation to cytosine and guanosine, respectively (figure 5).

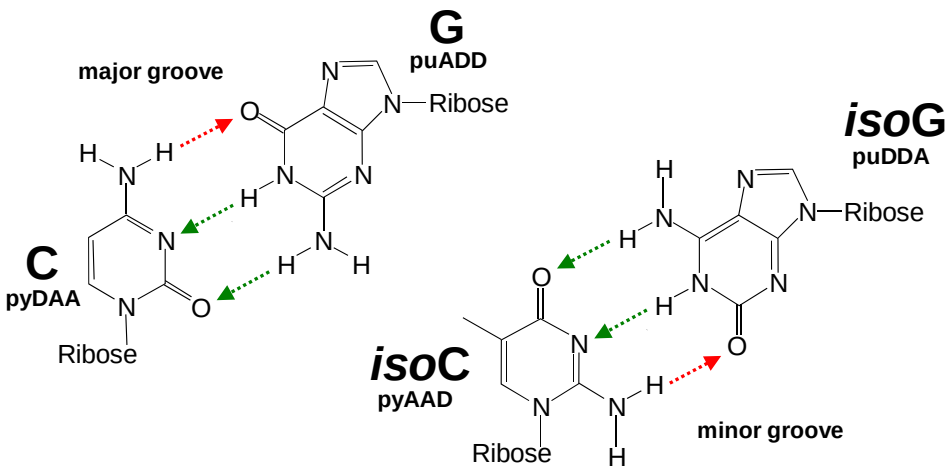

Figure 5: structural comparison of G-C and isoG-isoC Watson-Crick base pairs. Nucleobase pairs are complementary in both size (pyrimidine [py] opposite to purine [pu]) and hydrogen bond donors and acceptors. The upper case letters following py and pu designation indicate the hydrogen binding pattern of acceptor (A) and donor (D) groups moving from the major to the minor groove.

IsoC and isoG can form a Watson-Crick base pair with a standard geometry but with a hydrogen-binding pattern distinct from those occurring in the natural A-T(U) and G-C pairs (Roberts *et al.*, 1995; Switzer *et al.*, 1989). As the natural base pair G-C, isoG and isoC are base paired by three hydrogen bonds. The unnatural isoG-isoC pair was found to be as stable as a C-G Watson-Crick pair (Roberts *et al.*, 1995). *In vitro* polymerase experiments have demonstrated that both a Klenow fragment of DNA polymerase I and T7 RNA polymerase incorporate isoG and isoC into a growing oligonucleotide duplex under the direction of its counterpart in a template (Switzer *et al.*, 1989). Furthermore, the same authors have demonstrated that AMV reverse transcriptase incorporates isoC opposite to isoG in a template, but T4 DNA polymerase did not (Switzer *et al.*, 1993). It has been demonstrated that three of the major enzymes used in molecular diagnostics (polymerases, nucleases, and ligases) can recognize and carry out their specific activities on substrates that contain multiple isoG-isoC pairs (Moser & Prudent, 2003). In this study, a DNA duplex model system for complete DNA break repair, consisting of multiple isoG and isoC bases, was created. In brief, *Taq* DNA polymerase was able to recognize and remove a mispaired thymine with isoG and replaced it with an isoC. In addition, T4 DNA ligase recognized the nick formed by *Taq* DNA polymerase between isoC and isoG base pairs and covalently attached the two strands. Other study have also demonstrated that DNA targets containing both isoG and isoC could be correctly replicated and amplified with high fidelity under standard PCR amplification conditions (Johnson *et al.*, 2004b).

IsoC and isoG were first introduced to molecular diagnostics in 2002 in the form branched DNA (bDNA) diagnostics tools (Bayer Diagnostics) that quantitate the levels of human immunodeficiency, hepatitis B and C viruses in blood. Since then, isoC and isoG had been applied to different genetic analysis systems (Johnson *et al.*, 2004a; Marshall *et al.*, 2007; Moser *et al.*, 2006; Moser *et al.*, 2005; Nolte *et al.*, 2007; Prudent, 2006; Sherrill *et al.*, 2004).

1.7.1 Virus discovery methods using isoC and isoG

In this dissertation, we propose the establishment of a simple and feasible method for virus discovery of single-stranded, positive strain RNA viruses with the use of the unnatural nucleotides isoC and isoG. These nucleotides were used as constituents of the 5’ end of primers for PCR amplification, and chikungunya virus was used as a model.

1.8 Objectives of this dissertation

The objectives of this dissertation are:

Investigate the spectrum of respiratory known viruses and three atypical pneumonia-producing bacteria, especially in adults, in samples collected in Germany from patients with symptoms of respiratory infection after air travel, during the SARS epidemic in 2003;

Develop a simple and feasible approach to the detection of the full spectrum of coronaviruses with diagnostic sensitivity, combining generic RT-PCR and low-cost, low-density (LCD) DNA microarrays for species identification, which can be easily read and interpreted;

Implement and use the VIDISCA method to identify untypeable viruses growing in cell culture obtained from patients during a routine diagnostic work on enteritis in a municipal health service, the Institute of Hygiene and the Environment of Hamburg, Germany;

Develop the first broad-range real-time RT-PCR assay for all known HPeVs and investigate their prevalence in two large cohorts of patients with enteritis in Hamburg, Germany;

Establish a new method for discovery of viruses with single-stranded, positive strand RNA genomes, employing the use of the unnatural nucleotides isoC and isoG as constituents of the 5’ end of PCR primers, using chikungunya virus as a model organism.

2.0 Material and Methods

2.1 Spectrum of viruses and atypical bacteria in intercontinental air travellers with symptoms of acute respiratory infection

2.1.1 Clinical samples

A total of 214 respiratory samples from 172 patients were available. All samples had been sent to our institute for SARS-CoV testing. All patients fulfilled the WHO case definition of suspected or probable SARS as utilized during the 2003 epidemic (Anonymous, 2003) which required a combination of fever and lower-respiratory-tract symptoms such as cough or difficulty breathing, plus either a stay in an affected area during the preceding 10 days or close contact with suspected patients. Probable cases additionally had radiological evidence of respiratory distress syndrome without further reasons. All patients had travelled on long-haul flights into Germany from Canada or endemic Asian countries; this was in spring of 2003.

2.1.2 RNA extraction

RNA extraction protocol was performed according to Drosten *et al.* (2003). Sputum samples were pre-treated by shaking for 30 minutes with an equal volume of Sputum Lysis Buffer (10 g/l N-acetyl cysteine, 0.9% sodium chloride). The resulting homogenate, and resuspended swab samples (in 1 ml of phosphate-buffered saline) were extracted with QIAamp Viral RNA Mini Kit (QIAGEN, Hilden, Germany). DNA was also purified if present in the sample. For each sample, 560 µl of Lysis Buffer AVL containing carrier RNA was added into a 1.5 ml microcentrifuge tube and was then mixed with 140 µl of sample by vortexing for 15 seconds. The mixture was incubated at room temperature for 10 minutes. After lysis incubation, 560 µl of absolute ethanol was added; then the whole volume was mixed by vortexing for 15 seconds. Next, 630 µl of the solution was applied to the QIAamp Mini spin column (in a 2 ml collection tube) and was then centrifuged at 8,000 rpm for 1 minute. The spin column was placed into a clean collection tube; then, the rest of sample was passed through the column as described above. Once the spin column was placed in a new collection tube, it was washed with 500 µl of Buffer AW1 by centrifugation at 8,000 rpm for 1 minute. The spin column was washed a second time with 500 µl of Buffer AW2 by centrifugation at 13,000 rpm for 3 min. The dried spin column was placed in a clean 1.5 ml microcentrifuge tube; then 60 µl of Buffer AVE (0.04% sodium azide in RNase-free water) at 80°C was applied in the centre of the membrane column for RNA elution. The spin column was centrifuged at 8,000 rpm for 1 minute to collect the eluted RNA.

2.1.3 Panel of molecular diagnostic assays

Assays included human AdVs (Claas *et al.*, 2005); HRVs (Deffernez *et al.*, 2004); HBoV (Allander *et al.*, 2005); as well as formulations of new assays established for the present study for HCoV-229E, HCoV-NL63, HCoV-OC43, and Inf A. The combined assay, the Hexaplex Plus (Prodesse, USA), was also used for detection of Inf A and B viruses, human PIV 1, 2 and 3, RSV, and hMPV. The commercial system was chosen because several published respiratory multiplex assays gave instable results in our laboratory. We also tested for the presence of the three pneumonia-producing bacteria *Chlamydophila pneumoniae*, *Legionella ssp.*, and *Mycoplasma pneumoniae* (Raggam *et al.*, 2005). Detailed protocols you will find below.

2.1.3.1 Human AdV real-time PCR

This protocol was designed to amplify a region of the highly conserved hexon gene DNA of all relevant AdV serotypes.

The reaction was carried out with the HotStarTaq Master Mix Kit (QIAGEN, Hilden, Germany) in a 25 µl volume reaction containing Nuclease-free water, 1X HotStarTaq Master Mix (containing 1X PCR Buffer, 1.5 mM of $MgCl_2$, 200 µM of each deoxyribonucleoside triphosphate [dNTP] and 2,5 U of HotStarTaq DNA Polymerase), additional 3.0 mM $MgCl_2$, 300 nM of primes 372ADVs (5'-CAT GAC TTT TGA GGT GGA TC-3'), 346ADVas (5'-CCG GCC GAG AAG GGT GTG CGC AGG TA-3'), 423ADV4s (5'-CATG AAT TTC GAA GTC GAC C-3'), 424ADV31s (5'-TAT GAC ATT TGA AGT TGA CC-3'), 300 nM of probe 569ADV-MGB-FAM (FAM-AGC CCA CCC TKC TTT AT- MGBNFQ), and 5 µl of nucleic acid extract. Thermal cycling conditions in a ABI Prism 7000 Sequence Detection System (SDS) (Applied Biosystems, Germany) involved 15 minutes at 90°C for enzyme activation, followed by 50 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 72°C for 30 seconds.

2.1.3.2 Rhinovirus two-steps real-time RT-PCR

This assay was designed to amplify the most conserved region in the 5' UTR of HRVs.

Reverse transcription was performed using SuperScript II Reverse Transcriptase (Invitrogen, Karlsruhe, Germany) in a 20 µl volume reaction mixture containing RNase-free water, 1X First-Strand Buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 10 mM DTT, 500 µM of each dNTP (Applied Biosystems, Germany), 250 ng of random hexamer primers ($P[dN]_6$) (Roche, Germany), 20 U of RNaseOUT (Invitrogen, Karlsruhe, Germany), 40 U of SuperScript II RT and 5 µl of extracted RNA. The mixture was submitted to 25°C for 10 minutes, 42°C for 50 minutes, and 95°C for 5 minutes. The tubes were then placed on ice and immediately processed for PCR reaction, which was carried out with the TaqMan Universal PCR

Master Mix, No AmpErase UNG (Applied Biosystems, Germany) reagent in a 25 µl volume reaction containing Nuclease-free water, 1X TaqMan Universal PCR Master Mix (buffer containing dNTPs with dUTP, ROX passive reference, $MgCl_2$ and AmpliTaq Gold DNA polymerase), 900 nM of primer HRV-For (5'-GCA CTT CTG TTT CCC C-3'), 600 nM of primer HRV-Rev (5'-GGC AGC CAC GCA GGC T-3'), 200 nM of probe HRV-A (FAM-AGC CTC ATC TGC CAG GTC TA-TAMRA), 200 nM of probe HRV-B (FAM-AGC CTC ATC GAC CAA ACT A-TAMRA), and 5 µl of cDNA. Thermal cycling conditions in a ABI Prism 7000 SDS (Applied Biosystems, Germany) involved 2 minutes at 50ºC, 10 minutes at 95ºC for enzyme activation, followed by 55 cycles of 95ºC for 15 seconds and 55ºC for 1 minute.

2.1.3.3 Bocavirus PCR

This assay was designed to amplify a fragment of 354 base pairs from the NP-1 gene of HBoV.

The reaction was performed in a 25 µl reaction with Nuclease-free water, 1X GeneAmp PCR Buffer II (10 mM Tris-HCl, pH 8.3, 500 mM KCl) (Applied Biosystems, Germany), 2.5 mM of $MgCl_2$, 200 µM of each dNTP, 800 nM of primers 188F (5'-GAC CTC TGT AAG TAC TAT TAC-3') and 542R (5'-CTC TGT GTT GAC TGA ATA CAG-3'), 1.25 U of AmpliTaq Gold DNA polymerase (Applied Biosystems, Germany), and 5 µl of extracted DNA. Thermal cycling in a Mastercycler epgradient (Eppendorf) instrument involved 10 minutes at 94°C, followed by 35 cycles of 94°C for 1 minute, 54°C for 1 minute, and 72°C for 2 minutes.

2.1.3.4 Inf A virus 5'-nuclease RT-PCR (TaqMan) assay

The assay was designed to amplify a conserved region in the Matrix protein gene of Inf A viruses.

The reaction was performed with SuperScript One-Step RT-PCR with Platinum *Taq* (Invitrogen, Karlsruhe, Germany) reagents in a 25 µl volume reaction containing RNase-free water, 1X reaction mix (buffer containing 200 µM of each dNTP and 1.2 mM $MgSO_4$), additional 3.0 mM of $MgSO_4$, 1.0 µg of bovine serum albumin (BSA) (Sigma, Munich, Germany), 600 nM of primer M+25 (5'-GAT GAG TCT TCT AAC CGA GGT CG-3'), 600 nM of primer M-124 (5'-TGC AAA AAC ATC TTC AAG TCT CTG-3') and 200 nM of probe M+64 (FL-TCA GGC CCC CTC AAA GCC GAT-TAMRA), 0.5 µl of SuperScript II RT/Platinum *Taq* DNA polymerase mix (Invitrogen, Karlsruhe, Germany), and 5 µl of extracted RNA. Thermal cycling conditions in a ABI Prism 7000 SDS instrument (Applied Biosystems, Germany) involved reverse transcription at 45°C for 20 minutes, followed by 95°C for 3 minutes

for enzyme activation and 45 cycles of 95°C for 15 seconds, and 58°C for 35 seconds. Fluorescence was read in every cycle at 58°C.

2.1.3.5 HCoV-229E and HCoV-OC43 fluorescence resonance energy transfer (FRET) RT-PCR assay

For HCoV-OC43 and HCoV-229E detection, a common primer pair was placed in a highly conserved domain of the RdRp gene. A combination of FRET probes was selected which allowed the detection and differentiation of HCoV-OC43 and 229E by melting curve fluorescence analysis.

The Access RT-PCR System Kit (Promega, Karlsruhe, Germany) was used in a 25 µl volume reaction containing RNase-free water, 1X AMV/*Tfl* Reaction Buffer, 200 µM of each dNTP, 1 mM of $MgSO_4$, 1.0 µg of BSA, 400 nM of primer CproS (equimolar mixture of 5'-TAC TAT GAC TGG CAG AAT GTT TCA-3' and 5'-TAC TAT GAC TAC TAG ACA GTT TCA-3'), 400 nM of primer CproAs (5'-GGC ATA GCA CTA TCA CAC TTT GG-3'), 40 nM of probe Cpro5P (equimolar mixture of 5'-GTG TTC CTG TAG TTA TAG GCA CCA CTA A-3', 5'-ATG CCA CCG TTG TTA TCG GCA CTA CCA A-3', and 5'-GAG CTA CTG TGG TAA TTG GAA CAA GCA A-3', all 3'-labelled with fluorescein), 40 nM of probe Cpro3P (5'-TTT ATG GTG GCT GGG ATG ATA TGT T-3', 5'-labelled with LCRed640) (Tib-Molbiol, Berlin, Germany), 2.5 U of AMV-RT, 2.5 U of *Tfl* DNA polymerase and 5 µl of extracted RNA. Thermal cycling conditions in a LightCycler (Roche, Germany) involved reverse transcription at 42°C for 30 minutes, followed by 94°C for 2 minutes and 45 cycles of 94°C for 5 seconds, 50°C for 10 seconds and 72°C for 20 seconds. Melting curve fluorescence was started at 40°C, and the temperature was slowly raised, at 0.2°C per second, up to 80°C.

2.1.3.6 HCoV-NL63 5'-nuclease (TaqMan) real-time RT-PCR assay

The assay for HCoV-NL63 used standard 5'-nuclease oligonucleotides targeted to a conserved domain within the nucleocapsid gene.

The reaction was performed with SuperScript One-Step RT-PCR with Platinum *Taq* (Invitrogen, Karlsruhe, Germany) reagents in a 25 µl volume reaction containing RNase-free water, 1X reaction mix (buffer containing 200 µM of each dNTP and 1.2 mM MgSO4), additional 2.0 mM of $MgSO_4$, 1.0 µg of BSA (Sigma, Munich, Germany), 400 nM of primer 63NF2 (5'-AAA CCT CGT TGG AAG CGT GT-3'), 200 nM of primer 63NR1 (5'-TGT GGA AAA CCT TTG GCA TC-3'), 400 nM of probe 63NP (FL-ATG TTA TTC AGT GCT TTG GTC CTC GTG AT-TAMRA), 0.5 µl of SuperScript II RT / Platinum *Taq* DNA polymerase mix, and 5 µl of extracted RNA. Thermal cycling conditions in a LightCycler (Roche, Germany)

involved reverse transcription at 50°C for 20 minutes, 95°C for 3 minutes, followed by 45 cycles of 95°C for 10 seconds, 58°C for 30 seconds. Fluorescence was read in every cycle at 58°C.

2.1.3.7 Detection of human PV 1, 2 and 3; Inf A and B; RSV and hMPV

The assay was performed with the Hexaplex Plus Kit (Prodesse, USA). The assay enables detection of PIV 1, 2, and 3; Inf A and B viruses; RSV and hMPV in a multi-step procedure consisting of reverse transcription, multiplex PCR amplification, enzyme hybridization and colorimetric detection of DNA amplicon.

Reverse transcription was performed in a 20 µl volume reaction containing RNase-free water, 1X RT Mix (buffer containing random hexamer primers and each of dNTP) 5 mM of $MgCl_2$, 20 U of RNase Inhibitor, 50 U of MuLV Reverse Transcriptase, and 3 µl of extracted RNA. All reagents, apart from RT Mix, were manufactured by Applied Biosystems, Germany. The reaction mix was submitted for 10 minutes at 25°C, for 30 minutes at 42°C, as well as for 5 minutes at 95°C.

The PCR amplification reactions were performed in a 50 µl volume reaction containing Nuclease-free water, 1X Supermix (buffer containing oligonucleotide primers, each dNTP and $MgCl_2$), 2.5 U of AmpliTaq Gold DNA polymerase (Applied Biosystems, Germany) and 10 µL of cDNA. Thermal cycling conditions in a Mastercycler epgradient instrument (Eppendorf, Germany) involved 10 minutes at 95°C, followed by 2 Cycles of 95°C for 1 minute, 55°C for 30 seconds and 72°C for 45 seconds, and 38 Cycles of 94°C for 1 minute, 60°C for 30 seconds and 72°C for 30 seconds. The reaction was finished with a final extension at 72°C for 7 minutes.

PCR products were purified with QIAGEN QIAquick PCR Purification Kit (QIAGEN, Hilden, Germany) as follows: for every sample, 250 µl of Buffer PBI was mixed to 50 µl of the PCR sample. The mix was applied to the QIAquick spin column in a provided 2 ml collection tube and centrifuge at 13,000 rpm for 1 minute. The column was placed in a new collection tube and washed with 750 µl of Buffer PE by centrifugation at 13,000 rpm for 1 minute. In a new collection tube, the column was centrifuged for an additional 1 minute in order to completely dry the column membrane. After this, the column was placed in a clean 1.5 ml microcentrifuge tube; then, 50 µl of elution Buffer EB (10 mM Tris-Cl, pH 8.5) was applied to the centre of the QIAquick membrane. The column was incubated for 1 minute at room temperature for elution of bound DNA. The eluted DNA was collected by centrifugation at 13,000 rpm for 1 minute. After purification, the PCR samples were denatured twice, heated at 95ºC for 5 minutes and immediately chilled on ice for 5 minutes, to form ssDNA.

For enzyme hybridization assay, purified ssDNA was added to separate wells of a 96 well NeutrAvidin-coated microtitre strip plate; each well containing virus specific Horseradish peroxidase (HRP) labelled probes. The HRP-labelled oligonucleotide probes were bound to the biotin-labelled ssDNAs which in turn were bound to the plate. Following hybridization for 30 minutes at 42°C in a hybridization oven, the plate was washed several times with a 1X Wash Solution to remove unbound material. A substrate solution containing hydrogen peroxide and 3,3',5,5'-tetramethylbenzidine (TMB) was added to each well. In the presence of hydrogen peroxide, bound HRP catalyzed the oxidation of TMB to form a coloured complex. The reaction was stopped by the addition of Stop Solution provided with the kit and the absorbance was measured at 450 nm (A450) using a microwell plate reader.

2.1.3.8 Multiplex pneumonia-producing bacteria real-time PCR

This assay was designed for parallel detection of three bacteria, *Chlamydia pneumoniae*, *Legionella spp.*, and *Mycoplasma pneumoniae*, in clinical specimens.

The reaction was carried out with LightCycler FastStart DNA Master Hybridization Probes Kit (Roche, Mannheim, Germany) in a 20 µl volume reaction volume containing PCR-grade water, 1X LightCycler FastStart Reaction Mix HybProbe (buffer containing FastStart Taq DNA Polymerase, dNTP mix with dUTP instead of dTTP, and 1.0 mM $MgCl_2$), additional 3.0 mM of $MgCl_2$, primers and probes at concentrations according to Table 3, and 5 µl of extracted nucleic acid. Thermal cycling conditions in a LightCycler instrument (Roche, Germany) involved 7 minutes at 95°C, followed by 60 cycles of 95°C for 10 seconds, 60°C for 10 seconds and 72°C for 10 seconds. Melting curve fluorescence was started at 50°C, while temperature was slowly raised at 0.2°C per second, up to 82°C.

Table 3: primers and probes for the three pneumonia-producing bacteria assay.

bacterium	primers [500 nM]	Probe [200 nM]
C. pneumoniae	5'-AACTTGGGAATAACGGTTGGAAAC-3' 5'-AGTGTCTCAGTCCCAGTGTTGGC-3	5'-CCAACAAGCTGATATCGCATAAACTCTTCC-FL LCRed640-CAACCGAAAGGTCCGAAGATCCCCT-3'
L. ssp,	5'-AGGGTTGATAGGTTAAGAGC-3' 5'-CCAACAGCTAGTTGACATCG-3'	5'-GTGGCGAAGGCGGCTACCT-FL Red640-TACTGACACTGAGGCACGAAAGCGT-3'
M. pneumoniae	5'-CGCCAGCTTGTAAAAGTGAGC-3' 5'-TAGCAACACGTTTTTAAATATTAC-3'C	5'-CGGTGAATACGTTCTCGGGTCTTGTAC-FL LCRed640-CACCGCCCGTCAAACTATGAAAGC-3'

2.1.3.9 Limits of detection (LOD) of molecular assays

According to Claas *et al.* (2005), analytical sensitivity of AdV assay has shown to lie between 50 and 250 of viral particle copies per ml, which correlate to 1 to 5 copies in the reaction tube. HRV real-time RT-PCR (Deffernez *et al.*, 2004), the LOD was estimated at 1 copy of viral particle per 5 µl of reaction mixture. For HBoV PCR (Allander *et al.*, 2005) a LOD was not specified by the authors. For Hexaplex Plus Kit (Prodesse, USA), as described in the manufacture's handbook, the LODs of the assay were determined by analyzing 10-fold serial dilutions of 50% tissue culture infective dose ($TCID_{50}$) titrated stocks of PIV 1, 2 and 3; Inf A; RSV and 10-fold dilutions of non-specified initial concentrations of Inf B and hMPV. The limit of detection was defined as the concentration that was detected in more than 50% of the reactions, as follow: PV 1, 10^1 TCID50/ml; PV 2, 10^0 TCID50/ml; PV 3, 10^2 TCID50/ml; Inf A, 10^2 TCID50/ml; Inf B, 10^{-5} Dilution; RSV, 10^1 TCID50/ml; and hMPV, 10^{-6} Dilution. For the three pneumonia-producing bacteria real-time PCR (Raggam *et al.*, 2005), the LODs were described to be found between 0.5 and 5.0 infection-forming units (IFU) or colony-forming units (CFU) per PCR reaction for each of the bacteria.

The LODs for the new molecular assays were determined using *in vitro* transcribed RNAs. LOD for the Inf A real-time RT-PCR was estimated to be less than 5 copies of synthetic RNA genomes per reaction, corresponding to circa 50 RNA copies per swab sample. For HCoV-229E and OC43 FRET RT-PCR, LOD was less than 20 copies per reaction or circa 200 copies per swab sample. Finally, for the HCoV-NL63 TaqMan assay, the LOD was less than 5 copies of synthetic RNA genomes per reaction or circa 50 RNA copies per swab sample.

2.1.4 Cloning of PCR products

2.1.4.1 TOPO-TA cloning reaction and transformation of competent cells

For the new molecular assays formulated for this study, fragments including the PCR target regions and sufficient stretches of flanking sequence were cloned using the TOPO TA Cloning Kit (Invitrogen, Karlsruhe, Germany). For each sample, cloning reaction was made in a 6 µl volume reaction containing 2 µl of fresh RT-PCR products, 1 µl of Salt Solution, 2 µl of Nuclease-free water and 1 µl of pCR 2.1-TOPO vector. The mix was incubated at room temperature for 15 minutes and then chilled on ice. Next, 2 µl of cloning reaction mix was added into a vial of One Shot TOP10F' Chemically Competent *E. coli* for cell transformation and incubated on ice for 30 minutes. After incubation, the cell tubes were heat-shocked at 42°C for 30 seconds and immediately transferred to ice. 250 µl of room temperature S.O.C. medium was

added. The cells were incubated by shaking the tubes horizontally at 300 rpm in a Thermomixer Comfort (Eppendorf, Germany) for 1 hour at 37°C. Next, 50 µl of transformed cells were spread in LB (Luria Broth) agar (Carl Roth, Karlsruhe, Germany) plates, supplemented with 50 µg/ml of kanamycin (Sigma, Germany) and pre-treated with 40 µl of 40 mg/ml X-gal and 40 µl of 100 mM IPTG for blue/white screening of transformants containing inserts. The plates were incubated overnight at 37°C in a incubator oven.

2.1.4.2 Screening of transformants

For each cloned fragment, 10 white colonies were selected for PCR directional screening with plasmid-specific forward primer and virus-specific reverse primer for further *in vitro* transcription of sense RNA.

A very small portion of each colony was picked with a micropipette tip and inoculated directly into a 20 µl PCR reaction mix containing Nuclease-free water, 1X PCR Buffer (Invitrogen, Karlsruhe, Germany), 200 µM of each dNTP (Applied Biosystems, Germany), 2.0 mM of $MgCl_2$, 250 nM of M13 Forward (-20) primer (Invitrogen, Karlsruhe, Germany), 250 nM of virus-specific reverse primer, and 1 U of Platinum *Taq* DNA Polymerase (Invitrogen, Karlsruhe, Germany). Thermal cycling conditions in a Mastercycler epgradient (Eppendorf, Germany) involved 10 minutes at 94°C, followed by 30 cycles of 30 seconds for each temperature (94°C, 56°C and 72°C). After PCR, 10 µl of each reaction was analysed by 1,5% Tris-Borate-EDTA (TBE) agarose gel electrophoresis.

2.1.4.3 Growth of positive colonies and plasmid preparation

After PCR screening, a small portion was picked from 2 positive colonies for each cloned fragment and inoculated into 4 ml of LB medium supplemented with 50 µg/ml of kanamycin (Sigma, Germany). After overnight-growth at 37°C, bacterial cells were harvested twice by centrifugation at 10,000 rpm for 1 minute in a 2.0 ml microcentrifuge tube and discarded of the cleared medium. The cell pellets were then submitted to plasmid purification using the QIAprep Spin Miniprep Kit (QIAGEN, Hilden, Germany). The pelleted bacterial cells were resuspended in 250 µl Buffer P1 containing RNase A and then transferred to a clean 1.5 ml microcentrifuge tube. Next, 250 µl of Lysis Buffer P2 was added and mixed thoroughly by inverting the tube 6 times. In addition, 350 µl of Buffer N3 was added and mixed immediately and thoroughly by inverting the tube 6 times. The mix was centrifuged for 10 min at 13,000 rpm. The supernatants were then added to the QIAprep spin column in a 2.0 collection tube and centrifuge for 1 minute at 13,000 rpm. The column was placed in a new collection tube, was then washed with 500 µl of

Buffer PB by centrifugation for 1 minute at 13,000 rpm. The column was washed again with 750 µl of Buffer PE by centrifugation as described above. In a new collection tube, the column was centrifuged for an additional 1 minute to completely dry the column membrane. Next, the column was placed in a clean 1.5 ml microcentrifuge tube; 50 µl of elution Buffer EB was applied to the centre of each column membrane and incubated for 1 minute for elution of bound plasmid DNA Eluted plasmids were then collected by centrifugation at 13,000 rpm for 1 minute.

2.1.5 Sequencing of cloned PCR products

The cloned fragments were sequenced in both orientations using the GenomeLab™ Methods Development Kit (Beckman Coulter, Germany). The dye termination reaction was performed in 10 µl of volume containing Nuclease-free water, Sequencing Reaction Mix (buffer containing dNTP mix; dye terminators ddUTP, ddGTP, ddCTP, ddATP; and DNA polymerase enzyme), M13 sequencing primer, and about 130 ng of plasmid DNA template. Thermal cycling conditions in a Mastercycler epgradient (Eppendorf, Germany) involved 30 cycles of 96°C for 20 seconds, 50°C for 20 seconds and 60°C for 4 minutes.

After sequencing reaction, dye terminators were removed with the Agencourt CleanSEQ Sequencing Reaction Clean-Up system (Agencourt Bioscience, Germany). For each reaction tube, 10 µl of CleanSEQ magnetic particles solution were added to each sequencing reaction tube, followed by the addition of 42 µl of 85% ethanol. The sample was well mixed by pipetting thoroughly up and down 7 times. The tubes were placed onto a SPRIPlate 96R Ring Magnet Plate for 3 minutes to separate beads from solution. The cleared solution from the reaction was discarded and particles were washed with 100 µl of 85% ethanol by incubation at room temperature for 30 seconds. The ethanol was then aspirated out and discarded; then the tubes were air-dried for 10 minutes at room temperature, out of the ring magnet plate. Next, 40 µL of Sample Loading Solution (Beckman Coulter, Germany) was added for elution of the sequencing products and incubated for 5 minutes at room temperature. The tubes were placed back onto the ring magnet plate for 3 minutes to separate beads from solution. The cleared reaction was then transferred into a clean reaction plate, and each well was overlaid with one drop of light mineral oil provided with the kit. The sample plate was loaded into a CEQ 2000 DNA Analysis System (Beckman Coulter, Germany).

2.1.6 RNA *in vitro* transcription

For *in vitro* RNA transcription of cloned fragments, purified plasmids were re-amplified with the plasmid specific primers M13 Forward (-20) and M13 Reverse (Invitrogen, Karlsruhe, Germany) to lower the plasmid background in subsequent *in vitro* transcription. Reamplification was performed with PCR reagents (Invitrogen, Karlsruhe, Germany) in 50 µl volume reaction containing Nuclease-free water, 1X PCR Buffer, 200 µM of each dNTP (Applied Biosystems, Germany), 2.0 mM $MgCl_2$, 200 nM of each primer, 1 U of Platinum *Taq* DNA polymerase, and 1 µl of a 10^{-2} plasmid dilution. Thermal cycling conditions in a Mastercycler epgradient (Eppendorf, Germany) involved 3 minutes at 94°C, followed by 35 cycles of 30 seconds for each temperature (94°C, 56°C and 72°C).

The PCR products were purified with the AMPure PCR Purification Kit (Agencourt Bioscience, Germany). For each sample, 90 µl of AMPure magnetic particles solution was added to the PCR tube reaction and mixed thoroughly by pipetting 10 times. The mix was incubated for 5 minutes at room temperature for binding of PCR amplicons to magnetic beads and then placed onto a SPRIPlate 96R Ring Magnet Plate (Agencourt Bioscience, Germany) for 10 minutes to separate beads from solution, which was in turn aspirated out and discarded. Magnetic particles were washed twice with 200 µl of 70% ethanol by incubation for 30 seconds at room temperature and air dried for 15 minutes. PCR products were eluted with 40 µl of elution Buffer EB by pipette mixing the particles thoroughly 10 times. The tubes were placed in the Ring Magnet Plate for 10 minutes to collect the cleared eluted sample.

The purified PCR products originated from plasmid-cloned fragments were transcribed into RNA with the MEGAscript T7 Kit (Ambion, Austin, USA). Each plasmid-originated PCR product contains a T7 promoter region upstream from the sequence to be transcribed. The transcription reaction was performed in 20 µl volume reaction containing 2 µl of 10X Reaction Buffer, 8 µl of ribonucleotide solution mix (ATP, CTP, GTP, and UTP), 2 µl T7 RNA polymerase enzyme and 8 µl of purified PCR product. The reaction was incubated at 37°C for 4 hours. After transcription, the reaction was treated with 2 U of DNase 1, supplied with the kit, for an additional hour at 37°C. The transcribed RNA was purified with the RNeasy Mini Kit (QIAGEN, Hilden, Germany) by means of the RNA cleanup protocol. For each transcription reaction tube, the volume was adjusted to a volume of 100 µl with RNase-free water. 350 µl of Lysis Buffer RLT containing β-mercaptoethanol was added, and the sample was mixed well. In the following, 250 µl of absolute ethanol was added and mixed well by pipetting up and down. The entire volume was then transferred to an RNeasy spin column placed in a 2 ml collection tube and centrifuge for 15 seconds at 10,000 rpm. The spin column was placed in a new

collection tube and washed twice with 500 µl of Buffer RPE by centrifugation for 15 seconds at 10,000 rpm. The column was then placed in a new collection tube and centrifuged at 13,000 rpm for 1 minute to completely dry the column membrane. Next, the column was placed in a clean 1.5 ml microcentrifuge tube, and 50 µl of Buffer AVE at 80°C was applied in the centre of the column membrane for RNA elution. The eluted RNA was then collected by centrifugation at 10,000 rpm for 1 minute. The RNA transcripts were quantified in a BioPhotometer (Eppendorf, Germany).

2.2 Generic detection of coronaviruses and differentiation at the prototype strain level by reverse RT-PCR and non-fluorescent low-density DNA microarray

2.2.1 Samples and nucleic acids extraction

A total of 8 throat swabs, 2 sputum samples, and 1 bronchoalveolar lavage specimen was available from 11 patients with PCR-confirmed SARS in a previous study (Drosten *et al.*, 2003). In addition, 46 throat swabs from children with respiratory disease, in whom a CoV had previously been detected by routine PCR, were also used, including 1 sample from Hamburg, Germany, 36 samples from Switzerland, 7 samples from Greece and 2 samples from the Democratic Republic of Congo.

Additional samples that were available came from cell culture supernatants of prototype strains including some such as HCoV-229E, HCoV-OC43 and HCoV-HKU1, and animal CoVs, such as BtCoV, BCoV, FIPV, IBV, MHV, and TGEV. RNA was extracted from respiratory samples using the Viral RNA Mini Kit (QIAGEN, Hilden, Germany) as described in the section 2.1.2, with an elution volume of 50 µl for all samples. Sputum samples were pre-treated by 30 minutes shaking incubation in 2X Sputum Lysis Buffer (10 g/l N-acetyl cysteine, 0.9% sodium chloride). Throat swabs were introduced into QIAGEN Lysis Buffer AVL as a whole without prior solubilisation or resuspention. Up to 20 RNA extracts from samples testing negative for CoV RNA were pooled in large volumes and used as background nucleic acids for RT-PCR optimization.

2.2.2 Cloning of coronavirus RdRp gene fragments and synthesis of RNA standards

A fragment of about 670 base pairs including the PCR target regions and sufficient stretches of flanking sequence were cloned from several CoVs. The genome positions, GenBank accession number (in parenthesis) and pair of primers used for cloning each CoV RdRp fragment were as follows: SARS-CoV, positions 15008 to 15678 (NC_004718), primers Sar43Fwd (5'-TGC CAT TAG TGC AAA GAA TAG-3') and SarRev (5'-TGT TAT AGC ACA CAA CGG CAT CAT-3'); HCoV-OC43, positions 14942 to 15606 (NC_005147), primers Sar43Fwd and 43Rev (5'-AAT TAT AAC ACA CAA CCC CAT CAT-3'); HCoV-229E, positions 14118 to 14782 (NC_002645), primers 229Fwd (5'-CGC CAT ATC TGG TAA GGA ACG-3') and 229Rev (5'-TAT TAT AGC ACA CAA CAC TAT CAT-3'); HCoV-NL63, positions 14037 to 14701 (AY567487), primers NL63Fwd (5'-TGC TAT TAG TGG TAA AGA ACG-3') and NL63Rev (5'-TGT TAT AAC AGA CAA CAC CGT CAT-3'); HCoV-HKU1, positions 15201 to 15865 (AY597011), primers HKU1Fwd (5'-TGC TAT CAG TGC TAA GAA TAG AGC TCG C-3') and HKU1Rev (5'-AGT TAT AAC AGA CAA CAC CAT CAT CAC-3'); and IBV,

positions 13976 to 14637 (AJ311317), primers IBVFwd (5'-CAT ATC CGC GAA AAA TAG AGC G-3') and IBVRev (5'-ATC TGC TAC AAG ACC TTG TTT GGC-3').

The RT-PCR products for each CoV mentioned above were generated with the QIAGEN OneStep RT-PCR Kit reagents (QIAGEN, Hilden, Germany) in a 50 µl volume reaction containing RNase-free water, 1X of QIAGEN OneStep RT-PCR Buffer (containing 2.5 mM of $MgCl_2$), 400 µM of each dNTP, 200 nM of each specific CoV primer, 2 µl of QIAGEN OneStep RT-PCR Enzyme Mix and 5 µl of extracted RNA. Thermal cycling conditions in a Mastercycler epgradient (Eppendorf, Germany) involved 30 minutes at 50°C, 15 minutes at 95°C, followed by a touchdown protocol of 10 cycles at 94°C for 30 seconds, 62°C for 30 seconds (with a decrease of 1°C per cycle), 72°C for 45 seconds, and additional 35 cycles of 30 seconds at 94°C, 30 seconds at 52°C and 45 seconds at 72°C.

RT-PCR products were ligated into pCR 2.1-TOPO vector and introduced into chemically competent *E. coli* TOP10F' by means of the TOPO TA Cloning Kit (Invitrogen, Karlsruhe, Germany) according to the protocol described in section 2.1.4.1. Plasmids were then purified with the QIAprep Spin Miniprep Kit (Agencourt Bioscience, Germany), sequenced and re-amplified with plasmid-specific primers M13 Forward (-20) and M13 Reverse, provided with the TOPO TA kit, to lower the plasmid background in subsequent *in vitro* transcription.

The PCR reamplification reaction was performed in a 50 µl volume reaction containing Nuclease-free water, 1X PCR Buffer and 2.5 mM of $MgCl_2$ (Invitrogen, Karlsruhe, Germany), 200 µM of each dNTP (Applied Biosystems, Germany), 200 nM of primers M13 Forward (-20) and M13 Reverse, 1 U of Platinum *Taq* DNA polymerase (Invitrogen, Karlsruhe, Germany), and 1 µl of a 10^{-2} plasmid dilution. Thermal cycling conditions in a Mastercycler epgradient involved 3 minutes at 94°C, followed by 35 cycles of 30 seconds at 94°C, 30 seconds at 56°C and 1 minute at 72°C. After PCR, the products were purified with the AMPure PCR purification Kit (Beckman Coulter, Germany), according to the protocol described in section 2.1.6.

The purified PCR products were transcribed into RNA by T7 RNA polymerase from the MEGAscript T7 Kit (Ambion, Austin, USA) following the protocol described in section 2.1.6. After DNase 1 digestion, RNA transcripts were purified with QIAGEN RNeasy columns and quantified in a BioPhotometer (Eppendorf, Germany).

2.2.3 Design of the universal coronavirus RT-PCR and protocol optimization

Primers for universal RT-PCR for the genus CoV were designed after aligning all of the CoV RdRp sequences available in GenBank in May 2005. For primer targeting regions, the motifs A and C were chosen because they contain short amino acid patterns that are 100% identical with all CoVs (Xu *et al.*, 2003). Oligonucleotide hybridization regions corresponded to the patterns LMGWDYPKCD and MMILSDDAV, comprising domains essential for metal ion chelation and binding of the complex primer 3' end and template (Snijder *et al.*, 2003; Xu *et al.*, 2003).

The amplification of the universal RT-PCR was initially performed by means of three different commercial RT-PCR kits: SuperScript One-Step RT-PCR with Platinum *Taq* (Invitrogen, Karlsruhe, Germany); Access RT-PCR System (Promega, Mannheim, Germany) and QIAGEN OneStep RT-PCR Kit (QIAGEN, Hilden, Germany). For each RT-PCR kit, oligonucleotide and magnesium concentrations were optimized and the protocols were compared in an end-point dilution experiment to determine the best reaction kit.

To define the best magnesium concentration for each reaction kit, a titration matrix was established for titrations of magnesium and primers against each other. For magnesium titration, concentrations (in mM) corresponded to 0.5, 1.0, 1.5, 2.0, 2.5, and 3.0. For primer mix titration, equals concentrations (in nM) of forward and reverse primers corresponded to 100, 200, 400, and 800. The resulting 24 combination reactions were performed with a 10^{-9} dilution of SARS-CoV *in vitro* RNA transcript. Once the best magnesium concentration had been determined, a second matrix was established to determine the best concentration of forward and reverse primers. For primer titration, concentrations (in nM) for each primer corresponded to 100, 200, 400, and 800. The resulting 16 combination reactions were performed with a 10^{-9} dilution of SARS-CoV *in vitro* RNA transcript and the optimal magnesium concentration. After optimization of each reaction kit, a 10-fold dilution from 10^{-9} to 10^{-12} of RNA *in vitro* transcripts of SARS-CoV, HCoV-229E, and HCoV-OC43, both of them were submitted to RT-PCR reaction using the three optimized protocols. The final optimized protocols for each reaction kit and thermal conditions were as described below.

For QIAGEN Kit, 25 µl volume reaction containing RNase-free water, 1X QIAGEN OneStep RT-PCR Buffer (containing 2.5 mM of $MgCl_2$), 400 µM of each dNTP, 1.0 µg of BSA (Sigma, Munich, Germany), 200 nM of primer PC2S2 (equimolar mixture of 5'-TTA TGG GTT GGG ATT ATC-3' and 5'-TGA TGG GAT GGG ACT ATC-3'), 900 nM of primer PC2As1 (equimolar mixture of 5'-TCA TCA CTC AGA ATC ATC A-3', 5'-TCA TCA GAA AGA ATC

ATC A-3', and 5'-TCG TCG GAC AAG ATC ATC A-3'), 1 µl of QIAGEN OneStep RT-PCR Enzyme Mix and 3 µl of extracted RNA.

For Invitrogen kit, 25 µl volume reaction containing RNase-free water, 1X Reaction Mix (buffer containing 200 µM of each dNTP, 1.2 mM $MgSO_4$), additional 2.0 mM of $MgSO_4$, 1.0 µg of BSA (Sigma, Munich, Germany), 400 nM of primers PC2S2 and PC2As1, 0.5 µl of RT/ Platinum *Taq* Mix, and 3,0 of extracted RNA.

For Promega kit, 25 µl volume reaction containing RNase-free water, AMV/*Tfl* 1X Reaction Buffer, 200 µM of each dNTP, 2.5 mM of $MgSO_4$, 1.0 µg of BSA (Sigma, Munich, Germany), 200 nM of primer PC2S2, 800 nM of primer PC2As1, 2.5 U of AMV reverse transcriptase (RT), 2.5 U of *Tfl* DNA polymerase, and 3,0 extracted RNA.

Thermal cycling conditions in a Mastercycler epgradient (Eppendorf, Germany) were the same for each kit, involving reverse transcription at 48°C (50°C for QIAGEN kit) for 30 minutes, DNA polymerase activation at 95°C for 3 minutes (15 minutes for QIAGEN kit), followed by 45 cycles at 94°C for 20 seconds, 52°C for 30 seconds, and 72°C for 40 seconds.

The QIAGEN OneStep RT-PCR Kit was selected for further experiments as shown in results section.

2.2.4 The universal coronavirus nested RT-PCR

When CoV *in vitro* RNA transcripts were spiked in clinical samples, several Log10s of sensitivity were lost. Multiple unspecific amplification products appeared which could not be eliminated by any effort in subsequent assay optimization, such as modifications of temperature, salts, or denaturing agents, including the Q-Solution supplied with the QIAGEN OneStep RT-PCR Kit. The short primers used in the assay presumably caused random reactivity with human genomic DNA. Since only very short stretches of sequence were conserved among all CoVs, the use of longer primers was not an option. Rather than reducing the breadth of amplification, a nested PCR strategy was therefore adopted. With long nested amplification sense primers that bound to the short first round sense primer but reached inside the amplicon, CoV sequences were selectively amplified. This reaction was optimized in presence of a background of nucleic acids as encountered in routine operation.

After optimization, the first round amplification of the universal CoV nested RT-PCR was performed with the QIAGEN OneStep RT-PCR Kit (QIAGEN, Hilden, Germany) in 25 µl volume reaction containing RNase-free water, 1X QIAGEN OneStep RT-PCR Buffer (containing 2.5 mM of $MgCl_2$), 400 µM of each dNTP, 1.0 µg of BSA (Sigma, Munich, Germany), 200 nM of primer PC2S2 (equimolar mixture of 5'-TTA TGG GTT GGG ATT ATC-

3’ and 5’-TGA TGG GAT GGG ACT ATC-3’), 900 nM of primer PC2As1 (equimolar mixture of 5’-TCA TCA CTC AGA ATC ATC A-3’, 5’-TCA TCA GAA AGA ATC ATC A-3’, and 5’-TCG TCG GAC AAG ATC ATC A-3’), 1 µl of QIAGEN OneStep RT-PCR Enzyme Mix and 5 µl of extracted RNA. The thermal cycling conditions comprised 30 minutes at 50°C; 15 minutes at 95°C; a touchdown protocol of 10 cycles of 20 seconds at 94°C, 30 seconds starting at 62°C with a decrease of 1°C per cycle, and 40 seconds at 72°C; followed by 40 cycles of 20 seconds at 95°C, 30 seconds at 54°C, and 40 seconds at 72°C.

The second round or nested amplification was performed in a reaction volume of 50 µl containing Nuclease-free water, 1X PCR Buffer and 2.5 mM of $MgCl_2$ (Invitrogen, Karlsruhe, Germany), 200 µM of each dNTP (Applied Biosystems, Germany), 80 nM of primer PCS (equimolar mixture of 5’-CTT ATG GGT TGG GAT TAT CCT AAG TGT GA-3’ and 5’-CTT ATG GGT TGG GAT TAT CCC AAA TGT GA-3’), 400 nM of primer PCNAs (5’-biotin-CAC ACA ACA CCT TCA TCA GAT AGA ATC ATC A-3’), 1 U of Platinum *Taq* DNA polymerase (Invitrogen, Karlsruhe, Germany), and 1 µl of the first round RT-PCR mixture. The thermal cycling conditions involved 3 minutes at 94°C, followed by 30 cycles of 20 seconds at 94°C, 30 seconds at 60°C, and 30 seconds at 72°C. The primer PCNAs was synthesized by the supplier of the microarrays (Chipron, Berlin, Germany) using proprietary technology to increase its efficiency in array hybridization. For use as a plain nested RT-PCR, primer PCNAs could be replaced by an unmodified oligonucleotide at a working concentration of 300 nM and increasing the concentration of primer PCS to 300 nM. The thermal cycling conditions of the universal CoV nested RT-PCR were performed in a Mastercycler epgradient instrument (Eppendorf, Germany).

2.2.4.1 Specificity of the universal coronavirus nested RT-PCR

The specificity of the assay was verified on clinical samples tested in an earlier study (Luna *et al.*, 2007) and determined to contain inf A virus, inf B virus, PIV 1, 2 and 3, hMPV, HRV, AdV, and RSV. Three different samples were tested for each of the viruses mentioned above, except for RSV where only one sample was available.

2.2.4.2 Sensitivity of the universal coronavirus nested RT-PCR

To test the LOD of the assay and to challenge its robustness, quantified *in vitro* RNA transcripts of SARS-CoV, HCoV-OC43, HCoV-229E, HCoV-NL63, HCoV-HKU1 and IBV, were tested in the presence of high levels of background nucleic acids (human DNA, about 50 ng per reaction). The *in vitro* RNA transcripts containing background DNA were submitted to the

universal CoV nested RT-PCR. For each specific CoV, three replicate reactions were performed containing 90, 45, 15, 5 and 0 copies per reaction of RNA transcripts.

2.2.5 Specific human coronavirus RT-PCR assays

All samples containing HCoVs used in the universal CoV RT-PCR were also tested in specific real-time RT-PCRs. Protocol assays for HCoV-229E and HCoV-OC43 FRET real-time RT-PCR, and HCoV-NL63 TaqMan are described in the sections 2.1.3.5 and 2.1.3.6, respectively.

2.2.5.1 HCoV-HKU1 two steps real-time RT-PCR

This protocol was described by Garbino *et al.* (2006) and designed to amplify a partial fragment from the RdRp gene.

Reverse transcription was performed using the SuperScript II RT (Invitrogen, Karlsruhe, Germany) in a 20 µl volume reaction mixture containing RNase-free water, 1X First-Strand Buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 10 mM DTT, 500 µM of each dNTP (Applied Biosystems, Germany), 250 ng of random hexamer primers (Roche, Germany), 20 U of RNaseOUT and 100 U of SuperScript II RT enzyme (Invitrogen, Karlsruhe, Germany), and 5 µl of extracted RNA. The mixture was submitted to 42°C for 60 minutes, followed by 10 minutes at 95°C. After reverse transcription, the tubes were placed on ice.

The cDNA samples were immediately processed for PCR reaction, which was performed with the TaqMan Universal PCR Master Mix, No AmpErase UNG (Applied Biosystems, Germany) reagent in a 25 µl volume reaction containing Nuclease-free water, 1X TaqMan Universal PCR Master Mix (buffer containing dNTPs with dUTP, ROX passive reference, $MgCl_2$ and AmpliTaq Gold DNA polymerase), 600 nM of primer HKU1F (5'-GAA TTT TGT TGT TCA CAT GGT GAT AGA-3'), 600 nM of primer HKU1R (5'-GCA ACC GCC ACA CAT AAC TAT TT-3'), 200 nM of probe HKU1P (5'-FAM-TTT ATC GCC TTG CGA ATG AAT GTG CTC-TAMRA 3') and 5 µl of cDNA. Thermal cycling conditions in a ABI Prism 700 SDS (Applied Biosystems, Germany) involved 2 minutes at 50°C, 10 minutes at 95°C, followed by 55 cycles of 95°C for 15 seconds and 55°C for 1 minute. LOD for HCoV-HKU1 was described as less than 18 copies of viral particles per reaction, corresponding to less than 180 copies per swab sample.

2.2.6 Design and application of Coronavirus LCD array

The design of oligonucleotides for capturing PCR products on microarrays was based on an alignment of all CoV RdRp sequences as available in GenBank in May 2005. The exact positions of all probes on the LCD-chip are shown in Figure 6.

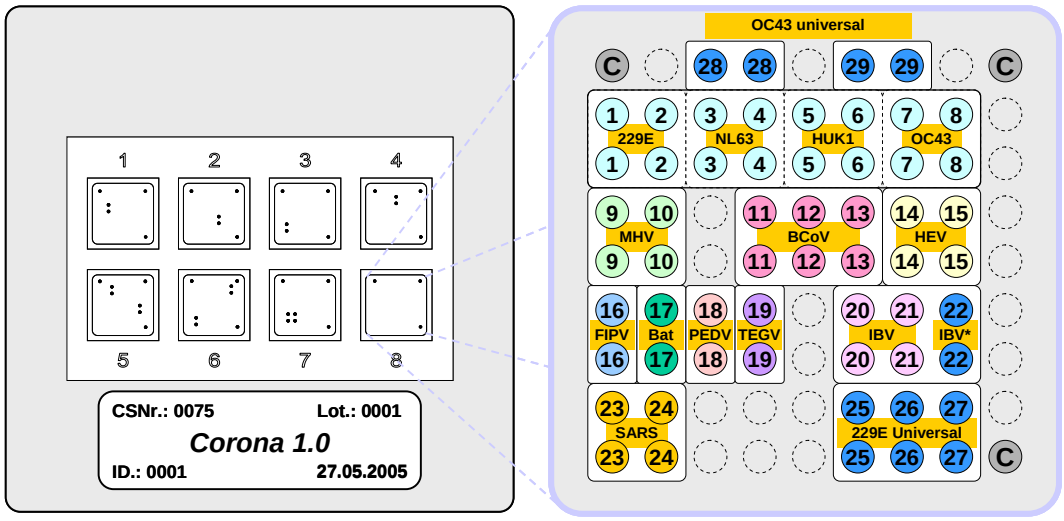

Figure 6: Schematic description of the LCD-chip and position of capture probes. Abbreviations: MHV, murine hepatitis virus; BCoV, bovine coronavirus; HEV, porcine hemagglutinating encephalomyelitis; FIPV, feline infectious peritonitis virus; PEDV, porcine epidemic diarrhoea virus; TGEV, transmissible gastroenteritis virus; IBV, avian bronchitis virus.

The oligonucleotide detection probes, listed in Table 4, were spotted on arrays presented in plastic LCD-chip, which resemble a photographic slide, using proprietary technology (Chipron, Berlin, Germany). As shown in Figure 6, each LCD-chip contains eight identical arrays in format of square reaction chambers which can be addressed individually. The capture probes were immobilized as duplicates (in vertical position). Functional controls to monitor hybridization and staining were located in three angles. The formed array is a 9 x 9 mm chamber containing the spots with detection probes, each one with average diameter of 300 µm. The LCD Array Corona 1.0 Kit (Chipron, Berlin, Germany) provides LCD-chips and all reagents needed for hybridization reaction.

The hybridization reactions were performed as follows: in addition to universal CoV nested RT-PCR, 10 µl of amplification product was mixed with 20 µl of Hybridization Buffer and 2 µl of Modulator solution. Then, 30 µl of this hybridization mix was applied to an array of the LCD-chip which was then placed in a simple moist chamber (Petri dish) and incubated at 37°C for 30 minutes in a cell culture incubator. After hybridization, the LCD-chip was washed by submersion in three bath chambers containing a 1X Wash Solution. In the 2 first baths, the LCD-chip was washed by moving it slowly 3 times backward and forward. In the third bath, the

LCD-chip was submerged and incubated for 1 minute. After the washing steps, the LCD-chip was placed into a Chipron transport container, provided with the kit, and then was dried in a centrifuge with swing out bucket by spin at 1,000 rpm for 1 minute. Once the arrays were completely dried, 30 µl of Labelling Mix Solution was applied to each array well and incubated for 5 minutes. After a second wash and further drying steps, each array was incubated with 30 µl of Stain Solution for 3 minutes. After a final wash and additional drying steps, hybridized spots were visible by naked eye as blue dots. A simple photo slide scanner, purchased from a department store, was used for electronic documentation of results. The entire procedure is illustrated in Figure 7.

Table 4: sequences of capture probes spotted in the LCD-chip.

Probe nr.	Name	Sequence	Tm Chip	G/C %
1	HCoV-229E A Mix-1	5'-ATGATTAGTATGTTCTCTGCCATGATAT-3'	34,8	32,1
	HCoV-229E A Mix-2	5'-ATGATTAGTATGTTCTCAGCCATGATAT-3'	34,8	32,1
	HCoV-229E A Mix-3	5'-ATGATTAGTATGTTCTCGGCCATG-3'	35,8	41,7
2	HCoV-229E B	5'-ATGTCACATGTTGTACGGCTAGTG-3'	35,2	45,8
3	HCoV-NL63 A	5'-AACAGATAGGTTTTATAGGCTTGGTAA-3'	34,6	33,3
4	HCoV-NL63 B	5'-TGGCACAAGTTTTAACAGAAGTTGT-3'	35,8	36,0
5	HCoV-HUK1 A	5'-CTATGCCAAATATTTTGCGTATTG-3'	34,7	33,3
6	HCoV-HUK1 B	5'-GTTTGTTCTCTTATGGCCTGTAATG-3'	35,0	40,0
7	HCoV-OC43 A	5'-AGTAGTTTGGTATTAGCCCGAAAAC-3'	35,3	40,0
8	HCoV-OC43 B	5'-CAATAAGATTGAAGATCTTAGTATACGTGC-3'	35,9	33,3
9	MHV/Rat A	5'-AACATGATTCGTGCTGTTCGC-3'	36,8	47,6
10	MHV/Rat B	5'-GCATGCAATGGACACAAAATTG-3'	37,1	40,9
11	BCoV A	5'-GTAGTCTGGTCTTGGCCCGAA-3'	37,3	57,1
12	BCoV B	5'-TAGTCTGGTTTTGGCTCGAAAAC-3'	36,1	43,5
13	BCoV C	5'-GTGCTTTAATGTCATGCAATGGTAA-3'	36,8	36,0
14	HEV A	5'-GTAGTCTGGTATTGGCCCGAA-3'	35.1	52,4
15	HCoV-OC43 + HEV	5'-CCTTAATGTCATGCAATGGCA-3'	35,2	42,9
16	CCoV + FIPV Mix-1	5'-AGAATGGCATCTGCCATGATATT-3'	36,0	39,1
	CCoV + FIPV Mix-2	5'-AGAATGGCCTCTGCCATGATAT-3'	36,0	45,5
	CCoV + FIPV Mix-3	5'-AGAATGGCTTCTGCCATGATATT-3'	35,4	39,1
17	BtCoV A	5'-GGGTTCTAAGCATGAGAATTGTTG-3'	35,6	41,7
18	PEDV A	5'-CACACCACATGCTGCAGTTCTAC-3'	36,4	52,2
19	Trans-Gastro A	5'-GTTGTACACATAATGATAGGTTCTACCG-3'	35,5	39,3
20	IBV A	5'-TGTTGCGCGTTTGCTTAGTG-3'	36,7	50,0
21	IBV B	5'-TGTTGCGCGTCTTTTGAGTG-3'	36,5	50,0
22	IBV universal A	5'-GCMAACAGTGTYTTYAAYATAATACAAGC-3'	37,9	34,5
23	SARS-CoV A	5'-GGATAATGGCCTCTCTTGTTCTTG-3'	36,9	45,8
24	SARS-CoV B	5'-CACACCGTTTCTACAGGTTAGCTAAC-3'	36,5	46,2
25	229E universal A	5'-GCTCAAGTACTCACAGAAGTTGTGCATTG-3'	42,2	44,8
26	229E universal B	5'-GGYGGKTTTTAYTTTAAACCTGGTGG-3'	41,6	44,2
27	229E universal C	5'-GGHGGKTTTTAYTTDAAGCCAGGTGG-3'	43,5	48,7
28	OC43 universal A-1	5'-TGAGTGAAATTGTTATGTGTGGTGG-3'	37,2	40,0
	OC43 universal A-2	5'-TGAGTGAAATAGTTATGTGTGGCGG-3'	38,7	44,0
29	OC43 universal B-1	GATGCAACTACTGCTTTTGCTAATTC-3'	36,7	38,5
	OC43 universal B-2	GCAACCACTGCCTTTGCTAATTC-3'	38,5	47,8

Abbreviations: HCoV, human coronavirus; BCoV, bovine coronavirus; CCoV, canine coronavirus; FIPV, feline infectious peritonitis virus; HEV, porcine hemagglutinating encephalomyelitis; IBV, avian infectious bronchitis virus; TCoV, turkey coronavirus; MHV, murine hepatitis virus; PEDV, porcine epidemic diarrhoea virus; SDAV, rat sialodacryoadenitis coronavirus; TGEV, transmissible gastroenteritis virus.

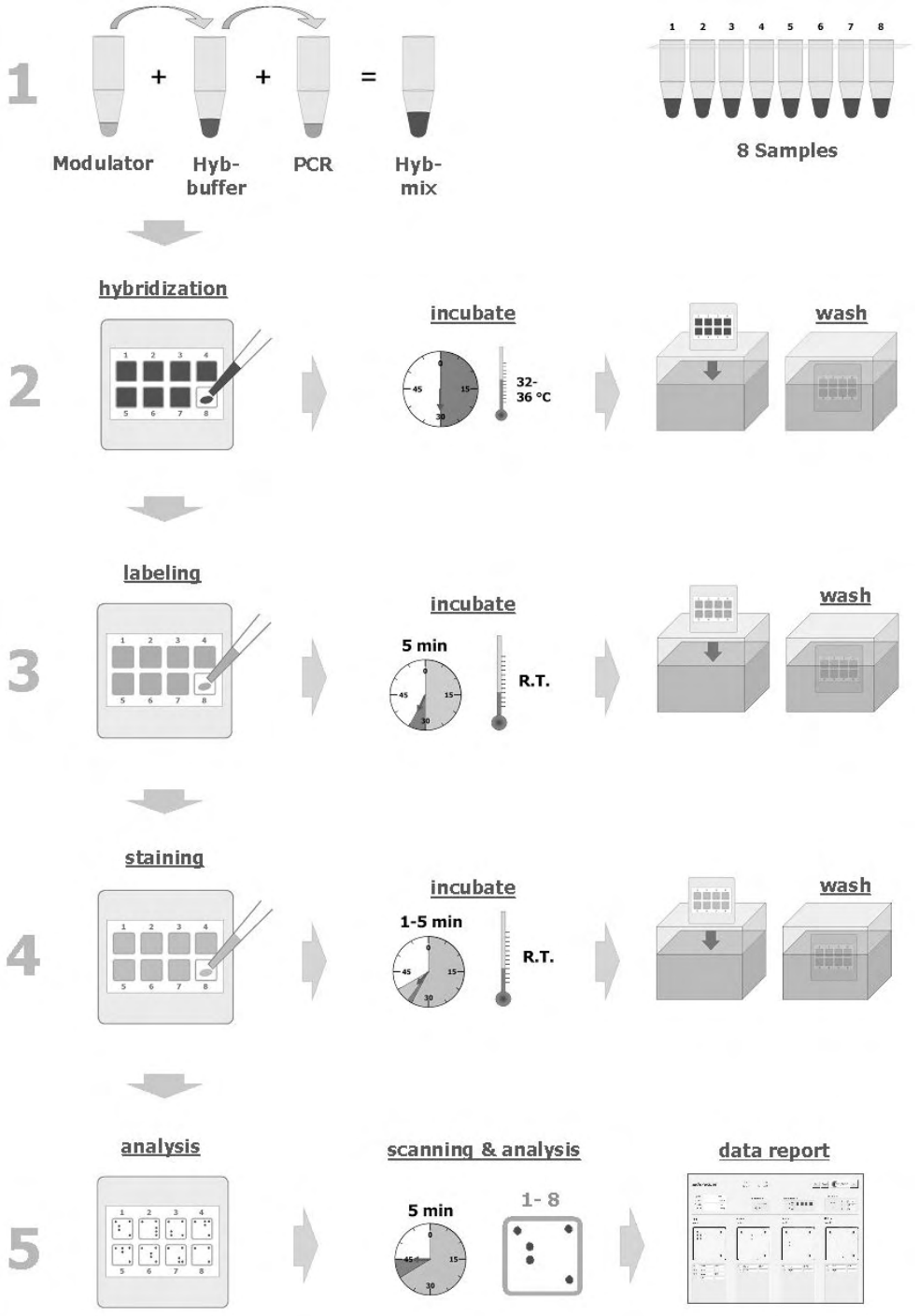

Figure 7: Workflow for hybridisation, washing and staining steps of the LCD-chip.

2.2.7 Sequencing of universal coronavirus PCR products and phylogenetic analysis

All universal CoV nested RT-PCR products were directly sequenced, in both orientations, after purification with AMPure beads (Agencourt Bioscience, Germany) according to the protocol described in section 2.1.6. The primers used for sequencing were PCSseq (5'-CTT ATG GGT TGG GAT TAT CC-3') and PCASseq (5'-TCA TCA GAT AGA ATC ATC A-3'). The sequences were analysed in a CEQ 2000 DNA Analysis System (Beckman Coulter, Germany), according to the procedure described in section 2.1.5. The generated sequences were phylogenetically analysed by the neighbour-joining method in the TREECON software version 1.3b.

2.3 Prevalence, types, and RNA concentrations of HPeVs in patients with acute enteritis, including a sixth parechovirus type

2.3.1 Patients and samples

A cell culture supernatant from MA-104 cells (African green monkey kidney cells) which showed a cytopathogenic effect (CPE) similar to that caused by enteroviruses was obtained from a municipal health service. Serotyping with reference serum pools A - H from the Statens Serum Institute, Copenhagen, was negative (Melnick and Wimberly, 1985). Enterovirus RT-PCR was negative.

Stool samples from patients with acute enteritis (n = 118) were obtained from the routine diagnostic laboratory of the same institution. Young children in this sub-cohort were hospitalized in municipal hospitals. All other patients were not hospitalized, but were subjected to enteritis investigations in outbreaks, and, depending on their age, these samples were originated from kindergartens, catering kitchens, and retirement homes.

Another set of 538 stool samples was collected from January 1 to December 31 of the year 2004, in a prospective study on acute, community-acquired diarrhoea. All patients were outpatients seen by general practitioners. Diarrhoea in these patients was defined as excretion of at least two loose and malodorous stools during 24 hours in breastfed infants, and at least two loose stools in a 24 hour period for all other patients. Patients were excluded if they had inflammatory bowel disease, celiac disease, cystic fibrosis, food intolerance, or else a known malignant disease. The cohort included stool samples from 39 control patients of compatible age distribution with conditions other than enteritis. Written informed consent was obtained from all patients or parents. The local ethics committee approved study protocol and data handling. All stool samples were stored without additives at -20°C for up to two years before analysis.

2.3.2 Cell culture

Subconfluent monolayers of MA-104 green monkey kidney cells or Vero cells were inoculated on 6-well plates with 100 µl of stool suspensions. Suspensions were prepared by diluting stool 1:10 in Phosphate Buffered Saline (PBS) by continuously vortexing for 1 minute, followed by centrifugation at 13,000 rpm for 1 minute, and filtration of supernatant through a 0.22 µm sterile syringe filter. Inoculated supernatants were taken off the cells after 3 hours, and fresh Dulbecco's Modified Eagle's Medium (DMEM) with 10% foetal calf serum (PAA Laboratories, Cölbe, Germany) was replaced. Cultures were observed for CPE over 10 days.

2.3.3 Preparation of stool samples for RT-PCR

RNA was extracted from stool samples stored at -20°C with the QIAamp DNA Stool Mini Kit (QIAGEN, Hilden, Germany). For each sample, an input of about 200 mg of solid stool or 200 µl of liquid stool was placed in a clean 2 ml microcentrifuge tube and 1.4 ml of Lysis Buffer ASL was added to the sample. The tube was vortexed continuously for 1 minute or else up to total homogenization of the stool sample. The suspension was then incubated for 5 minutes at 70°C, vortexed for 15 seconds and centrifuge at 14,000 rpm for 1 minute to pellet stool particles. Then, 1.2 ml of the supernatant was placed into a new 2 ml microcentrifuge tube, following the addition of 1 InhibitEX tablet and immediately vortexing continuously for 1 minute or until the tablet had been completely suspended. The suspension was incubated for 1 minute at room temperature to allow inhibitors to adsorb to the InhibitEX matrix. The tube was then centrifuged at 14,000 rpm for 3 minutes to pellet inhibitors bound to InhibitEX. All the supernatant was placed in a clean 1.5 ml microcentrifuge tube and centrifuged again at 14,000 rpm for 3 minutes. Next, 200 µl of the supernatant was placed in a clean 1.5 ml microcentrifuge tube containing 15 µl of Proteinase K, following the addition of 200 µl of Buffer AL and vortexing for 15 seconds. The tube was then incubated at 70°C for 10 minutes. After incubation, 200 µl of absolute ethanol was added to the lysate and mixed by vortexing for 10 seconds. The mix was applied to the QIAamp spin column in a 2 ml collection tube and centrifuged at 14,000 rpm for 1 minute. The column was placed in a new 2 ml collection tube and washed with 500 µl of Buffer AW1 by centrifugation at 14,000 rpm for 1 minute. The column was washed again with 500 µl of Buffer AW2 by centrifugation at 14,000 rpm for 3 minutes, to completely dry the column membrane. The column was then placed in a clean 1.5 ml microcentrifuge and 200 µl of elution Buffer AE was applied directly onto the column membrane for nucleic acid elution. The column was incubated at room temperature for 1 minute. The eluted nucleic acid was then collected by centrifugation at 14,000 rpm for 1 minute. Effectively, 1/10th of the input, about 20 mg, was brought into the elution volume. It should be noted that the kit was specified for recovery of DNA, but in our hands viral RNA was co-purified with good efficiency (Drosten *et al.*, 2003).

In a later phase of the study, we used the QIAampViral RNA Mini Kit (QIAGEN, Hilden, Germany) in order to save costs. It was noticed that this kit extracted viral RNA with high efficiency from 100 mg or 100 µl stool samples pre-diluted 1:10 in PBS. The suspensions were vortexed, centrifuged, and 140 µl of supernatant was extracted according to protocol described in section 2.1.2. Elution volume was 100 µl. Efficiently, RNA from 10 mg of stool was brought into the elution volume.

2.3.4 Methods used for pre-testing of stool samples

Pre-testing of stool samples was performed in the municipal health service at the Hamburg Institute of Hygiene and Environment. IDEIA^{TM} Rotavirus, AdV and Astrovirus antigen enzyme immunoassays (Dako Cytomation, Cambridgeshire, UK) were used as recommended by the manufacturer. Nested RT-PCR for Norovirus was done as previously described (Oh *et al.*, 2003).

Nested RT-PCR for enteroviruses followed a very similar protocol described by Nijhuis *et al.* (2002). Some stool samples were analysed in our lab, in collaboration with the Hamburg Institute of Hygiene and Environment. For enterovirus nested RT-PCR, the QIAGEN OneStep RT-PCR Kit (QIAGEN, Hilden, Germany) was used in a 25 µl volume reaction containing RNase-free water, 1X QIAGEN OneStep RT-PCR Buffer (containing 2.5 mM of $MgCl_2$), 400 µM of each dNTP, 1 µg of BSA (Sigma, Munich, Germany), 400 nM of primers RHI-1 (5'-CGG TAA YTT TGT ACG CCA GTT-3'), RHI-2 (5'-ACA CGG ACA CCC AAA GTA-3') and MET21C (5'-ATG TAC YTT TGT ACG CCT GTT-3'), 1 µl of QIAGEN OneStep RT-PCR Enzyme Mix and 5 µl of extracted RNA. Thermal cycling conditions involved 30 minutes at 50°C; 15 minutes at 95°C; followed by 35 cycles of 30 seconds at 95°C, 30 seconds at 52°C, and 30 seconds at 72°C. The nested PCR was prepared in a volume reaction of 50 µl, containing Nuclease-free water, 1X PCR Buffer and 2.5 mM $MgCl_2$ (Invitrogen, Karlsruhe, Germany), 200 µM of each dNTP (Applied Biosystems), 400 nM of primers RHI-3 (5'-CAA GCA CTT CTG TTT CCC CGG-3), RHI-4 (5'-CAT TCA GGG GCC GGA GGA-3'), 1 U of Platinum *Taq* DNA polymerase (Invitrogen, Karlsruhe, Germany), and 1 µl of the first amplification reaction mix. Thermal cycling conditions involved 3 minutes at 94°C and 35 cycles of 30 seconds at 94°C, 30 seconds at 64°C, and 30 seconds at 72°C. The nested RT-PCR thermal cycling conditions were performed in a Mastercycler epgradient instrument (Eppendorf).

2.3.5 Virus-discovery-cDNA-AFLP (VIDISCA) analysis

VIDISCA was performed as described by van der Hoek *et al.* (2004), with minor modifications. Figure 8 shows a schematic illustration of the entire procedure.

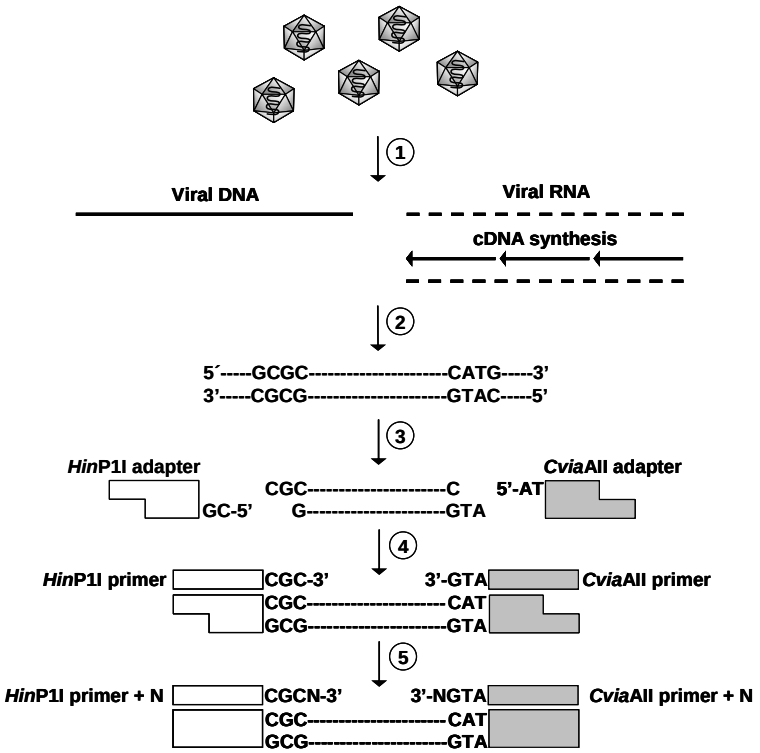

Figure 8: Schematic view of VIDISCA, according to van der Hoek *et al.* (2004). 1) enrichment of viral particles, genomic and mitochondrial DNA digestion from lysed cells and viral nucleic acid extraction; 2) double-stranded cDNA synthesis; 3) ligation of adaptors; 4) preamplification; 5) selective amplification with 16 primer combinations.

2.3.5.1 Viral nucleic acid extraction

A total of 10 ml of supernatant from the MA-104 cells showing CPE was cleared by centrifugation at 8,000 rpm for 10 minutes in a centrifuge and subsequently filtered through a 0.22 µm sterile filter. For enrichment of viral particles, supernatant thereof was centrifuged at 38,000 g for 4 hours in an ultracentrifuge Optima™ L-90 K BioSafe Centrifuge System (Beckman Coulter, Germany). The supernatant was discarded, and pellet was resuspended in 100 µl of 1X DNase 1 buffer (Ambion, Austin, USA). To remove genomic and mitochondrial DNA from lysed cells, the resuspention was treated with 2 U of DNase 1 (Ambion, Austin, USA) at 37°C for 1 hour. The viral nucleic acid is protected by the nucleocapsid during DNase digestion. After digestion, the volume reaction was placed in a clean 1.5 ml microcentrifuge tube and mixed with 2 volumes of 1:1 Roti-Phenol and Chloroform (Carl Roth, Karlsruhe, Germany) solution and vigorously vortexed for 10 seconds. The homogenized solution was incubated at

room temperature for 5 minutes to permit the complete dissociation of nucleoprotein complexes. The solution was centrifuged at 13,000 rpm for 5 minutes in a microcentrifuge. The upper aqueous phase was transferred to a new 1.5 ml microcentrifuge tube and mixed with 1 volume chloroform solution by vortexing for 10 seconds and centrifuged at 13,000 rpm for 5 minutes. The aqueous layer was then transferred to a new 1.5 ml microcentrifuge tube. For nucleic acid precipitation, 1/10th volume of 3M sodium acetate buffer solution (pH 5.2) (Sigma, Germany) was added to the aqueous phase plus 2.5 volumes of absolute ethanol. The solution was homogenized by vortexing for 10 seconds and centrifuged at 13,000 rpm for 30 minutes at room temperature. The supernatant was carefully discarded, and pellet was washed in 500 µl of 70% ethanol by vortexing and centrifugation at 13,000 rpm for 5 minutes. The supernatant was carefully discarded, and the pellet was air-dried for 10 minutes. The pellet was resuspended in 50 µl of Diethyl pyrocarbonate (DEPC)-treated water (Invitrogen, Karlsruhe, Germany).

2.3.5.2 Reverse transcription and double-stranded cDNA synthesis

The extracted viral nucleic acid was submitted to double stranded cDNA synthesis with the SuperScript Double-Stranded cDNA Synthesis Kit (Invitrogen, Karlsruhe, Germany) as described below. If the investigated virus has a DNA genome, it is not affected by this reaction step.

2.3.5.2.1 First strand cDNA synthesis

For the first strand cDNA synthesis, 1 µl (250 ng) of random hexamer primers (Roche, Mannheim, Germany) was added to 10 µl of resuspended viral nucleic acid. The mix was heated at 70°C for 10 minutes and rapidly chilled on ice. The reaction was then performed in a final volume of 20 µl containing additional DEPC-treated water, 1X First Strand Reaction Buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 10 mM of DTT, 500 µM of each dNTP, and 200 U of SuperScript II RT enzyme. The reaction mix was submitted at 25°C for 15 minutes, followed by 45°C for 1 hour. The reaction was terminated placing the tube on ice.

2.3.5.2.2 Second strand cDNA synthesis

The 20 µl reaction volume from the first strand synthesis was submitted to second strand synthesis of cDNA in a final volume of 150 µl containing Nuclease-free water, 1X Second Strand Reaction Buffer (20 mM Tris-HCl pH 6.9, 90 mM KCl, 4,6 mM $MgCl_2$, 0.15 mM β-NAD+, 10 mM $(NH4)_2SO_4$), 200 µM of each dNTP, 10 U of *E. coli* DNA ligase, 40 U of *E. coli* DNA polymerase, and 2 U of *E. coli* RNase H. All reagents were supplied by the kit. The

reaction mix was incubated for 2 hours at 16°C and immediately afterwards was placed on ice. Next, 10 U of T4 DNA polymerase was added; the reaction mix continued to incubate at 16°C for more than 5 minutes. The reaction was then placed on ice for DNA purification.

2.3.5.3 Purification of double-stranded cDNA

The double-stranded cDNA purification was performed with the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). The 150 µl volume from second strand cDNA synthesis reaction was increased with PBS up to 200 µl and transferred to a clean 1.5 ml microcentrifuge tube containing 20 µl of Proteinase K. Next, 200 µl of Lysis Buffer AL was added to the sample and mixed by vortexing for 15 seconds. The mix was incubated at 56°C for 10 minutes. After incubation, 200 µl of absolute ethanol was added to the sample and mixed again by vortexing for 15 seconds. After mixing, the sample was applied to the QIAamp spin column (in a 2 ml collection tube) and centrifuged at 8,000 rpm for 1 minute. The column was placed in a clean 2 ml collection tube and washed with 500 µl of Buffer AW1 by centrifugation at 8,000 rpm for 1 minute. A second wash step was made with 500 µl of Buffer AW2 by centrifugation at 13,000 rpm for 3 minutes. The completely dried column was then placed in a clean 1.5 ml microcentrifuge tube, and 60 µl of Nuclease-free water was applied at the centre of the column membrane for DNA elution. After incubation at room-temperature for 1 minute, the cDNA elution was collected by centrifugation at 8,000 rpm for 1 minute.

2.3.5.4 Enzymatic digestion

The purified double-stranded cDNA was submitted to a double digestion with the restriction enzymes *Hin*P1I, as described in the original protocol (van der Hoek *et al.*, 2004), and *Cvi*AII instead of *Mse*I in order to optimise the 3' end of the primer used for amplification later on. Both enzymes were supplied by New England Biolabs (Frankfurt, Germany).

Double digestion was performed in a 70 µl volume reaction containing the 60 µl volume of the purified double-stranded cDNA, additional Nuclease-free water, 1X NEBuffer 4 (New England Biolabs, Frankfurt, Germany), and 5 U of *Cvi*AII enzyme. The reaction mix was incubated at 25°C for 1 hour. Then, 10 U of *Hin*P1I was added and the reaction was incubated at 37°C for more than 1 hour. The digestion was stopped by inactivation of both enzymes raising the temperature to 65°C for 20 minutes. The digested DNA was phenol-chloroform purified as described in section 2.3.5.1. Then it was resuspended in 38 µl of Nuclease-free water.

2.3.5.5 Ligation of adapters to the double-digested cDNA

The adapters were prepared by mixing 200 nM of the oligonucleotides U-*Hin*P1I (5'-GAC GAT GAG TCC TGA T-3') and D-*Hin*P1I (Phosphate-CGA TCA GGA CTC AT-3') for *Hin*P1I-site; and 200 nM of U-*Cvi*AII (5'-CTC GTA GAC TGC GTA CG-3') and D-*Cvi*AII (Phosphate-ATC GTA CGC AGT C-3') for *Cvi*AII-site. The adapters were prepared in different tubes for each adapter enzyme site by heating the mix to 70°C for 10 minutes and then placed at room temperature for complete hybridization. The ligation of adapters to double-digested cDNA was performed in a 50 µl reaction containing the 38 µl volume of the purified double-digested cDNA, additional Nuclease-free water, 600 nM of each *Hin*P1I and *Cvi*AII adapters, 1X DNA Ligase Buffer and 5 U of T4 DNA Ligase enzyme (Roche, Mannheim, Germany). The reaction was incubated at 37°C for 2 hours. The ligation products were then purified with the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany), according to the protocol described in the section 2.3.5.3, and eluted in a volume of 50 µl with Buffer EB (10 mM Tris-Cl, pH 8,5).

2.3.5.6 Amplification of adapter-ligated double-stranded cDNA fragments

The preamplification reaction was performed in a 50 µl volume reaction containing Nuclease-free water, 1X PCR Buffer and 2.5 mM $MgCl_2$ (Invitrogen, Karlsruhe, Germany), 200 µM of each dNTP (Applied Biosystems, Germany), 300 nM of primers *Cvi*AII (5'-CTC GTA GAC TGC GTA CGA TG-3') and *Hin*P1I (5'-GAC GAT GAG TAC TGA TCG C-3'), 1 U of Platinum *Taq* DNA polymerase (Invitrogen, Karlsruhe, Germany) and 5 µl of adapters-cDNA fragments. Thermal cycling conditions involved 3 minutes at 94°C, followed by 20 cycles of 94°C for 30 seconds, 56°C for 1 minute and 72°C for 1 minute.

The second amplification, or selective amplification, used four variants of each of the aforementioned primers, containing single nucleotide extensions of A, T, G, or C, respectively, at their 3' ends. The resulting 16 different combinations of forward and reverse primers were each used in a 50 µl volume reaction containing the same reagents used in the preamplification reaction, with addition of 300 nM of each forward and reverse primers, and 2 µl of the preamplification PCR mix. Thermal cycling conditions involved 3 minutes at 94°C, followed by a touchdown protocol of 10 cycles of 94°C for 30 seconds, 65°C for 30 seconds (temperature decreasing 1°C per cycle), 72°C for 1 minute, and 25 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 72°C for 1 minute.

2.3.5.7 Analysis of VIDISCA products

The amplified products were analysed by 1.5% Tris-Acetate-EDTA (TAE) buffer agarose gel electrophoresis. The selected fragment was gel purified with the QIAEX II Gel Extraction Kit (QIAGEN, Hilden, Germany) as follows: approximately 100 mg of agarose gel containing a 188 base pairs VIDISCA product was excised from the gel with a clean scalpel and placed in a clean 1.5 ml microcentrifuge tube. Next, 350 µl of Buffer QX1 was added, followed by the addition of 10 µl of QIAEX II silica particles for DNA adsorption. The mix was incubated at 50°C for 10 minutes to solubilise the agarose gel and bind the VIDISCA product to the silica particles. During incubation, the tube was vortexed every 2 minutes to keep QIAEX II particles in suspension. After incubation, the sample was centrifuged at 13,000 rpm for 30 seconds, and supernatant was removed. The pellet was washed with 500 µl of Buffer QX1 by vortexing the sample for 5 seconds. Pellet was collected again, by centrifugation at 13,000 rpm for 30 seconds and removal of supernatant, and washed twice with 500 µl of Buffer PE by mix vortexing, centrifugation and removal of supernatant as described in the first wash step. The pellet was then air-dried for 15 minutes, resuspended by vortexing in 20 µl of elution Buffer EB (10 mM Tris-Cl, pH 8.5) and incubated at room temperature for 5 minutes. The tube was centrifuged at 13,000 rpm for 30 seconds; the supernatant containing the eluted VIDISCA product was placed in a new clean tube.

The sequencing analysis was performed directly with the purified VIDISCA product in both orientations, with the respective selective primers *Cvi*AII-G and *Hin*P1I-A. The dye termination reaction as well as the subsequent purification was performed according to the protocols described in section 2.1.5. The samples were loaded into a CEQ 2000 DNA Analysis System (Beckman Coulter, Germany). The generated sequences were assembled with the sequence analysis software SeqMan II version 5.08 (Lasergene DNASTAR Software Pack), and analysed by the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1997), with the blastn algorithm.

2.3.6 HPeV 5' nuclease (TaqMan) broad-range real-time RT-PCR assay

The primers and probes of the assay were designed to hybridize to the most conserved region found in the 5' UTR of HPeVs, after an alignment of all four genotypes available at GenBank in November 2005.

The reaction was performed with the QIAGEN OneStep RT-PCR Kit (QIAGEN, Hilden, Germany). The reaction was carried out in 50 µl of volume reaction containing RNase-free water, 1X QIAGEN OneStep RT-PCR Buffer (containing 2.5 mM of $MgCl_2$), 400 µM of each

dNTP, 1 µg of BSA (Sigma, Munich, Germany), 600 nM of primers HPS (5'-GTG CCT CTG GGG CCA AAA G-3'), HPA (5'-TCA GAT CCA TAG TGT CGC TTG TTA C), and 200 nM of probe HPP (5'-FAM-CGA AGG ATG CCC AGA AGG TAC CCG T-TAMRA-3'), 1 µl of QIAGEN OneStep RT-PCR Enzyme Mix and 5 µl of extracted RNA. Thermal cycling conditions in a ABI 7700 SDS instrument (Applied Biosystems, Germany) involved 30 minutes at 50°C, 15 minutes at 95°C, followed by 45 cycles of 95°C for 15 seconds and 58°C for 30 seconds.

For calculation of approximate virus RNA concentrations in stool samples, we assumed 100% efficiency of RNA recovery in the purification step. The projected amount of stool tested per PCR vial, receiving 5 µl of RNA eluate, was 0.5 µg or 0.5% µl.

2.3.7 Synthesis of parechovirus *in vitro* RNA transcript

The PCR fragment of the real-time RT-PCR screening assay as obtained from the HPeV strain BNI-788St was ligated into a pCR 2.1-TOPO plasmid and cloned in TOP10F' *E. coli* by means of the TOPO TA Cloning Kit (Invitrogen, Karlsruhe, Germany). The plasmids were purified, sequenced and re-amplified with plasmid-specific primers M13 Forward (-20) and M13 Reverse, from the kit, according to the protocols described in the subtopics of section 2.1.4.

Reamplification products were transcribed into RNA with the MEGAscript T7 Kit (Ambion, Austin, USA). After DNase 1 digestion, RNA transcripts were purified with the RNeasy Mini Kit (QIAGEN, Karlsruhe, Germany), according to the protocols described in section 2.1.6, and quantified in a BioPhotometer (Eppendorf, Germany). The *in vitro* RNA transcripts were used as RNA standards for PCR quantification. By RT-PCR with and without RT enzyme, the RNA/DNA ratio in the preparation was determined to be 10^6.

2.3.8 Sequencing of P1 and 3D genes from detected parechoviruses

The P1 gene sequences were generated directly from RT-PCR products, using upstream primers in the 5' UTR region and a downstream primer in the end of VP1 region. The 3C/3D sequences were obtained by first amplifying the highly conserved distal segment of the 3D gene from all viruses. The obtained fragments served as template for the design of specific reverse primers for each strain. These were combined with sets of candidate upstream primers as derived from an alignment of all prototype strains available in GenBank in November 2005. Candidate primer pairs were designed to allow amplification of about 3,000 base pair products.

The cDNA were synthesised using the SuperScript III Reverse Transcriptase reagents (Invitrogen, Karlsruhe, Germany) in a 20 µl volume reaction containing RNase-free water, 1X

First-Strand Buffer (50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM $MgCl_2$), 500 µM of each dNTP, 10 mM of DTT, 50 µM of oligo(dT) primer and 20 U of RNaseOUT (Invitrogen, Karlsruhe, Germany), 200 U of SuperScript III RT and 3 µl of extracted RNA. The reaction mix was incubated at 50°C for 1 hour, followed by 70°C for 15 minutes. After reverse transcription, 2 U of *E. coli* RNase H (Invitrogen, Karlsruhe, Germany) was added to the mix and incubated at 37°C for 20 minutes.

Long-range PCR was done with the Platinum *Taq* DNA Polymerase High Fidelity reagents (Invitrogen, Karlsruhe, Germany), in a 50 µl volume reaction containing Nuclease-free water, 1X High Fidelity PCR Buffer, 200 µM of each dNTP (Applied Biosystems, Germany), 2 mM of $MgSO_4$, 200 nM of specific forward and reverse primers, 1 U of Platinum *Taq* DNA polymerase High Fidelity and 2 µl of cDNA. Thermal cycling conditions involved 95°C for 2 minutes, followed by a touchdown protocol of 10 cycles at 94°C for 30 seconds, 70°C for 30 seconds (decrease of 1°C per cycle), 72°C for 3,4 minutes, and 35 cycles at 94°C for 30 seconds, 60°C for 30 seconds, and 72°C for 3,4 minutes.

The obtained products from successful long-range amplifications were cloned into pCR4-TOPO vector (Invitrogen, Karlsruhe, Germany), in a very similar protocol as described in section 2.1.4., and sequenced by primer walking technique.

2.3.9 Full genome sequencing of HPeVs BNI-788St and BNI-67

The identified HPeV BNI-788St by VIDISCA method and BNI-67 by the broad range real-time PCR have been submitted for sequencing of the entire genome. The extracted RNAs were submitted to RT with oligo(dT) primer using the SuperScript III Reverse Transcriptase reagents (Invitrogen, Karlsruhe, Germany) as described above.

Based on the full genome information of published HPeVs available at GenBank in November 2005, a combination of consensus primers were designed to amplify fragments covering the complete genome. The first determined sequences from RT-PCR fragments served as template for the design of new specific primers for each strain, and to allow the amplification of fragments of about 3,000 base pairs. Long-range PCR reactions were performed as described above; remaining genome sequences were obtained by the primer walking method.

2.3.9.1 Amplification and sequencing of the 5’ UTR

Two strategies were used to sequence the 5’ end of the BNI-788St and BNI-67 viruses. First, the 5’ end was sequenced from cloned RT-PCR products using specific reverse primers and different combinations of consensus forward primers composed of the first 13, 16, 19, 24 and 27 nucleotides of all HPeV sequences deposited in GenBank. Second, the 3’ end of newly synthesized first-strand cDNA from 5’ UTR end was extended by homopolymeric tailing reaction using dATP, to form a poly(A) anchor, with the use of a terminal deoxynucleotidyl transferase enzyme or terminal transferase (TdT).

The cDNA from 5’ ends of both HPeVs were synthesised using the SuperScript III Reverse Transcriptase reagents (Invitrogen, Karlsruhe, Germany) in a similar protocol described above, using the conserved primer 562Rev (5’-CGG GTA CCT TCT GGG CAT CCT TCG-3’), and reverse transcription temperature of 55 °C. The cDNA was purified as described in section 2.3.5.1, and resuspended in 30 µl of Nuclease-free water.

The tailing reaction was performed with New England Biolabs reagents (Frankfurt, Germany) in a 50 µl volume reaction containing Nuclease-free water, 1X NEBuffer 4, 250 µM of $CoCl_2$, 100 µM of dATP (Applied Biosystems, Germany), 10 U of TdT enzyme, and the 30 µl volume of purified cDNA. The mix was incubated at 37°C for 30 minutes, and 70°C for 10 minutes. The tailed cDNA was then phenol/chloroform purified and resuspended in 30 µl of Nuclease-free water.

The purified tailed cDNA was submitted to a hemi-nested PCR using the Platinum *Taq* DNA polymerase reagents (Invitrogen, Karlsruhe, Germany), and 50 µl of volume reaction in both amplification steps. The reaction mixes contained Nuclease-free water, 1X PCR Buffer, 2.5 mM of $MgCl_2$, 200 µM of each dNTP, and 1 U of Platinum *Taq* DNA polymerase. The thermal cycling conditions in a Mastercycler epgradient instrument (Eppendorf, Germany) involved 3 minutes at 94°C, followed by 30 cycles of 90°C, 50°C and 70°C for 30 seconds to each temperature. The first round amplification contained 200 nM of primers 5'UTR6608Fwd (5’-AGG CCG TGG AAG TCT TGG CTT ATT GTC TTT TTT TTT TTT TTT TTT-3’) and 562Rev (5’-CGG GTA CCT TCT GGG CAT CCT TCG-3’), and 5 µl of purified and tailed cDNA. The second round amplification contained 200 nM of primers 6608Fwd (5’-AGG CCG TGG AAG TCT TGG CTT ATT GTC-3’) and 562Rev (5’-CGG GTA CCT TCT GGG CAT CCT TCG-3’), and 1 µl of first round amplification mix.

2.3.9.2 Amplification and sequencing of the 3’ UTR

The 3’ ends of BNI-788St and BNI-67 viruses were amplified using a 3’ RACE strategy according to instructions of the GeneRacer Kit (Invitrogen, Karlsruhe, Germany). The reverse transcription was performed using the GeneRacer oligo(dT) primer and Superscript RT reagents as described above. The first-strand cDNAs were submitted to a hemi-nested PCR amplification using the Platinum *Taq* DNA polymerase reagents (Invitrogen, Karlsruhe, Germany) in a 50 µl volume reaction in both amplification steps. The reactions mixes contained Nuclease-free water, 1X PCR Buffer, 2.0 mM of $MgCl_2$, 200 µM of each dNTP, and 1 U of Platinum *Taq* DNA polymerase. The thermal cycling conditions for both amplification steps in a Mastercycler instrument involved 3 minutes at 94°C, followed by a touchdown protocol of 10 cycles at 94°C, 70°C (with decrease of 1°C per cycle), and 72°C, during 30 seconds for each temperature, followed by 35 cycles of 90°C, 60°C and 70°C, 30 seconds for each temperature. The first round amplification contained 200 nM of strain-specific forward primer, which hybridize to the terminal portion of 3D gene, 600 nM of GeneRacer 3’ primer and 2 µl of cDNA. The second round amplification contained 200 nM of strain-specific forward primer and GeneRacer 3’ nested primer, and 1 µl of first round amplification mix. The amplicons from second round amplification were directly sequenced in both orientations with respective strain-specific forward primers and GeneRacer 3’ nested primer.

2.4 Universal amplification of single-stranded, positive RNA viruses with modified PCR primers aimed to the conserved GDD motif of RdRp

2.4.1 The +RNA GDD virus discovery method

The +RNA GDD virus discovery method was designed to amplify fragments selected by the nucleotide sequences of the conserved glycine-aspartic acid-aspartic acid or GDD motif found in the RdRp of almost all viruses with single-stranded and positive polarity genomes, in a two steps nested RT-PCR reaction. The primers used in the reaction are composed of standard nucleotides, and the non-standard nucleotides isoG and isoC. The reverse primer +RNA (5'-CGG tGt CCC aGt CaC Gat tgt cgt ckc c-3') was designed to hybridize preferentially to the nucleotides of the GDD motif. The non-standard isoG and isoC nucleotides in the +RNA primer sequence are indicated in upper-case letters. In this primer, the first 13 nucleotides of the 3' end are composed of standard nucleotides, and are followed by a sequence rich in the non-standard nucleotides isoG and isoC (~69% of isoG/C content). During the reverse transcription reaction, the 3' end of +RNA primer hybridizes to the nucleotides of GDD motif, whereas the isoG/C-rich tract remains non-hybridized by base pair mismatch. After cDNA synthesis, the +RNA primer is used again in the first round PCR amplification together with other modified, forward primer, called Back (5'-tGC Cta GCa CtG aCC CtG ttt agc ggc cgc ggg-3'). Back primer has the same conformation of the reverse primer +RNA. Their first 15 nucleotides of 3' end are composed by standard nucleotides in a G/C-rich sequence (~73% of G/C content), followed by an isoG/C-rich tract (~61% of isoG/C content). Back primer was designed to hybridize elsewhere upstream from +RNA primer biding site and may hybridize non-specifically in low stringency conditions. The second round PCR amplification was performed initially with the forward primer Uni1 (5'-CGG tGt CCC aGT CaC Gat-3') and reverse primer Uni2 (5'-tGC Cta GCa CtG aCC CtG t-3'), which are manly composed of isoG/C nucleotides, that hybridize to the isoG/C portion of the +RNA and Back primers, respectively. Finally, in the second round or nested amplification the primers Uni1 and Uni2 were substituted for the standard forward primer +RNA-*Not*I (5'-TTA GCG GCC GCA TTG TCG TCK CC-3') and reverse primer Back-*Not*I (5'-ACT GAC TGT TTA GCG GCC GCG GG-3'). The primer +RNA-*Not*I have the same nucleotide sequence in the first 12 nucleotides of the modified primer +RNA, followed by a nucleotide recognition site for *Not*I restriction enzyme, whereas the primer Back-*Not*I have the first 18 nucleotides in common with the modified primer Back, including the recognition site for *Not*I enzyme. This substitution was important to enable cloning of amplicons in TOPO TA-based vectors or after digestion with *Not*I restriction enzyme.

2.4.2 Optimization of the two steps nested RT-PCR +RNA GDD virus discovery method

2.4.2.1 DNA polymerase for isoG-isoC base pair

To implement the GDD method PCR procedure, we have tested two available *Taq* DNA polymerases in our laboratory, the Titanium *Taq* DNA polymerase (Clontech, USA), a nuclease-deficient, N-terminal truncated mutant or KlenTaq DNA polymerase, and the Platinum *Taq* DNA polymerase (Invitrogen, Karlsruhe, Germany). The KlenTaq DNA polymerase was described to improve fidelity catalysis during PCR (Barnes, 1992) and also to improve the fidelity of isoG:isoC hybridization (Johnson *et al.*, 2004b).

2.4.2.1.1 First strand cDNA synthesis

The first strand cDNA synthesis reaction was performed using RNA extracts from cell cultured HPeV BNI-788St supernatant and a clinical sample previously tested to contain HCoV-OC43 RNA. Both RNA extracts were submitted to reverse transcription with random hexamer primers (P[dN]$_6$) (Roche, Germany) or +RNA reverse primer in two separated reactions, according to the Table 5.

Table 5: reverse transcription of RNA extracts with primers +RNA and P[dN]$_6$.

Genomic RNA	Primer	
	+RNA	P[dN]$_6$
HPeV BNI-788St	1	3
HCoV-OC43	2	4

Reverse transcription reactions were carried out with SuperScript III Reverse Transcriptase reagents (Invitrogen, Karlsruhe, Germany) in a 20 µl volume reaction. Each reaction tube contained RNase-free water, 1X First-Strand Buffer, 500 µM of each dNTP, 10 mM of DTT, 20 U of RNaseOUT (Invitrogen, Karlsruhe, Germany), 200 U of SuperScript III RT and 4 µl of extracted RNA. The HPeV BNI-788St and HCoV-OC43 RNA extracts were reverse transcribed with 250 ng of P[dN]$_6$ or 500 nM of +RNA primer in separated tubes. Prior to reverse transcription, a pre-mix containing RNase-free water, dNTP mix, +RNA reverse primer or P[dN]$_6$, and RNA template was incubated at 65°C for 5 minutes and immediately chilled on ice for at least 1 minute, before the addition of buffer and enzymes. The final volume reaction mixes were incubated at 25°C for 5 minutes, followed by 50°C for 1 hour, and 70°C for 15 minutes. The four different cDNA samples were immediately processed for PCR.

2.4.2.1.2 Nested PCR

The nested PCR reactions were performed with Titanium *Taq* PCR Kit (Clontech, USA) and Platinum *Taq* DNA polymerase reagents (Invitrogen, Karlsruhe, Germany) in 25 µl of total volume reaction.

For Titanium *Taq* PCR Kit, the four reaction mixes of each round amplification step contained Nuclease-free water, 1X Titanium *Taq* PCR Buffer (containing 3.5 mM of $MgCl_2$), 200 µM of each dNTP (Applied Biosystems, Germany), 50 µM of each iso-dGTP and iso-dCTP (EraGen, Wisconsin, USA), and 1 U of Titanium *Taq* DNA polymerase.

For Platinum *Taq* DNA polymerase reagents, the four reaction mixes of each round amplification step contained Nuclease-free water, 1X PCR Buffer, 2.5 mM of $MgCl_2$, 200 µM of each dNTP (Applied Biosystems, Germany), 50 µM of each iso-dGTP and iso-dCTP (EraGen, Wisconsin, USA), and 1 U of Platinum *Taq* DNA polymerase.

The first round amplification mixes contained 400 nM of primers +RNA and Back, and 2 µl of cDNA. Thermal cycling conditions in a Mastercycler epgradient instrument (Eppendorf, Germany) involved 3 minutes at 95°C, followed by 40 cycles of 94°C for 30 seconds, 40° to 68°C (increase of 0.7°C per cycle) for 30 seconds, and 70°C for 90 seconds.

The second round amplification mixes contained 400 nM of primers Uni1 and Uni2, and 2 µl of first round amplification mix. Thermal cycling conditions involved 3 minutes at 95°C, followed by 40 cycles of 94°C for 30 seconds, 52°C for 30 seconds, and 70°C for 90 seconds. The amplified products were subsequently analysed by 1.5% agarose gel electrophoresis.

2.4.2.2 Chikungunya virus as a model organism for the +RNA GDD virus discovery method

The +RNA GDD virus discovery method was next optimized with the use of a high pure and full length *in vitro* transcribed RNA of the chikungunya virus, strain Reunion, synthesised from a clinical sample previously tested in our laboratory to contain the viral chikungunya genomic RNA. The chikungunya viruses have a genome length about 11,800 base pairs. Considering the concentration of the chikungunya Reunion full length *in vitro* RNA transcript being about 0.3 µg/µl, the genome equivalents per µl corresponded to about 4.5 10^{10}. The *in vitro* transcribed RNA was 10-fold diluted, from 10^{-1} to 10^{-7}, and submitted to reverse transcription with primer +RNA.

The reverse transcription reactions were performed with SuperScript III RT reagents (Invitrogen, Karlsruhe, Germany). A pre-mix containing RNase free water, +RNA primer, dNTP mix and RNA template was incubated at 65°C for 5 minutes and immediately chilled on ice. The

final volume reaction was 20 µl and contained RNase-free water, 1X First-Strand Buffer, 500 µM of each dNTP, 10 mM of DTT, 500 nM of +RNA primer, 20 U of RNaseOUT (Invitrogen, Karlsruhe, Germany), 200 U of SuperScript III RT and 4 µl of *in vitro* transcribed RNA. The reaction tubes were incubated at 25°C for 5 minutes, followed by 50°C for 1 hour and 70°C for 15 minutes. The reaction tubes were then placed on ice and immediately processed for nested PCR.

The first and second round amplification reactions were performed with Platinum *Taq* DNA polymerase, in 25 µl of volume reaction. Each tube contained Nuclease-free water, 1X PCR Buffer, 2.5 mM of $MgCl_2$, 200 µM of each dNTP (Applied Biosystems, Germany), 50 µM of each iso-dGTP and iso-dCTP (EraGen, Wisconsin, USA), 1 U of Platinum *Taq* DNA polymerase, and 2 µl of respective cDNA.

The first round amplification contained 400 nM of primers +RNA and Back. Thermal cycling conditions involved 3 minutes at 95°C, followed by 45 cycles of 94°C for 30 seconds, 40° to 67°C (increase of 0.6°C per cycle) for 30 seconds, and 70°C for 90 seconds.

The second round amplification contained 400 nM of primers Uni1 and Uni2. Thermal cycling conditions involved 3 minutes at 95°C, followed by 45 cycles of 90°C for 30 seconds, 50°C for 30 seconds, and 70°C for 90 seconds. The 28 PCR products were analysed by 1.5% agarose gel electrophoresis.

Later on, the products from the first and second round amplification steps were reamplified with standard primers +RNA-NotI and Back-NotI. The reamplification reaction was performed in 50 µl of volume reaction and contained Nuclease-free water, 1X PCR Buffer, 2.5 mM of $MgCl_2$, 200 µM of each dNTP (Applied Biosystems, Germany), 100 nM of primers +RNA-NotI and Back-NotI, 1 U of Platinum *Taq* DNA polymerase, and 2 µl of first or second round amplification mix. Thermal cycle conditions involved 3 minutes at 95°C, followed by 45 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 72°C for 90 seconds. The amplicons were then analysed in 1.5% agarose gel electrophoresis.

Following agarose gel electrophoresis, a fragment of about 500 base pairs, common to all positive reamplified dilutions, was excised from the gel and purified according to the protocol described in section 2.3.5.7. Sequencing analysis was performed directly with the purified product in both orientations, with the primers +RNA-NotI and Back-NotI. The purified products from dye termination reaction were loaded into a CEQ 2000 DNA Analysis System (Beckman Coulter, Germany). The obtained sequences were then assembled in a single contig with the SeqMan II sequence analysis software, version 5.08 (Lasergene DNASTAR), and aligned with the full genomic sequence of the chikungunya Reunion strain.

2.4.2.3 +RNA GDD virus discovery method with cell cultured chikungunya virus

2.4.2.3.1 Genomic RNA extraction

A total of 2 ml of cell cultured chikungunya supernatant containing about 10^9 viral particles per ml was cleared by centrifugation at 8,000 rpm for 10 minutes in a microcentrifuge and filtered through a 0.22 µm sterile filter. Viral particles were than pelleted in a refrigerated ultracentrifuge SIGMA 3-18K (SIGMA, Germany) at 30,000 g for 4 hours at 4°C. After centrifugation, the supernatant was discarded and pellet was resuspended in 100 µl of 1X DNase 1 buffer (Ambion, Austin, USA). For removal of mitochondrial and genomic DNA from lysed cells, resuspended viral particles were treated with 2 U of DNase 1 (Ambion, Austin, USA) at 37°C for 1 hour. After DNA digestion, the volume reaction was placed in a clean 1.5 ml microcentrifuge tube, and the volume was adjusted to 250 µl with PBS buffer. For RNA extraction, 750 µl of TRIzol LS Reagent (Invitrogen, Karlsruhe, Germany), a monophasic solution of phenol and guanidine isothiocyanate, was added and the suspension was homogenized by pipetting up and down several times. The homogenized was then incubated at room temperature for 5 minutes, to permit the complete dissociation of nucleoprotein complexes. Then, 200 µl of chloroform (Carl Roth, Karlsruhe, Germany) were added and the suspension was mixed vigorously by hand for 15 seconds and incubated at room temperature for 15 minutes. The sample was then centrifuged at 13,000 rpm for 15 minutes at 4°C in a refrigerated microcentrifuge. Following centrifugation, the aqueous phase was transferred to a clean tube and mixed with 500 µl of isopropyl alcohol (Carl Roth, Karlsruhe, Germany) by vortexing and was then incubated at room temperature for 10 minutes. After incubation, the mix was centrifuged at 13,000 rpm for 10 minutes at 4°C. Next, the supernatant was carefully discarded and the pellet washed with 1 ml of 75% ethanol by vortexing and centrifugation at 8,000 rpm for 5 minutes at 4°C. Following the last centrifugation, the supernatant was discarded and the RNA pellet was air-dried for 10 minutes at room temperature. The RNA was then resuspended in 20 µl of RNase-free water by pipetting up and down a few times, and incubating the suspension at 55°C for 10 minutes.

2.4.2.3.2 First strand cDNA synthesis

The extracted genomic RNA of chikungunya virus was 10-fold diluted up to 10^{-3}. Five RNA templates, including the original RNA extract, the three dilutions, and a 10^{-5} dilution of full length *in vitro* transcribed RNA were submitted to first strand cDNA synthesis with SuperScript III RT reagents as described above. Following the pre-mix heat, at 65°C for 5 minutes, containing RNase-free water, dNTP mix, +RNA primer and RNA template, the tubes were

immediately chilled on ice for at least 1 minute. Buffer and enzymes were added next and the final 20 µl volume reaction contained RNA-free water, 1X First-Strand Buffer, 500 µM of each dNTP, 10 mM of DTT, 500 nM of +RNA primer, 20 U of RNaseOUT, 200 U of SuperScript III RT and 4 µl of RNA template. The tubes were incubated at 25°C for 5 minutes, followed by 50°C for 1 hour, and 75°C for 15 minutes. After first strand cDNA synthesis, 2 U of RNase H (Invitrogen, Karlsruhe, Germany) was added to each tube and mixes were incubated at 37°C for 20 minutes.

2.4.2.3.3 Nested PCR

The first round amplification reactions were performed in 25 µl volume reaction. Each tube contained Nuclease-free water, 1X PCR Buffer, 2.5 mM of $MgCl_2$, 200 µM of each dNTP (Applied Biosystems, Germany), 50 µM of each iso-dGTP and iso-dCTP (EraGen, Wisconsin, USA), 400 nM of primers +RNA and Back, 1 U of Platinum *Taq* DNA polymerase, and 2 µl of respective cDNA. Thermal cycle conditions involved 3 minutes at 95°C, followed by 45 cycles of 94°C for 30 seconds, 40° to 67°C (increase of 0.6°C per cycle) for 30 seconds, and 72°C for 90 seconds.

The nested amplification reactions were performed in 50 µl of volume reaction. Each tube contained Nuclease-free water, 1X PCR Buffer, 2.5 mM of $MgCl_2$, 200 µM of each dNTP (Applied Biosystems, Germany), 100 nM of primers +RNA-NotI and Back-NotI, 1 U of Platinum *Taq* DNA polymerase, and 1 µl of respective first round amplification mix. Thermal cycle conditions involved 3 minutes at 95°C, followed by 45 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 72°C for 90 seconds. The amplicons were then analysed in 1.5% agarose gel electrophoresis.

2.4.2.4 +RNA GDD virus discovery method with cell cultured HPeV BNI-788St virus

A second virus with single-stranded, positive polarity RNA genome, the HPeV BNI-788St virus, was also tested by the +RNA GDD virus discovery method. The cell cultured supernatant containing the viral particles was cleared, enriched by ultracentrifugation and treated with DNase 1. RNA extraction, reverse transcription and nested PCR procedures were the same as described above, in section 2.4.2.3, for chikungunya virus. The products were then analysed by agarose gel electrophoresis and fragments ranging from 200 to 1,000 base pairs were gel purified and submitted to sequence analysis.

2.4.2.5 Alternative forward primers for the +RNA GDD virus discovery method

In order to avoid unspecific hybridization of forward primer with bacterial G/C-rich sequences as observed with HPeV BNI-788St RNA extracted from cell culture, the primers PC2S2a (5'-TTA TGG GTT GGG ATT ATC-3') and PC2S3a (5'-TTA TGG GTT GGG ATT AT-3') previously designed to hybridize to the conserved motif A of CoVs, which is also found in all RdRp and located upstream from GDD motif, were used. In addition, the primer RV1678F (5'-TCG ACG TGG AAC AGC TTG-3'), specific for chikungunya virus, was used as a control for specific amplification.

For first strand cDNA synthesis, 10-fold dilution, from 10^{-1} to 10^{-8}, of full length chikungunya *in vitro* transcribed RNA were used following the same protocol as described in section 2.4.2.3.2. After reverse transcription reaction, each first-stranded cDNA was submitted to three different hemi-nested PCR amplification mixes. Table 6 shows the forward and reverse primers used in each amplification mix reaction.

Table 6: forward and reverse primers used in each step and amplification mix reaction.

primers	1st round amplification			2nd round amplification		
	Mix 1	Mix 2	Mix 3	Mix 1'	Mix 2'	Mix 3'
Forward	RV1678F	PC2S2a	PC2S3a	RV1678F	PC2S2a	PC2S3a
Reverse	+RNA	+RNA	+RNA	+RNA-*Not*I	+RNA-*Not*I	+RNA-*Not*I

All first round amplification mixes contained, in 25 µl of volume reaction, Nuclease-free water, 1X PCR Buffer, 2.5 mM of $MgCl_2$, 200 µM of each dNTP (Applied Biosystems, Germany), 50 µM of each iso-dGTP and iso-dCTP (EraGen, Wisconsin, USA), 400 nM of forward and reverse primers, 1 U of Platinum *Taq* DNA polymerase, and 2 µl of respective cDNA. Thermal cycle conditions involved 3 minutes at 95°C, followed by 45 cycles of 94°C for 30 seconds, 40° to 67°C (increase of 0.6°C per cycle) for 30 seconds, and 72°C for 90 seconds.

The second round amplification mixes contained, in 50 µl of volume reaction, Nuclease-free water, 1X PCR Buffer, 2.5 mM of $MgCl_2$, 200 µM of each dNTP (Applied Biosystems, Germany), 100 nM of forward and reverse primers, 1 U of Platinum *Taq* DNA polymerase, and 1 µl of respective first round amplification mix. Thermal cycle conditions involved 3 minutes at 95°C, followed by 45 cycles of 94°C for 30 seconds, 54°C for 30 seconds, and 72°C for 90 seconds. The amplicons were then analysed in 1.5% agarose gel electrophoresis. The amplicons were then analysed in 1.5% agarose gel electrophoresis and fragments ranging from 200 to 1,000 base pairs were gel purified and submitted to sequence analysis.

The same protocol described in this section was used next with genomic RNAs from chikungunya and HPeV BNI-788St viruses extracted from cell culture supernatants. The RNA extraction followed the same protocol described in section 2.4.2.3.1. The genomic RNA of both viruses was 10-fold diluted and specimens from 10^0 to 10^{-3} were used for reverse transcription and hemi-nested amplification. The amplicons were then analysed by agarose gel electrophoresis and selected fragments were gel purified and sequenced.

2.5 *In silico* methods and statistical analysis

The Statgraphics Plus Package software version 5.0 (Manugistics, Dresden, Germany) was used for all statistical analysis. Probit analysis was used as its input dataset the cumulative hit rates in parallel reactions and their respective target RNA concentrations. It determined a continuous dose-response relationship (response rate of the assay dependent of the dose of RNA per reaction) with 95% confidence intervals (CI).

The design of oligonucleotides for TaqMan and FRET assays was done with Primer Express software version 2.0 (Applied Biosystems, Weiterstadt, Germany). The stable non-Watson-Crick base pair G:T was not strictly adjusted for (Peyret *et al.*, 1999), in order to enable usage of optimal oligonucleotide annealing sites (Wolfel *et al.*, 2007).

The computational analyses of sequenced nucleotides were performed with specific programs included in the DNASTAR package software version 5.08. Assemble of contigs was performed with SeqMan program. Amino acid and nucleotide alignments were performed with MegAlign program.

Simplot analyses were done with SimPlot software version 3.5.1 (Lole *et al.*, 1999).

Phylogenetic analysis of the partial RdRp nucleotide sequences from amplified human CoVs, P1, 3C/3D from HPeVs, as well as complete genomes of BNI-788St and BNI-67 strains were conducted by the neighbour-joining method in the TREECON software version 1.3b.

3.0 Results

3.1 Spectrum of viruses and atypical bacteria in intercontinental air travellers with symptoms of acute respiratory infection

3.1.1 Patient groups

Travel histories of patients were reconstructed by telephone interviews with hospitals, family physicians, or patients themselves. For 164 patients, additional information was retrieved. This included 124 patients for whom the exact dates and destinations of departure to Germany were known and 22 of them for whom only destinations could be reconstructed. Seventeen patients did not travel, as initially reported, but had been in contact with patients who had suspected SARS. For 8 patients, no additional information could be retrieved. They were included in the cohort, because a travel history had been confirmed before samples were received. Only those patients who had reportedly not travelled were evaluated separately. They are hereafter called “contact patients”; the remaining 155 patients are called “flight patients”.

The average age of the flight patients was 42.2 years, ranging between 1 to 79 years. The average age of contact patients was 41.8 years, ranging between 4 to 79 years. Neither the average nor the median ages nor the standard deviations in both groups were significantly different. Figure 9 shows the age distributions in both groups.

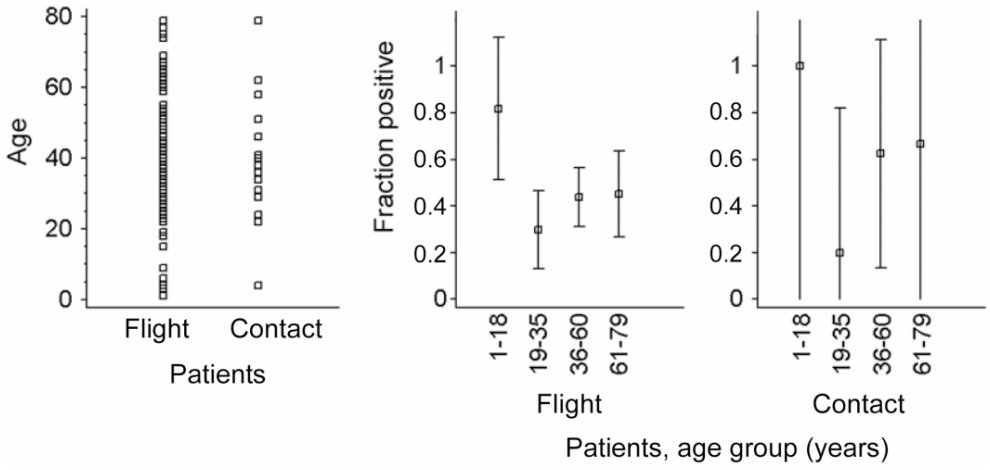

Figure 9: Age distribution of patients (left) and rates of detection of any tested agent in 4 different age groups (middle and right). The age of 15 of 172 patients could not be recorded because these were coded requests forwarded to us by other laboratories. Range lines in the middle and right panels depict 95% confidence intervals (CIs). Note that large CIs in the contacts panel are due to the small number of patients in this group (n = 17). Sample sizes in the different age groups were as follow: In the middle panel (flight patients), 1-18 years, 8 patients; 19-35 years, 37 patients; 36-60 years, 64 patients; and 61-79 years, 31 patients. In the right panel (contact patients), 1-18 years, 1 patient; 19-35 years, 5 patients; 36-60 years, 8 patients; and 61-79 years, 3 patients.

Some of the flight patients (71%) were between 19 and 60 years old, as it were 76% of the contact patients (P = 0.57). Only 12 patients were below 19 years old, 11 of whom were flight patients. They contributed to 13.4% of all positive findings in flight patients, which was significantly more than the mean positivity rate in all age groups (P < .012 Student's t test).

3.1.2 Detection of pathogens

At least 1 pathogen was detected in 67 flight patients (43.2%), and in 8 contact patients (47%) (P = 0.7, Student's t test). Table 7 shows the global pathogen detection rates by age group, as well as the relative and absolute prevalence of pathogens in flight patients.

Table 7: Absolute detection rates of agents in flight patients.

Agent	Detection rate no. (%)
Human parainfluenza virus	24 (15.5)
Inf A	11 (7.1)
Inf B	11 (7.1)
Adenovirus	6 (3.9)
HCoV-OC43/HKU1	5 (3.2)
HCoV-229E	0
HCoV-NL63	0
Human Rhinovirus	4 (2.6)
Human metapneumovirus	4 (2.6)
Respiratory syncytial virus	1 (0.6)
Human boca virus	0
Mycoplasma pneumoniae	1 (0.6)
Legionella ssp.	1 (0.6)
Chlamydophila pneumoniae	0
Total	68 (43.8)

In the 17 contact patients, AdV, PIV, Inf B, and hMPV were detected in 2 patients each and Inf A in 1 patient. Double infections occurred in only 2 patients: HRV and PIV in a flight patient, and AdV and PIV in a contact patient.

Inf and PIV were clearly the most prevalent agents in flight patients, at 14.2% and 15.5%, respectively, without significant differences between age groups (One-way analysis of variance [ANOVA], 95% significance level). Equal distribution of these viruses was also seen in contact patients. CoVs were more frequent than one would expect in a mostly adult cohort. Detection rates did not differ between age groups at the 95% significance level. RSV and hMPV appeared to be significantly more frequent in flight patients under 18 years old than in the other age groups (P = 0.006 and 0.01, F test). For hMPV, this was also seen in contact patients (P = 0.02, F test).

The novel HBoV was not detected in any patient, which indicates that this agent may be restricted to children. Indeed, all data available so far about HBoV have been derived from cohorts of young infants (Allander *et al.*, 2005; Ma *et al.*, 2006; Sloots *et al.*, 2006). A complete absence was also observed of *Chlamydophila pneumoniae*, but it cannot be derived from our data what fraction of patients had already received pre-emptive antibiotic treatment at the time of sampling. Figure 10 shows the relative detection rates of agents in travellers.

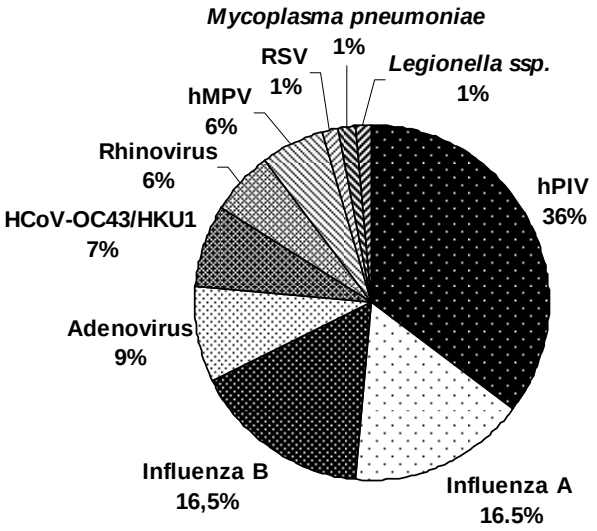

Figure 10: Relative detection rates of agents in travellers. HCoV, human coronavirus; hMPV, human metapneumovirus; PIV, human parainfluenza virus; RSV, respiratory syncytial virus.

3.1.3 Flight departures

Flight destinations were analyzed next. Destinations of departure to Germany included 24 from Beijing, 19 from Hong Kong, 17 from Bangkok, 14 from Shanghai, 13 from Hanoi, 13 from Singapore, 6 from Toronto, 3 from Taiwan, and 2 from Kuala Lumpur; single patients travelled from other airports in China, Southeast Asia, and the United States. Flights were distributed over 69 different days. For 12 of 146 flights evaluated, there was at least 1 other patient who had departed from the same city on the same day. Due to the fact that flight numbers were not available, results from all of the 25 patients who had travelled on non-unique flight dates were compared. Only in 2 cases, 2 patients had the same diagnosis, which was PIV. There was no association between the airport of departure and the detection of pathogens in general ($P < 0.3$, One-way ANOVA) nor of any pathogen in particular.

3.1.4 Samples and detection rates

To identify any possible influence of sampling on detection rates, we analyzed common categories of clinical samples separately for both flight and contact patients. Samples were categorized as follows: upper-respiratory-tract samples (category 1), throat-wash fluids (category 2), and lower-respiratory-tract samples (sputum and bronchoalveolar lavage [BAL] fluids; category 3). A comparison of age versus sample type yielded a significant lack of lower-respiratory-tract samples in patients under 19 years old (One-way ANOVA, 95% significance level). Due to the fact that a significantly higher detection rate of viruses had been identified in these patients, they were eliminated from the analyses, in order to avoid a bias. After elimination, no significant age differences in the three categories of samples remained ($P = 0.3$, F test). The average ages for categories 1, 2, and 3 were 42.1, 43.9, and 47.0 years, respectively.

Upper-respiratory-tract samples analyzed included 39 pharyngeal, nasal, and nasopharyngeal swabs and 99 throat-wash fluids. Lower-respiratory-tract samples included 50 sputum specimens and 4 BAL fluids. The detection frequencies of every pathogen in the three categories were compared by separate ANOVA analyses. Global detection frequencies did not differ significantly between categories for any pathogen, which were 51.3%, 34.3%, and 46.3% in category 1, 2, and 3, respectively. Only for Influenza was the detection frequency in swabs significantly higher than that for other samples, which was 30.1% in category 1; 10.1% in category 2; and 14.8% in category 3 ($P = 0.01$, F test). The same was observed for Inf A only, which was 21% in category 1; 5% in category 2; and 9% in category 3, respectively ($P = 0.01$), but was not observed for Inf B alone.

3.2 Generic detection of coronaviruses and differentiation at the prototype strain level by RT-PCR and non-fluorescent low-density DNA microarray

3.2.1 Universal coronavirus nested RT-PCR assay

Oligonucleotides for a universal CoV-genus nested RT-PCR assay were designed on the basis of an up-to-date alignment of all CoV RdRp gene segments available in GenBank in May 2005. The motifs A and C of RdRp were chosen because they contained amino acid patterns conserved among all CoVs. The 3'OH terminus of primers avoided the third codon position. Wobble positions in primers were avoided in order to guarantee reliable re-synthesis of primers. When necessary, mismatched base pairings were adjusted by mixing of defined oligonucleotides. The stable non-Watson-Crick base pairing T:G was not strictly adjusted for (Peyret *et al.*, 1999).

Primer hybridization sites with all of the important representative members of the CoV genus are shown in Figure 11.

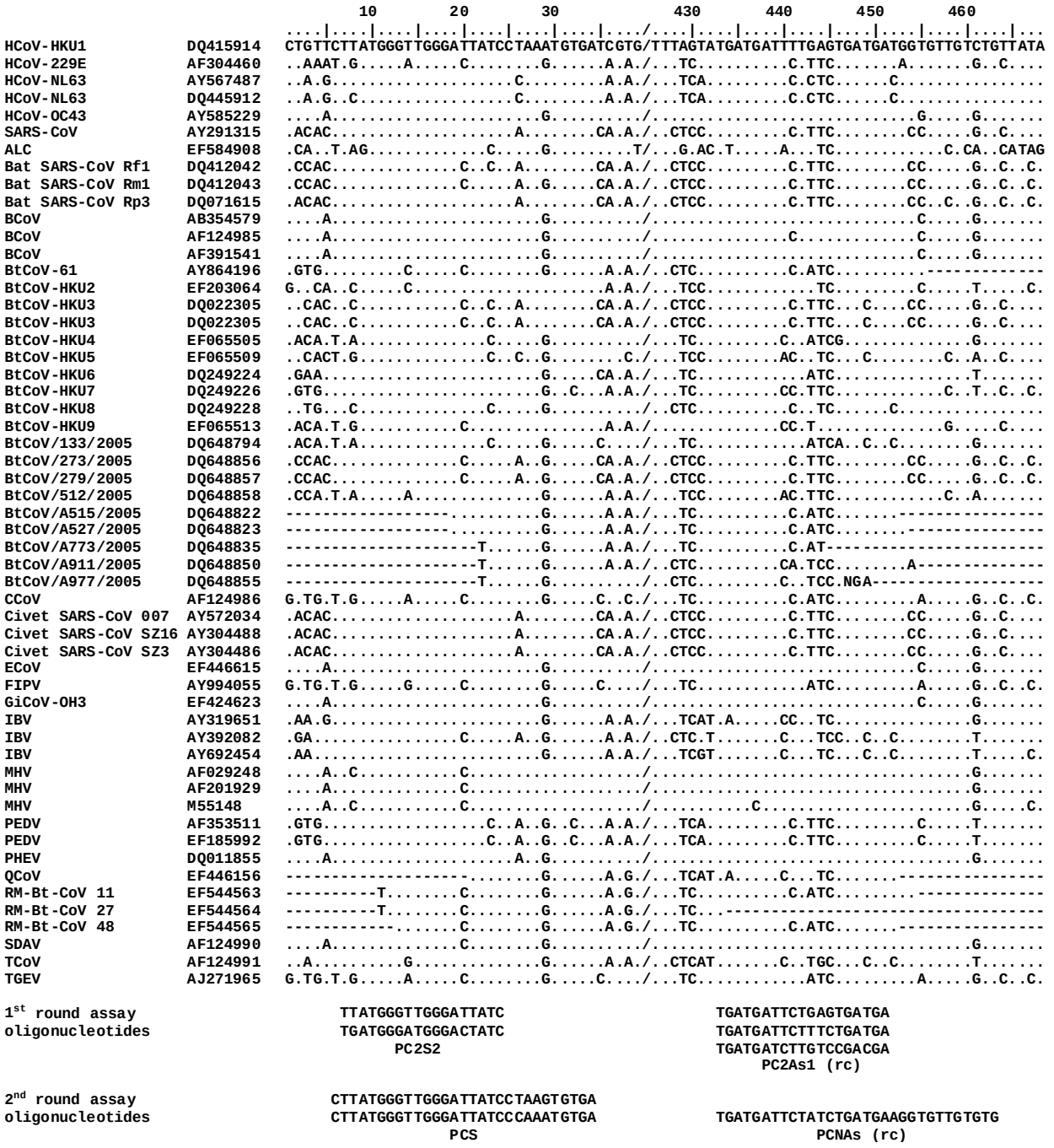

Figure 11: Nucleotide alignment of CoV prototypes and hybridization sites of the universal coronavirus nested RT-PCR oligonucleotides. The second column shows the GenBank accession number of respective coronavirus species. Assay oligonucleotides are shown below the alignment panel. Base count in the top line is based on HCoV-HKU1, which also serves as the comparison sequence in the alignment. Dots represent identical bases in compared sequences; deviations are spelled out. A slash (/) represents a gap in the alignment; (rc) means that the reverse complementary sequence is shown for the antisense primer. Abbreviations: HCoV, human coronavirus; BCoV, bovine coronavirus; CCoV, canine coronavirus; FIPV, feline infectious peritonitis virus; HEV, porcine hemagglutinating encephalomyelitis; IBV, avian infectious bronchitis virus; TCoV, turkey coronavirus; MHV, murine hepatitis virus; PEDV, porcine epidemic diarrhoea virus; SDAV, rat sialodacryoadenitis coronavirus; TGEV, transmissible gastroenteritis virus. If more than one sequence variant occurred at either primer binding site, all possible variants are listed (HCoV-NL63, BCoV, IBV, and MHV).

3.2.1.1 Universal coronavirus RT-PCR optimization

The universal CoV RT-PCR assay was first optimized using *in vitro* transcribed RNA from SARS-CoV, HCoV-OC43 and HCoV-229E CoVs. Different RT-PCR kits available in our laboratory were also used, such as SuperScript One-Step RT-PCR with Platinum *Taq* (Invitrogen, Karlsruhe, Germany), Access RT-PCR System (Promega, Mannheim, Germany) and QIAGEN OneStep RT-PCR Kit (QIAGEN, Hilden, Germany). After optimization by titrating magnesium and primers concentration for each available RT-PCR kit, an end-point dilution experiment was performed with RNA standards and high sensitivities could be achieved. However, the RT-PCR assay performed with QIAGEN OneStep RT-PCR Kit yielded better results when compared with the other kits, particularly as to HCoV-OC43 and HCoV-229E (Figure 12). Thus, QIAGEN kit was used for subsequent experiments.

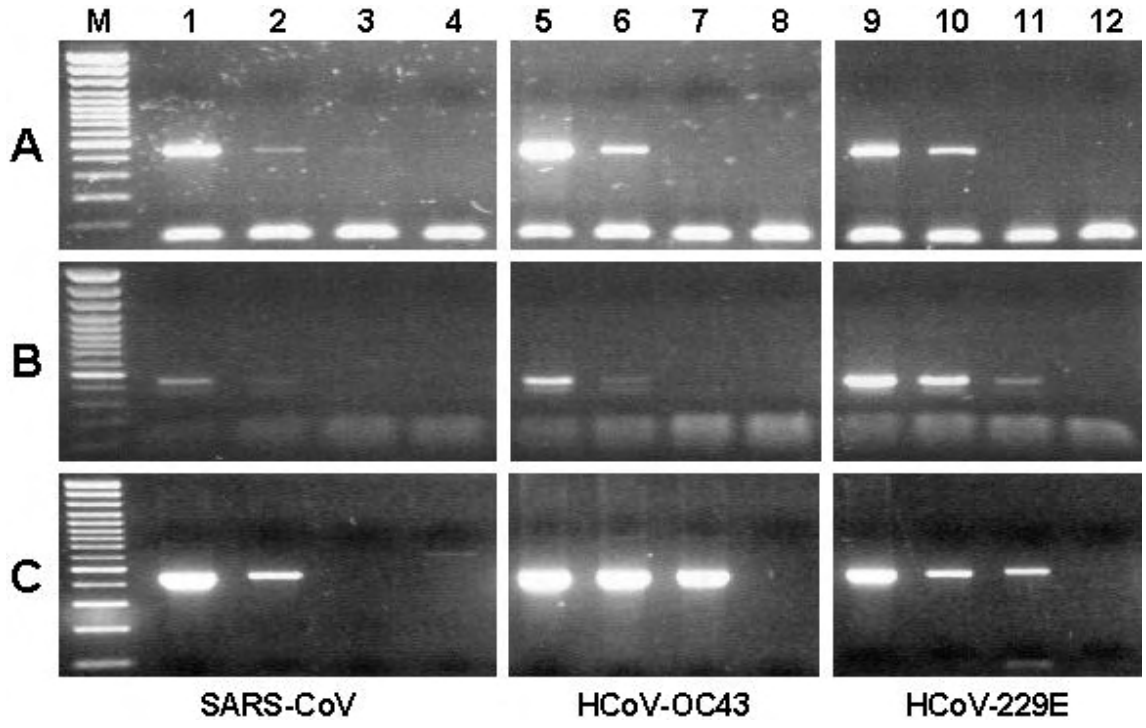

Figure 12: Agarose gel electrophoresis analysis of an end-point dilution experiment of the universal CoV assay with three commercial RT-PCR kits after optimization. Each row represents a 10-fold dilution of *in vitro* RNA transcripts, from 10^{-9} to 10^{-12} between intervals 1-4 (SARS-CoV), 5-8 (HCoV-OC43) and 9-12 (HCoV-229E). A, SuperScript One-Step RT-PCR with Platinum *Taq* (Invitrogen); B, Access RT-PCR System (Promega); C, QIAGEN OneStep RT-PCR Kit (QIAGEN). M, 100 bp DNA ladder.

When CoV *in vitro* RNA transcripts were spiked in clinical samples, however, several Log10s of sensitivity were lost. Multiple unspecific amplification products appeared which could not be eliminated by any effort in subsequent assay optimization, such as modifications of temperature, salts, or denaturing agents, including the Q-Solution supplied with the QIAGEN OneStep RT-PCR Kit, which often enables or improves suboptimal RT-PCR protocols. The short primers used in the assay obviously caused random reactivity with human genomic DNA.

Since only very short stretches of sequence were conserved among all CoVs, the use of longer primers was not an option. Rather than reducing the breadth of amplification, for this purpose, a nested PCR strategy was adopted instead. With long nested amplification primers that bound to the short first round forward primer but reached inside the amplicon, CoV sequences were selectively amplified. This reaction was optimized in presence of a background of nucleic acids as encountered in routine operation. Side products were largely eliminated with this formulation.

3.2.1.2 Sensitivity of the universal coronavirus nested RT-PCR assay

To determine the sensitivity of the universal CoV nested RT-PCR assay, the target regions including sufficient stretches of flanking sequence were cloned from SARS-CoV, HCoV-OC43, HCoV-229E, HCoV-NL63, HCoV-HKU1, that are representative CoVs from the genetic groups 1 and 2, and IBV, from group 3. The clones were converted into RNA by *in vitro* transcription with MEGAscript T7 kit (Ambion, Austin, USA), as described in section 2.1.6. The transcripts were photometrically quantified and tested in presence of high levels of background nucleic acids (human DNA, about 50 ng per reaction). As shown in Table 8, statistically constant detection could be achieved with as few as 45 copies of RNA per reaction for all three CoV groups. Sporadic detection was possible down to five copies per reaction. The cumulative hit rates for all viruses were subjected to probit analysis, showing a 50% detection chance at 15.7 copies per assay (95% CI, 14-27 copies per assay). A 95% chance of detection required 34 copies per assay.

Table 8: Detection of quantified RNA from representative strains of all three coronavirus groups.

No. of copies per reaction	Detection of indicated viruses (No. of strains correctly detected/total)*					
	SARS-CoV	HCoV-OC43	HCoV-229E	HCoV-NL63	HCoV-HKU1	IBV
90	3/3	3/3	3/3	3/3	3/3	3/3
45	3/3	3/3	3/3	3/3	3/3	3/3
15	1/3	1/3	2/3	0/3	3/3	0/3
5	0/3	1/3	1/3	0/3	3/3	1/3
0	0/3	0/3	0/3	0/3	3/3	0/3

*Each datum point summarizes the results of three replicate tests. The *in vitro* RNA transcripts were cloned from genome positions described in section 2.2.2.

3.2.1.3 Specificity of the universal coronavirus nested RT-PCR assay

The specificity of the assay was confirmed on 19 clinical samples containing nucleic acids of inf A, inf B, PIV 1, 2 and 3, hMPV, HRV, AdV, and RSV. None of these samples yielded PCR products.

3.2.2 LCD array hybridization for coronavirus species identification

CoVs comprise a very diverse spectrum of pathogens of humans and animals and throughput RT-PCR amplification of positive samples and subsequent strain identification require DNA sequencing, limiting the applicability in the clinical routine. To overcome this problem, we developed a simple and feasible approach to detect the full spectrum of CoVs with diagnostic sensitivity, combining the generic RT-PCR assay and low-cost, low-density (LCD) DNA microarrays.

Oligonucleotide detection probes, as described in the “material and methods” part, were spotted on plastic microarrays using proprietary technology (Chipron, Berlin, Germany). The anti-sense primers PCNAs were biotinylated and modified by use of proprietary technology to allow efficient hybridization. The PCR products were taken directly from the tube and hybridized to LCD arrays in a 45 minute procedure requiring no technical equipment except pipettes and a 37°C incubator.

It was determined how well different CoV strains could be discriminated by array hybridization. RNA was extracted from cultured virus or directly from patient material, amplified, and hybridized on LCD-chip arrays. As expected, all PCRs yielded amplification products, and all PCR products gave hybridization patterns on the arrays that matched the expected virus strains (Figure 13).

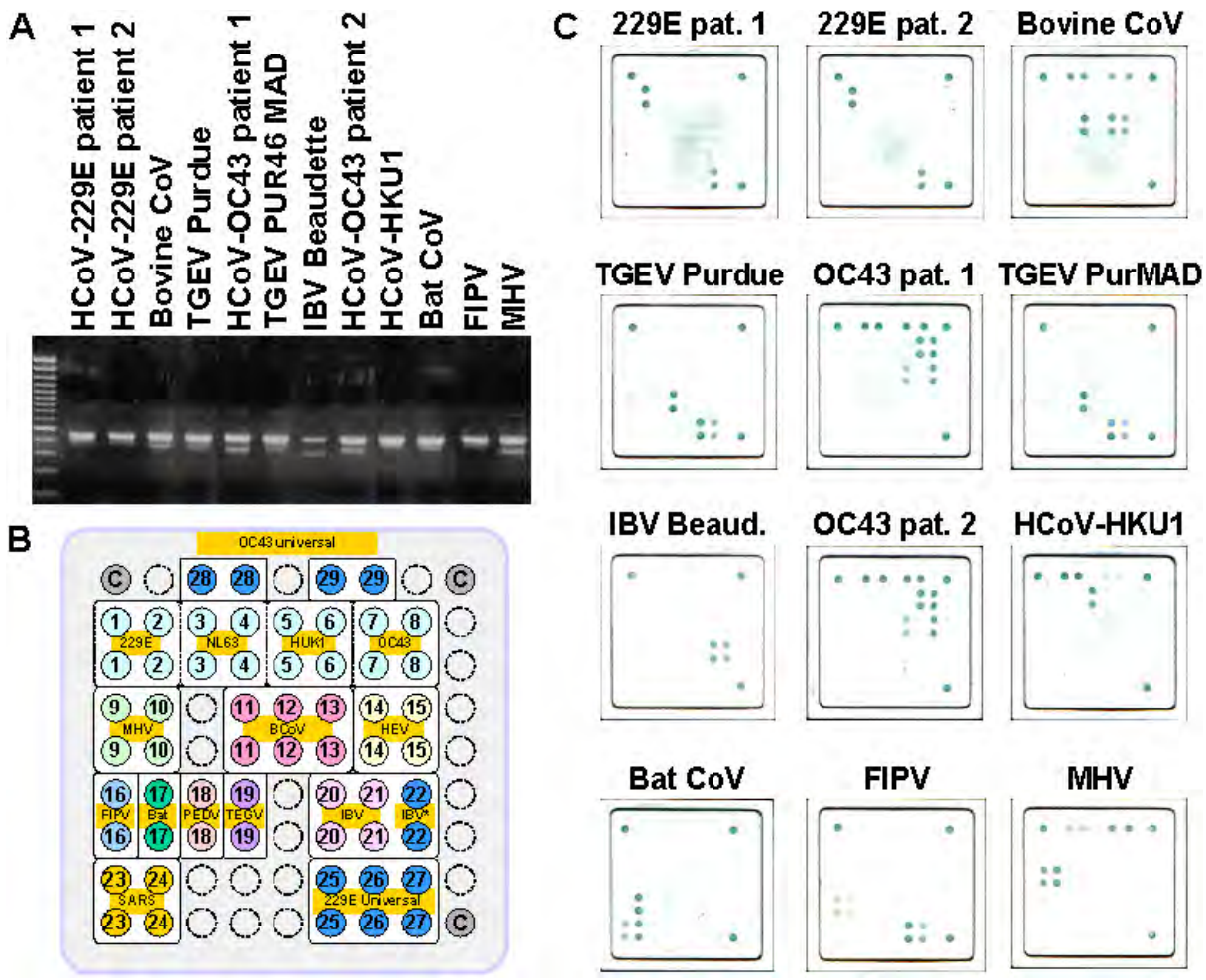

Figure 13: Agarose gel electrophoresis and hybridization patterns of prototype CoVs in the LCD array. A, prototype CoVs as indicated above each agarose gel lane were amplified and the depicted PCR products hybridised to oligonucleotide arrays as shown in the panel on the right (C). B, spotting pattern of oligonucleotides on array. Each coronavirus group (I to III) is represented by a set of universal probes (in blue). Species-specific probes are depicted in different colours. Each probe is represented in duplicate spots.

The specificity of the array hybridization was confirmed by reactions that contained about 0.5 µg of background DNA from human leukocytes. PCR inhibition was ruled out by spiking of parallel reactions with low concentrations of feline infectious peritonitis virus (FIPV) or murine hepatitis virus (MHV) RNA. They did neither generate any unspecific amplification signals, nor hybridization signals on the arrays (Figure 14).

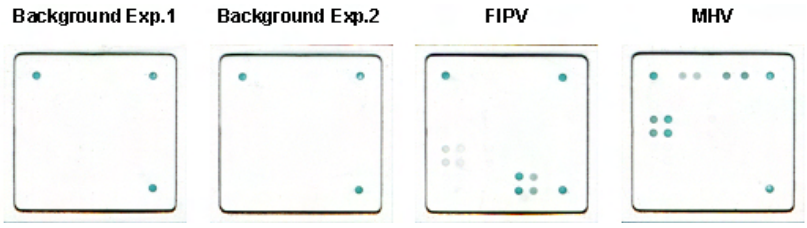

Figure 14: Amplification of background nucleic acids from human leukocytes (Background exp. 1 and 2). As a control, the functionality of RT-PCR and arrays was confirmed by co-amplifying spiked FIPV or MHV RNA, respectively, in the same reaction.

To determine whether array hybridization provided the same sensitivity as gel detection, amplification products from HCoV-229E, HCoV-NL63, HCoV-OC43 and SARS-CoV were gel-purified and diluted to decreasing concentrations in PCR reaction buffer. The same amount of PCR product was then analyzed by standard gel electrophoresis and array hybridization. As shown in Figure 15, even very faint gel bands yielded corresponding hybridization signals on the array. We thus concluded that the sensitivity of array detection was at least equal to that of a gel.

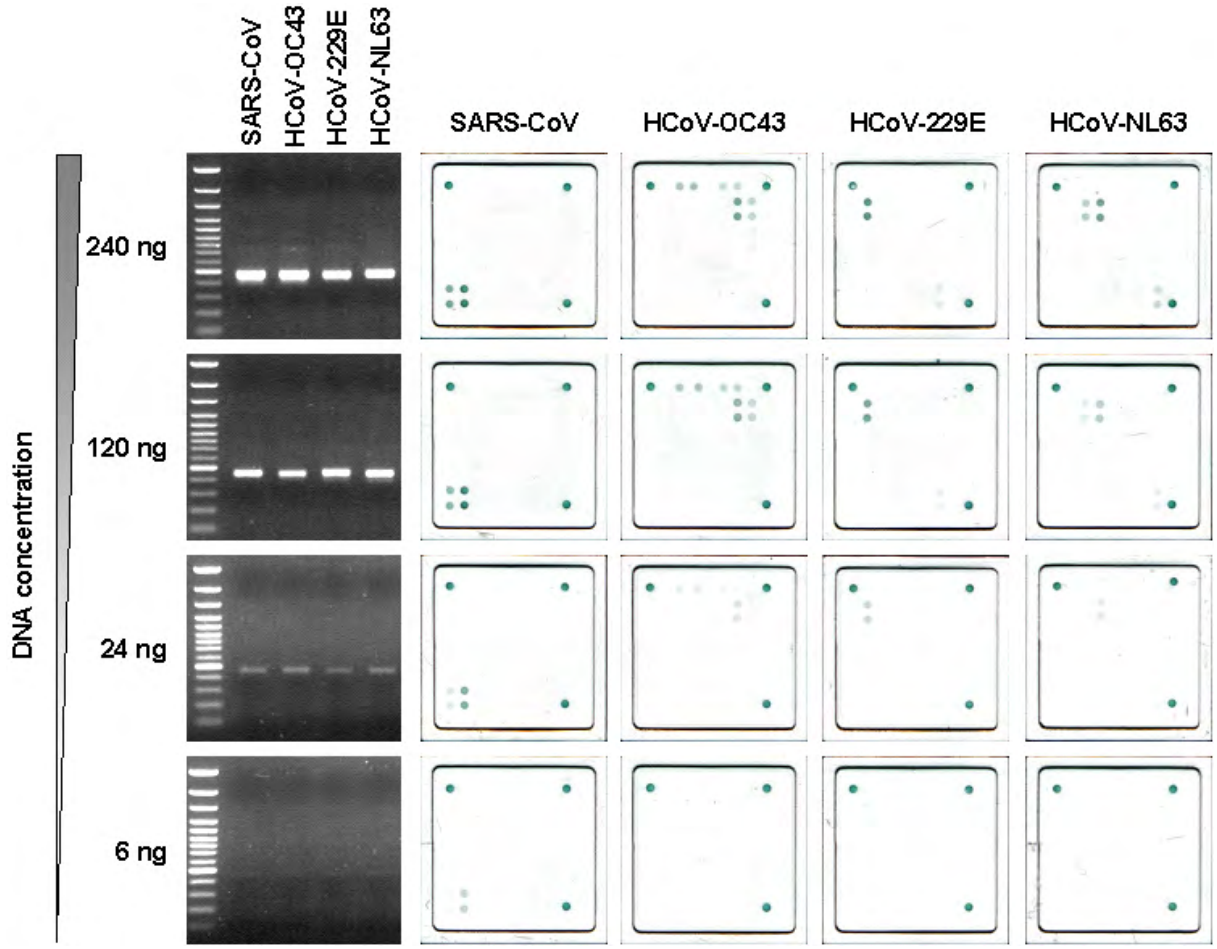

Figure 15: Agarose gel electrophoresis and hybridization patterns of four HCoVs. The species are indicated above each gel and array panels. The amounts of DNA given in the left column were subjected to gel analysis or array hybridisation, respectively. Blue dots on arrays represent hybridisation signals

3.2.3 Clinical detection limit of the universal coronavirus nested RT-PCR assay

To determine a clinical limit of detection, we retested 11 original RNA preparations from throat swab samples of SARS patients collected during the 2003 epidemic. The material was tested and quantified by a commercial real-time RT-PCR assay as previously described (Drosten *et al.*, 2004). As shown in Figure 16, sensitivity of the universal assay was virtually equivalent to that of the real-time assay. Of the 11 swabs, no detection occurred with only 3, all of which contained less than 100 copies of virus RNA per swab.

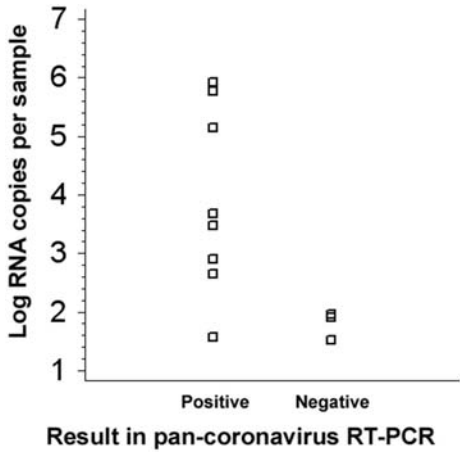

Figure 16: Comparison of sensitivity between the universal CoV RT-PCR assay and the SARS real-time RT-PCR. Eleven different respiratory tract samples (eight throat swabs, one sputum, one bronchoalveolar lavage sample) were quantified by real-time RT-PCR (viral loads on the y-axis) and tested by the broad-range coronavirus assay (results on the x-axis).

The assay was next applied to stored clinical samples. All samples had been determined in other laboratories to contain human CoVs by different real-time PCR protocols. CoVs types were determined either by sequencing or by separate, virus-specific real-time RT-PCRs (Garbino *et al.*, 2006). Due to the fact that some of the samples had been stored for a long time, the material was retested in parallel with specific real-time RT-PCRs for HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1, respectively (Garbino *et al.*, 2006; Luna *et al.*, 2007). As shown in Table 9, no sample positive by individual virus-specific real-time RT-PCR was missed by the universal CoV RT-PCR/LCD array. All PCR products were analysed on LCD arrays and then sequenced. The generated sequences were submitted to phylogenetic tree analyses. As shown in Figure 17, results of typing and phylogenesis were concordant in all cases at the CoV group and strain levels.

Table 9: Results of real-time RT-PCR and universal coronavirus assays in stored clinical samples previously reported positive for human CoVs.

Assay	No. of positive results per no. of samples			
	HCoV-229E	HCoV-NL63	HCoV-OC43	HCoV-HKU1
Real-time RT-PCR	3/8	7/12	11/13	4/6
Universal CoV assay	3/8	8/12	11/13	5/6

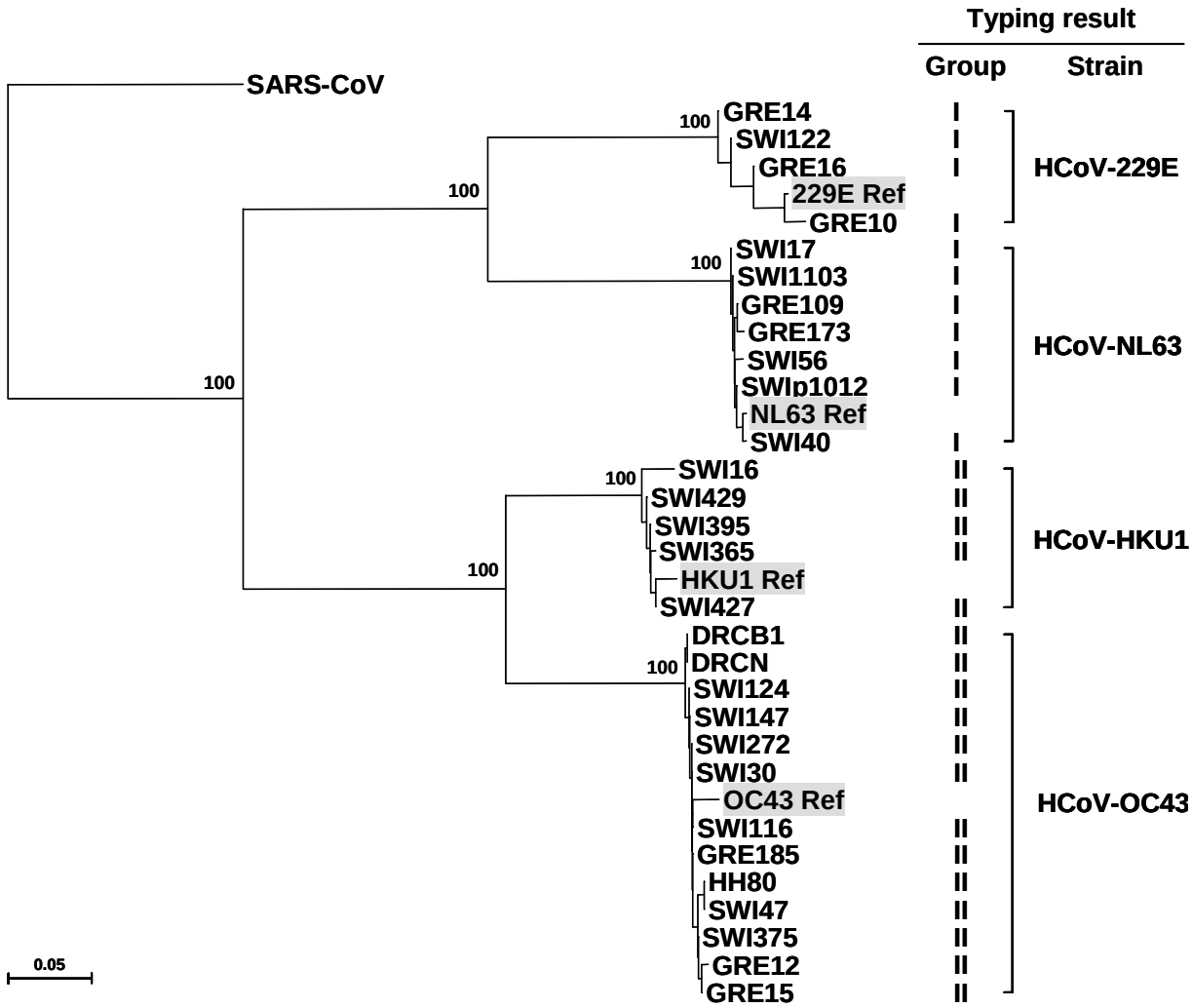

Figure 17: Phylogenetic tree based on the nucleotide sequences of the partial RdRp for amplified CoVs, showing the relationship between samples and reference CoVs HCoV-229E (accession number AF304460), HCoV-NL63 (AY567487), HCoV-HKU1 (DQ415914), and HCoV-OC43 (AY585229). The tree was generated by the neighbour-joining method in the TREECON version 1.3b program. Numbers at the nodes represent bootstrap values (percent) calculated from 1,000 replicates. The tree was rooted to SARS-CoV (AY291315). Abbreviations: DRC (Democratic Republic of Congo), GRE (Greece), HH (Hamburg, Germany), SWI (Switzerland).

3.3 Prevalence, types, and RNA concentrations of HPeVs in patients with acute enteritis, including a sixth parechovirus type

3.3.1 Isolation and characterization of a new HPeV strain

During routine diagnostic work on enteritis patients in a municipal health service (Hamburg Institute of Hygiene and Environment), a stool sample displayed a CPE on cell culture. The CPE resembled that of enterovirus, but the isolate was not typeable using a bank of polyclonal anti-enterovirus sera. Enterovirus RT-PCR and AdV PCR were negative. The original stool samples stemmed from a 30-year-old female kitchen worker with acute enteritis. It had been collected along with samples from several other patients in a confirmed food borne norovirus outbreak. Norovirus RT-PCR in this particular patient was negative.

In order to identify a rare or possibly unknown agent in culture, supernatant was subjected to VIDISCA method. In the second amplification or selective stage, one of the 16 gel lanes yielded a single distinct DNA band (Figure 18). The VIDISCA PCR product of about 188 base pairs was cloned in pCR 2.0-TOPO plasmid vector and nucleotide sequence was determined. BLAST nucleotide search at GenBank website (www.ncbi.nlm.nih.gov/BLAST) showed that the sequence presented high similarity with the capsid P1 protein region of human parechovirus (HPeV) strains. Closest identity was with type 1 HPeVs (Harris strain).

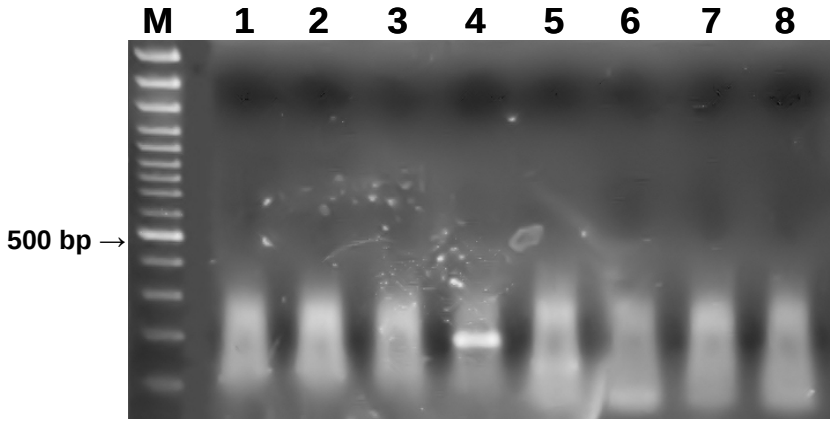

Figure 18: Agarose gel electrophoresis of VIDISCA PCR product showing the amplified fragment in lane 4, with primers *Hin*P1I-A and *Cvi*AII-G. M, 100 bp DNA ladder.

3.3.1.1 Complete genome sequencing of the isolated HPeV

To determine the full genome nucleotide sequence of the isolated HPeV, hereafter called BNI-788St virus, the RNA was extracted from cell culture supernatant. The sequences were determined by RT-PCR and sequencing of cDNA fragments of the genome. Based on the complete genome nucleotide sequences of published HPeVs on GenBank, combinations of consensus primers were used to amplify fragments covering the complete genome. RT-PCR products were directly sequenced. Specific primers were then synthesized based on the nucleotide sequences near the ends of amplicons; remaining sequences were obtained according to a primer walking strategy.

It was not possible to detect the sequence of the extreme 5' end with absolute certainty, despite employing different strategies such as 5' RACE by homopolymeric tailing reaction using deoxyadenosine triphosphate (dATP), to form a poly(A) anchor to the 3' hydroxyl terminus of cDNA, with Terminal Transferase (TdT) enzyme (New England Biolabs), or consensus primers of the first nucleotides conserved for all published HPeVs. The final sequence of the 5' terminus was obtained with a primer composed of the first 19 nucleotides of published consensus HPeV sequences.

The extreme 3' end of the genome was amplified with GeneRacer Oligo dT primer and sequenced according to the GeneRacer kit (Invitrogen, Karlsruhe, Germany) as described in material and methods, section 2.3.9.2.

The complete genome sequence of BNI-788St virus has been submitted to the GenBank database under accession number EF051629.

3.3.1.2 Genetic analysis of BNI-788St virus

Analysis of the complete nucleotide sequences showed that the length of the BNI-788St genome, excluding the poly(A) tract, was 7337 nucleotides. The genome organization was similar to that of other parechoviruses and consists of a 5' untranslated region (UTR) of 709 nucleotides, followed by a large ORF of 6537 nucleotides (that encodes the putative polyprotein precursor of 2179 amino acids); a 3' UTR of 91 nucleotides and a poly(A) tail.

The 5' UTR of BNI-788St was most similar to that of HPeV4 viruses, showing nucleotide identities of 88.9% and 90.8% to the type 4 strains T75-4077 and K251176-02, respectively. The secondary RNA structure elements (stem-loops) of the 5' UTR of BNI-788St virus could be determined by computer prediction of the Mfold program version 3.2 (Zuker, 2003). The structures depicted in Figure 19 were the most similar of the alternative energetically possible structures in comparison to the already proposed structures of HPeV1 Harris and

HPeV2 Williamson (Ghazi *et al.*, 1998; Oberste *et al.*, 1998). The predicted folding of this region was very similar to that of HPeV1 Harris and HPeV2 Williamson. Differences were observed in the sequences forming stem-loops B, C, G and H. In BNI-788St virus, these sequences are forming only 2 stem-loop structures, designed B-C and G-H. The other stem-loops are well conserved, especially those from I to L, that are known to form the type II IRES in other picornaviruses, as described for cardioviruses and aphthoviruses (Duke *et al.*, 1992; Jang & Wimmer, 1990; Pilipenko *et al.*, 1990).

The polypyrimidine-rich tract, found in all studied picornaviruses, was also observed in the BNI-788St virus and is located 17 nucleotides upstream from the initiation codon, which is structured as a Kozak sequence (ANNAUGG).

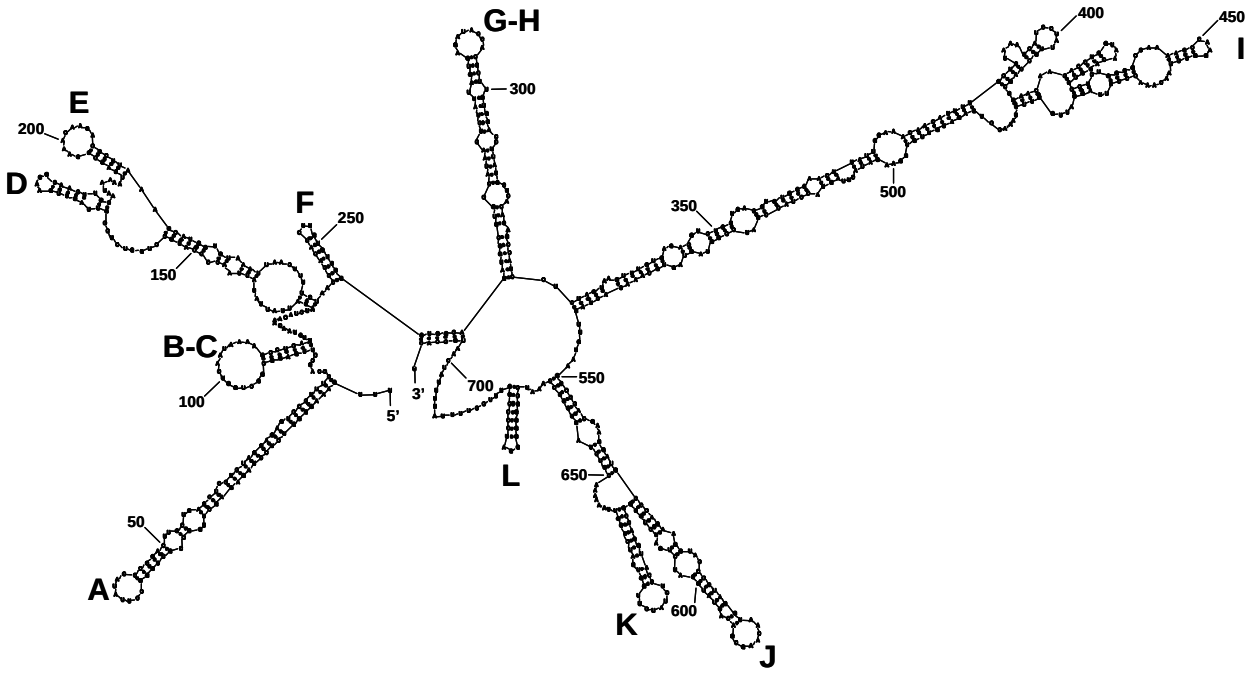

Figure 19: Predicted secondary structure of the 5' UTR of BNI-788St virus. Stem-loop domains (A to L) are labelled. Nucleotide positions are indicated.

Phylogenetic tree analysis with complete P1 amino acid sequences of published HPeVs showed that the BNI-788St virus was most similar to HPeV1 Harris and both viruses fall in the same genetic group (Figure 20).

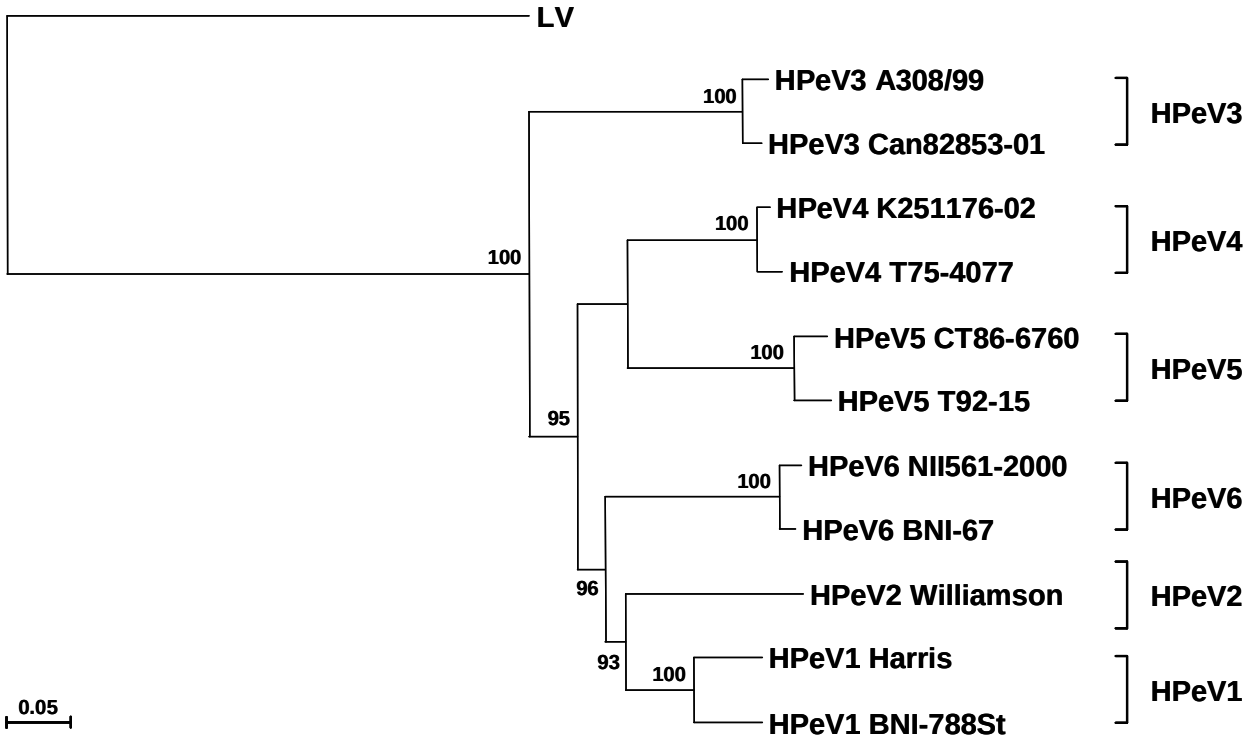

Figure 20: Phylogenetic tree based on the amino acid sequence of P1 gene for capsid proteins, showing the relationship between HPeV BNI-788St and representative HPeVs. HPeVs cluster into 6 genotypes. The tree was generated by the neighbour-joining method in the TREECON version 1.3b program. Numbers at the nodes represent bootstrap values (percent) calculated from 1,000 replicates. The tree was rooted to Ljungan virus (LV) 87-012 (AF327920).

The similarities between nucleotide and deduced amino acid sequences of BNI-788St virus and complete genome sequences of HPeVs were summarized in Table 10.

The structural gene region P1, composed by the genes VP0, VP3 and VP1, of BNI-788St virus showed 79.0% and 91.5% of similarity to HPeV1 Harris in nucleotide and amino acid sequences, respectively. For VP0, similarity was 82.6% and 94.1%; for VP3, 76.9% and 89.3%; and VP1, 76.6% and 89.6%. Since VP1 amino acid similarity has been used extensively for the classification of new genotypes in the Enterovirus genus (Oberste *et al.*, 1999), the BNI-788St virus was classified as a new member of the HPeV genotype 1 (HPeV1) based on the same criteria as for the classification of new genotypes in the Enterovirus genus.

Table 10: Percentage similarities of the BNI-788St virus nucleotide and predicted amino acid sequences to all HPeVs.

	Nucleotide similarity (amino acid similarity)									
Region	**HPeV1[a]**	**HPeV2[b]**	**HPeV3[c]**	**HPeV3[d]**	**HPeV4[e]**	**HPeV4[f]**	**HPeV5[g]**	**HPeV5[h]**	**HPeV6[i]**	**HPeV6[j]**
5'UTR	86.6	84.9	84.9	85.2	88.9	90.8	ND*	ND	88.3	ND
VP0	82.6 (94.1)	75.7 (84.1)	70.6 (73.4)	71.0 (73.4)	72.9 (78.9)	72.9 (79.2)	72.3 (79.2)	72.3 (79.2)	72.7 (78.9)	72.3 (79.2)
VP3	76.9 (89.3)	72.8 (82.2)	68.1 (76.3)	67.9 (76.7)	71.7 (79.8)	70.2 (80.2)	68.9 (76.7)	68.5 (77.1)	73.1 (83.4)	72.1 (83.8)
VP1	76.6 (89.6)	71.1 (79.2)	64.2 (70.6)	64.6 (71.0)	70.7 (74.0)	71.7 (77.1)	69.6 (74.0)	65.7 (71.4)	69.7 (74.9)	69.3 (74.9)
2A	76.7 (89.3)	77.1 (87.3)	87.1 (88.0)	88.7 (87.3)	76.2 (89.3)	88.2 (94.7)	79.1 (88.7)	76.2 (88.7)	79.8 (88.7)	78.2 (88.7)
2B	82.2 (96.7)	78.4 (95.9)	91.5 (100.0)	91.0 (100.0)	83.1 (100.0)	85.2 (100.0)	83.9 (98.4)	83.3 (98.4)	82.0 (97.5)	82.8 (98.4)
2C	79.4 (92.7)	77.4 (86.6)	92.8 (99.1)	92.0 (98.8)	82.1 (97.3)	88.1 (98.2)	81.8 (95.1)	79.3 (93.9)	78.6 (91.8)	80.1 (96.4)
3A	75.8 (87.2)	76.9 (83.8)	91.2 (91.5)	92.6 (94.0)	81.8 (93.2)	83.2 (92.3)	80.1 (81.2)	77.5 (88.9)	78.6 (88.0)	80.3 (91.5)
3B	71.7 (90.0)	70.0 (90.0)	86.7 (85.0)	86.7 (90.0)	78.3 (95.0)	73.3 (85.0)	80.0 (90.0)	66.7 (90.0)	78.3 (90.0)	76.7 (90.0)
3C	81.0 (97.0)	81.7 (97.0)	91.0 (98.0)	91.3 (98.5)	86.8 (98.0)	87.0 (97.5)	79.5 (99.0)	80.5 (96.0)	84.2 (98.0)	81.3 (97.0)
3D	82.8 (95.9)	82.9 (94.2)	91.0 (97.0)	90.8 (97.0)	88.7 (97.0)	87.9 (97.7)	82.3 (95.1)	82.4 (93.8)	83.9 (95.7)	82.3 (95.3)
3' UTR	83.3	87.9	94.5	92.3	89.0	94.5	84.6	89.0	83.5	84.6
P1	79.0 (91.5)	73.4 (82.5)	68.2 (73.2)	68.5 (73.6)	71.8 (77.7)	71.8 (79.0)	70.5 (77.1)	69.1 (76.6)	71.9 (79.4)	71.3 (79.3)
P2	79.3 (92.7)	77.5 (88.7)	91.1 (96.5)	91.0 (96.2)	80.9 (95.8)	87.6 (97.7)	81.5 (94.2)	79.4 (93.5)	79.6 (92.2)	80.2 (94.8)
P3	81.1 (94.8)	81.5 (93.3)	90.9 (96.2)	91.1 (96.8)	87.0 (96.7)	86.6 (96.5)	81.2 (93.9)	80.8 (93.5)	83.1 (95.0)	81.6 (95.0)
P2-P3	80.3 (93.9)	79.8 (91.3)	91.0 (96.3)	91.0 (96.5)	84.4 (96.3)	87.0 (97.0)	81.4 (94.0)	80.2 (93.5)	81.6 (93.8)	81.0 (95.0)
ORF	79.7 (93.0)	77.5 (88.2)	83.1 (88.1)	83.2 (88.3)	80.0 (89.7)	81.7 (90.6)	77.6 (88.0)	76.3 (87.5)	78.2 (88.7)	77.6 (89.4)

*ND, not determined; [a]Harris, [b]Williamson, [c]A308/99, [d]Can82853-01, [e]T75-4077, [f]K251176-02, [g]CT86-6760, [h]T92-15, [i]NII561-2000, [j]BNI-67

In the non-structural region, composed by P2 and P3 gene regions, the HPeV1 BNI-788St virus showed best-match nucleotide and amino acid similarities to HPeV3 viruses, particularly with the strain Can82853-01 (91.0% and 96.5% respectively). To investigate the differential similarities along the highly variable P1 region and the more conserved non-structural region (P2 and P3) of BNI-788St virus, SimPlot analysis was performed on the complete HPeV genomic RNA sequences available (Figure 21).

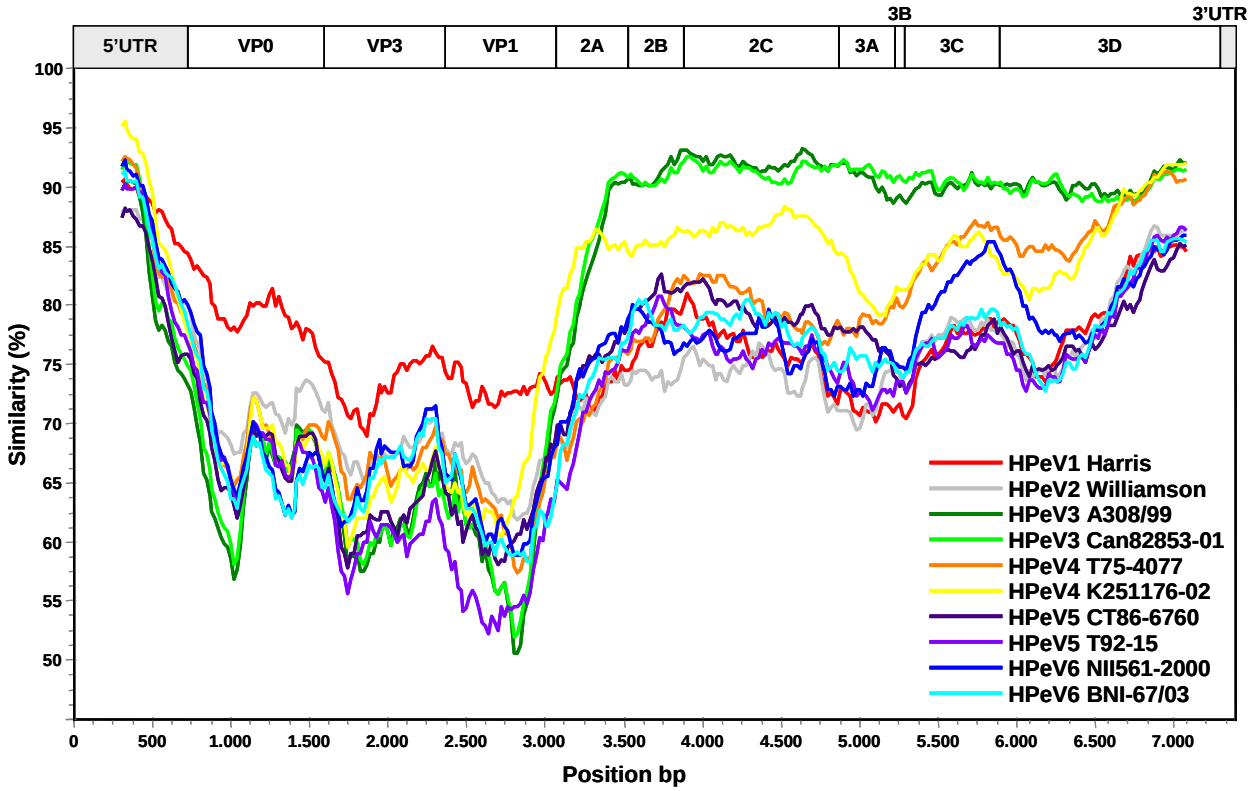

Figure 21: Full-length genome similarity plot of HPeV1 BNI-788St against HPeVs calculated by SimPlot 3.5.1. Each point represents the similarity between the BNI-788St and a given HPeV, within a sliding window of 600 nucleotides centred on the position plotted, with a step side of 20 residues between points. Positions containing gaps were excluded from the analysis by gap stripping. Jukes and Cantor correction was applied.

We observed that BNI-788St virus showed more nucleotide similarity to Harris virus in the P1 region; in the non-structural region, however, high degree of similarity to HPeV3 viruses was observed. These results suggest that recombination events possibly have occurred between HPeV genotypes 1 and 3, in the VP1-2A junction.

A comparison of the complete ORF of BNI-788St with other HPeVs showed an amino acid identity ranging from 87.5% to 93.0% and from 77.5% to 83.2% for nucleotides. These values are in the same range of identity as for those obtained between known HPeVs.

A comparison of the predicted amino acid sequences of the VP0, VP3 and VP1 genes between BNI-788St and HPeV prototypes showed that the deduced sequences forming the β-sheets and α-helices in the β-barrel structure (Ghazi *et al.*, 1998; Ito *et al.*, 2004; Stanway *et al.*, 1994) were well conserved (Figure 22).

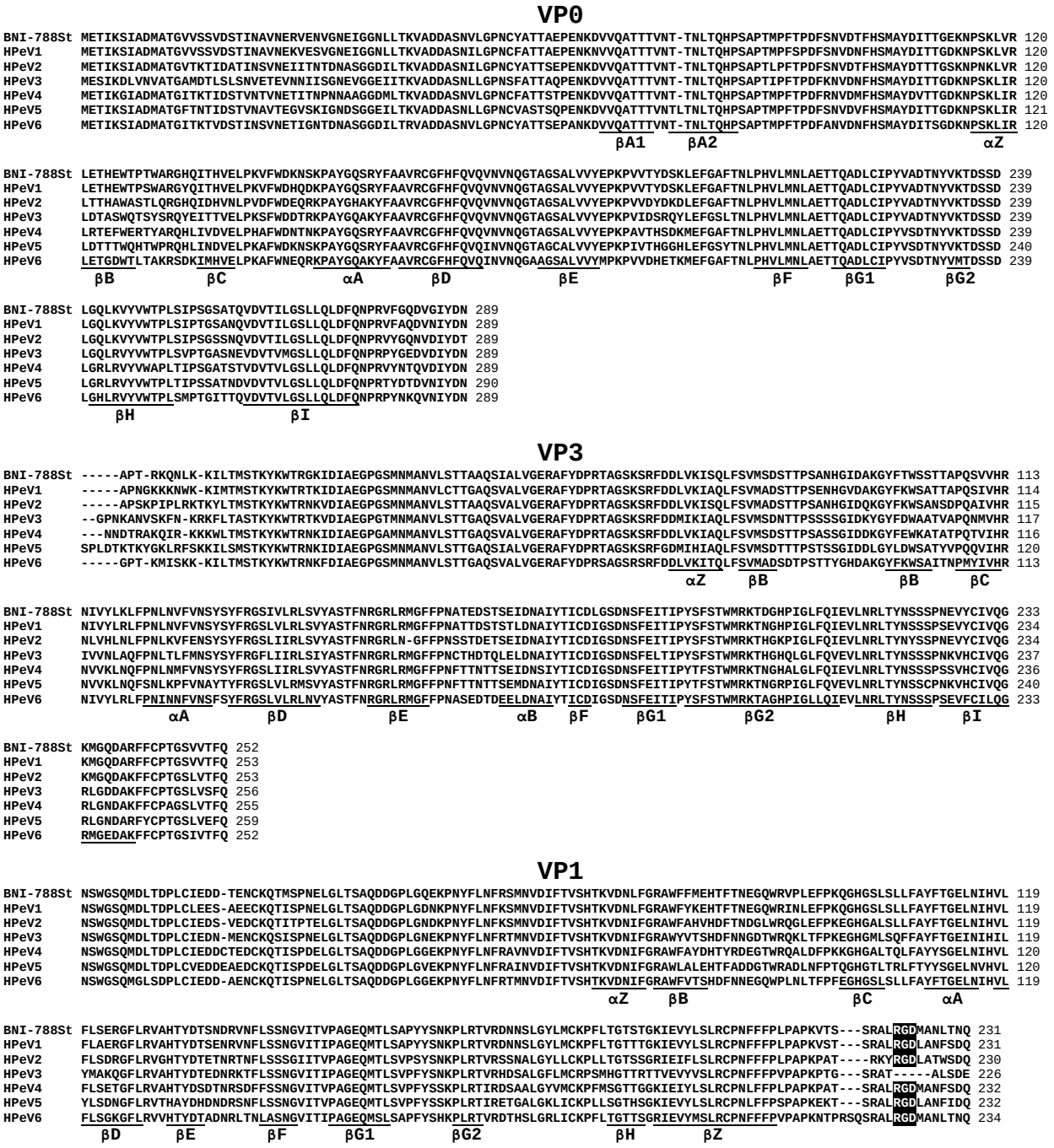

Figure 22: Alignment of the BNI-788St predicted amino acid sequences of capsid proteins, VP0, VP3 and VP1, with the corresponding polypeptides of HPeVs prototypes whose three-dimensional virus particle structures are known. The regions of secondary structure are underlined. The RGD motif of VP1 C-terminal is in white highlighted in black. HPeV prototypes: HPeV1 Harris (accession number S45208), HPeV2 Williamson (AJ005695), HPeV3 A308/99 (AB084913), HPeV4 K251176-02 (DQ315670), HPeV5 CT86-6760 (AF055846), and HPeV6 NII561-2000 (AB252582).

The BNI-788St virus has presented the RGD motif at the C-terminus of VP1, which is not present in the HPeV3 viruses (Boivin *et al.*, 2005; Ito *et al.*, 2004). The RGD motif is known to play an important role in the attachment and entry into host cells in other picornaviruses (Fox *et al.*, 1989; Roivainen *et al.*, 1991) and has shown to be essential for infectivity in HPeV1 (Boonyakiat *et al.*, 2001; Stanway *et al.*, 1994).

Some of the non-structural proteins of parechoviruses, particularly 2C, 3C and 3D, are known to contain common motifs (Ghazi *et al.*, 1998; Hyypia *et al.*, 1992; Stanway & Hyypia, 1999; Stanway *et al.*, 1994). The BNI-788St virus showed to contain the 2C motifs GXXGXGK(S/T) and DDLXQ, which are found in all picornaviruses and are predicted to have a helicase function. In the 3C protease, the active-site cysteine GXCG was found. In the 3D RNA-dependent RNA polymerase, the active site YGDD was observed together with the other known motifs KDELR, PSG and FLKR.

BNI-788St virus, such as all of the other HPeVs, has 9 polyprotein cleavage sites; the related amino acid sequences for each site were as follows: VP0-VP3 (N/A),VP3-VP1 (Q/N),VP1-2A (Q/S), 2A-2B (Q/G), 2B-2C (Q/G), 2C-3A (Q/T), 3A-3B (E/R), 3B-3C (Q/R), and 3C-3D (Q/G). All known HPeVs share the same conserved polyprotein cleavage sites at junctions VP3/VP1, 2A/2B, 2B/2C, 3A/3B, 3B/3C, and 3C/3D. The BNI-788St virus had the same cleavage site of Harris virus at junction VP0/VP3, but different sites to that of HPeV3 viruses at VP1/2A (E/S), HPeV4 T92-15 and VI2177-67 viruses at 2C/3A (Q/S).

The predicted secondary structure of the 3’ UTR of BNI-788St virus was also determined by Mfold program version 3.2 (Zuker, 2003) and is depicted in Figure 23. The virus has the same one stem-loop organization as it has previously been described for other parechoviruses (Abed & Boivin, 2005); moreover, the same two highly conserved repeats arranged in tandem were also observed (Al-Sunaidi *et al.*, 2007).

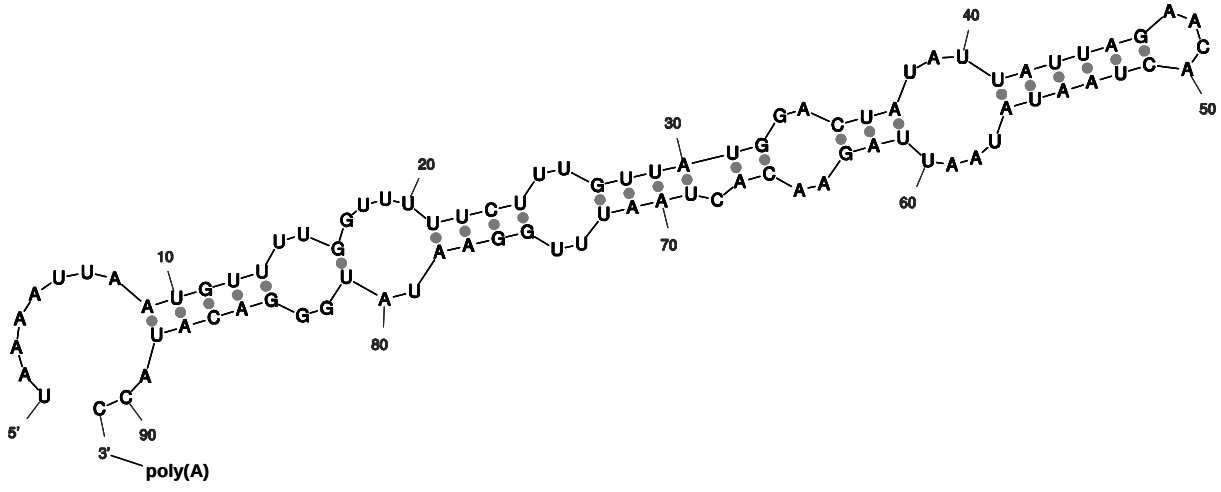

Figure 23: Predicted secondary structure of the 3’ UTR of HPeV1 BNI-788St.

3.3.2. Screening and detection of HPeVs and identification of a new type six

As described above, hundreds of stool samples were collected from patients with acute enteritis in Hamburg. The samples were submitted to the diagnosis of common gastroenteritis ethological agents and cell culture. Whenever a CPE producing agent was isolated, the same diagnostic procedures were performed. After isolation and characterization of the HPeV1 BNI-788St virus, at that point of time, only four genotypes were described; moreover, recombination events between genotypes were not well investigated. In the case of HPeV1 BNI-788St, SimPlot analysis showed that possibly recombination had occurred between HPeV genotypes 1 and 3. In order to appreciate the relevance of parechoviruses in patients with acute enteritis, a real-time RT-PCR assay was established.

3.3.2.1 The HPeV broad-range real-time RT-PCR assay

The HPeV real-time RT-PCR assay was developed with the aim of amplifying the highly conserved region of the 5' UTR. It was designed to amplify all four of the HPeV types available at GenBank in November 2005. Since then, two novel HPeVs have been discovered and described (Al-Sunaidi *et al.*, 2007; Benschop *et al.*, 2006a). Figure 24 shows an alignment of amplicon sequences with prototype strains of all six parechovirus types known at the time of writing.

The sensitivity of the assay was evaluated by *in-vitro* transcribed RNA of BNI-788St, confirming that single copies of RNA could be amplified (The data are not shown). The *in vitro* transcribed RNA was also used as a quantification standard to determine virus RNA concentrations in stool samples.

```
Strain type            440       450       560       570       580       590       600
                      |....|....|....|..../....|....|....|....|....|....|....|....|....|...
HPeV1 Harris          ATGCCTCTGGGGCCAAAAGC/ACGAAGGATGCCCAGAAGGTACCCGCAGGTAACAAGAGACACTGTGGATCTGA
HPeV1 BNI-788St       G..................../.........................T.........C......A.........
HPeV2 Williamson      G....C...............//////////////////////////////////////////////////////
HPeV3 A308/99         ...................../.........................T.........T.G....A.........
HPeV3 Can82853-01     ...................../.........................T.........T......A.........
HPeV4 K251176-02      G..................../.........................T.........T......A.........
HPeV4 T75-4077        G..................../.........................T.........T......A.........
HPeV5 CT86-6760       ...................../G........................T.........T......A.........
HPeV5 T92-15          G..................../.........................T.........T......A.........
HPeV6 NII561-2000     G....C...............//////////////////////////T.........T......A.........

Assay                 GTGCCTCTGGGGCCAAAAG   CGAAGGATGCCCAGAAGGTACCCGT  GTAACAAGCGACACTATGGATCTGA
Oligonucleotides             HPS          /          HPP           \           HPA (rc)
                                    FAM                              TAMRA
```

Figure 24: Nucleotide alignment of HPeV prototypes and hybridization sites of diagnostic real-time RT-PCR oligonucleotides. The assay oligonucleotides are shown below the alignment panel. Base count in the top line is based on HPeV1 reference strain Harris, which also serves as the comparison sequence in the alignment. Dots represent identical bases in compared sequences; deviations are spelled out. A slash (/) represents a gap in the alignment; (rc) means that the reverse complementary sequence is shown for the antisense primer. Genomes of available HPeVs at GenBank in November 2005 are highlighted in grey.

3.3.2.2 Identification of HPeVs in stool samples.

In total, 656 stool samples from outpatients of all age groups were tested for parechovirus by the broad-range real-time RT-PCR assay (Figure 25).

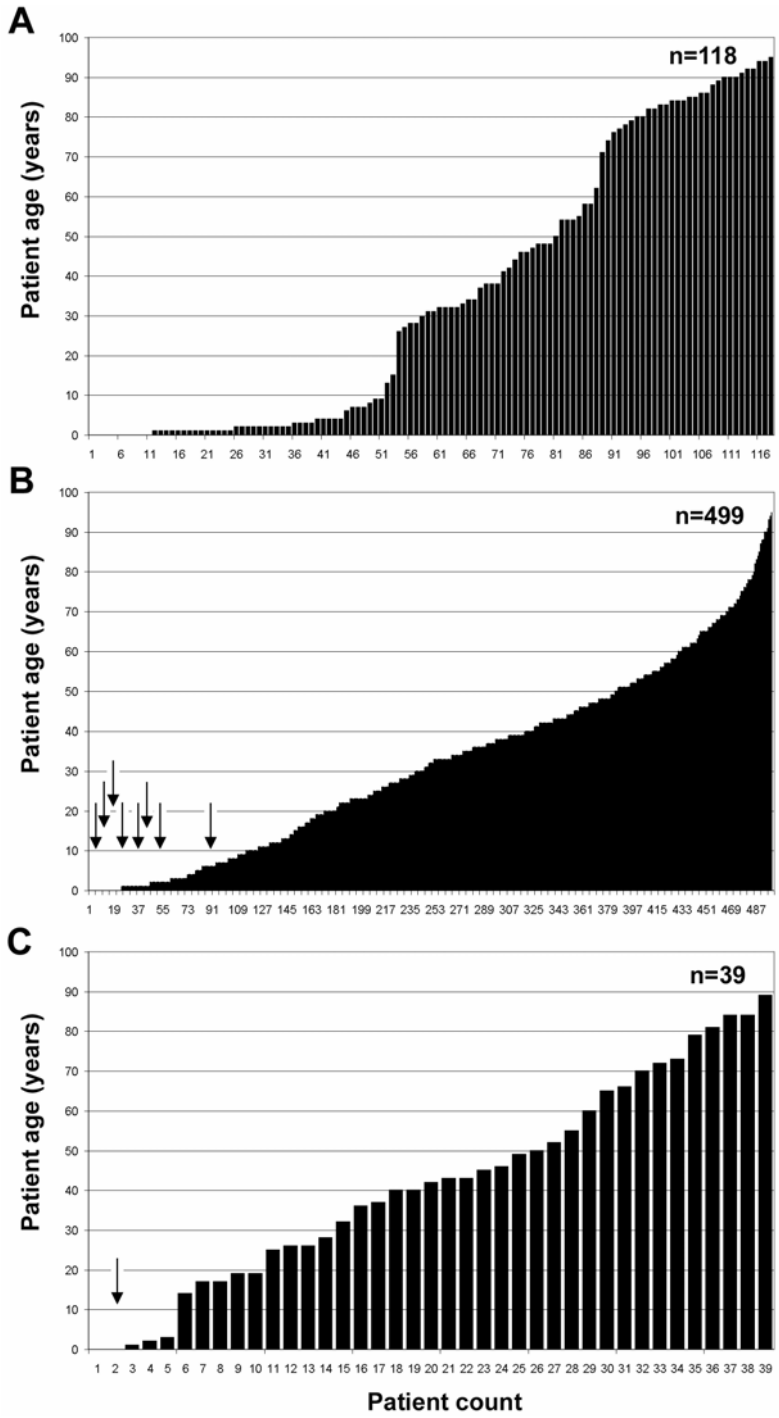

Figure 25: Age distribution of patient cohorts tested in this study. The y-axis in each panel shows patient age, the x-axis shows patient count. A, patients tested in routine diagnostic laboratory. Children in this cohort were mostly hospitalised; adults were mostly tested in the context of investigations of food-borne enteritis outbreaks or in outbreaks in retirement homes. B, outpatients seen for enteritis by practitioners. Samples in both cohorts were pretested negative for astro-, rota-, adeno-, and noroviruses. Arrows point to patients tested positive for parechovirus by real-time RT-PCR. C, non-enteritis control patients for the study cohort shown in B.

The samples were taken throughout the year. All of the samples had been pre-tested by antigen EIA for rotavirus, astrovirus, and AdV, as well as by RT-PCR for norovirus. Samples from the first 118 patients were obtained from the diagnostic laboratory of the municipal health service. This cohort comprised hospitalised young children, as well as patients of all age groups who were tested in the context of investigations on the outbreak of enteritis. No parechoviruses were found.

The second cohort stemmed from a study on community-acquired enteritis. Only patients seen by general practitioners were tested. Nine parechoviruses were found in 499 patients suffering from acute enteritis, and one in 39 control patients was found without symptoms of enteritis. The prevalence of parechoviruses in outpatients was thus 1.6% (8 of 499 patients; 95% confidence interval = 1-3%), but detection rates in patients and controls were not significantly different at the 95% confidence level (two-tailed *t*-test). All positive patients, except for one, were below 2 years of age, and the gender ratio was neutral. If the cohort would had been restricted to young children, for example, children less than 2 years of age, the detection rate of parechoviruses would have been 11.6% (7 of 60).

All detected HPeVs were listed as follows: BNI-67, BNI-90, BNI-R04, BNI-R09, BNI-R15, BNI-R17, BNI-R21, BNI-R30, and BNI-R32.

Clinical information and viral load data on all of the HPeV positive patients are shown in Table 11 In analogy with enteroviruses, positive samples only occurred in summer and autumn. Only one patient had accompanying respiratory symptoms. A very large range of viral loads was observed (3170 - 503,377,290 copies/ml), and one of the highest viral loads occurred in a control patient without symptoms.

Table 11: Clinical information and viral load data of HPeV positive patients.

Sample code	Sampling month	Virus load RNA (copies per g or mL of stool)	Patient age (months)	Gender	Diarrhoea	Fever	Other symptoms	Animal contact	Symptoms in contact persons	Recent travel history
BNI-R21	August	502,380	14	Male	Yes	-	-	-	-	-
BNI-R90	August	502,377,290	6	Female	Yes	-	-	-	-	-
BNI-R32	September	251,790	28	Male	Yes	Yes	Cough, rhinorrhoea	Pet mouse	Father	
BNI-R09	September	15,886,560	4	Male	Yes	-	-	-	-	-
BNI-R30	October	502,380	9	Female	Yes	-	Vomiting, 5 times daily	-	-	-
BNI-R17	October	3,170	21	Female	Yes	-	-	-	-	Turkey
BNI-67	October	3,169,790	82	Female	Yes	-	-	Budgie	-	-
BNI-R15	November	5,020	11	Male	Yes	Yes	Vomiting, 2 times daily	-	Both parents	Mallorca
BNI-R04	November	158,865,650	5	Female	-	-	-	Cat	-	-

3.3.2.3 Analysis and genotyping of detected HPeVs

All positive samples were submitted to culture in MA-104 cells, for virus isolation, but none of them have produced CPE. The subsequent genetic analysis of the detected HPeVs was made on the viral RNA extracted directly from the stool samples.

In order to appreciate the types of parechoviruses present in our patients the whole P1 protein (VP0, VP3, and VP1) was sequenced in all of the positive samples for HPeV. For sequencing of the entire P1 region, first strand cDNA was synthesized with GeneRacer Oligo dT primer (Invitrogen, Karlsruhe, Germany). For PCR amplification, primers where designed to anneal in the highly conserved terminal portion of the 5' UTR and to the end of VP1 gene. The PCR products were sequenced directly; amplicons were fully sequenced with the synthesis of specific primers according to a primer walking strategy. Additionally, we have sequenced the 3D gene regions to verify recombination. These genomic regions are the largest ones found in the highly conserved region of the non-structural proteins.

All detected HPeVs had the P1 and 3D amplified and sequenced, except for the sample BNI-R17, which showed the lowest virus concentration. The nearly full genome sequence of the sample BNI-67 was submitted to GenBank under accession number EU024629. The P1 and 3C/3D sequences of the other samples were submitted to GenBank according to the respective genome region (Table 12).

Table 12: HPeV samples and respective GenBank accession number according to sequenced genomic region.

HPeV sample	GenBank accession number	
	P1 gene	3C/3D genes
BNI-90	EU024630	EU024637
BNI-R04	EU024631	EU024638
BNI-R09	EU024632	EU024639
BNI-R15	EU024633	EU024640
BNI-R21	EU024634	EU024641
BNI-R30	EU024635	EU024642
BNI-R32	EU024636	EU024643

The genotyping of samples was based on the VP1 amino acid sequence identity. This methodology has extensively been used to correlate serotypes and to classify new genotypes among enteroviruses (Oberste *et al.*, 1999). The same criterion has recently been proposed for HPeV type determination (Al-Sunaidi *et al.*, 2007).

As shown in Figure 20, the majority of the analysed HPeVs samples clustered with a group of contemporary HPeV1 strains, including virus isolate BNI-788St. In concordance with earlier observations (Al-Sunaidi *et al.*, 2007), the prototype strain HPeV1 Harris had only a basal relationship with the contemporary group. The amino acid identity of HPeV samples with HPeV1 Harris prototype varied from 88.4% to 89.6%. In relation to the recent isolated HPeV1 BNI-788St, the similarity ranged from 94.8% to 97.8%.

The BNI-R30 from our cohort segregated as a separate lineage between the historical type 1 Harris strain and the contemporary group. Another virus, BNI-67, could not be associated with any type. It had basal relationship with types 1 and 2. A 207 amino acid fragment from the VP1 protein of the BNI-67 virus was compared with established reference genomes. As shown in Table 13, the amino acid genetic distance between BNI-67 virus and the most similar related HPeV type, the HPeV1 Harris, was 0.22. This value exceeded the distance between the two most similar established HPeV types, which was 0.21 between HPeV1 Harris and HPeV2 Williamson. It was thus assumed that BNI-67 constituted a novel HPeV type. During preparation of the manuscript, a sixth parechovirus type was identified (Watanabe *et al.*, 2007); it was denominated as strain NII561-2000. When this strain was included in the P1 phylogenesis, it clustered with BNI-67 with high bootstrap support (Figure 20). VP1 amino acid distance between HPeV6 BNI-67 and NII561-2000 was 0.04.

Table 13: Amino acid distance matrix for HPeV VP1 protein.

HPeV type*	1	2	3	4	5
BNI-67	0.22	0.27	0.27	0.29	0.28
1	-	0.21	0.28	0.28	0.24
2	-	-	0.29	0.29	0.25
3	-	-	-	0.35	0.30
4	-	-	-	-	0.23
5	-	-	-	-	-

*The following reference strains were used to define HPeV type: Type 1, Harris (accession number S45208), and A1086-99 (AB112485); Type 2, Williamson (AJ005695); Type 3, A308/99 (AB084913) and Can82853-01 (AJ889918); Type 4, K251176-02 (DQ315670) and T75-4077 (AM235750); Type 5, CT86-6760 (AF055846) (formerly classified as Type 2) and T92-15 (AM235749). Calculation was based on a 207 amino acid fragment of the VP1 gene as described by Al-Sunaidi, 2007, using the p-distance model in MEGA4 program. Distance values for types represented by more than one reference strain are based on average distances.

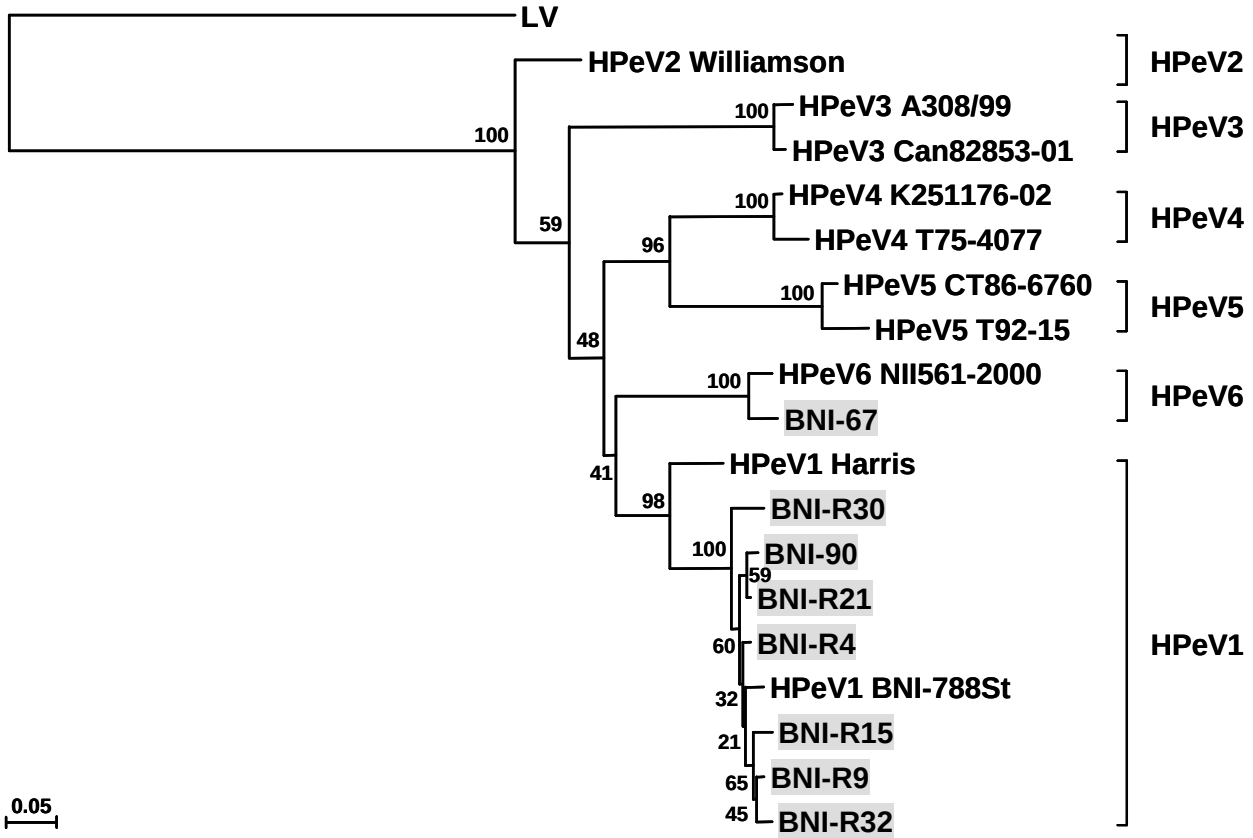

Figure 26: Phylogenetic tree based on the partial amino acid sequence of VP1, showing the relationship between detected HPeVs (highlighted in grey) and representative HPeVs. The tree was generated by the neighbour-joining method in the TREECON version 1.3b program. Numbers at the nodes represent the percentage of bootstrap values calculated from 1,000 replicates. Ljungan virus (LV) strain 87-012 (AF327920) was used as out-group.

3.3.2.4 Analysis of the 3D gene from detected HPeVs

The 3D gene is well conserved in HPeVs. The known prototypes and detected HPeVs described in this paper have a percentage identity ranging from 81.0% (between HPeV5 T92-15 and BNI-R32 virus) to 96.5% (between HPeV3 viruses) for nucleotide sequences, and from 93.5% (between HPeV5 T92-15 and BNI-R21) to 98.5% (between HPeV3 viruses) for amino acid sequences.

A phylogenetic tree based on the 3D nucleotide sequences of detected HPeVs and prototypes was generated and depicted in Figure 27. The sequences have shown a different pattern of clustering when compared with VP1 phylogenetic analysis (Figure 26). In the 3D phylogenetic tree, the majority of the HPeVs from the same genotypes are neither grouped together nor form polyphyletic groups, with the exception of genotype 3 HPeVs. The detected HPeVs presented different clustering patterns in relation to the prototype HPeVs; the most distinct sequences were found in BNI-R4 and BNI-90 viruses. The 3D sequence of BNI-R15 virus was most similar to an HPeV prototype and had 95.1% of nucleotide identity to HPeV6 NII561-2000. The discrepancy between VP1 and 3D phylogenetic analysis is an evidence of possible recombination events during the evolution history of HPeV.

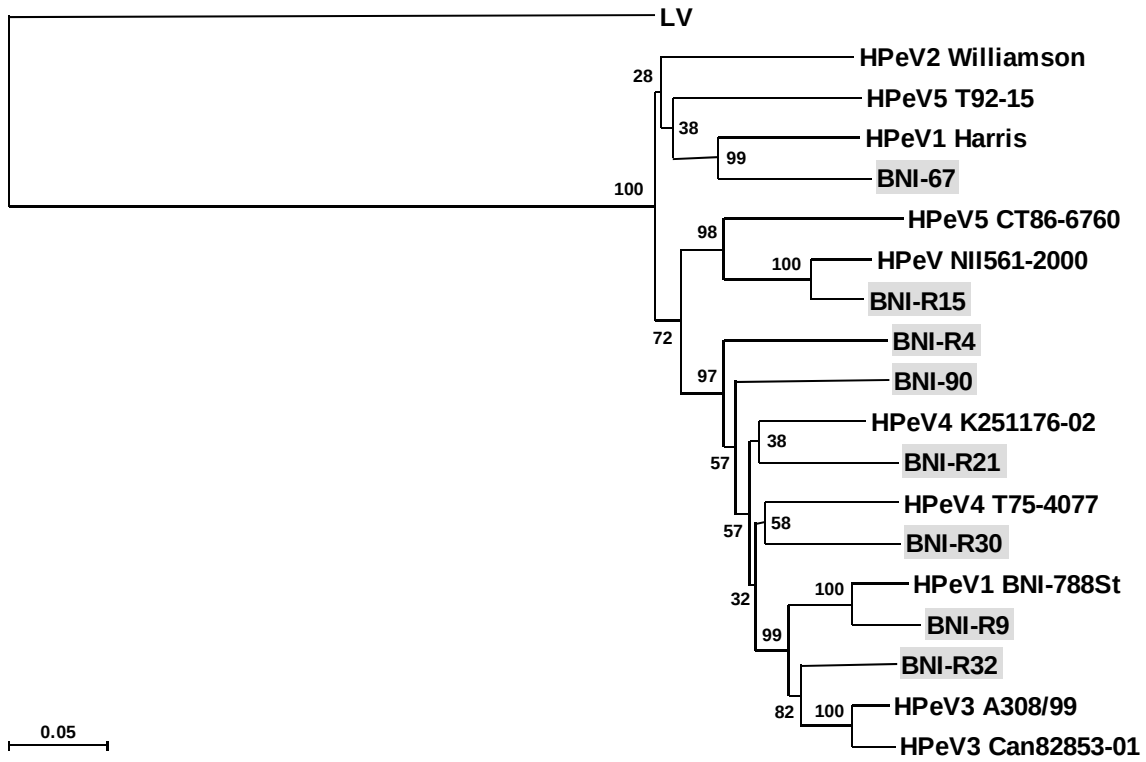

Figure 27: Phylogenetic tree based on the nucleotide sequence of 3D, showing the relationship between detected HPeVs (highlighted in grey) and representative HPeVs. The tree was generated by the neighbour-joining method in the TREECON version 1.3b program. Numbers at the nodes represent the percentage of bootstrap values calculated from 1,000 replicates. Ljungan virus (LV) strain 87-012 (AF327920) was used as out-group.

3.3.2.5 Genome sequencing of the new HPeV BNI-67

During genotyping of detected HPeVs, the BNI-67 virus was classified as a new genotype, and later on as a new member of HPeV type 6. The nearly complete genome sequence of the HPeV6 BNI-67 virus was determined by RT-PCR and sequencing of cDNA genome fragments. The sequencing strategy was the same as the one used for HPeV1 BNI-788St, as it was described above.

The initial portion of the 5' UTR sequence was not determined. It is true that different consensus primers composed of the first 13, 16, 19, 24 and 27 nucleotides based on published HPeV sequences were used, but have failed to amplify the expected fragment.

The extreme 3' end of the genome was reverse transcribed with GeneRacer Oligo dT primer (Invitrogen, Karlsruhe, Germany), followed by amplification with gene specific primers and GeneRacer 3' nested primer as described in the "material and methods"-section numbered 2.3.9.2. The 3' UTR nucleotide sequences were obtained by direct sequencing of PCR amplicons.

The virtually complete genome sequence of HPeV6 BNI-67 virus has been submitted to the GenBank database under accession number EU024629.

3.3.2.5.1 Genetic analysis of HPeV6 BNI-67

A total of 7090 nucleotides comprising almost the complete genome of the HPeV6 BNI-67 were determined. The genome organization was similar to that of other HPeVs and consists of a 5'UTR (only the last 436 nucleotides have been sequenced), followed by a large ORF of 6546 nucleotides that encodes a putative polyprotein precursor of 2182 amino acids, a 3' UTR of 90 nucleotides and a poly(A) tail.

The partially sequenced 5' UTR has shown to contain the polypyrimidine-rich tract found in all studied picornaviruses and is located 17 nucleotides upstream from the initiation codon. As observed in all HPeVs, the initiation codon is found to form a Kozak sequence (ANNAUGG).

Phylogenetic tree analysis with complete P1 amino acid sequences of HPeV prototypes showed that the HPeV6 BNI-67 was most similar to HPeV6 NII561-2000 so that both strains are clustered in the same genetic group (Figure 26).

In order to complete genome sequences of HPeVs, the nucleotide and amino acid sequence identities of HPeV6 BNI-67 were summarized in Table 14.

Table 14: Percentage identities of HPeV6 BNI-67 nucleotide and deduced amino acid sequences to all known HPeVs.

	Nucleotide similarity (amino acid similarity)									
Region	**HPeV1[a]**	**HPeV1[b]**	**HPeV2[c]**	**HPeV3[d]**	**HPeV3[e]**	**HPeV4[f]**	**HPeV4[g]**	**HPeV5[h]**	**HPeV5[i]**	**HPeV6[j]**
VP0	73.7	72.3	75.5	69.4	69.9	72.0	72.9	72.1	72.1	92.0
	(78.5)	(79.2)	(81.0)	(73.4)	(73.7)	(80.6)	(81.3)	(76.9)	(76.8)	(99.0)
VP3	74.2	72.1	71.7	65.1	67.3	68.1	68.4	65.8	67.3	92.2
	(83.8)	(81.3)	(78.3)	(72.3)	(73.0)	(76.5)	(76.9)	(71.0)	(71.0)	(97.6)
VP1	73.0	69.3	67.8	68.4	68.9	69.3	69.8	68.0	69.0	90.7
	(76.2)	(74.0)	(72.3)	(73.0)	(73.0)	(69.4)	(68.5)	(68.5)	(68.1)	(95.7)
2A	79.3	78.2	76.9	77.2	77.9	78.9	77.8	78.9	78.7	89.1
	(89.3)	(88.7)	(87.3)	(83.9)	(83.2)	(89.3)	(90.7)	(88.7)	(87.3)	(96.0)
2B	80.1	82.8	77.6	81.7	83.3	83.6	80.9	80.3	80.1	93.2
	(98.4)	(98.4)	(96.7)	(98.4)	(98.4)	(98.4)	(98.4)	(100)	(100)	(99.2)
2C	78.0	80.1	75.2	80.5	80.7	82.0	80.4	82.5	80.2	80.2
	(92.4)	(96.4)	(85.7)	(97.3)	(97.6)	(96.7)	(97.6)	(95.4)	(93.3)	(92.4)
3A	77.2	80.3	76.6	78.6	80.3	79.5	77.2	82.3	76.6	76.9
	(87.2)	(91.5)	(83.8)	(88.9)	(91.5)	(90.6)	(90.6)	(81.2)	(92.3)	(88.9)
3B	78.3	76.7	75.0	75.0	78.3	75.0	68.3	78.3	66.7	76.7
	(95.0)	(90.0)	(95.0)	(90.0)	(95.0)	(85.0)	(90.0)	(90.0)	(85.0)	(80.0)
3C	83.3	81.3	82.5	80.5	79.8	80.7	80.0	82.2	81.8	82.0
	(97.5)	(97.0)	(97.5)	(96.5)	(97.0)	(96.5)	(96.5)	(97.5)	(96.5)	(98.5)
3D	87.8	82.3	83.4	81.7	81.4	82.9	82.2	82.4	84.0	82.7
	(97.4)	(95.3)	(94.9)	(94.5)	(94.0)	(94.7)	(94.9)	(95.5)	(94.2)	(95.7)
3' UTR	87.8	84.6	90.0	85.7	85.7	80.6	85.7	82.6	88.9	81.5
P1	73.7	71.3	71.9	67.9	69.0	69.9	70.4	68.9	69.8	91.7
	(81.0)	(79.3)	(78.0)	(73.2)	(73.7)	(76.5)	(76.7)	(73.5)	(73.7)	(97.5)
P2	78.8	80.2	76.1	79.9	80.6	81.6	79.9	81.1	79.8	85.1
	(92.8)	(94.8)	(88.4)	(94.2)	(94.2)	(95.2)	(96.0)	(94.7)	(93.2)	(94.7)
P3	84.8	81.6	81.8	80.8	80.8	81.6	80.6	82.1	81.9	81.6
	(95.9)	(95.0)	(93.9)	(94.0)	(94.4)	(94.3)	(94.5)	(93.8)	(94.3)	(95.0)
P2-P3	82.2	81.0	79.4	80.4	80.7	81.6	80.3	81.7	81.0	83.1
	(94.6)	(95.0)	(91.5)	(94.1)	(94.3)	(94.7)	(95.2)	(94.2)	(93.8)	(94.9)
ORF	79.2	77.6	76.7	76.0	76.5	77.4	76.8	77.1	77.0	86.1
	(89.8)	(89.4)	(86.7)	(86.7)	(87.0)	(88.2)	(88.5)	(86.7)	(86.6)	(95.8)

[a]Harris, [b]BNI-788St, [c]Williamson, [d]A308/99, [e]Can82853-01, [f]T75-4077, [g]K251176-02, [h]CT86-6760, [i]T92-15, [j]NII561-2000

As observed in the P1 phylogenetic tree, HPeV6 BNI-67 showed a very high degree of nucleotide similarity to that of the HPeV6 NII561-2000 virus, i.e. 91.7% and 97.5% of nucleotide and amino acid identities, respectively (table 3.7). In order to verify the relationship between the completely sequenced HPeVs and HPeV6 BNI-67 based on nucleotide similarities, we used SimPlot to further analyze these two strains (Figure 28).

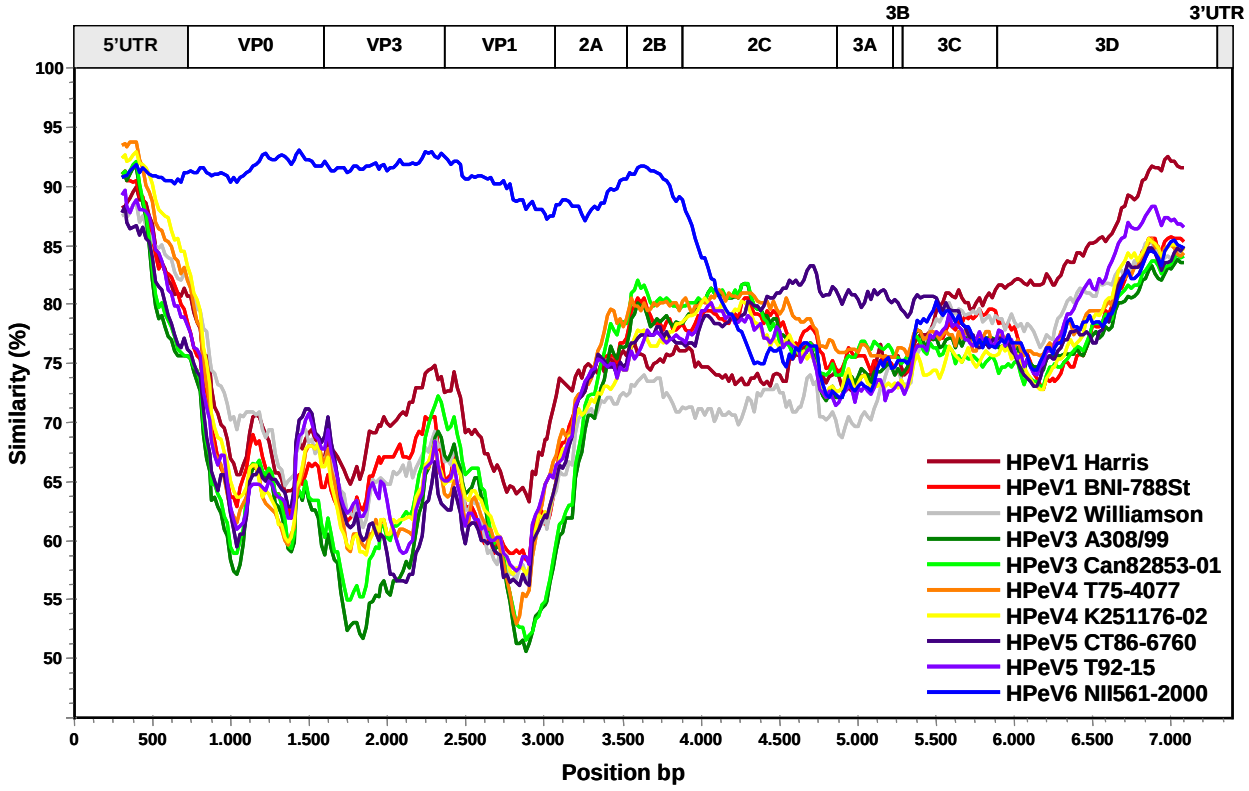

Figure 28: Full-length genome similarity plot of HPeV6 BNI-67 against HPeVs calculated by SimPlot 3.5.1. Each point represents the similarity between the BNI-67 and a given HPeV, within a sliding window of 600 nucleotides centred on the position plotted, with a step of 20 residues between points. Positions containing gaps were excluded from the analysis by gap stripping. Jukes and Cantor correction was applied.

The high nucleotide similarity between BNI-67 and NII561-2000 was observed across the P1 region and began to cease in the initial portion of the 2C gene. From the position 4144 to 4380 in the 2C region, BNI-67 virus showed highest similarity with HPeV4 K251176-02. From position 4380 to 5517, corresponding from the middle portions of 2C to 3C, the observed similarity was to HPeV5 CT86-6760. Beginning from position 5517 and across the 3D gene, the BNI-67 virus was most related to HPeV1 Harris. These results are suggestive of past recombination events among HPeVs. The breakpoints of possible intergenotypic recombination could better be visualized by a BootScan analysis between the BNI-67 virus and the HPeVs (Figure 29).

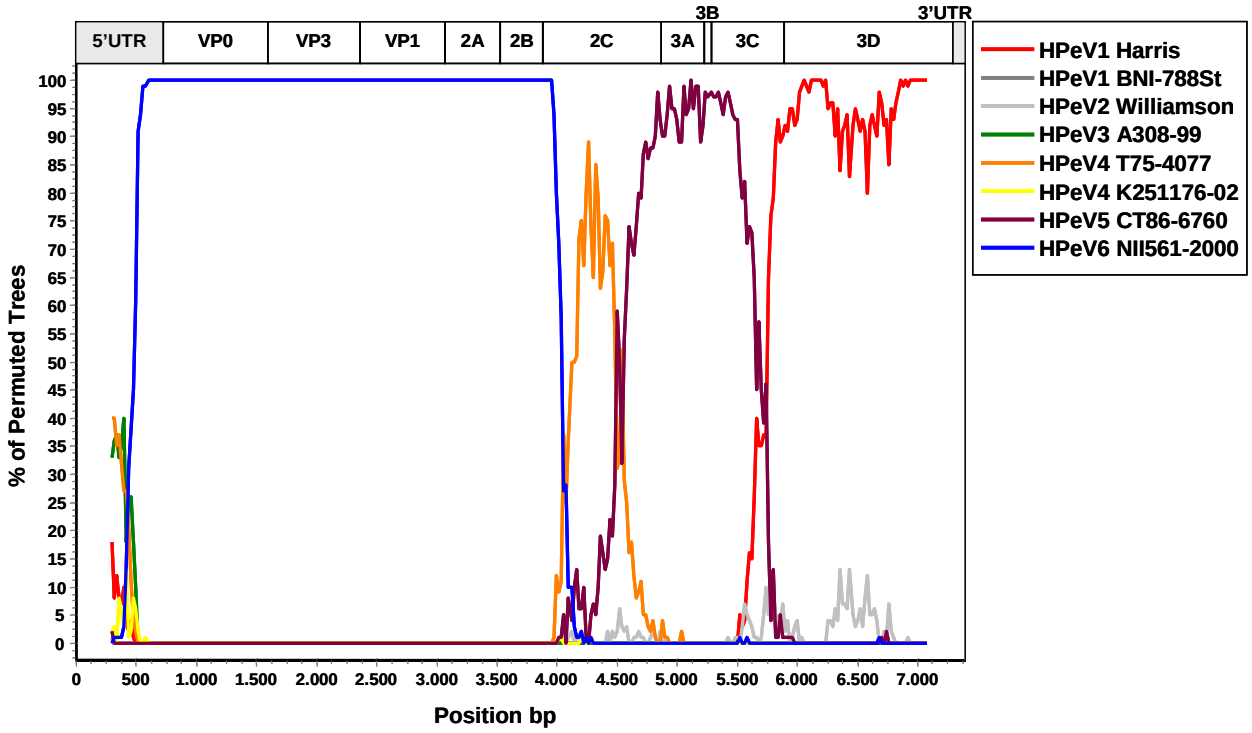

Figure 29: BootScan analysis plot, calculated from 100 bootstrap replicates, of HPeV6 BNI-67 against HPeVs calculated by SimPlot 3.5.1., showing recombination break points. Each point represents the similarity between the BNI-67 and a given HPeV, within a sliding window of 600 nucleotides centred on the position plotted, with a step side of 20 residues between points. Positions containing gaps were excluded from the analysis by gap stripping. Jukes and Cantor correction was applied.

When the complete ORF of BNI-67 virus was compared to the complete sequenced HPeVs, nucleotide and amino acid similarities ranged from 76.0% to 86.1% and from 86.7% to 95.8%, respectively. These values are in the same range of similarity as those obtained for other HPeVs.

With the exception of viruses from genotype 3, as it was observed in other HPeVs, the BNI-67 virus presented the RGD motif at the C-terminus of VP1 gene. Like with the HPeV6 NII561-2000, an insertion of three amino acids residues was likewise observed in the BNI-67 virus, 7 amino acids upstream to the RGD motif, which is not found in none of the known HPeVs (Figure 30).

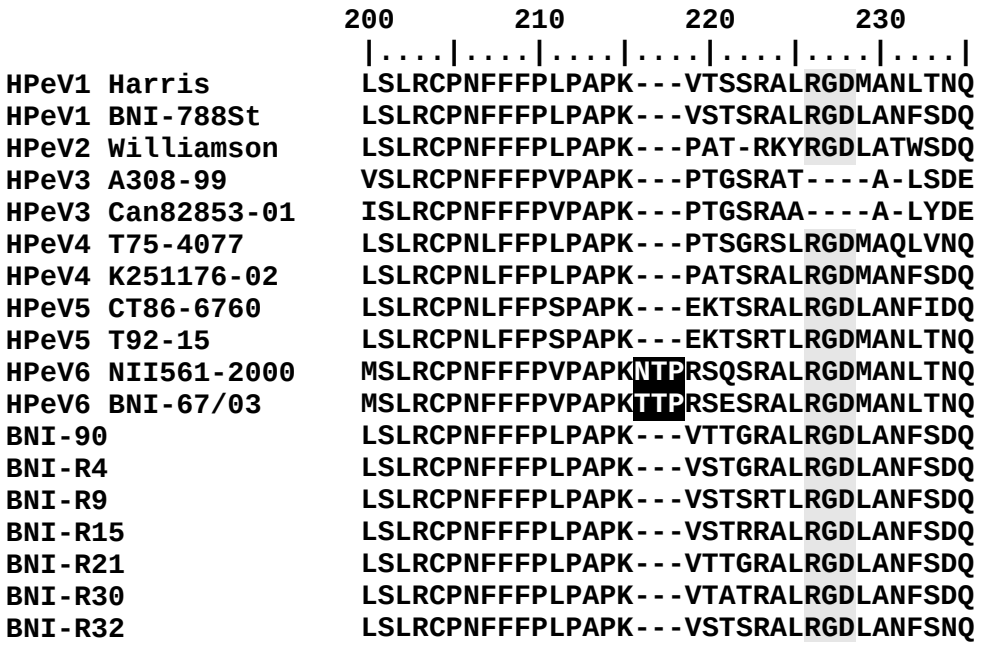

Figure 30: Amino acid alignment of VP1 C-terminus of HPeVs strains and samples. Insertion of three amino acids (in white and highlighted in black) found in the known HPeV6 viruses located in the C-terminus of VP1 gene. The RGD motif is in black, highlighted in grey.

The Common motifs found in the non-structural proteins of HPeVs, particularly in the 2C, 3C and 3D proteins, were observed in the HPeV6 BNI-67 virus. In the predicted helicase protein 2C, the motifs GXXGXGK(S/T) and DDLXQ where observed, as well as the active-site cysteine motif GXCG in the 3C protease, and the active site YGDD of the RdRp, or 3D gene, together with the motifs KDELR, PSG and FLKR.

HPeV6 BNI-67 virus, like all others HPeVs, had nine polyprotein cleavage sites. As observed in Table 15, all known HPeVs share the same conserved polyprotein cleavage sites at VP3/VP1, 2A/2B, 2B/2C, 3A/3B, 3B/3C, and 3C/3D. The more variable cleavage site between HPeVs is found in the VP0/VP3 junction, and HPeV6 BNI-67 had the same cleavage amino acid sequence (N/G) of the related HPeV6 NII561-200 and HPeV3. As observed in the SimPlot and BootScan analyses (Figures 28 and 29), the HPeV6 BNI-67 virus is more closely related to HPeV5 CT86-6760 at the 2C/3A junction, although it presents the same amino acid cleavage sequence (Q/S) of HPeV5 T92-15.

Table 15: Putative amino acid sequences of cleavage sites between proteins of HPeVs.

Virus*	VP0/VP3	VP3/VP1	VP1/2A	2A/2B	2B/2C	2C/3A	3A/3B	3B/3C	3C/3D
HPeV1[a]	N/A	Q/N	Q/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV1[b]	N/A	Q/N	Q/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV2[c]	T/A	Q/N	Q/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV3[d]	N/G	Q/N	E/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV3[e]	N/G	Q/N	E/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV4[f]	N/N	Q/N	Q/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV4[g]	N/N	Q/N	Q/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV5[h]	N/S	Q/N	Q/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV5[i]	N/N	Q/N	Q/S	Q/G	Q/G	Q/S	E/R	Q/R	Q/G
HPeV6[j]	N/G	Q/N	Q/S	Q/G	Q/G	Q/T	E/R	Q/R	Q/G
HPeV6[k]	N/G	Q/N	Q/S	Q/G	Q/G	Q/S	E/R	Q/R	Q/G

* The analysed HPeVs: [a]Harris, [b]BNI-788St, [c]Williamson, [d]A308-99, [e]Can82853-01, [f]T75-4077, [g]K251176-02, [h]CT86-6760, [i]T92-15, [j]NII561-2000 and [k]BNI-67.

The predicted secondary structure of the 3' UTR of HPeV6 BNI-67 was determined by Mfold program version 3.2 (Zuker, 2003) and is depicted in Figure 31. Different from what has previously been described for other HPeVs (Abed & Boivin, 2005), the virus is composed of two stem-loops, instead of just one, and the same time, two highly conserved repeats arranged in tandem were also observed (Al-Sunaidi *et al.*, 2007).

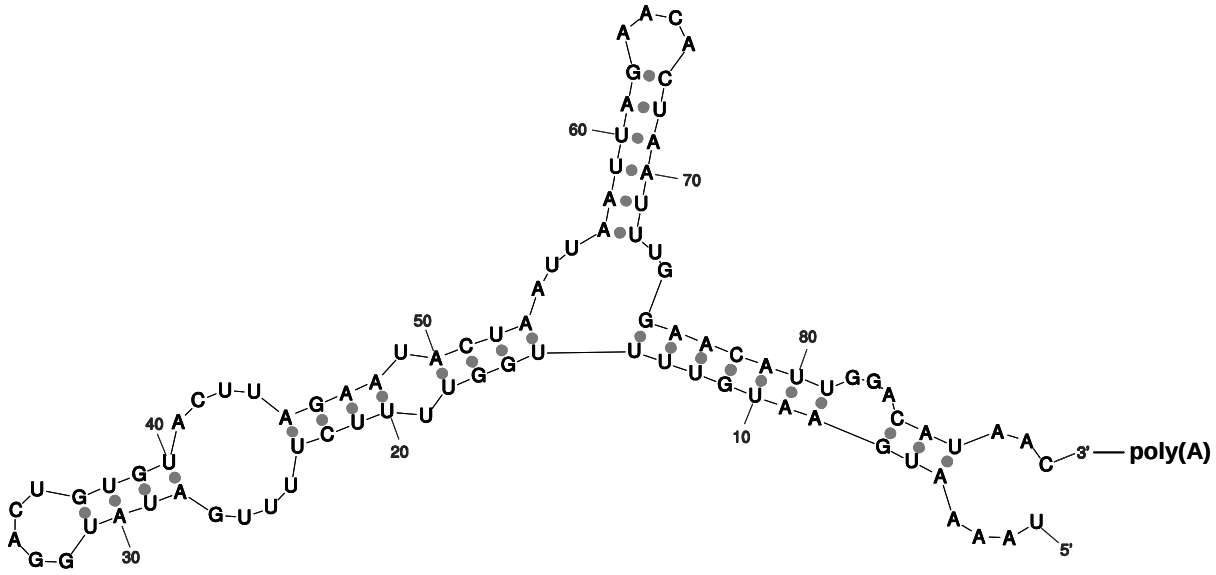

Figure 31: Predicted secondary structure of the 3' UTR of HPeV BNI-67.

3.4 Universal amplification of single-stranded, positive RNA viruses with modified PCR primers aimed to the conserved GDD motif of RdRp

3.4.1 Optimization of the two steps nested RT-PCR +RNA GDD virus discovery method

To implement the +RNA GDD virus discovery method, we have tested two *Taq* DNA polymerases: Titanium *Taq* DNA polymerase (Clontech, USA), a nuclease-deficient, N-terminal truncated mutant or KlenTaq DNA polymerase, and the Platinum *Taq* DNA polymerase (Invitrogen, Karlsruhe, Germany). The KlenTaq DNA polymerase was described to improve fidelity catalysis during PCR (Barnes, 1992) as well as fidelity of isoG-isoC hybridization (Johnson *et al.*, 2004b). Reverse transcription (RT) was performed using RNA extracts from cell cultured HPeV BNI-788St supernatant and a clinical sample containing HCoV-OC43 RNA. Each RNA extract was reverse transcribed with $P[dN]_6$ or +RNA primer in separated reactions. The nested PCRs were performed with the Titanium *Taq* PCR Kit (Clontech, USA) and Platinum *Taq* DNA polymerase reagents (Invitrogen, Karlsruhe, Germany) in separated reactions for each cDNA template. The products from the first and the second round amplification steps were analysed by 1.5% agarose gel electrophoresis as show in Figure 32

As observed in Figure 32, Platinum *Taq* DNA polymerase produced higher yields of fragments when compared with Titanium *Taq* DNA polymerase, particularly to HCoV-OC43. Amplification catalysed by Titanium *Taq* enzyme also increased the formation of primer-dimers.

When the products from first and second round amplification were compared, the fragment patterns were almost the same. Differences were observed in the products yielded by first-stranded cDNA primed with +RNA and $P[dN]_6$. For HPeV BNI-788St RNA extract, amplification only occurred when RT reaction was primed with the modified +RNA primer. In contrast, for HCoV-OC43 RNA extract, the first stranded cDNA synthesized with $P[dN]_6$ yielded more fragments than with +RNA primer (Figure 32B). However, fragments could be amplified in both RNA templates when RT reaction was primed with +RNA primer.

Based on these first results described above, all further first strand cDNA syntheses were performed with the modified +RNA primer and PCR experiments carried out with *Taq* DNA polymerase.

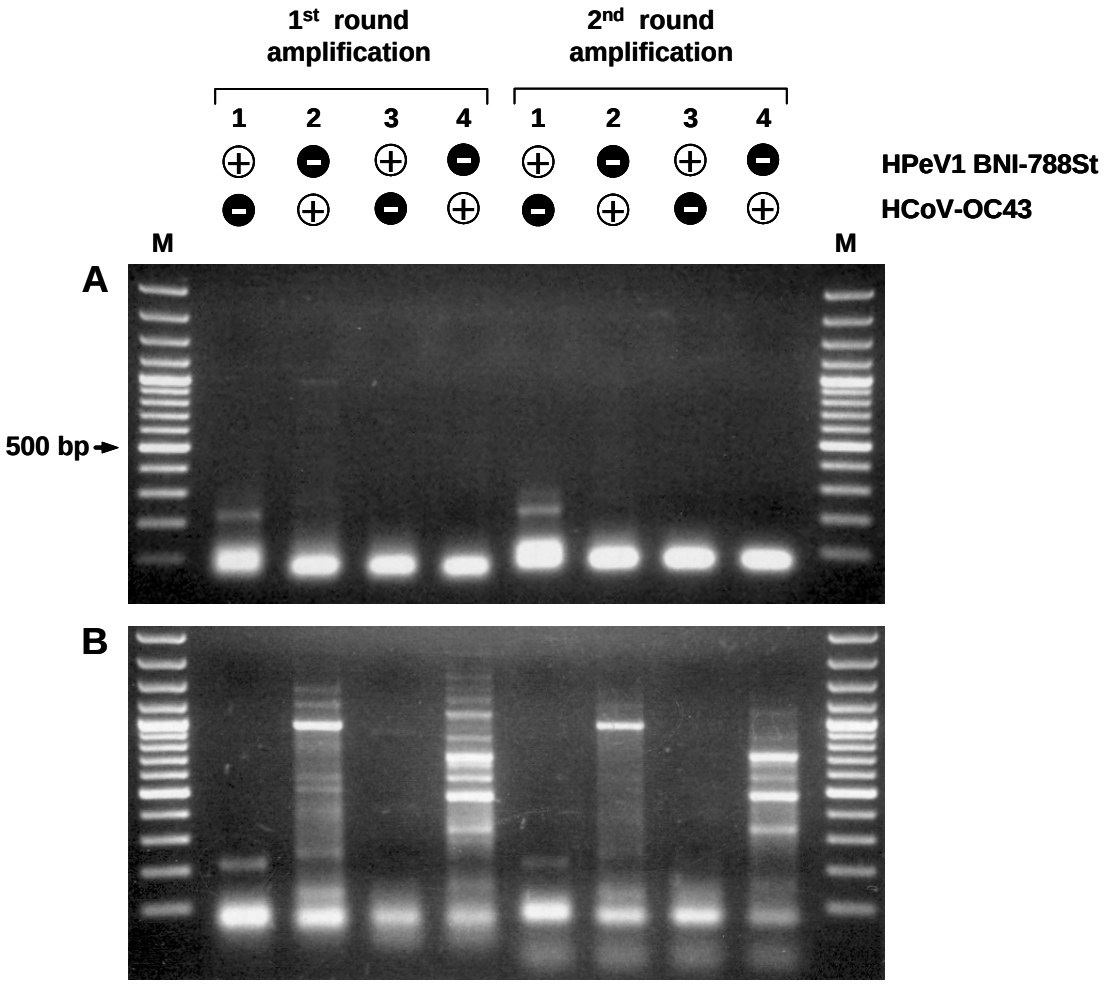

Figure 32: Agarose gel electrophoresis of +RNA GDD virus discovery method, with HPeV1 BNI-788St and HCoV-OC43 genomic RNA. Lanes 1 and 2, RT primed with +RNA primer. Lanes 3 and 4, RT primed with $P[dN]_6$. A, amplification with Titanium *Taq* DNA PCR kit. B, amplification with Platinum *Taq* DNA polymerase. M, 100 bp DNA ladder.

3.4.1.1 The +RNA GDD virus discovery method with full length *in vitro* transcribed RNA of Chikungunya virus

The +RNA GDD virus discovery method was then optimized by using a high pure and full length *in vitro* transcribed RNA of chikungunya virus, strain Reunion. The *in vitro* transcribed RNA was 10-fold diluted, from 10^{-1} to 10^{-7}, and then submitted to RT reaction with primer +RNA. Following RT reaction, first-stranded cDNAs were submitted to the nested PCR. Amplicons from first and second round amplification steps were analysed by 1.5% agarose gel electrophoresis as shown in Figure 33. Only weak fragments could be observed in the first round amplification step. In the second round amplification step, fragments could be observed as very faint bands; an increased formation of primer-dimers could also be observed.

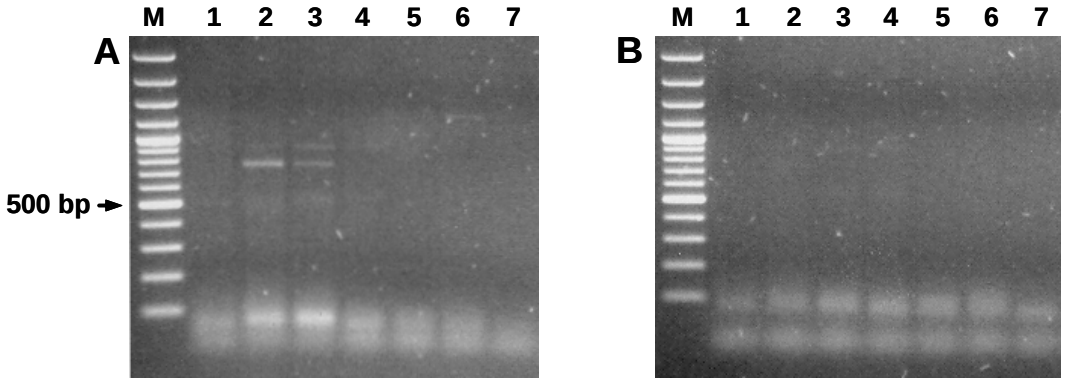

Figure 33: Agarose gel electrophoresis of a 10-fold dilution of chikungunya full length *in vitro* transcribed RNA, from 10^{-1} to 10^{-7}. A, first round amplification. B, second round amplification. M, 100 bp DNA ladder.

Furthermore, 1 µl of reaction mix from each tube of the first and second round amplification steps were reamplified individually with the standard primers +RNA-*Not*I and Back-*Not*I. After reamplification, fragments could be easily discriminated as shown in Figure 34.

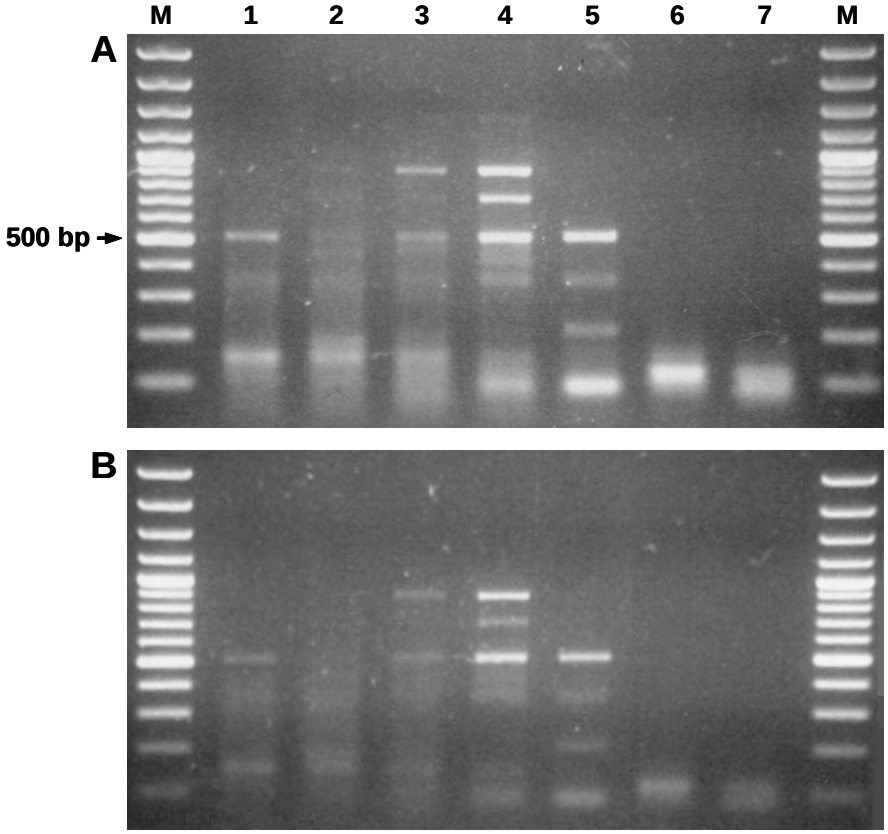

Figure 34: Agarose gel electrophoresis of products reamplified from the +RNA GDD virus discovery method. A, reamplification of products from the first amplification step. B, reamplification of products from the second amplification step.

Furthermore, there was no difference in the amplification pattern between reamplified products derived from first or second round amplification reactions. Thus, in the subsequent optimization experiments, the modified primers Uni1 and Uni2 were substituted for the standard +RNA-*Not*I and Back-*Not*I primers in the second round amplification step. Moreover, this modification enables the direct cloning of amplicons, when necessary, into TOPO TA-based vectors or by digestion with *Not*I restriction enzyme.

The common fragment of about 500 base pairs observed in all positive reamplified dilutions (Figure 34) was gel purified and directly submitted to sequencing analysis in both orientations. The obtained sequences were then assembled in a single contig with the use of SeqMan II software and aligned with the complete genome sequence of the chikungunya Reunion strain. The 5' and 3' terminal nucleotides of the product could be sequenced and primer-biding sequences were determined, as shown in Figure 35.

In the sequenced product, the first 9 nucleotides (GGMGACGAC) of the 5' end of primer +RNA hybridized in a region of chikungunya genome located in the position 2090 to 2098 (nucleotide sequence GGCGACTTG), distinct from the GDD motif. The GDD motif nucleotides of chikungunya virus Reunion was found to be located in the genomic position 7007 to 7015 of ORF1 and coded by the nucleotide sequence GGCGACGAC.

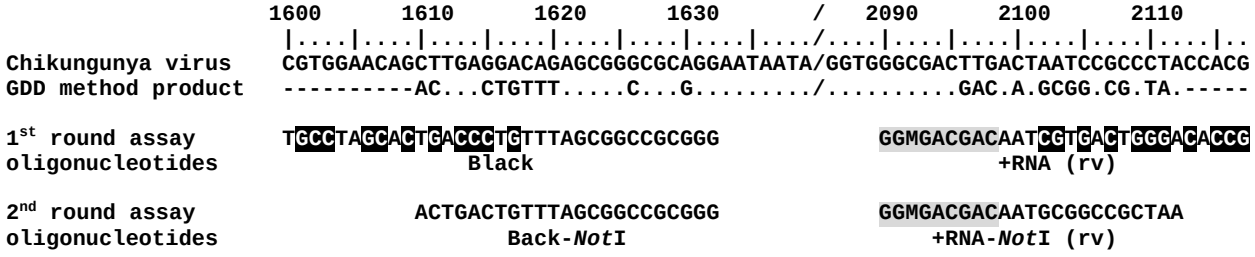

Figure 35: Nucleotide alignment of chikungunya virus Reunion strain and +RNA GDD virus discovery method product. Assay oligonucleotides are shown below the alignment panel. Base count in the top line is based on chikungunya virus, which also serves as the comparison sequence in the alignment. Dots represent identical bases in compared sequences; deviations are spelled out. A slash (/) represents a gap in the alignment; (rc) means that the reverse complementary sequence is shown for the reverse primer. The GDD motif nucleotide sequences of reverse primers are shown in black, highlighted in grey. IsoG and isoC non-standard nucleotides in the first round assay primers are shown in white and highlighted in black.

During reverse transcription, the +RNA primer has the tendency to hybridize with homologue sequences and especially so with GDD sequence. However, during first round amplification, the annealing time favoured amplification of fragments up to 1,500 base pairs. In the case of chikungunya virus, amplification of the 500 base pair fragment was strongly favoured by the forward Back primer. The hybridization site of this primer was in the position 1621 to 1632 (nucleotide sequence AGCGGGCGCAGG), from about 500 base pairs upstream of the +RNA primer hybridization site.

3.4.1.2 +RNA GDD virus discovery method with cell cultured chikungunya virus

Viral particles from cell cultured chikungunya virus supernatant were enriched by ultracentrifugation. The pellet was resuspended and submitted to DNase digestion for removal of mitochondrial and genomic DNA from lysed cells. Following RNA extraction, genomic RNA was 10-fold diluted up to 10^{-3} and submitted to the +RNA GDD virus discovery method, including a 10^{-5} dilution of the full length *in vitro* transcribed chikungunya RNA. After reverse transcription with +RNA primer, nested PCR was followed by first round amplification with the modified Back and +RNA primers, and second round amplification with +RNA-*Not*I and Back-*Not*I primers. The products were analysed by 1.5% agarose gel electrophoresis. As shown in Figure 36, the same fragment of about 500 base pairs obtained previously when using *in vitro* RNA transcript could be obtained. Interestingly, higher yields of fragments were obtained inversely proportional to RNA dilution.

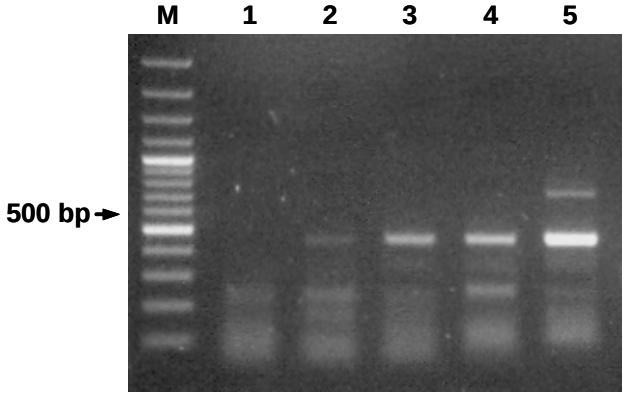

Figure 36: Agarose gel electrophoresis of a 10-fold dilution of chikungunya genomic RNA. Lanes 1 to 4, dilutions from 10^{0} to 10^{-3}. Lane 5, a 10^{-5} dilution of full length *in vitro* transcribed chikungunya RNA. M, 100 bp DNA ladder.

Following the same protocol for viral particles enrichment and RNA extraction, HPeV1 BNI-788St genomic RNA obtained from cell cultured supernatants was tested next. A 10-fold dilution, from 10^{0} to 10^{-2}, was submitted to the +RNA GDD virus discovery method together

with a 10^{-5} dilution of chikungunya genomic RNA as a control. As shown in Figure 37, the RNA used in the original extraction concentration (10^{0}) presented the same pattern of amplification from the 10^{-2} dilution, which in turn was different from the pattern obtained by the 10^{-1} dilution. As expected, the chikungunya RNA dilution has amplified the fragment of 500 base pairs.

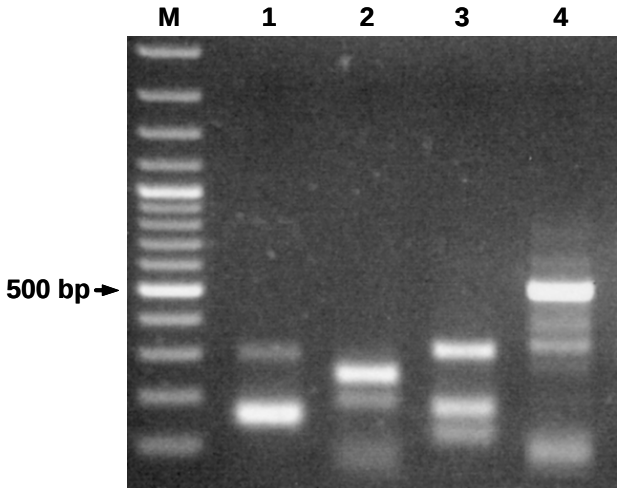

Figure 37: Agarose gel electrophoresis of a 10-fold dilution of HPeV1 BNI-788St genomic RNA. Lanes 1 to 4, dilutions from 10^{0} to 10^{-3}. Lane 5, a 10^{-5} dilution of full length *in vitro* transcribed chikungunya RNA. M, 100 bp DNA ladder.

The fragments with more than 200 base pairs, obtained from HPeV1 BNI-788St RNA, were gel purified and directly sequenced. The obtained sequences were analysed with BLAST nucleotide search and showed high similarity with the 16S ribosomal RNA gene of *Mycoplasma hyorhinis*, a common bacterium contaminant of cell culture (Timenetsky *et al.*, 2006). Thus, the presence of amplified *M. hyorhinis* DNA indicated contamination of the cell culture medium. In this case, when cell culture contamination occurs, the bacteria cells are generally collected through the 0.22 µl filter with viral particles. During DNase digestion, bacterial nucleic acid is protected by the cell membrane and is consequently extracted. Another reason for selective amplification of *Mycoplasma* is the fact that bacterial DNA has more GC content than RNA viruses. Consequently, the GC-rich content of Back primer has favoured the amplification of bacterial DNA.

3.4.1.3 Alternative forward primers for the +RNA GDD virus discovery method

In order to overcome the problem of selective amplification of contaminant nucleic acid from *M. hyorhinis* in cell culture, as observed for HPeV BNI-788St supernatant extracts, alternative forward primers were tested. It is important of note that appropriate antibiotic supplementation in the cell culture have to be done. The primers PC2S2a and PC2S3a,

previously designed to hybridize to the conserved motif A of the RdRp from CoVs and located upstream from GDD motif, were used. The first strand cDNA synthesis reactions were performed with a 10-fold dilution, from 10^{-1} to 10^{-8}, of the full length chikungunya *in vitro* transcribed RNA and primed by the +RNA primer. The reaction volume and thermal cycle conditions for the entire procedure followed the protocol described in section 2.4.2.3.2. After RT reaction, each first-stranded cDNA was submitted to three different semi-nested PCR amplification mixes. The first mix contained the primer PC2S2a; the second one, primer PC2S3a and the third one, primer RV1678F, which is specific for chikungunya virus and was used as a control for specific amplification. In addition, all mixes from the first amplification step contained the reverse primer +RNA; and the second amplification step contained the reverse primer +RNA *Not*I. The amplified products from the three different semi-nested PCR mixes were analysed by 1.5 agarose gel electrophoresis as shown in Figure 38.

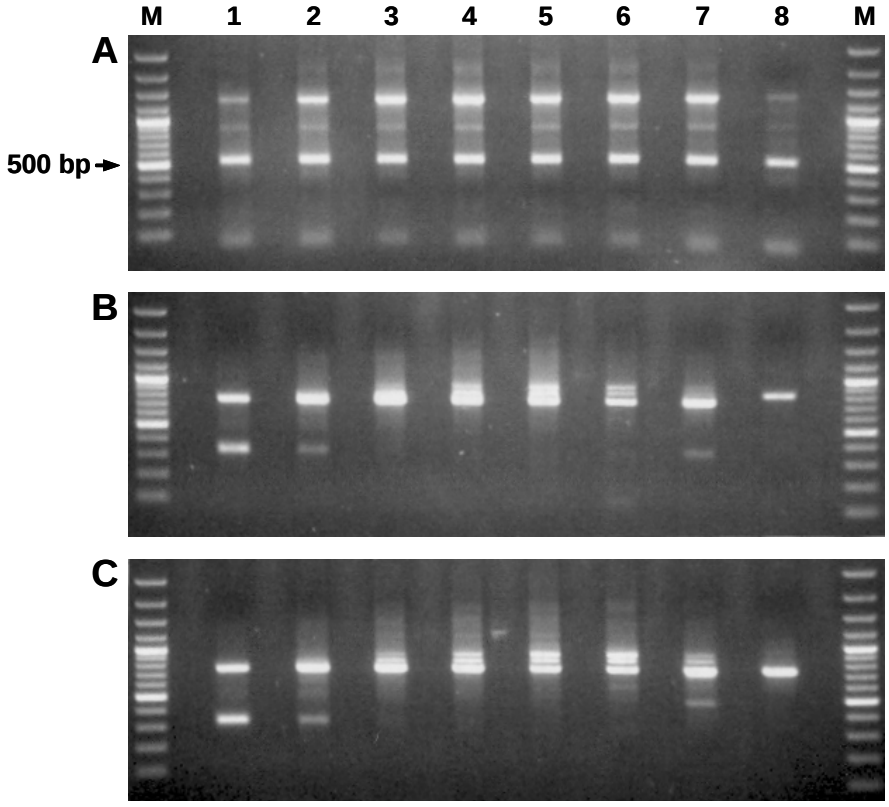

Figure 38: Agarose gel electrophoresis of the second amplification step of the semi-nested RT-PCR using a 10-fold dilution of the full length *in vitro* transcribed chikungunya RNA. A, amplification with forward primer RV1678F. B, amplification with forward primer PC2S2a. C, amplification with forward primer PC2S3a.

As observed in Figure 38, all RNA dilutions yielded fragments when submitted to the three different semi-nested PCR, resembling the same sensitivity as the specific amplification with primer RV1678F. Amplification with primers PC2S2a and PC2S3a yielded nearly the same

pattern of fragments. Furthermore, the same protocol was tested with chikungunya and HPeV1 BNI-788St genomic RNAs extracted from cell culture supernatants. The genomic RNA of both viruses was 10-fold diluted and specimens from 10^0 to 10^{-3} were used for reverse transcription. The respective cDNAs were then submitted to the three semi-nested amplifications described above. The amplicons were analysed by 1.5% agarose gel electrophoresis as show in Figure 39.

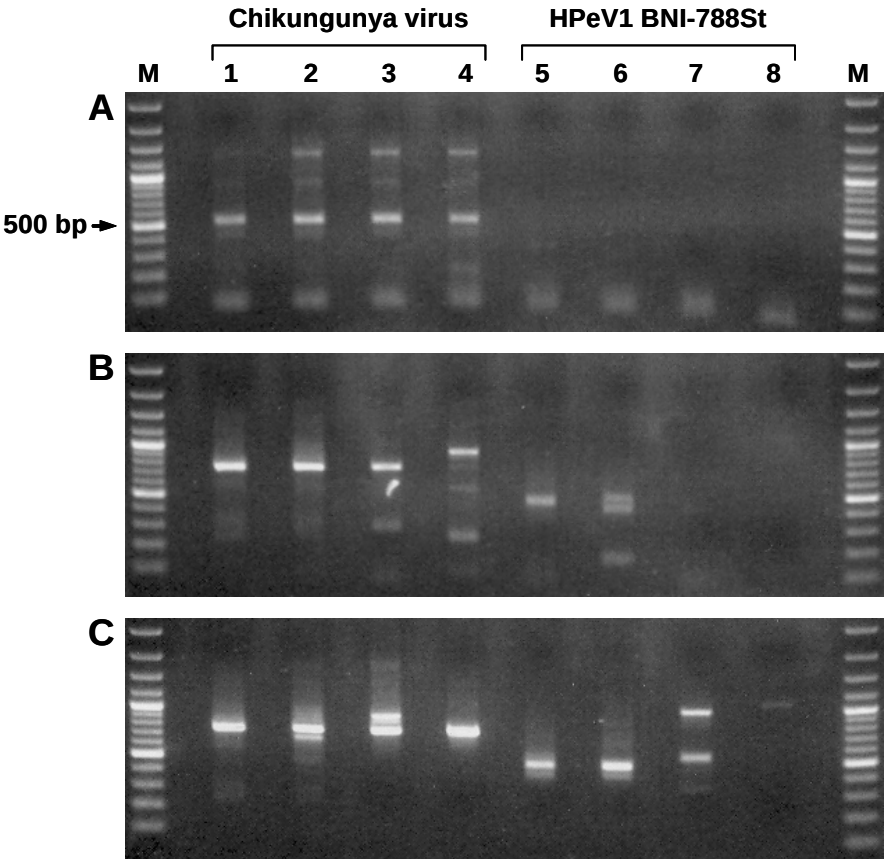

Figure 39: Agarose gel electrophoresis of the second amplification step of the semi-nested RT-PCR of a 10-fold dilution, from 10^0 to 10^{-3}, of chikungunya and HPeV genomic RNA. Lanes 1 to 4, chikungunya genomic RNA. Lanes 5 to 8, HPeV1 BNI-788St genomic RNA. A, amplification with forward primer RV1678F. B, amplification with forward primer PC2S2a. C, amplification with forward primer PC2S3a.

As expected, the semi-nested PCR with primer RV1678F yield products only when chikungunya RNA was used (Figure 39A). Amplification with primers PC2S2a and PC2S3a again showed nearly the same pattern of fragments, with exception of lanes 4 and 7 (Figures 39B and 39C). From the semi-nested RT-PCR with forward primer PC2S3a (Figure 39C), fragments from lanes 4 to 8 were gel purified, cloned into pCR 2.0 TOPO plasmid and sequenced as described in the Material and Methods section. The obtained sequences of each fragment were assembled in a single contig with the SeqMan II software and analysed by BLAST nucleotide search at GenBank website.

As expected, the 747 base pairs fragments of lane 4 corresponded to chikungunya virus. The fragment was then aligned with the nearly complete genome sequence of chikungunya Reunion strain and primer biding sites were determined, as shown in Figure 40.

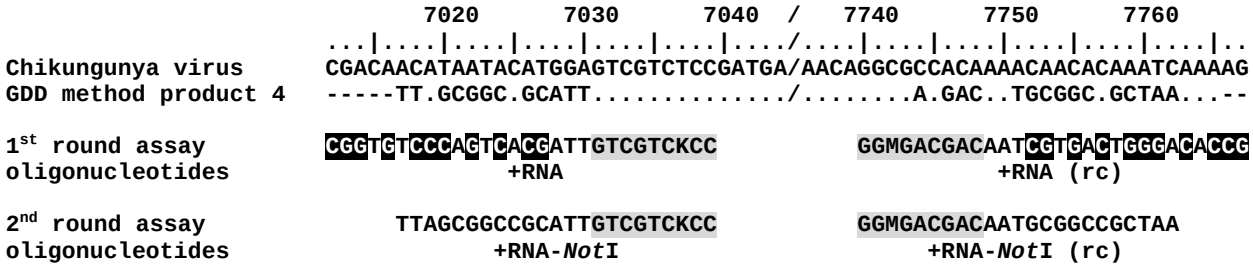

Figure 40: Nucleotide alignment of chikungunya virus Reunion strain and +RNA GDD virus discovery method product 4. Assay oligonucleotides are shown below the alignment panel. Base count in the top line is based on chikungunya virus, which also serves as comparison sequence in the alignment. Dots represent identical bases in compared sequences; deviations are spelled out. A slash (/) represents a gap in the alignment; (rc) means that the reverse complementary sequence is shown for the reverse primer. The GDD motif nucleotide sequences of primers are shown in black, highlighted in grey. IsoG and isoC non-standard nucleotides in the first round assay primers are shown in white and highlighted in black.

Interestingly, amplification was flanked in both directions by the primers +RNA and +RNA-*Not*I during first and second amplification steps, respectively. The amplified genomic region of chikungunya Reunion was also distinct from the previously amplified region when primers Back and Back-*Not*I were used as forward primers. The GDD nucleotide sequence of primer +RNA, have matched entirely with the nucleotide sequence GTCGTCTCC, located in the genomic position 7031 to 7039 of chikungunya virus. This position was 701 nucleotides upstream from the reverse primer biding site.

The results of the BLAST search of fragments 5 to 8 have shown homology with bacterial sequences rather than HPeVs. All cloned fragments could be entirely sequenced, excepted for fragment 8. Fragments 5 and 6 have shown identical nucleotide sequences. When these sequences were analysed by BLAST nucleotide search, they have shown high identity with the 16S ribosomal RNA gene of *M. hyorhinis*. Homology with fragments 7 and 8 could be obtained only by tBLASTx analysis, which searches translated nucleotide database using a translated nucleotide query. The 555 base pairs of fragment 7 have shown homology with bacterial pyruvate kinase. From fragment 8, only two short sequences could be obtained. A 233 base pairs sequence has shown homology with bacterial putative NAD-dependent formate dehydrogenase, whereas the 346 base pairs sequence have shown homology with bacterial glutamate-ammonia-ligase adenylyltransferase. Obviously, those homologies reflected the

amplification of related protein genes from *M. hyorhinis*. Once again, as observed in the experiments with primers Back and Back-*Not*I, nucleic acid of *M. hyorhinis* was preferentially amplified.

For the entirely sequenced fragments, sequences of primer binding sites could be determined. In the case of fragments 5 and 6, their sequences were aligned with those of 16S ribosomal RNA gene of *M. hyorhinis*, which corresponded to the three most significant matches in BLAST search analysis (Figure 41).

Figure 41: Nucleotide alignment of products 5 and 6 from +RNA GDD virus discovery method and three sequences of the 16S ribosomal RNA gene of *M. hyorhinis* (GenBank accession numbers are shown beside scientific name). Assay oligonucleotides are shown below the alignment panel. Base count in the top line is based on *M. hyorhinis* (GenBank accession number AF412982) which also serves as comparison sequence in the alignment. Dots represent identical bases in compared sequences; deviations are spelled out. A slash (/) represents a gap in the alignment; (rc) means that the reverse complementary sequence is shown for the reverse primer. The GDD motif nucleotide sequences of primers are shown in black, highlighted in grey. IsoG and isoC non-standard nucleotides in the first round assay oligonucleotides are shown in white and highlighted in black.

As observed in Figure 40, the amplicons were flanked by primers +RNA and +RNA-*Not*I in the first and second amplification steps, respectively. The +RNA primer working as a reverse primer have their GDD nucleotide sequences hybridized with a homologue genomic region. In contrast, different from the amplification of chikungunya RNA (Figure 35), the GDD nucleotide sequences at 3' end of primer +RNA, as a forward primer, hybridized non-specifically with a GC-rich location of *M. hyorhinis* 16S ribosomal RNA gene.

For fragment 7, the same amplification pattern with primers +RNA and +RNA-*Not*I was observed (Figure 42), but homologue nucleotide sequences could not be retrieved from GenBank to perform nucleotide alignment. Thus, the hybridization pattern of primer +RNA could not be analysed with certainty. In the case of both sequences obtained from fragment 8, the quality of 5' and 3' end sequences was poor, and primer biding sites could not be determined.

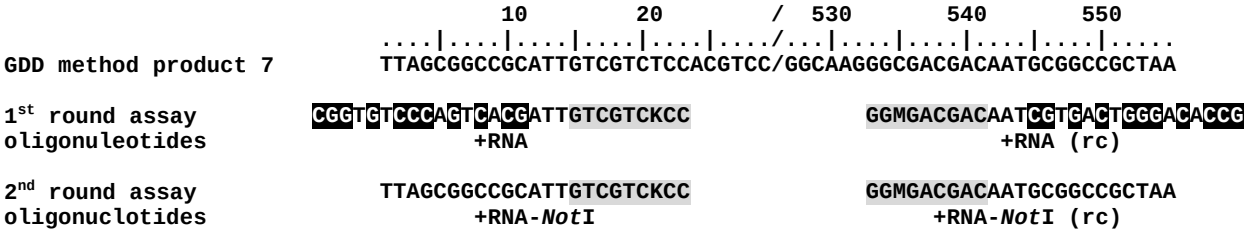

Figure 42: Nucleotide sequence of the +RNA GDD virus discovery method product 7. Assay oligonucleotides are shown below the sequence. A slash (/) represents a gap in the alignment; (rc) means that the reverse complementary sequence is shown for the reverse primer. The GDD motif nucleotide sequences of primers are shown in black, highlighted in grey. IsoG and isoC non-standard nucleotides in the first round assay primers are shown in white and highlighted in black.

4.0 Discussion

4.1 Respiratory infections in patients after intercontinental air travel during SARS epidemic

Acute respiratory infections are frequently experienced after air travel. However, baseline data on the prevalence of respiratory viruses and atypical bacteria after air travel are not currently available. Because most patients do not see a doctor, current knowledge about incidence and aetiology is imprecise and anecdotal (Leder & Newman, 2005). Even though antibiotic therapy against common bacterial respiratory disease is often administered, laboratory data on the spectrum of causative agents are actually not available (File, 2003; Leder & Newman, 2005; Mangili & Gendreau, 2005). The very few systematic studies available suggested that up to 20% of passengers may develop respiratory infections within 1 week after air travel and that flight attendants show significantly higher incidence rates of respiratory infections than do control groups (Whelan *et al.*, 2003; Zitter *et al.*, 2002). These patients constitute a specific sub cohort of patients with community-acquired respiratory disease. Unlike general cohorts, children and elderly persons are underrepresented, and adults of working age constitute the majority of patients. In our study, 71% of the flight patients were between 19 and 60 years old.

Normally, respiratory infections in adults are mild and infrequent, but this may change in the context of travel. There is evidence that air travel increases the incidence of respiratory disease in general and that significant transmission in modern aircraft has occurred of influenza, SARS, and other agents of respiratory infection (Klontz *et al.*, 1989; Mangili & Gendreau, 2005; Marsden, 2003; Moser *et al.*, 1979; Olsen *et al.*, 2003; WHO, 2003a; Yu *et al.*, 2004; Zitter *et al.*, 2002). There is also evidence about the aircrafts acting as vector for spread of influenza worldwide (Laurel *et al.*, 2001; Perz *et al.*, 2001; Sato *et al.*, 2000).

Because morbidity during working life is economically relevant and new therapeutic options are at hand, investigation into the aetiology of travel-associated respiratory disease seems to be well justified. The epidemic of SARS in 2003 involved a period of heightened awareness of respiratory infections after air travel and mobilized the WHO to apply travelling restrictions to certain parts of the world, in addition to establish the SARS case definition of suspected or probable SARS. The WHO SARS case definition required a combination of fever and lower-respiratory-tract symptoms such as cough or difficulty breathing, plus either a stay in an affected area during the preceding 10 days or close contact with suspected patients. The criteria of the WHO SARS case definition, including fever, possibly caused an under representation of milder

infections in our cohort, in the case of RSV and hMPV viruses. However, this made our patients representative of those who are likely to be absent from work.

We used up-to-date diagnostic assays to determine the spectrum of viral or atypical agents in these patients. The rate of resolved aetiologies (43.8%) was markedly higher than that in studies of community-acquired respiratory infections, including recent collective analyses and some of the latest original studies (File, 2003; Kaye *et al.*, 2006; Lee *et al.*, 2006; Louie *et al.*, 2005). This may be due to the extended spectrum of highly sensitive assays applied. Our main finding is that the spectrum of agents in returning travellers is broad. Almost one-half of all patients were infected with respiratory viruses. Agents with treatment options, such as influenza or atypical bacteria, were present in only 15.4% of flight patients and 15.7% of all patients, making even up-to-date diagnostics unrewarding. In this context, it is interesting that lower-respiratory-tract samples did not yield increased detection rates. The general assumption that nasopharyngeal washes are more sensitive than swabs in patients with community-acquired respiratory disease could not be confirmed in our study (File, 2003). Good detection rates and a low risk of aerosolization suggest that swabs should be the preferred in this type of patients.

It cannot be told whether the high prevalence and diversity of respiratory viruses seen in our study is specific to patients with recent intercontinental air travel. Because otherwise healthy adults of working age are usually not tested for respiratory viruses, comparison data from similar cohorts are not available. Our contact patients provided a small control group, and the observed prevalence of viruses confirms the findings seen in flight patients. However, it should be noted that all contact patients fulfilled the case definition of SARS, requiring recent contact with a suspected patient (Drosten *et al.*, 2003). In Germany, this means that all of them most likely had contact with a recent intercontinental traveller.

It is unclear why and how patients acquire viral respiratory disease in the context of air travel (Mangili & Gendreau, 2005; Zitter *et al.*, 2002). In our study, it was interesting that significant clusters of patients with the same diagnosis, who could have been on the same flight, did not exist. Similarly, no association was detected between any pathogen and a particular airport. It would thus be likely that viruses are picked up during prior travel rather than being acquired in flight. In any case, data from this unique cohort suggest that the prevalence of known, emerging, and potentially novel respiratory viruses in adults must be carefully studied.

Understanding the spectrum and the etiological contribution of viruses in adult respiratory disease will be of growing importance in the future. With new antiviral drugs at hand, more consultations will occur for respiratory diseases. Targeted therapy will require broad-spectrum pathogen detection with rapid results, most likely by PCR. Given the large range of

assays required for this study and considering current reagent costs, it is obvious that developing more efficient diagnostic assays is as essential as developing drugs themselves. Today, diagnostic technology is far from attaining this aim.

4.2 The universal CoV RT-PCR/LCD assay

CoVs infect many species of animals, including humans, and are important causes of diseases that include upper and lower respiratory infections, gastroenteritis, hepatic and neurological disorders (Weiss & Navas-Martin, 2005).

After the discovery of SARS-CoV, many systematic investigations have been made in the search for new CoVs. This have led to the discovery of novel HCoVs (NL63 and HKU1), as well as a great diversity of animal CoVs. Interestingly, bats are found to harbour the most genetic diversity of CoVs among the novel discovered species. The CoV diversity associated with bats also seems to be older than all other CoV recognized from any other animal species (Vijaykrishna *et al.*, 2007). Consequences of these hypotheses are that bats are the natural hosts for all CoV lineages and that all CoVs recognized in other species were derived from viruses residing in bats. Therefore, continued investigation in this field seems to be justified.

Viral agents are best diagnosed by direct assays and the most sensitive of them is PCR. In order to identify successfully known and unknown CoVs, generic RT-PCR assays with a very broad detection range are required, but few such assays are available and none of them has been previously validated in a diagnostic setting (Moes *et al.*, 2005; Stephensen *et al.*, 1999). In this dissertation, it was described a novel RT-PCR assay for universal detection of viruses across the genus Coronavirus, combined with a LCD DNA microarray for species identification. Primers were designed to hybridize in the most conserved gene among all CoVs, the RdRp. Motifs A and C were selected because they contain amino acid patterns 100% identical in all coronaviruses (Xu *et al.*, 2003). The selected binding regions correspond to patterns LMGWDYPKCD and MMILSDDAV, which include domains essential for metal ion chelation and binding of the primer 3'-end/template complex (Snijder *et al.*, 2003; Xu *et al.*, 2003). It could be shown experimentally that the whole spectrum of CoVs known up to May 2005 are covered by this assay with diagnostic sensitivity. However, a new alignment including all new CoV RdRp sequences available in GenBank at the time of writing this dissertation was performed. It is important of note that the designed primers are sufficiently conserved to hybridize with the conserved motifs A and C of the RdRp from the new characterized CoVs (figure 11). In addition, the universal CoV RT-PCR assay was successfully used to amplify novel CoVs identified in bats

that were collected recently in Germany and the Republic of Ghana (Drosten *et al.*, data not published).

Although the genetic diversity of coronaviruses is extraordinarily high, the universal RT-PCR/LCD assay provided a simple method of species identification, obviating the need for initial sequencing. Once a virus was amplified, a simple plastic DNA microarray could be utilized for further analysis. After a manual and low-cost hybridisation procedure, the array could be read out by the naked eye in order to determine phylogenetic group membership and species identity. Only when unexpected hybridisation patterns are observed, sequencing of samples would have to be undertaken in order to characterize possibly novel coronavirus species.

Array hybridisation reduces sequence analysis times by one day, and it also decreases costs. While the price of DNA sequencing is still above 3 Euros, array hybridisation identifies any known coronavirus species for about 1.25 Euros per reaction. The reliability of differentiation has been proven by testing a large panel of virus species, as well as a sufficient number of original clinical samples. No discordant result was obtained in any sample tested.

Although the universal CoV RT-PCR/LCD assay will mostly be applied in virus identification studies, its analytical properties also qualify it for application in the clinical diagnostic setting. The formulation has been developed on clinical specimens, including an authentic background of human DNA in all optimisation experiments. Because some of the samples have been stored for a long time, the material was retested in parallel with specific real-time RT-PCRs for HCoV-229E, HCoV-OC43, HCoV-HKU1, and HCoV-NL63 (Garbino *et al.*, 2006; Luna *et al.*, 2007). As shown in Table 9, the sensitivity of the universal CoV RT-PCR/LCD assay was comparable to that of individual virus-specific real-time RT-PCRs. All PCR products were analyzed on LCD arrays and sequenced. All typing results were correct at the group and species levels. Its appropriateness on the clinical level has been proven by testing a large panel of virus species as well as a sufficient number of original patient samples. Sensitivity in clinical samples ranged around 100 copies of RNA per throat swab, which is equivalent to the sensitivities of diagnostic assays that are targeted to one specific virus only, including highly optimized commercial kits (Drosten *et al.*, 2004).

There is only one feature in the universal CoV RT-PCR/LCD assay that may limit its applicability in the clinical setting, which is the nested PCR formulation. However, such a protocol had to be chosen to achieve reliable broad-range amplification across the very diverse spectrum of coronaviruses. Like with all nested PCR protocols, carry-over contamination can occur when PCR product is transferred from the first to the second round of amplification. Enhanced anti-contamination measures must be taken, but this should be acceptable for the

benefit of covering all known coronaviruses, including potential novel strains, with a single assay. In routine use, the application of one-step RT-PCR in the first amplification round reduces the processing time to a manageable level in spite of the nested step (Drosten *et al.*, 2002). Samples received until noon time can be prepared and pre-amplified within three hours on the day of reception. If the second round of amplification is run overnight, followed by an agarose gel and a 15-minute array hybridization of positive samples, results can be issued within less than 24 hours following reception of samples.

Due to the generic features of the LCD array technology, it should be applicable to many other fields in clinical microbiology, for example, for detecting whole virus families or for differentiating bacteria and fungi. Two examples, for genotyping *Mycobacterium tuberculosis* drug resistance (Aragon *et al.*, 2006) and for subtyping of human papilloma virus in squamous-cell cancer tissue (Will *et al.*, 2006), have recently been described. In addition, this is the first application of LCD array technology to a whole viral genus. An expansion of this practical and affordable technology can be expected.

4.3 Prevalence and types of HPeVs in two different cohorts from Hamburg, Germany

Gastroenteritis is one of the most common illnesses in humans worldwide and it has a great impact on people, particularly in infants and the elderly, leading to millions of deaths annually worldwide. Many different viruses have been recognized as important causes of gastroenteritis, including Noroviruses, Rotaviruses, Astroviruses, enteric AdVs, EVs, and HPeVs. However, the spectrum of aetiological viruses are most likely not fully appreciated. When testing stool samples on cell culture, virus isolates are sometimes obtained which cannot be typed by current methods.

In the last years, improved molecular diagnostic methods and better clinical surveillance have led to the identification of novel HPeVs associated with a variety of clinical symptoms similar to some EVs, including mostly enteritis with diarrhoea and respiratory disease. In this dissertation, in order to investigate the HPeV types causing gastroenteritis, the first broad-range real-time RT-PCR assay for HPeVs was developed, the detection frequency in two different cohorts of patients with acute enteritis was determined, and a novel HPeV type was identified and characterized independently.

The assay presented in this dissertation facilitates broad-range testing with high sensitivity and with all advantages of real-time RT-PCR, compatible with modern clinical laboratory settings. Broad applicability is demonstrated by the fact that HPeV6 was readily identified in a stool sample even though this type was not known when the assay was designed.

Good concordance with all known HPeV types is demonstrated by nucleic acid alignment of the oligonucleotide target regions. Of note, mismatches observed in the alignment are almost exclusively G:T base pairs, which are thermodynamically nearly as stable as regular Watson-Crick base pairs (Peyret *et al.*, 1999). The concept of this primer design strategy has recently been proved on another RNA virus (Wolfel *et al.*, 2007). Importantly, most HPeVs in this study were detectable by RT-PCR only. Virus could be cultured only from two samples that had extraordinarily high virus concentrations. It can thus be emphasized that specific, broad-range RT-PCR assays for HPeVs should be applied in the clinical laboratory to ensure complete detection (Abed & Boivin, 2005; Benschop *et al.*, 2006b).

Due to the fact that agents of enteritis especially in outpatients are often not tested for, the HPeVs RT-PCR assay was applied in this type of patient. In concordance with other studies, HPeVs were mostly restricted to young children under 2 years of age (Abed & Boivin, 2005; Benschop *et al.*, 2006b; Tauriainen *et al.*, 2007). Interestingly, HPeVs seemed to be detectable at higher frequencies in outpatient children seen by practitioners than in hospitalized children and those involved in outbreaks of enteritis ($P < 0.05$, comparing detection rates in children in the first and second cohorts tested). Moreover, secondary cases in contact persons were only rarely observed. It could be therefore speculated that HPeVs may not be the primarily cause of epidemics of viral gastroenteritis. They would rather be observed on a sporadic, endemic basis in children who normally remain undiagnosed. Associations with travel or animal exposure could not be identified in this study.

The detection frequency of 11.6% in outpatient children under 2 years of age should lead to a consideration of HPeV in the differential diagnosis of enteritis. The true population prevalence of HPeVs is probably somewhat lower because patients with other common enteric infections were eliminated from the study group. However, such elimination minimizes the rate of patients with double infections and emphasizes those in whom symptoms are caused exclusively by HPeV. On the basis of the data showed in this study, it would be advisable to include RT-PCR for HPeVs in routine viral enteritis testing, at least as a second-line screening assay.

On the other hand, it must not be ignored that an HPeV was also found in a healthy control child. Importantly, the child showed one of the highest viral loads of all patients. As this study is the first to determine virus concentrations in HPeV patients, it is difficult to derive conclusions regarding the relevance of HPeVs from this finding. It should be noted that there were no clinical records about the immediate course of this patient. It cannot be excluded that this patient may have developed clinical signs shortly after the sample was taken. Future studies

should put emphasis on the inclusion of proper control groups and possibly determine viral loads in larger numbers of patients, correlating these with clinical courses. Some recent targeted virus discovery approaches yielded novel viruses which disease association is still unclear (Allander *et al.*, 2007; Allander *et al.*, 2005; Gaynor *et al.*, 2007). This underlies the importance of clinical studies with appropriate control groups to be conducted in a timely manner along with the identification of a potential agent. As with other diseases, the availability of viral load data from real-time RT-PCR will add an additional dimension to the appreciation of disease association.

The independent identification of a sixth parechovirus type, the characterization of another recombinant parechovirus, as well as the confirmation of a predicted diverging group (Al-Sunaidi *et al.*, 2007) between the historical and contemporary type 1 HPeVs underlines the lack of information currently available on the ecology of these agents. Recombination seems to be an important issue in HPeVs, and the current methods of HPeV typing, using VP1 sequence typing or serotyping, are probably not sufficient to describe the true diversity and composition of HPeV populations present in humans (Al-Sunaidi *et al.*, 2007; Benschop *et al.*, 2006a). More-detailed genotyping is probably necessary to understand transmission patterns and strain-specific disease associations.

4.4 Characterization of the HPeV1 BNI-788St

The technical search directed toward the discovery of novel viral agents has become a focus in virology, triggered by the identification of important new agents such as HHV-8 (Chang *et al.*, 1994), hMPV (van den Hoogen *et al.*, 2001), and SARS-CoV (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Peiris *et al.*, 2003; Rota *et al.*, 2003). More recently identified agents include the HCoV-NL63 (van der Hoek *et al.*, 2004), HCoV-HKU1 (Woo *et al.*, 2005), HBoV (Allander *et al.*, 2005), polyomaviruses WU (Gaynor *et al.*, 2007) and KI (Allander *et al.*, 2007), HPV102 and 106 (Chen *et al.*, 2007), as well as a new Arenavirus (Palacios *et al.*, 2008). Different technical approaches have been described to find novel viruses, including SISPA, RDA, random PCR, degenerate PCR, microarray-based detection and ultra-high-throughput sequencing. These methods are sometimes time-consuming or too sophisticated and costly for routine application.

In this dissertation, an alternative approach to identifying unknown viruses called VIDISCA was used. This method combines protected nuclease digestion and AFLP-PCR and has shown to be efficient at least to identify viruses growing in cell culture (van der Hoek *et al.*, 2004). By applying VIDISCA independently, it was proved that the method could be reproduced easily from the literature. The procedure employed a combination of ready-to-use molecular

biology reagents that could be used without technical difficulties. The entire procedure including virus particle enrichment, nuclease digestion, nucleic acids preparation, double-stranded cDNA synthesis, restriction digestion, adapter ligation and two stages of PCR amplification took two full working days to be completed.

Using VIDISCA, an isolated virus that could not be identified by serologic and molecular diagnostic methods could be successfully identified, which turned out to be an HPeV type 1. This new HPeV1 strain was designated as BNI-788St and its genome was complete determined. Evidences based on genome analysis suggested that HPeV1 BNI-788St was the result of a recombination event between HPeV3 non-structural protein genes and HPeV1 structural protein genes. These findings are very interesting because the first strains of HPeVs found, the HPeV1 Harris and HPeV2 Williamson, formerly classified as echovirus types 22 and 23 (Wigand & Sabin, 1961), were described in the 1960s. Recent intensified molecular surveillance has revealed new HPeVs strains such as HPeV3 (Benschop *et al.*, 2006b; Boivin *et al.*, 2005; Ito *et al.*, 2004), HPeV4 (Al-Sunaidi *et al.*, 2007; Benschop *et al.*, 2006a), HPeV5 (Al-Sunaidi *et al.*, 2007), and HPeV6 (Baumgarte *et al.*, 2008; Watanabe *et al.*, 2007) in very short chronological sequence. However, no further studies of full genomes of currently circulating isolates of HPeV1 have been conducted. A finding of recombination, in principle, seems to be not surprising given the propensity of picornaviruses, including parechoviruses, to recombine (Al-Sunaidi *et al.*, 2007; Lukashev *et al.*, 2005; Oberste *et al.*, 2004; Santti *et al.*, 1999; Simmonds, 2006; Simmonds & Welch, 2006).

Following an accepted hypothesis that recombination frequencies is higher between more related genotypes, genomes of enteroviruses form a pool of circulating non-structural gene elements that maintain themselves in the host population by recombination with structural protein genes from other members of their genus (Lukashev *et al.*, 2005; Oberste *et al.*, 2004). The contemporary strain BNI-788St was probably derived from a recombinant ancestor, with a breakpoint at the junction between structural and non-structural genes. Most parts of the structural genes were similar to HPeV1, while the non-structural genes were more similar to that of HPeV3. The non-structural protein genes of BNI-788St were most similar to those of HPeV3, and it is interesting that only BNI-788St and both HPeV3 prototype strains did not show co-segregation of their non-structural genes with that of other prototype strains in BootScan analysis.

Considering the hypothesis mentioned above, it would be conceivable that HPeV3 non-structural protein genes could form elements that are more inert within the pool of HPeVs that may not easily recombine with non-structural genes of other HPeVs. Together with the analysis

of phylogeny and recombination patterns, this particular feature makes it possible to reconstruct likely events in the formation of BNI-788St. Phylogenetic analysis of the complete non-structural gene portion placed BNI-788St and both HPeV3 strains behind a common ancestor with 88% bootstrap support. This common ancestor would have accepted a complete set of structural protein genes by recombination in the 5' proximal part of the non-structural protein genes, close to the VP1/2A junction.

Due to the fact that the VP1 portion of BNI-788St and its group of relatives is directly derived from the common ancestor of VP1 proteins of all contemporary strains, this recombination would have been a basal, non-recent event. It should nevertheless be mentioned that BNI-R30 seems to have taken its VP1 protein from an even older ancestor that is not preserved in other contemporary type 1 strains and has also been lost in BNI-R30 in the other structural protein portions. As a more recent event in the formation of BNI-788St, the common ancestor of the BNI-788St-related group would have acquired its VP0 region from a contemporary type 1 strain. Such intra-capsid recombination in picornaviruses is an uncommon event (Simmonds, 2006). However, it has been described for several picornaviruses including Foot-and-Mouth Disease Virus, poliomyelitis virus type 1, human enterovirus species B, and hepatitis A virus (Costa-Mattioli *et al.*, 2003; Oberste *et al.*, 2004; Tosh *et al.*, 2002; Yang *et al.*, 2005).

The 5' UTR of BNI-788St seems to have been originated from HPeV4 as suggested from the analysis of its predicted structural properties. Such recombination is frequently observed in other picornaviruses (Simmonds, 2006).

The HPeV1 BNI-788St contemporary strain provides limited but plausible evidence of recombination between type 3 non-structural protein genes and type 1 structural protein genes, which has not been observed before. The surprising absence of mosaic recombination in the non-structural protein genes of BNI-788St and of HPeV3 prototype strains underlines the lack of knowledge on the genesis and ecology of human parechoviruses. More research into HPeVs genome diversity is necessary in order to connect virus ecology with disease patterns in humans.

4.5 Characterization of the HPeV6 BNI-67

The new HPeV6 was successfully identified using the first broad-range HPeV real-time RT-PCR assay. Initial phylogenetic analysis based on the amino acid sequences of P1 gene, the precursor of capsid proteins, have shown a basal relationship with HPeV types 1 and 2. Further analysis of the VP1 amino acid genetic distances between BNI-67 and HPeV prototypes have shown that BNI-67 is more related to HPeV1 Harris. The genetic distance between them was 0.22. This value exceeded the genetic distance between HPeV1 Harris and HPeV2 Williams, which was 0.21. It was thus assumed that BNI-67 constituted a novel genotype. During the writing of the manuscript publication, a group in Japan have described a new HPeV type 6. Further analysis including the new HPeV6 has shown a genetic distance of 0.04 with BNI-67, demonstrating that BNI-67 was an HPeV6. Interestingly, BNI-67 seems to be a recombinant of HPeV types 6, 4, 5 and 1, based on BootScan analysis. HPeV6 BNI-67 has conserved the structural protein and the non-structural proteins 2A and 2B from the related HPeV6. Recombination has occurred from non-structural proteins 2C to the end of 3D with HPeVs 4, 5 and 1 (figure 29). As observed for HPeV1 BNI-788St and HPeV6 BNI-67, recombination within the genus Human parechovirus seems to be frequent.

4.6 The + RNA GDD discovery method

In view of the time demanding, or sophisticated and high-cost methodologies applied in the last years for virus discovery, it was suggested in this dissertation a feasible alternative method for nucleic acid amplification of unknown viruses. The first focus on this method was to design primers for recognition of conserved motifs found in viruses with single-stranded, positive RNA genomes. The target of choice was the GDD motif found in almost all RdRp polymerases. In the context of the RdRp amino acid sequences available in the Conserve Domain Database (CDD) at NCBI homepage, the GDD motif seems to be optimal.

The innovative approach of this method is the introduction of the unnatural nucleotides isoC and isoG in the 5’ end of the primers used for amplification. It has been demonstrated that isoC and isoG nucleotides could be recognized by the major enzymes used in molecular diagnostics (polymerases, nucleases, and ligases) and their specific catalytic reactions could be carried out in the presence of these unnatural nucleotides (Moser & Prudent, 2003).

The reason for using such nucleotides in primer design is due to the unique hydrogen-binding pattern of isoC and isoG, which is not found in natural nucleotides. The orientation of donors and acceptors in the hydrogen bond interactions between isoC and isoG are directionally altered in comparison to their isomeric forms cytosine and guanosine, respectively. In

consequence, isoC and isoG should not hybridize with natural nucleotides. Consequently, It allows the 3' end of modified primers to hybridize more efficiently to their specific targets without the interference of the remaining primer nucleotides at the 5' end, composed mainly of isoC/G-rich sequences, even though under low stringency PCR conditions.

Sample preparation for GDD method followed the same initial procedures described for VIDISCA, such as virus particle enrichment, nuclease digestion and nucleic acid preparation. Further procedures include reverse transcriptase and nested PCR.

The reverse primer (+RNA) was designed to specifically hybridize with the conserved GDD motif sequences, whereas the forward primer (Back) was designed to hybridize elsewhere upstream from the reverse primer.

The first approach to the nested PCR was to use +RNA and Back primers in the first round amplification, and primers Uni1 and Uni2 in the second round. Primers Uni1 and Uni2 have the sequences corresponding to the 5' end of +RNA and Back, respectively, which are constituted almost entirely by unnatural nucleotides. In the first experiments, this approach yielded only faint fragments and high concentration of primer-dimers. This is probably due to the short length and high percentage of unnatural nucleotides in the primers Uni1 and Uni2. To overcome these problems, alternative standard primers (+RNA-*Not*I and Back-*Not*I) were designed to hybridize to the portion constituted by natural nucleotides of the first round primers. In addition, sequences corresponding to the binding site of restriction enzyme *Not*I were introduced. With these modifications, fragments from the second round amplification are able to be reamplified and , if necessary, be cloned into TOPO-TA-based plasmids or by digestion with the restriction enzyme *Not*I.

The choice for chikungunya virus as a model organism was the availability, in our lab, of a full-length *in vitro* transcript RNA and a cell cultured strain, in addition to the fact that chikungunya virus has a single-stranded RNA genome of positive polarity. First experiments demonstrated that GDD method could amplify chikungunya RNA both from *in vitro* transcript and extracted from cell culture supernatants. Interestingly, the primer biding site of +RNA primer in the chikungunya genome was not the GDD motif sequences as expected. However, sequence analysis demonstrated that the sequences recognized by +RNA primer in chikungunya genome presented high homology with GDD sequences. One hypothesis for this result is that alternative hybridization of +RNA primer within homologue sequences during reverse transcription should occur and amplification pattern should be determined by hybridization biding site of the forward primer. This could be demonstrated with the use of a forward primer

specific for chikungunya and targeted to a position upstream from the initial biding site of reverse primer +RNA.

The amplification capacity of GDD method was tested in further experiments with HPeV1 BNI-788St cell cultured supernatants. The obtained products, however, have demonstrated high homology with sequences of 16S ribosomal RNA gene of *Mycoplasma hyorhinis*, a common contaminant of cell culture. The selective amplification of bacterial DNA could be explained by the fact that such organisms are usually find to have genes with high CG content in their genomes (Ikehara *et al.*, 1996). This feature could be favoured the forward primer Back to hybridize more efficiently with bacterial DNA.

A second alternative for PCR amplification was to substitute both the modified and standard forward primers by others with reduced GC content. It was then used primers previously designed to recognize the motif A of RdRp in CoVs. After new reaction using the same extracted RNA from HPeV1 BNI-788St, common fragments about 750 base pairs could be obtained from a 10-fold serial dilution. Some selected PCR products were cloned and the sequence could be determined, as well as the primer biding sites. For the second time, selective amplification of bacterial sequences has taken place. More interesting, none of the forward primers was detected flanking the PCR products. This time, +RNA and +RNA-*Not*I have been hybridized as forward and reverse primers in the first and second round amplifications, respectively. It seems that somehow the modified primers have advantages over standard primers when PCR reactions are performed in low stringency conditions.

In the view of the above mentioned results, it seems obvious that cell cultured viruses have to be controlled and tested for common contaminating agents, such as *Mycoplasma* species, to avoid interference in subsequent analysis. However, we have demonstrated, at least for chikungunya virus, that GDD method has the potential to be introduced as a new viral discovery methodology. It is also clear that further optimization experiments should be performed in order to improve the reliability of GDD method.

5.0 References

Anonymous (2003). Severe acute respiratory syndrome (SARS). *Wkly Epidemiol Rec* **78**(21), 181-183.

Abed, Y. & Boivin, G. (2005). Molecular characterization of a Canadian human parechovirus (HPeV)-3 isolate and its relationship to other HPeVs. *J Med Virol* **77**(4), 566-570.

Agirre, A., Barco, A., Carrasco, L. & Nieva, J. L. (2002). Viroporin-mediated membrane permeabilization. Pore formation by nonstructural poliovirus 2B protein. *J Biol Chem* **277**(43), 40434-40441.

Al-Sunaidi, M., Williams, C. H., Hughes, P. J., Schnurr, D. P. & Stanway, G. (2007). Analysis of a new human parechovirus allows the definition of parechovirus types and the identification of RNA structural domains. *J Virol* **81**(2), 1013-1021.

Albuquerque, M. C., Rocha, L. N., Benati, F. J., Soares, C. C., Maranhao, A. G., Ramirez, M. L., Erdman, D. & Santos, N. (2007). Human bocavirus infection in children with gastroenteritis, Brazil. *Emerg Infect Dis* **13**(11), 1756-1758.

Aldabe, R., Barco, A. & Carrasco, L. (1996). Membrane permeabilization by poliovirus proteins 2B and 2BC. *J Biol Chem* **271**(38), 23134-23137.

Aldabe, R. & Carrasco, L. (1995). Induction of membrane proliferation by poliovirus proteins 2C and 2BC. *Biochem Biophys Res Commun* **206**(1), 64-76.

Ali, I. K., McKendrick, L., Morley, S. J. & Jackson, R. J. (2001). Truncated initiation factor eIF4G lacking an eIF4E binding site can support capped mRNA translation. *EMBO J* **20**(15), 4233-4242.

Allander, T., Andreasson, K., Gupta, S., Bjerkner, A., Bogdanovic, G., Persson, M. A., Dalianis, T., Ramqvist, T. & Andersson, B. (2007). Identification of a third human polyomavirus. *J Virol* **81**(8), 4130-4136.

Allander, T., Emerson, S. U., Engle, R. E., Purcell, R. H. & Bukh, J. (2001). A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A* **98**(20), 11609-11614.

Allander, T., Tammi, M. T., Eriksson, M., Bjerkner, A., Tiveljung-Lindell, A. & Andersson, B. (2005). Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci U S A* **102**(36), 12891-12896.

Almazan, F., Gonzalez, J. M., Penzes, Z., Izeta, A., Calvo, E., Plana-Duran, J. & Enjuanes, L. (2000). Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. *Proc Natl Acad Sci U S A* **97**(10), 5516-5521.

Alto, W. A. (2004). Human metapneumovirus: a newly described respiratory tract pathogen. *J Am Board Fam Pract* **17**(6), 466-469.

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**(17), 3389-3402.

Anderson, E. J. & Weber, S. G. (2004). Rotavirus infection in adults. *Lancet Infect Dis* **4**(2), 91-99.

Andino, R., Rieckhof, G. E., Achacoso, P. L. & Baltimore, D. (1993). Poliovirus RNA synthesis utilizes an RNP complex formed around the 5'-end of viral RNA. *EMBO J* **12**(9), 3587-3598.

Andino, R., Rieckhof, G. E. & Baltimore, D. (1990). A functional ribonucleoprotein complex forms around the 5' end of poliovirus RNA. *Cell* **63**(2), 369-380.

Aragon, L. M., Navarro, F., Heiser, V., Garrigo, M., Espanol, M. & Coll, P. (2006). Rapid detection of specific gene mutations associated with isoniazid or rifampicin resistance in Mycobacterium tuberculosis clinical isolates using non-fluorescent low-density DNA microarrays. *J Antimicrob Chemother* **57**(5), 825-831.

Bachem, C. W., van der Hoeven, R. S., de Bruijn, S. M., Vreugdenhil, D., Zabeau, M. & Visser, R. G. (1996). Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development. *Plant J* **9**(5), 745-753.

Banner, L. R., Keck, J. G. & Lai, M. M. (1990). A clustering of RNA recombination sites adjacent to a hypervariable region of the peplomer gene of murine coronavirus. *Virology* **175**(2), 548-555.

Barco, A. & Carrasco, L. (1998). Identification of regions of poliovirus 2BC protein that are involved in cytotoxicity. *J Virol* **72**(5), 3560-3570.

Baric, R. S., Fu, K., Schaad, M. C. & Stohlman, S. A. (1990). Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. *Virology* **177**(2), 646-656.

Barnes, W. M. (1992). The fidelity of Taq polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* **112**(1), 29-35.

Basavappa, R., Syed, R., Flore, O., Icenogle, J. P., Filman, D. J. & Hogle, J. M. (1994). Role and mechanism of the maturation cleavage of VP0 in poliovirus assembly: structure of the empty capsid assembly intermediate at 2.9 A resolution. *Protein Sci* **3**(10), 1651-1669.

Baumgarte, S., de Souza Luna, L. K., Grywna, K., Panning, M., Drexler, J. F., Karsten, C., Huppertz, H. I. & Drosten, C. (2008). Prevalence, types, and RNA concentrations of human parechoviruses, including a sixth parechovirus type, in stool samples from patients with acute enteritis. *J Clin Microbiol* **46**(1), 242-248.

Bedard, K. M. & Semler, B. L. (2004). Regulation of picornavirus gene expression. *Microbes Infect* **6**(7), 702-713.

Belsham, G. J. & Sonenberg, N. (1996). RNA-protein interactions in regulation of picornavirus RNA translation. *Microbiol Rev* **60**(3), 499-511.

Benschop, K. S., Schinkel, J., Luken, M. E., van den Broek, P. J., Beersma, M. F., Menelik, N., van Eijk, H. W., Zaaijer, H. L., VandenBroucke-Grauls, C. M., Beld, M. G. & Wolthers, K. C. (2006a). Fourth human parechovirus serotype. *Emerg Infect Dis* **12**(10), 1572-1575.

Benschop, K. S., Schinkel, J., Minnaar, R. P., Pajkrt, D., Spanjerberg, L., Kraakman, H. C., Berkhout, B., Zaaijer, H. L., Beld, M. G. & Wolthers, K. C. (2006b). Human parechovirus infections in Dutch children and the association between serotype and disease severity. *Clin Infect Dis* **42**(2), 204-210.

Bishop, R. F., Davidson, G. P., Holmes, I. H. & Ruck, B. J. (1973). Virus particles in epithelial cells of duodenal mucosa from children with acute non-bacterial gastroenteritis. *Lancet* **2**(7841), 1281-1283.

Boivin, G., Abed, Y. & Boucher, F. D. (2005). Human parechovirus 3 and neonatal infections. *Emerg Infect Dis* **11**(1), 103-105.

Boonyakiat, Y., Hughes, P. J., Ghazi, F. & Stanway, G. (2001). Arginine-glycine-aspartic acid motif is critical for human parechovirus 1 entry. *J Virol* **75**(20), 10000-10004.

Boriskin, Y. S., Rice, P. S., Stabler, R. A., Hinds, J., Al-Ghusein, H., Vass, K. & Butcher, P. D. (2004). DNA microarrays for virus detection in cases of central nervous system infection. *J Clin Microbiol* **42**(12), 5811-5818.

Boursnell, M. E., Brown, T. D., Foulds, I. J., Green, P. F., Tomley, F. M. & Binns, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J Gen Virol* **68 (Pt 1)**, 57-77.

Bredenbeek, P. J., Pachuk, C. J., Noten, A. F., Charite, J., Luytjes, W., Weiss, S. R. & Spaan, W. J. (1990). The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59; a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res* **18**(7), 1825-1832.

Brian, D. A. & Baric, R. S. (2005). Coronavirus genome structure and replication. *Curr Top Microbiol Immunol* **287**, 1-30.

Brian, D. A., Dennis, D. E. & Guy, J. S. (1980). Genome of porcine transmissible gastroenteritis virus. *J Virol* **34**(2), 410-415.

Brownlee, J. W. & Turner, R. B. (2008). New developments in the epidemiology and clinical spectrum of rhinovirus infections. *Curr Opin Pediatr* **20**(1), 67-71.

Cammack, N., Phillips, A., Dunn, G., Patel, V. & Minor, P. D. (1988). Intertypic genomic rearrangements of poliovirus strains in vaccinees. *Virology* **167**(2), 507-514.

Cavanagh, D. (1997). Nidovirales: a new order comprising Coronaviridae and Arteriviridae. *Arch Virol* **142**(3), 629-633.

Cavanagh, D. (2007). Coronavirus avian infectious bronchitis virus. *Vet Res* **38**(2), 281-297.

Cavanagh, D., Mawditt, K., Welchman Dde, B., Britton, P. & Gough, R. E. (2002). Coronaviruses from pheasants (Phasianus colchicus) are genetically closely related to coronaviruses of domestic fowl (infectious bronchitis virus) and turkeys. *Avian Pathol* **31**(1), 81-93.

Chang, H. G., Glass, R. I., Smith, P. F., Cicirello, H. G., Holman, R. C. & Morse, D. L. (2003). Disease burden and risk factors for hospitalizations associated with rotavirus infection among children in New York State, 1989 through 2000. *Pediatr Infect Dis J* **22**(9), 808-814.

Chang, Y., Cesarman, E., Pessin, M. S., Lee, F., Culpepper, J., Knowles, D. M. & Moore, P. S. (1994). Identification of herpesvirus-like DNA sequences in AIDS-associated Kaposi's sarcoma. *Science* **266**(5192), 1865-1869.

Chen, Z., Schiffman, M., Herrero, R. & Burk, R. D. (2007). Identification and characterization of two novel human papillomaviruses (HPVs) by overlapping PCR: HPV102 and HPV106. *J Gen Virol* **88**(Pt 11), 2952-2955.

Cho, M. W., Richards, O. C., Dmitrieva, T. M., Agol, V. & Ehrenfeld, E. (1993). RNA duplex unwinding activity of poliovirus RNA-dependent RNA polymerase 3Dpol. *J Virol* **67**(6), 3010-3018.

Cho, M. W., Teterina, N., Egger, D., Bienz, K. & Ehrenfeld, E. (1994). Membrane rearrangement and vesicle induction by recombinant poliovirus 2C and 2BC in human cells. *Virology* **202**(1), 129-145.

Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W. & Houghton, M. (1989). Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. *Science* **244**(4902), 359-362.

Chouljenko, V. N., Lin, X. Q., Storz, J., Kousoulas, K. G. & Gorbalenya, A. E. (2001). Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. *J Gen Virol* **82**(Pt 12), 2927-2933.

Circella, E., Camarda, A., Martella, V., Bruni, G., Lavazza, A. & Buonavoglia, C. (2007). Coronavirus associated with an enteric syndrome on a quail farm. *Avian Pathol* **36**(3), 251-258.

Claas, E. C., Schilham, M. W., de Brouwer, C. S., Hubacek, P., Echavarria, M., Lankester, A. C., van Tol, M. J. & Kroes, A. C. (2005). Internally controlled real-time PCR monitoring of adenovirus DNA load in serum or plasma of transplant recipients. *J Clin Microbiol* **43**(4), 1738-1744.

Clewley, J. P., Lewis, J. C., Brown, D. W. & Gadsby, E. L. (1998). A novel simian immunodeficiency virus (SIVdrl) pol sequence from the drill monkey, Mandrillus leucophaeus. *J Virol* **72**(12), 10305-10309.

Compton, S. R., Vivas-Gonzalez, B. E. & Macy, J. D. (1999). Reverse transcriptase polymerase chain reaction-based diagnosis and molecular characterization of a new rat coronavirus strain. *Lab Anim Sci* **49**(5), 506-513.

Conejero-Goldberg, C., Wang, E., Yi, C., Goldberg, T. E., Jones-Brando, L., Marincola, F. M., Webster, M. J. & Torrey, E. F. (2005). Infectious pathogen detection arrays: viral detection in cell lines and postmortem brain tissue. *Biotechniques* **39**(5), 741-751.

Cornelissen, L. A., Wierda, C. M., van der Meer, F. J., Herrewegh, A. A., Horzinek, M. C., Egberink, H. F. & de Groot, R. J. (1997). Hemagglutinin-esterase, a novel structural protein of torovirus. *J Virol* **71**(7), 5277-5286.

Costa-Mattioli, M., Ferre, V., Casane, D., Perez-Bercoff, R., Coste-Burel, M., Imbert-Marcille, B. M., Andre, E. C., Bressollette-Bodin, C., Billaudel, S. & Cristina, J. (2003). Evidence of recombination in natural populations of hepatitis A virus. *Virology* **311**(1), 51-59.

Cotmore, S. F. & Tattersall, P. (1984). Characterization and molecular cloning of a human parvovirus genome. *Science* **226**(4679), 1161-1165.

Crawford, N. M. & Baltimore, D. (1983). Genome-linked protein VPg of poliovirus is present as free VPg and VPg-pUpU in poliovirus-infected cells. *Proc Natl Acad Sci U S A* **80**(24), 7452-7455.

Cunliffe, N. A., Dove, W., Gondwe, J. S., Molyneux, M. E. & Hart, C. A. (2002). Detection of enteric adenoviruses in children with acute gastro-enteritis in Blantyre, Malawi. *Ann Trop Paediatr* **22**(3), 267-269.

de Groot, R. J., Luytjes, W., Horzinek, M. C., van der Zeijst, B. A., Spaan, W. J. & Lenstra, J. A. (1987). Evidence for a coiled-coil structure in the spike proteins of coronaviruses. *J Mol Biol* **196**(4), 963-966.

de Haan, C. A., Masters, P. S., Shen, X., Weiss, S. & Rottier, P. J. (2002). The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. *Virology* **296**(1), 177-189.

De Jong, J. C., Wermenbol, A. G., Verweij-Uijterwaal, M. W., Slaterus, K. W., Wertheim-Van Dillen, P., Van Doornum, G. J., Khoo, S. H. & Hierholzer, J. C. (1999). Adenoviruses from human immunodeficiency virus-infected individuals, including two strains that represent new candidate serotypes Ad50 and Ad51 of species B1 and D, respectively. *J Clin Microbiol* **37**(12), 3940-3945.

de Vries, M., Pyrc, K., Berkhout, R., Vermeulen-Oost, W., Dijkman, R., Jebbink, M. F., Bruisten, S., Berkhout, B. & van der Hoek, L. (2008). Human parechovirus type 1, 3, 4, 5, and 6 detection in picornavirus cultures. *J Clin Microbiol* **46**(2), 759-762.

Deffernez, C., Wunderli, W., Thomas, Y., Yerly, S., Perrin, L. & Kaiser, L. (2004). Amplicon sequencing and improved detection of human rhinovirus in respiratory samples. *J Clin Microbiol* **42**(7), 3212-3218.

Delmas, B. & Laude, H. (1990). Assembly of coronavirus spike protein into trimers and its role in epitope expression. *J Virol* **64**(11), 5367-5375.

Doedens, J. R. & Kirkegaard, K. (1995). Inhibition of cellular protein secretion by poliovirus proteins 2B and 3A. *EMBO J* **14**(5), 894-907.

Domingo, E. & Holland, J. J. (1997). RNA virus mutations and fitness for survival. *Annu Rev Microbiol* **51**, 151-178.

Dominguez, S. R., O'Shea, T. J., Oko, L. M. & Holmes, K. V. (2007). Detection of group 1 coronaviruses in bats in North America. *Emerg Infect Dis* **13**(9), 1295-1300.

Donehower, L. A., Bohannon, R. C., Ford, R. J. & Gibbs, R. A. (1990). The use of primers from highly conserved pol regions to identify uncharacterized retroviruses by the polymerase chain reaction. *J Virol Methods* **28**(1), 33-46.

Dong, B. Q., Liu, W., Fan, X. H., Vijaykrishna, D., Tang, X. C., Gao, F., Li, L. F., Li, G. J., Zhang, J. X., Yang, L. Q., Poon, L. L., Zhang, S. Y., Peiris, J. S., Smith, G. J., Chen, H. & Guan, Y. (2007). Detection of a novel and highly divergent coronavirus from asian leopard cats and Chinese ferret badgers in Southern China. *J Virol* **81**(13), 6920-6926.

Donnelly, M. L., Gani, D., Flint, M., Monaghan, S. & Ryan, M. D. (1997). The cleavage activities of aphthovirus and cardiovirus 2A proteins. *J Gen Virol* **78 (Pt 1)**, 13-21.

Drosten, C., Chiu, L. L., Panning, M., Leong, H. N., Preiser, W., Tam, J. S., Gunther, S., Kramme, S., Emmerich, P., Ng, W. L., Schmitz, H. & Koay, E. S. (2004). Evaluation of advanced reverse transcription-PCR assays and an alternative PCR target region for detection of severe acute respiratory syndrome-associated coronavirus. *J Clin Microbiol* **42**(5), 2043-2047.

Drosten, C., Gottig, S., Schilling, S., Asper, M., Panning, M., Schmitz, H. & Gunther, S. (2002). Rapid detection and quantification of RNA of Ebola and Marburg viruses, Lassa virus, Crimean-Congo hemorrhagic fever virus, Rift Valley fever virus, dengue virus, and yellow fever virus by real-time reverse transcription-PCR. *J Clin Microbiol* **40**(7), 2323-2330.

Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H. R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R. A., Berger, A., Burguiere, A. M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J. C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H. D., Osterhaus, A. D., Schmitz, H. & Doerr, H. W. (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **348**(20), 1967-1976.

Duke, G. M., Hoffman, M. A. & Palmenberg, A. C. (1992). Sequence and structural elements that contribute to efficient encephalomyocarditis virus RNA translation. *J Virol* **66**(3), 1602-1609.

Dye, C. & Siddell, S. G. (2005). Genomic RNA sequence of Feline coronavirus strain FIPV WSU-79/1146. *J Gen Virol* **86**(Pt 8), 2249-2253.

East, M. L., Moestl, K., Benetka, V., Pitra, C., Honer, O. P., Wachter, B. & Hofer, H. (2004). Coronavirus infection of spotted hyenas in the Serengeti ecosystem. *Vet Microbiol* **102**(1-2), 1-9.

Echeverri, A., Banerjee, R. & Dasgupta, A. (1998). Amino-terminal region of poliovirus 2C protein is sufficient for membrane binding. *Virus Res* **54**(2), 217-223.

Ehrnst, A. & Eriksson, M. (1993). Epidemiological features of type 22 echovirus infection. *Scand J Infect Dis* **25**(3), 275-281.

Ehrnst, A. & Eriksson, M. (1996). Echovirus type 23 observed as a nosocomial infection in infants. *Scand J Infect Dis* **28**(2), 205-206.

Enjuanes, L., Sanchez, C., Mendez, A. & Ballesteros, M. L. (1995). Tropism and immunoprotection in transmissible gastroenteritis coronaviruses. *Dev Biol Stand* **84**, 145-152.

Escutenaire, S., Isaksson, M., Renstrom, L. H., Klingeborn, B., Buonavoglia, C., Berg, M., Belak, S. & Thoren, P. (2007). Characterization of divergent and atypical canine coronaviruses from Sweden. *Arch Virol* **152**(8), 1507-1514.

Evans, D. J. & Almond, J. W. (1998). Cell receptors for picornaviruses as determinants of cell tropism and pathogenesis. *Trends Microbiol* **6**(5), 198-202.

Figueroa, J. P., Ashley, D., King, D. & Hull, B. (1989). An outbreak of acute flaccid paralysis in Jamaica associated with echovirus type 22. *J Med Virol* **29**(4), 315-319.

File, T. M. (2003). Community-acquired pneumonia. *Lancet* **362**(9400), 1991-2001.

Fischer, F., Peng, D., Hingley, S. T., Weiss, S. R. & Masters, P. S. (1997). The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. *J Virol* **71**(2), 996-1003.

Fouchier, R. A., Hartwig, N. G., Bestebroer, T. M., Niemeyer, B., de Jong, J. C., Simon, J. H. & Osterhaus, A. D. (2004). A previously undescribed coronavirus associated with respiratory disease in humans. *Proc Natl Acad Sci U S A* **101**(16), 6212-6216.

Foulongne, V., Olejnik, Y., Perez, V., Elaerts, S., Rodiere, M. & Segondy, M. (2006). Human bocavirus in French children. *Emerg Infect Dis* **12**(8), 1251-1253.

Fox, G., Parry, N. R., Barnett, P. V., McGinn, B., Rowlands, D. J. & Brown, F. (1989). The cell attachment site on foot-and-mouth disease virus includes the amino acid sequence RGD (arginine-glycine-aspartic acid). *J Gen Virol* **70 (Pt 3)**, 625-637.

Gallagher, T. M., Escarmis, C. & Buchmeier, M. J. (1991). Alteration of the pH dependence of coronavirus-induced cell fusion: effect of mutations in the spike glycoprotein. *J Virol* **65**(4), 1916-1928.

Garbino, J., Crespo, S., Aubert, J. D., Rochat, T., Ninet, B., Deffernez, C., Wunderli, W., Pache, J. C., Soccal, P. M. & Kaiser, L. (2006). A prospective hospital-based study of the clinical impact of non-severe acute respiratory syndrome (Non-SARS)-related human coronavirus infection. *Clin Infect Dis* **43**(8), 1009-1015.

Gaynor, A. M., Nissen, M. D., Whiley, D. M., Mackay, I. M., Lambert, S. B., Wu, G., Brennan, D. C., Storch, G. A., Sloots, T. P. & Wang, D. (2007). Identification of a novel polyomavirus from patients with acute respiratory tract infections. *PLoS Pathog* **3**(5), e64.

Ghazi, F., Hughes, P. J., Hyypia, T. & Stanway, G. (1998). Molecular analysis of human parechovirus type 2 (formerly echovirus 23). *J Gen Virol* **79 (Pt 11)**, 2641-2650.

Gonzalez, J. M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A. E. & Enjuanes, L. (2003). A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch Virol* **148**(11), 2207-2235.

Gough, R. E., Drury, S. E., Culver, F., Britton, P. & Cavanagh, D. (2006). Isolation of a coronavirus from a green-cheeked Amazon parrot (Amazon viridigenalis Cassin). *Avian Pathol* **35**(2), 122-126.

Gramberg, T., Hofmann, H., Moller, P., Lalor, P. F., Marzi, A., Geier, M., Krumbiegel, M., Winkler, T., Kirchhoff, F., Adams, D. H., Becker, S., Munch, J. & Pohlmann, S. (2005). LSECtin interacts with filovirus glycoproteins and the spike protein of SARS coronavirus. *Virology* **340**(2), 224-236.

Greenberg, S. B. (2003). Respiratory consequences of rhinovirus infection. *Arch Intern Med* **163**(3), 278-284.

Griffiths, D. J., Venables, P. J., Weiss, R. A. & Boyd, M. T. (1997). A novel exogenous retrovirus sequence identified in humans. *J Virol* **71**(4), 2866-2872.

Grist, N. R., Bell, E. J. & Assaad, F. (1978). Enteroviruses in human disease. *Prog Med Virol* **24**, 114-157.

Gu, J. & Korteweg, C. (2007). Pathology and pathogenesis of severe acute respiratory syndrome. *Am J Pathol* **170**(4), 1136-1147.

Guan, Y., Zheng, B. J., He, Y. Q., Liu, X. L., Zhuang, Z. X., Cheung, C. L., Luo, S. W., Li, P. H., Zhang, L. J., Guan, Y. J., Butt, K. M., Wong, K. L., Chan, K. W., Lim, W., Shortridge, K. F., Yuen, K. Y., Peiris, J. S. & Poon, L. L. (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**(5643), 276-278.

Hamming, I., Cooper, M. E., Haagmans, B. L., Hooper, N. M., Korstanje, R., Osterhaus, A. D., Timens, W., Turner, A. J., Navis, G. & van Goor, H. (2007). The emerging role of ACE2 in physiology and disease. *J Pathol* **212**(1), 1-11.

Hamming, I., Timens, W., Bulthuis, M. L., Lely, A. T., Navis, G. J. & van Goor, H. (2004). Tissue distribution of ACE2 protein, the functional receptor for SARS coronavirus. A first step in understanding SARS pathogenesis. *J Pathol* **203**(2), 631-637.

Hansen, J. L., Long, A. M. & Schultz, S. C. (1997). Structure of the RNA-dependent RNA polymerase of poliovirus. *Structure* **5**(8), 1109-1122.

Hasoksuz, M., Alekseev, K., Vlasova, A., Zhang, X., Spiro, D., Halpin, R., Wang, S., Ghedin, E. & Saif, L. J. (2007). Biologic, antigenic, and full-length genomic characterization of a bovine-like coronavirus isolated from a giraffe. *J Virol* **81**(10), 4981-4990.

He, R., Dobie, F., Ballantine, M., Leeson, A., Li, Y., Bastien, N., Cutts, T., Andonov, A., Cao, J., Booth, T. F., Plummer, F. A., Tyler, S., Baker, L. & Li, X. (2004). Analysis of multimerization of the SARS coronavirus nucleocapsid protein. *Biochem Biophys Res Commun* **316**(2), 476-483.

Herrewegh, A. A., Smeenk, I., Horzinek, M. C., Rottier, P. J. & de Groot, R. J. (1998). Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. *J Virol* **72**(5), 4508-4514.

Hogue, B. G., Kienzle, T. E. & Brian, D. A. (1989). Synthesis and processing of the bovine enteric coronavirus haemagglutinin protein. *J Gen Virol* **70 (Pt 2)**, 345-352.

Horsburgh, B. C., Brierley, I. & Brown, T. D. (1992). Analysis of a 9.6 kb sequence from the 3' end of canine coronavirus genomic RNA. *J Gen Virol* **73 (Pt 11)**, 2849-2862.

Hynes, R. O. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* **69**(1), 11-25.

Hyypia, T., Horsnell, C., Maaronen, M., Khan, M., Kalkkinen, N., Auvinen, P., Kinnunen, L. & Stanway, G. (1992). A distinct picornavirus group identified by sequence analysis. *Proc Natl Acad Sci U S A* **89**(18), 8847-8851.

Ikehara, K., Amada, F., Yoshida, S., Mikata, Y. & Tanaka, A. (1996). A possible origin of newly-born bacterial genes: significance of GC-rich nonstop frame on antisense strand. *Nucleic Acids Res* **24**(21), 4249-4255.

Ito, M., Yamashita, T., Tsuzuki, H., Takeda, N. & Sakae, K. (2004). Isolation and identification of a novel human parechovirus. *J Gen Virol* **85**(Pt 2), 391-398.

Ito, N., Mossel, E. C., Narayanan, K., Popov, V. L., Huang, C., Inoue, T., Peters, C. J. & Makino, S. (2005). Severe acute respiratory syndrome coronavirus 3a protein is a viral structural protein. *J Virol* **79**(5), 3182-3186.

Ivanov, K. A., Hertzig, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A. E. & Ziebuhr, J. (2004). Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc Natl Acad Sci U S A* **101**(34), 12694-12699.

Jacobson, M. F. & Baltimore, D. (1968). Morphogenesis of poliovirus. I. Association of the viral RNA with coat protein. *J Mol Biol* **33**(2), 369-378.

Jacobson, S. J., Konings, D. A. & Sarnow, P. (1993). Biochemical and genetic evidence for a pseudoknot structure at the 3' terminus of the poliovirus RNA genome and its role in viral RNA amplification. *J Virol* **67**(6), 2961-2971.

Jakab, F., Walter, J. E., Berke, T., Matson, D. O., Mitchell, D. K. & Szucs, G. (2003). Molecular characterization and sequence analysis of human astroviruses circulating in Hungary. *FEMS Immunol Med Microbiol* **39**(2), 97-102.

Jang, S. K. & Wimmer, E. (1990). Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Genes Dev* **4**(9), 1560-1572.

Jarvis, T. C. & Kirkegaard, K. (1992). Poliovirus RNA recombination: mechanistic studies in the absence of selection. *EMBO J* **11**(8), 3135-3145.

Johnson, S. C., Marshall, D. J., Harms, G., Miller, C. M., Sherrill, C. B., Beaty, E. L., Lederer, S. A., Roesch, E. B., Madsen, G., Hoffman, G. L., Laessig, R. H., Kopish, G. J., Baker, M. W., Benner, S. A., Farrell, P. M. & Prudent, J. R. (2004a). Multiplexed genetic analysis using an expanded genetic alphabet. *Clin Chem* **50**(11), 2019-2027.

Johnson, S. C., Sherrill, C. B., Marshall, D. J., Moser, M. J. & Prudent, J. R. (2004b). A third base pair for the polymerase chain reaction: inserting isoC and isoG. *Nucleic Acids Res* **32**(6), 1937-1941.

Joki-Korpela, P. & Hyypia, T. (1998). Diagnosis and epidemiology of echovirus 22 infections. *Clin Infect Dis* **27**(1), 129-136.

Joki-Korpela, P. & Hyypia, T. (2001). Parechoviruses, a novel group of human picornaviruses. *Ann Med* **33**(7), 466-471.

Joki-Korpela, P., Marjomaki, V., Krogerus, C., Heino, J. & Hyypia, T. (2001). Entry of human parechovirus 1. *J Virol* **75**(4), 1958-1967.

Jonassen, C. M., Kofstad, T., Larsen, I. L., Lovland, A., Handeland, K., Follestad, A. & Lillehaug, A. (2005). Molecular identification and characterization of novel coronaviruses infecting graylag geese (Anser anser), feral pigeons (Columbia livia) and mallards (Anas platyrhynchos). *J Gen Virol* **86**(Pt 6), 1597-1607.

Jones, M. S., Kapoor, A., Lukashov, V. V., Simmonds, P., Hecht, F. & Delwart, E. (2005). New DNA viruses identified in patients with acute viral infection syndrome. *J Virol* **79**(13), 8230-8236.

Jurgens, C. & Flanegan, J. B. (2003). Initiation of poliovirus negative-strand RNA synthesis requires precursor forms of p2 proteins. *J Virol* **77**(2), 1075-1083.

Kaye, M., Skidmore, S., Osman, H., Weinbren, M. & Warren, R. (2006). Surveillance of respiratory virus infections in adult hospital admissions using rapid methods. *Epidemiol Infect* **134**(4), 792-798.

Kazi, L., Lissenberg, A., Watson, R., de Groot, R. J. & Weiss, S. R. (2005). Expression of hemagglutinin esterase protein from recombinant mouse hepatitis virus enhances neurovirulence. *J Virol* **79**(24), 15064-15073.

Keck, J. G., Matsushima, G. K., Makino, S., Fleming, J. O., Vannier, D. M., Stohlman, S. A. & Lai, M. M. (1988). In vivo RNA-RNA recombination of coronavirus in mouse brain. *J Virol* **62**(5), 1810-1813.

Kesson, A. M. (2007). Respiratory virus infections. *Paediatr Respir Rev* **8**(3), 240-248.

Khanna, N., Goldenberger, D., Graber, P., Battegay, M. & Widmer, A. F. (2003). Gastroenteritis outbreak with norovirus in a Swiss university hospital with a newly identified virus strain. *J Hosp Infect* **55**(2), 131-136.

Kirkegaard, K. & Baltimore, D. (1986). The mechanism of RNA recombination in poliovirus. *Cell* **47**(3), 433-443.

Kistler, A., Avila, P. C., Rouskin, S., Wang, D., Ward, T., Yagi, S., Schnurr, D., Ganem, D., DeRisi, J. L. & Boushey, H. A. (2007). Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *J Infect Dis* **196**(6), 817-825.

Klausegger, A., Strobl, B., Regl, G., Kaser, A., Luytjes, W. & Vlasak, R. (1999). Identification of a coronavirus hemagglutinin-esterase with a substrate specificity different from those of influenza C virus and bovine coronavirus. *J Virol* **73**(5), 3737-3743.

Klontz, K. C., Hynes, N. A., Gunn, R. A., Wilder, M. H., Harmon, M. W. & Kendal, A. P. (1989). An outbreak of influenza A/Taiwan/1/86 (H1N1) infections at a naval base and its association with airplane travel. *Am J Epidemiol* **129**(2), 341-348.

Klumperman, J., Locker, J. K., Meijer, A., Horzinek, M. C., Geuze, H. J. & Rottier, P. J. (1994). Coronavirus M proteins accumulate in the Golgi complex beyond the site of virion budding. *J Virol* **68**(10), 6523-6534.

Kocherhans, R., Bridgen, A., Ackermann, M. & Tobler, K. (2001). Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* **23**(2), 137-144.

Korimbocus, J., Scaramozzino, N., Lacroix, B., Crance, J. M., Garin, D. & Vernet, G. (2005). DNA probe array for the simultaneous identification of herpesviruses, enteroviruses, and flaviviruses. *J Clin Microbiol* **43**(8), 3779-3787.

Koskiniemi, M., Paetau, R. & Linnavuori, K. (1989). Severe encephalitis associated with disseminated echovirus 22 infection. *Scand J Infect Dis* **21**(4), 463-466.

Kottier, S. A., Cavanagh, D. & Britton, P. (1995). Experimental evidence of recombination in coronavirus infectious bronchitis virus. *Virology* **213**(2), 569-580.

Krogerus, C., Egger, D., Samuilova, O., Hyypia, T. & Bienz, K. (2003). Replication complex of human parechovirus 1. *J Virol* **77**(15), 8512-8523.

Krogerus, C., Samuilova, O., Poyry, T., Jokitalo, E. & Hyypia, T. (2007). Intracellular localization and effects of individually expressed human parechovirus 1 non-structural proteins. *J Gen Virol* **88**(Pt 3), 831-841.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A. E., Humphrey, C. D., Shieh, W. J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., DeRisi, J., Yang, J. Y., Cox, N., Hughes, J. M., LeDuc, J. W., Bellini, W. J. & Anderson, L. J. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348**(20), 1953-1966.

Kusov, Y. Y., Probst, C., Jecht, M., Jost, P. D. & Gauss-Muller, V. (1998). Membrane association and RNA binding of recombinant hepatitis A virus protein 2C. *Arch Virol* **143**(5), 931-944.

Kusters, J. G., Niesters, H. G., Lenstra, J. A., Horzinek, M. C. & van der Zeijst, B. A. (1989). Phylogeny of antigenic variants of avian coronavirus IBV. *Virology* **169**(1), 217-221.

Lai, M. M. (1992). RNA recombination in animal and plant viruses. *Microbiol Rev* **56**(1), 61-79.

Lai, M. M. & Cavanagh, D. (1997). The molecular biology of coronaviruses. *Adv Virus Res* **48**, 1-100.

Lai, M. M. & Stohlman, S. A. (1978). RNA of mouse hepatitis virus. *J Virol* **26**(2), 236-242.

Lai, M. M. & Stohlman, S. A. (1981). Comparative analysis of RNA genomes of mouse hepatitis viruses. *J Virol* **38**(2), 661-670.

Lai, M. M. C., Perlman, S. & J., A. L. (2007). *Coronaviridae*. In *Fields - Virology*, 5th edn, pp. 1305-1335. Edited by K. D. M. & P. M. Howley. Philadelphia: Lippincott Williams & Wilkins.

Lama, J., Sanz, M. A. & Carrasco, L. (1998). Genetic analysis of poliovirus protein 3A: characterization of a non-cytopathic mutant virus defective in killing Vero cells. *J Gen Virol* **79 (Pt 8)**, 1911-1921.

Lambden, P. R., Cooke, S. J., Caul, E. O. & Clarke, I. N. (1992). Cloning of noncultivatable human rotavirus by single primer amplification. *J Virol* **66**(3), 1817-1822.

Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Tsoi, H. W., Wong, B. H., Wong, S. S., Leung, S. Y., Chan, K. H. & Yuen, K. Y. (2005). Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* **102**(39), 14040-14045.

Lau, S. K., Woo, P. C., Li, K. S., Huang, Y., Wang, M., Lam, C. S., Xu, H., Guo, R., Chan, K. H., Zheng, B. J. & Yuen, K. Y. (2007). Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology* **367**(2), 428-439.

Laurel, V. L., De Witt, C. C., Geddie, Y. A., Yip, M. C., Dolan, D. M., Canas, L. C., Dolan, M. J. & Walter, E. A. (2001). An outbreak of influenza a caused by imported virus in the United States, July 1999. *Clin Infect Dis* **32**(11), 1639-1642.

Leder, K. & Newman, D. (2005). Respiratory infections during air travel. *Intern Med J* **35**(1), 50-55.

Lee, B. E., Robinson, J. L., Khurana, V., Pang, X. L., Preiksaitis, J. K. & Fox, J. D. (2006). Enhanced identification of viral and atypical bacterial pathogens in lower respiratory tract samples with nucleic acid amplification tests. *J Med Virol* **78**(5), 702-710.

Legay, V., Chomel, J. J., Fernandez, E., Lina, B., Aymard, M. & Khalfan, S. (2002). Encephalomyelitis due to human parechovirus type 1. *J Clin Virol* **25**(2), 193-195.

Leparc-Goffart, I., Hingley, S. T., Chua, M. M., Jiang, X., Lavi, E. & Weiss, S. R. (1997). Altered pathogenesis of a mutant of the murine coronavirus MHV-A59 is associated with a Q159L amino acid substitution in the spike protein. *Virology* **239**(1), 1-10.

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., Somasundaran, M., Sullivan, J. L., Luzuriaga, K., Greenough, T. C., Choe, H. & Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* **426**(6965), 450-454.

Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J. H., Wang, H., Crameri, G., Hu, Z., Zhang, H., Zhang, J., McEachern, J., Field, H., Daszak, P., Eaton, B. T., Zhang, S. & Wang, L. F. (2005). Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**(5748), 676-679.

Lin, B., Vora, G. J., Thach, D., Walter, E., Metzgar, D., Tibbetts, C. & Stenger, D. A. (2004). Use of oligonucleotide microarrays for rapid detection and serotyping of acute respiratory disease-associated adenoviruses. *J Clin Microbiol* **42**(7), 3232-3239.

Lin, B., Wang, Z., Vora, G. J., Thornton, J. A., Schnur, J. M., Thach, D. C., Blaney, K. M., Ligler, A. G., Malanoski, A. P., Santiago, J., Walter, E. A., Agan, B. K., Metzgar, D., Seto, D., Daum, L. T., Kruzelock, R., Rowley, R. K., Hanson, E. H., Tibbetts, C. & Stenger, D. A. (2006). Broad-spectrum respiratory tract pathogen identification using resequencing DNA microarrays. *Genome Res* **16**(4), 527-535.

Linnen, J., Wages, J., Jr., Zhang-Keck, Z. Y., Fry, K. E., Krawczynski, K. Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, S., Karayiannis, P., Fung, K., Nakatsuji, Y., Shih, J. W., Young, L., Piatak, M., Jr., Hoover, C., Fernandez, J., Chen, S., Zou, J. C., Morris, T., Hyams, K. C., Ismay, S., Lifson, J. D., Hess, G., Foung, S. K., Thomas, H., Bradley, D., Margolis, H. & Kim, J. P. (1996). Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* **271**(5248), 505-508.

Lisitsyn, N. & Wigler, M. (1993). Cloning the differences between two complex genomes. *Science* **259**(5097), 946-951.

Liu, D. X., Yuan, Q. & Liao, Y. (2007a). Coronavirus envelope protein: a small membrane protein with multiple functions. *Cell Mol Life Sci* **64**(16), 2043-2048.

Liu, Q., Bai, Y., Ge, Q., Zhou, S., Wen, T. & Lu, Z. (2007b). Microarray-in-a-tube for detection of multiple viruses. *Clin Chem* **53**(2), 188-194.

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W. & Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**(1), 152-160.

Longtin, J., Bastien, M., Gilca, R., Leblanc, E., de Serres, G., Bergeron, M. G. & Boivin, G. (2008). Human bocavirus infections in hospitalized children and adults. *Emerg Infect Dis* **14**(2), 217-221.

Lopman, B., Vennema, H., Kohli, E., Pothier, P., Sanchez, A., Negredo, A., Buesa, J., Schreier, E., Reacher, M., Brown, D., Gray, J., Iturriza, M., Gallimore, C., Bottiger, B., Hedlund, K. O., Torven, M., von Bonsdorff, C. H., Maunula, L., Poljsak-Prijatelj, M., Zimsek, J., Reuter, G., Szucs, G., Melegh, B., Svennson, L., van Duijnhoven, Y. & Koopmans, M. (2004). Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. *Lancet* **363**(9410), 682-688.

Louie, J. K., Hacker, J. K., Gonzales, R., Mark, J., Maselli, J. H., Yagi, S. & Drew, W. L. (2005). Characterization of viral agents causing acute respiratory infection in a San Francisco University Medical Center Clinic during the influenza season. *Clin Infect Dis* **41**(6), 822-828.

Lukashev, A. N., Lashkevich, V. A., Ivanova, O. E., Koroleva, G. A., Hinkkanen, A. E. & Ilonen, J. (2005). Recombination in circulating Human enterovirus B: independent evolution of structural and non-structural genome regions. *J Gen Virol* **86**(Pt 12), 3281-3290.

Luna, L. K., Panning, M., Grywna, K., Pfefferle, S. & Drosten, C. (2007). Spectrum of viruses and atypical bacteria in intercontinental air travelers with symptoms of acute respiratory infection. *J Infect Dis* **195**(5), 675-679.

Luytjes, W., Bredenbeek, P. J., Noten, A. F., Horzinek, M. C. & Spaan, W. J. (1988). Sequence of mouse hepatitis virus A59 mRNA 2: indications for RNA recombination between coronaviruses and influenza C virus. *Virology* **166**(2), 415-422.

Ma, X., Endo, R., Ishiguro, N., Ebihara, T., Ishiko, H., Ariga, T. & Kikuta, H. (2006). Detection of human bocavirus in Japanese children with lower respiratory tract infections. *J Clin Microbiol* **44**(3), 1132-1134.

Macneughton, M. R. & Davies, H. A. (1978). Ribonucleoprotein-like structures from coronavirus particles. *J Gen Virol* **39**(3), 545-549.

Maller, H. M., Powars, D. F., Horowitz, R. E. & Portnoy, B. (1967). Fatal myocarditis associated with ECHO virus, type 22, infection in a child with apparent immunological deficiency. *J Pediatr* **71**(2), 204-210.

Mallia, P. & Johnston, S. L. (2006). How viral infections cause exacerbation of airway diseases. *Chest* **130**(4), 1203-1210.

Mangili, A. & Gendreau, M. A. (2005). Transmission of infectious diseases during commercial air travel. *Lancet* **365**(9463), 989-996.

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., Jando, S. C., Alenquer, M. L., Jarvie, T. P., Jirage, K. B., Kim, J. B., Knight, J. R., Lanza, J. R., Leamon, J. H., Lefkowitz, S. M., Lei, M., Li, J., Lohman, K. L., Lu, H., Makhijani, V. B., McDade, K. E., McKenna, M. P., Myers, E. W., Nickerson, E., Nobile, J. R., Plant, R., Puc, B. P., Ronan, M. T., Roth, G. T., Sarkis, G. J., Simons, J. F., Simpson, J. W., Srinivasan, M., Tartaro, K. R., Tomasz, A., Vogt, K. A., Volkmer, G. A., Wang, S. H., Wang, Y., Weiner, M. P., Yu, P., Begley, R. F. & Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**(7057), 376-380.

Marsden, A. G. (2003). Outbreak of influenza-like illness [corrected] related to air travel. *Med J Aust* **179**(3), 172-173.

Marshall, D. J., Reisdorf, E., Harms, G., Beaty, E., Moser, M. J., Lee, W. M., Gern, J. E., Nolte, F. S., Shult, P. & Prudent, J. R. (2007). Evaluation of a multiplexed PCR assay for detection of respiratory viral pathogens in a public health laboratory setting. *J Clin Microbiol* **45**(12), 3875-3882.

Mason, P. W., Bezborodova, S. V. & Henry, T. M. (2002). Identification and characterization of a cis-acting replication element (cre) adjacent to the internal ribosome entry site of foot-and-mouth disease virus. *J Virol* **76**(19), 9686-9694.

Masters, P. S. (2006). The molecular biology of coronaviruses. *Adv Virus Res* **66**, 193-292.

Matsui, S. M., Kim, J. P., Greenberg, H. B., Young, L. M., Smith, L. S., Lewis, T. L., Herrmann, J. E., Blacklow, N. R., Dupuis, K. & Reyes, G. R. (1993). Cloning and characterization of human astrovirus immunoreactive epitopes. *J Virol* **67**(3), 1712-1715.

Mendez-Toss, M., Griffin, D. D., Calva, J., Contreras, J. F., Puerto, F. I., Mota, F., Guiscafre, H., Cedillo, R., Munoz, O., Herrera, I., Lopez, S. & Arias, C. F. (2004). Prevalence and genetic diversity of human astroviruses in Mexican children with symptomatic and asymptomatic infections. *J Clin Microbiol* **42**(1), 151-157.

Minskaia, E., Hertzig, T., Gorbalenya, A. E., Campanacci, V., Cambillau, C., Canard, B. & Ziebuhr, J. (2006). Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc Natl Acad Sci U S A* **103**(13), 5108-5113.

Mirmomeni, M. H., Hughes, P. J. & Stanway, G. (1997). An RNA tertiary structure in the 3' untranslated region of enteroviruses is necessary for efficient replication. *J Virol* **71**(3), 2363-2370.

Moes, E., Vijgen, L., Keyaerts, E., Zlateva, K., Li, S., Maes, P., Pyrc, K., Berkhout, B., van der Hoek, L. & Van Ranst, M. (2005). A novel pancoronavirus RT-PCR assay: frequent detection of human coronavirus NL63 in children hospitalized with respiratory tract infections in Belgium. *BMC Infect Dis* **5**(1), 6.

Mohandas, D. V. & Dales, S. (1991). Endosomal association of a protein phosphatase with high dephosphorylating activity against a coronavirus nucleocapsid protein. *FEBS Lett* **282**(2), 419-424.

Morris, D. R. & Geballe, A. P. (2000). Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* **20**(23), 8635-8642.

Moser, M. J., Christensen, D. R., Norwood, D. & Prudent, J. R. (2006). Multiplexed detection of anthrax-related toxin genes. *J Mol Diagn* **8**(1), 89-96.

Moser, M. J. & Prudent, J. R. (2003). Enzymatic repair of an expanded genetic information system. *Nucleic Acids Res* **31**(17), 5048-5053.

Moser, M. J., Ruckstuhl, M., Larsen, C. A., Swearingen, A. J., Kozlowski, M., Bassit, L., Sharma, P. L., Schinazi, R. F. & Prudent, J. R. (2005). Quantifying mixed populations of drug-resistant human immunodeficiency virus type 1. *Antimicrob Agents Chemother* **49**(8), 3334-3340.

Moser, M. R., Bender, T. R., Margolis, H. S., Noble, G. R., Kendal, A. P. & Ritter, D. G. (1979). An outbreak of influenza aboard a commercial airliner. *Am J Epidemiol* **110**(1), 1-6.

Nichol, S. T., Spiropoulou, C. F., Morzunov, S., Rollin, P. E., Ksiazek, T. G., Feldmann, H., Sanchez, A., Childs, J., Zaki, S. & Peters, C. J. (1993). Genetic identification of a hantavirus associated with an outbreak of acute respiratory illness. *Science* **262**(5135), 914-917.

Nijhuis, M., van Maarseveen, N., Schuurman, R., Verkuijlen, S., de Vos, M., Hendriksen, K. & van Loon, A. M. (2002). Rapid and sensitive routine detection of all members of the genus enterovirus in different clinical specimens by real-time PCR. *J Clin Microbiol* **40**(10), 3666-3670.

Niklasson, B., Kinnunen, L., Hornfeldt, B., Horling, J., Benemar, C., Hedlund, K. O., Matskova, L., Hyypia, T. & Winberg, G. (1999). A new picornavirus isolated from bank voles (Clethrionomys glareolus). *Virology* **255**(1), 86-93.

Nishizawa, T., Okamoto, H., Konishi, K., Yoshizawa, H., Miyakawa, Y. & Mayumi, M. (1997). A novel DNA virus (TTV) associated with elevated transaminase levels in posttransfusion hepatitis of unknown etiology. *Biochem Biophys Res Commun* **241**(1), 92-97.

Nolte, F. S., Marshall, D. J., Rasberry, C., Schievelbein, S., Banks, G. G., Storch, G. A., Arens, M. Q., Buller, R. S. & Prudent, J. R. (2007). MultiCode-PLx system for multiplexed detection of seventeen respiratory viruses. *J Clin Microbiol* **45**(9), 2779-2786.

Novak, J. E. & Kirkegaard, K. (1991). Improved method for detecting poliovirus negative strands used to demonstrate specificity of positive-strand encapsidation and the ratio of positive to negative strands in infected cells. *J Virol* **65**(6), 3384-3387.

Nugent, C. I. & Kirkegaard, K. (1995). RNA binding properties of poliovirus subviral particles. *J Virol* **69**(1), 13-22.

Nygard, K., Torven, M., Ancker, C., Knauth, S. B., Hedlund, K. O., Giesecke, J., Andersson, Y. & Svensson, L. (2003). Emerging genotype (GGIIb) of norovirus in drinking water, Sweden. *Emerg Infect Dis* **9**(12), 1548-1552.

O'Regan, S., Robitaille, P., Mongeau, J. G. & McLaughlin, B. (1980). The hemolytic uremic syndrome associated with ECHO 22 infection. *Clin Pediatr (Phila)* **19**(2), 125-127.

Oberste, M. S., Maher, K., Kilpatrick, D. R. & Pallansch, M. A. (1999). Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. *J Virol* **73**(3), 1941-1948.

Oberste, M. S., Maher, K. & Pallansch, M. A. (1998). Complete sequence of echovirus 23 and its relationship to echovirus 22 and other human enteroviruses. *Virus Res* **56**(2), 217-223.

Oberste, M. S., Maher, K. & Pallansch, M. A. (2004). Evidence for frequent recombination within species human enterovirus B based on complete genomic sequences of all thirty-seven serotypes. *J Virol* **78**(2), 855-867.

Oh, D. Y., Gaedicke, G. & Schreier, E. (2003). Viral agents of acute gastroenteritis in German children: prevalence and molecular diversity. *J Med Virol* **71**(1), 82-93.

Olsen, S. J., Chang, H. L., Cheung, T. Y., Tang, A. F., Fisk, T. L., Ooi, S. P., Kuo, H. W., Jiang, D. D., Chen, K. T., Lando, J., Hsu, K. H., Chen, T. J. & Dowell, S. F. (2003). Transmission of the severe acute respiratory syndrome on aircraft. *N Engl J Med* **349**(25), 2416-2422.

Opstelten, D. J., Raamsman, M. J., Wolfs, K., Horzinek, M. C. & Rottier, P. J. (1995). Envelope glycoprotein interactions in coronavirus assembly. *J Cell Biol* **131**(2), 339-349.

Osterhaus, A. D., Pedersen, N., van Amerongen, G., Frankenhuis, M. T., Marthas, M., Reay, E., Rose, T. M., Pamungkas, J. & Bosch, M. L. (1999). Isolation and partial characterization of a lentivirus from talapoin monkeys (Myopithecus talapoin). *Virology* **260**(1), 116-124.

Palacios, G., Druce, J., Du, L., Tran, T., Birch, C., Briese, T., Conlan, S., Quan, P. L., Hui, J., Marshall, J., Simons, J. F., Egholm, M., Paddock, C. D., Shieh, W. J., Goldsmith, C. S., Zaki, S. R., Catton, M. & Lipkin, W. I. (2008). A New Arenavirus in a Cluster of Fatal Transplant-Associated Diseases. *N Engl J Med.*

Palmenberg, A. C. (1982). In vitro synthesis and assembly of picornaviral capsid intermediate structures. *J Virol* **44**(3), 900-906.

Parashar, U. D., Hummelman, E. G., Bresee, J. S., Miller, M. A. & Glass, R. I. (2003). Global illness and deaths caused by rotavirus disease in children. *Emerg Infect Dis* **9**(5), 565-572.

Park, S. J., Jeong, C., Yoon, S. S., Choy, H. E., Saif, L. J., Park, S. H., Kim, Y. J., Jeong, J. H., Park, S. I., Kim, H. H., Lee, B. J., Cho, H. S., Kim, S. K., Kang, M. I. & Cho, K. O. (2006). Detection and characterization of bovine coronaviruses in fecal specimens of adult cattle with diarrhea during the warmer seasons. *J Clin Microbiol* **44**(9), 3178-3188.

Parker, M. M. & Masters, P. S. (1990). Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. *Virology* **179**(1), 463-468.

Pata, J. D., Schultz, S. C. & Kirkegaard, K. (1995). Functional oligomerization of poliovirus RNA-dependent RNA polymerase. *RNA* **1**(5), 466-477.

Paul, A. V., Mugavero, J., Molla, A. & Wimmer, E. (1998). Internal ribosomal entry site scanning of the poliovirus polyprotein: implications for proteolytic processing. *Virology* **250**(1), 241-253.

Paul, A. V., Rieder, E., Kim, D. W., van Boom, J. H. & Wimmer, E. (2000). Identification of an RNA hairpin in poliovirus RNA that serves as the primary template in the in vitro uridylylation of VPg. *J Virol* **74**(22), 10359-10370.

Peiris, J. S., Lai, S. T., Poon, L. L., Guan, Y., Yam, L. Y., Lim, W., Nicholls, J., Yee, W. K., Yan, W. W., Cheung, M. T., Cheng, V. C., Chan, K. H., Tsang, D. N., Yung, R. W., Ng, T. K. & Yuen, K. Y. (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**(9366), 1319-1325.

Perz, J. F., Craig, A. S. & Schaffner, W. (2001). Mixed outbreak of parainfluenza type 1 and influenza B associated with tourism and air travel. *Int J Infect Dis* **5**(4), 189-191.

Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J., Jr. (1999). Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry* **38**(12), 3468-3477.

Pfeiffer, J. K. & Kirkegaard, K. (2005). Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* **1**(2), e11.

Pilipenko, E. V., Blinov, V. M. & Agol, V. I. (1990). Gross rearrangements within the 5'-untranslated region of the picornaviral genomes. *Nucleic Acids Res* **18**(11), 3371-3375.

Poon, L. L., Chu, D. K., Chan, K. H., Wong, O. K., Ellis, T. M., Leung, Y. H., Lau, S. K., Woo, P. C., Suen, K. Y., Yuen, K. Y., Guan, Y. & Peiris, J. S. (2005). Identification of a novel coronavirus in bats. *J Virol* **79**(4), 2001-2009.

Prudent, J. R. (2006). Using expanded genetic alphabets to simplify high-throughput genetic testing. *Expert Rev Mol Diagn* **6**(2), 245-252.

Raggam, R. B., Leitner, E., Berg, J., Muhlbauer, G., Marth, E. & Kessler, H. H. (2005). Single-run, parallel detection of DNA from three pneumonia-producing bacteria by real-time polymerase chain reaction. *J Mol Diagn* **7**(1), 133-138.

Rasschaert, D., Duarte, M. & Laude, H. (1990). Porcine respiratory coronavirus differs from transmissible gastroenteritis virus by a few genomic deletions. *J Gen Virol* **71 (Pt 11)**, 2599-2607.

Redshaw, N., Wood, C., Rich, F., Grimwood, K. & Kirman, J. R. (2007). Human bocavirus in infants, New Zealand. *Emerg Infect Dis* **13**(11), 1797-1799.

Reyes, G. R. & Kim, J. P. (1991). Sequence-independent, single-primer amplification (SISPA) of complex DNA populations. *Mol Cell Probes* **5**(6), 473-481.

Reyes, G. R., Purdy, M. A., Kim, J. P., Luk, K. C., Young, L. M., Fry, K. E. & Bradley, D. W. (1990). Isolation of a cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* **247**(4948), 1335-1339.

Risco, C., Anton, I. M., Enjuanes, L. & Carrascosa, J. L. (1996). The transmissible gastroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. *J Virol* **70**(7), 4773-4777.

Roberts, C., Bandaru, R. & Switzer, C. (1995). Synthesis of Oligonucleotides Bearing the Nonstandard Bases Iso-C and Iso-G - Comparison of Iso-C-Iso-G, C-G and U-a Base-Pair Stabilities in Rna/DNA Duplexes. *Tetrahedron Letters* **36**(21), 3601-3604.

Rodriguez, P. L. & Carrasco, L. (1993). Poliovirus protein 2C has ATPase and GTPase activities. *J Biol Chem* **268**(11), 8105-8110.

Rodriguez, P. L. & Carrasco, L. (1995). Poliovirus protein 2C contains two regions involved in RNA binding activity. *J Biol Chem* **270**(17), 10105-10112.

Roivainen, M., Hyypia, T., Piirainen, L., Kalkkinen, N., Stanway, G. & Hovi, T. (1991). RGD-dependent entry of coxsackievirus A9 into host cells and its bypass after cleavage of VP1 protein by intestinal proteases. *J Virol* **65**(9), 4735-4740.

Roman, E., Wilhelmi, I., Colomina, J., Villar, J., Cilleruelo, M. L., Nebreda, V., Del Alamo, M. & Sanchez-Fauquier, A. (2003). Acute viral gastroenteritis: proportion and clinical relevance of multiple infections in Spanish children. *J Med Microbiol* **52**(Pt 5), 435-440.

Rose, T. M., Schultz, E. R., Henikoff, J. G., Pietrokovski, S., McCallum, C. M. & Henikoff, S. (1998). Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Res* **26**(7), 1628-1635.

Rose, T. M., Strand, K. B., Schultz, E. R., Schaefer, G., Rankin, G. W., Jr., Thouless, M. E., Tsai, C. C. & Bosch, M. L. (1997). Identification of two homologs of the Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8) in retroperitoneal fibromatosis of different macaque species. *J Virol* **71**(5), 4138-4144.

Rota, P. A., Oberste, M. S., Monroe, S. S., Nix, W. A., Campagnoli, R., Icenogle, J. P., Penaranda, S., Bankamp, B., Maher, K., Chen, M. H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J. L., Chen, Q., Wang, D., Erdman, D. D., Peret, T. C., Burns, C., Ksiazek, T. G., Rollin, P. E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A. D., Drosten, C., Pallansch, M. A., Anderson, L. J. & Bellini, W. J. (2003). Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**(5624), 1394-1399.

Ruoslahti, E. & Pierschbacher, M. D. (1987). New perspectives in cell adhesion: RGD and integrins. *Science* **238**(4826), 491-497.

Russell, S. J. & Bell, E. J. (1970). Echoviruses and carditis. *Lancet* **1**(7650), 784-785.

Sampath, R., Hofstadler, S. A., Blyn, L. B., Eshoo, M. W., Hall, T. A., Massire, C., Levene, H. M., Hannis, J. C., Harrell, P. M., Neuman, B., Buchmeier, M. J., Jiang, Y., Ranken, R., Drader, J. J., Samant, V., Griffey, R. H., McNeil, J. A., Crooke, S. T. & Ecker, D. J. (2005). Rapid identification of emerging pathogens: coronavirus. *Emerg Infect Dis* **11**(3), 373-379.

Samuilova, O., Krogerus, C., Poyry, T. & Hyypia, T. (2004). Specific interaction between human parechovirus nonstructural 2A protein and viral RNA. *J Biol Chem* **279**(36), 37822-37831.

Santti, J., Hyypia, T., Kinnunen, L. & Salminen, M. (1999). Evidence of recombination among enteroviruses. *J Virol* **73**(10), 8741-8749.

Sasseville, A. M., Boutin, M., Gelinas, A. M. & Dea, S. (2002). Sequence of the 3'-terminal end (8.1 kb) of the genome of porcine haemagglutinating encephalomyelitis virus: comparison with other haemagglutinating coronaviruses. *J Gen Virol* **83**(Pt 10), 2411-2416.

Satija, N. & Lal, S. K. (2007). The molecular biology of SARS coronavirus. *Ann N Y Acad Sci* **1102**, 26-38.

Sato, K., Morishita, T., Nobusawa, E., Suzuki, Y., Miyazaki, Y., Fukui, Y., Suzuki, S. & Nakajima, K. (2000). Surveillance of influenza viruses isolated from travellers at Nagoya International Airport. *Epidemiol Infect* **124**(3), 507-514.

Sawicki, S. G., Sawicki, D. L. & Siddell, S. G. (2007). A contemporary view of coronavirus transcription. *J Virol* **81**(1), 20-29.

Schochetman, G., Stevens, R. H. & Simpson, R. W. (1977). Presence of infectious polyadenylated RNA in coronavirus avian bronchitis virus. *Virology* **77**(2), 772-782.

Schultheiss, T., Emerson, S. U., Purcell, R. H. & Gauss-Muller, V. (1995). Polyprotein processing in echovirus 22: a first assessment. *Biochem Biophys Res Commun* **217**(3), 1120-1127.

Schultz, E. R., Rankin, G. W., Jr., Blanc, M. P., Raden, B. W., Tsai, C. C. & Rose, T. M. (2000). Characterization of two divergent lineages of macaque rhadinoviruses related to Kaposi's sarcoma-associated herpesvirus. *J Virol* **74**(10), 4919-4928.

Sherrill, C. B., Marshall, D. J., Moser, M. J., Larsen, C. A., Daude-Snow, L., Jurczyk, S., Shapiro, G. & Prudent, J. R. (2004). Nucleic acid analysis using an expanded genetic alphabet to quench fluorescence. *J Am Chem Soc* **126**(14), 4550-4556.

Shibata, I., Tsuda, T., Mori, M., Ono, M., Sueyoshi, M. & Uruno, K. (2000). Isolation of porcine epidemic diarrhea virus in porcine cell cultures and experimental infection of pigs of different ages. *Vet Microbiol* **72**(3-4), 173-182.

Simmonds, P. (2006). Recombination and selection in the evolution of picornaviruses and other Mammalian positive-stranded RNA viruses. *J Virol* **80**(22), 11124-11140.

Simmonds, P. & Welch, J. (2006). Frequency and dynamics of recombination within different species of human enteroviruses. *J Virol* **80**(1), 483-493.

Simons, J. N., Pilot-Matias, T. J., Leary, T. P., Dawson, G. J., Desai, S. M., Schlauder, G. G., Muerhoff, A. S., Erker, J. C., Buijk, S. L., Chalmers, M. L. & et al. (1995). Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proc Natl Acad Sci U S A* **92**(8), 3401-3405.

Simpson, R., Aliyu, S., Iturriza-Gomara, M., Desselberger, U. & Gray, J. (2003). Infantile viral gastroenteritis: on the way to closing the diagnostic gap. *J Med Virol* **70**(2), 258-262.

Singh, A. M., Moore, P. E., Gern, J. E., Lemanske, R. F., Jr. & Hartert, T. V. (2007). Bronchiolitis to asthma: a review and call for studies of gene-virus interactions in asthma causation. *Am J Respir Crit Care Med* **175**(2), 108-119.

Sloots, T. P., McErlean, P., Speicher, D. J., Arden, K. E., Nissen, M. D. & Mackay, I. M. (2006). Evidence of human coronavirus HKU1 and human bocavirus in Australian children. *J Clin Virol* **35**(1), 99-102.

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., Guan, Y., Rozanov, M., Spaan, W. J. & Gorbalenya, A. E. (2003). Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol* **331**(5), 991-1004.

Song, H. D., Tu, C. C., Zhang, G. W., Wang, S. Y., Zheng, K., Lei, L. C., Chen, Q. X., Gao, Y. W., Zhou, H. Q., Xiang, H., Zheng, H. J., Chern, S. W., Cheng, F., Pan, C. M., Xuan, H., Chen, S. J., Luo, H. M., Zhou, D. H., Liu, Y. F., He, J. F., Qin, P. Z., Li, L. H., Ren, Y. Q., Liang, W. J., Yu, Y. D., Anderson, L., Wang, M., Xu, R. H., Wu, X. W., Zheng, H. Y., Chen, J. D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L. Y., Li, H., Chen, F., Di, B., He, L. J., Lin, J. Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X. J., Spiga, O., Guo, Z. M., Pan, H. Y., He, W. Z., Manuguerra, J. C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y. X., Wu, C. I. & Zhao, G. P. (2005). Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci U S A* **102**(7), 2430-2435.

Stang, A., Korn, K., Wildner, O. & Uberla, K. (2005). Characterization of virus isolates by particle-associated nucleic acid PCR. *J Clin Microbiol* **43**(2), 716-720.

Stanway, G. & Hyypia, T. (1999). Parechoviruses. *J Virol* **73**(7), 5249-5254.

Stanway, G., Joki-Korpela, P. & Hyypia, T. (2000). Human parechoviruses--biology and clinical significance. *Rev Med Virol* **10**(1), 57-69.

Stanway, G., Kalkkinen, N., Roivainen, M., Ghazi, F., Khan, M., Smyth, M., Meurman, O. & Hyypia, T. (1994). Molecular and biological characteristics of echovirus 22, a representative of a new picornavirus group. *J Virol* **68**(12), 8232-8238.

Stephensen, C. B., Casebolt, D. B. & Gangopadhyay, N. N. (1999). Phylogenetic analysis of a highly conserved region of the polymerase gene from 11 coronaviruses and development of a consensus polymerase chain reaction assay. *Virus Res* **60**(2), 181-189.

Stertz, S., Reichelt, M., Spiegel, M., Kuri, T., Martinez-Sobrido, L., Garcia-Sastre, A., Weber, F. & Kochs, G. (2007). The intracellular sites of early replication and budding of SARS-coronavirus. *Virology* **361**(2), 304-315.

Strauss, D. M., Glustrom, L. W. & Wuttke, D. S. (2003). Towards an understanding of the poliovirus replication complex: the solution structure of the soluble domain of the poliovirus 3A protein. *J Mol Biol* **330**(2), 225-234.

Switzer, C., Moroney, S. E. & Benner, S. A. (1989). Enzymatic Incorporation of a New Base Pair into DNA and Rna. *Journal of the American Chemical Society* **111**(21), 8322-8323.

Switzer, C. Y., Moroney, S. E. & Benner, S. A. (1993). Enzymatic Recognition of the Base-Pair between Isocytidine and Isoguanosine. *Biochemistry* **32**(39), 10489-10496.

Tam, A. W., Smith, M. M., Guerra, M. E., Huang, C. C., Bradley, D. W., Fry, K. E. & Reyes, G. R. (1991). Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length viral genome. *Virology* **185**(1), 120-131.

Tan, K., Zelus, B. D., Meijers, R., Liu, J. H., Bergelson, J. M., Duke, N., Zhang, R., Joachimiak, A., Holmes, K. V. & Wang, J. H. (2002). Crystal structure of murine sCEACAM1a[1,4]: a coronavirus receptor in the CEA family. *EMBO J* **21**(9), 2076-2086.

Tan, W. C. (2005). Viruses in asthma exacerbations. *Curr Opin Pulm Med* **11**(1), 21-26.

Tang, R. S., Barton, D. J., Flanegan, J. B. & Kirkegaard, K. (1997). Poliovirus RNA recombination in cell-free extracts. *RNA* **3**(6), 624-633.

Tang, X. C., Zhang, J. X., Zhang, S. Y., Wang, P., Fan, X. H., Li, L. F., Li, G., Dong, B. Q., Liu, W., Cheung, C. L., Xu, K. M., Song, W. J., Vijaykrishna, D., Poon, L. L., Peiris, J. S., Smith, G. J., Chen, H. & Guan, Y. (2006). Prevalence and genetic diversity of coronaviruses in bats from China. *J Virol* **80**(15), 7481-7490.

Tauriainen, S., Martiskainen, M., Oikarinen, S., Lonnrot, M., Viskari, H., Ilonen, J., Simell, O., Knip, M. & Hyoty, H. (2007). Human parechovirus 1 infections in young children--no association with type 1 diabetes. *J Med Virol* **79**(4), 457-462.

Teterina, N. L., Gorbalenya, A. E., Egger, D., Bienz, K. & Ehrenfeld, E. (1997). Poliovirus 2C protein determinants of membrane binding and rearrangements in mammalian cells. *J Virol* **71**(12), 8962-8972.

Thiel, V., Herold, J., Schelle, B. & Siddell, S. G. (2001). Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J Gen Virol* **82**(Pt 6), 1273-1281.

Timenetsky, J., Santos, L. M., Buzinhani, M. & Mettifogo, E. (2006). Detection of multiple mycoplasma infection in cell cultures by PCR. *Braz J Med Biol Res* **39**(7), 907-914.

Tosh, C., Hemadri, D. & Sanyal, A. (2002). Evidence of recombination in the capsid-coding region of type A foot-and-mouth disease virus. *J Gen Virol* **83**(Pt 10), 2455-2460.

Triantafilou, K., Triantafilou, M., Takada, Y. & Fernandez, N. (2000). Human parechovirus 1 utilizes integrins alphavbeta3 and alphavbeta1 as receptors. *J Virol* **74**(13), 5856-5862.

Tu, C., Crameri, G., Kong, X., Chen, J., Sun, Y., Yu, M., Xiang, H., Xia, X., Liu, S., Ren, T., Yu, Y., Eaton, B. T., Xuan, H. & Wang, L. F. (2004). Antibodies to SARS coronavirus in civets. *Emerg Infect Dis* **10**(12), 2244-2248.

Vainionpaa, R. & Hyypia, T. (1994). Biology of parainfluenza viruses. *Clin Microbiol Rev* **7**(2), 265-275.

van den Hoogen, B. G., de Jong, J. C., Groen, J., Kuiken, T., de Groot, R., Fouchier, R. A. & Osterhaus, A. D. (2001). A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat Med* **7**(6), 719-724.

van der Hoek, L., Pyrc, K., Jebbink, M. F., Vermeulen-Oost, W., Berkhout, R. J., Wolthers, K. C., Wertheim-van Dillen, P. M., Kaandorp, J., Spaargaren, J. & Berkhout, B. (2004). Identification of a new human coronavirus. *Nat Med* **10**(4), 368-373.

van Kuppeveld, F. J., Galama, J. M., Zoll, J., van den Hurk, P. J. & Melchers, W. J. (1996). Coxsackie B3 virus protein 2B contains cationic amphipathic helix that is required for viral RNA replication. *J Virol* **70**(6), 3876-3886.

van kuppeveld, F. J., Melchers, W. J., Kirkegaard, K. & Doedens, J. R. (1997). Structure-function analysis of coxsackie B3 virus protein 2B. *Virology* **227**(1), 111-118.

Vijaykrishna, D., Smith, G. J., Zhang, J. X., Peiris, J. S., Chen, H. & Guan, Y. (2007). Evolutionary insights into the ecology of coronaviruses. *J Virol* **81**(8), 4012-4020.

Vijgen, L., Keyaerts, E., Moes, E., Thoelen, I., Wollants, E., Lemey, P., Vandamme, A. M. & Van Ranst, M. (2005). Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J Virol* **79**(3), 1595-1604.

Villena, C., Gabrieli, R., Pinto, R. M., Guix, S., Donia, D., Buonomo, E., Palombi, L., Cenko, F., Bino, S., Bosch, A. & Divizia, M. (2003). A large infantile gastroenteritis outbreak in Albania caused by multiple emerging rotavirus genotypes. *Epidemiol Infect* **131**(3), 1105-1110.

Vlasak, R., Luytjes, W., Leider, J., Spaan, W. & Palese, P. (1988a). The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetylesterase activity. *J Virol* **62**(12), 4686-4690.

Vlasak, R., Luytjes, W., Spaan, W. & Palese, P. (1988b). Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc Natl Acad Sci U S A* **85**(12), 4526-4529.

Wang, D., Coscoy, L., Zylberberg, M., Avila, P. C., Boushey, H. A., Ganem, D. & DeRisi, J. L. (2002). Microarray-based detection and genotyping of viral pathogens. *Proc Natl Acad Sci U S A* **99**(24), 15687-15692.

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

Wang, L., Junker, D. & Collisson, E. W. (1993). Evidence of natural recombination within the S1 gene of infectious bronchitis virus. *Virology* **192**(2), 710-716.

Wang, L., Junker, D., Hock, L., Ebiary, E. & Collisson, E. W. (1994). Evolutionary implications of genetic variations in the S1 gene of infectious bronchitis virus. *Virus Res* **34**(3), 327-338.

Wang, M., Yan, M., Xu, H., Liang, W., Kan, B., Zheng, B., Chen, H., Zheng, H., Xu, Y., Zhang, E., Wang, H., Ye, J., Li, G., Li, M., Cui, Z., Liu, Y. F., Guo, R. T., Liu, X. N., Zhan, L. H., Zhou, D. H., Zhao, A., Hai, R., Yu, D., Guan, Y. & Xu, J. (2005). SARS-CoV infection in a restaurant from palm civet. *Emerg Infect Dis* **11**(12), 1860-1865.

Watanabe, K., Oie, M., Higuchi, M., Nishikawa, M. & Fujii, M. (2007). Isolation and characterization of novel human parechovirus from clinical samples. *Emerg Infect Dis* **13**(6), 889-895.

Wege, H., Muller, A. & ter Meulen, V. (1978). Genomic RNA of the murine coronavirus JHM. *J Gen Virol* **41**(2), 217-227.

Weiss, S. R. & Navas-Martin, S. (2005). Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev* **69**(4), 635-664.

Whelan, E. A., Lawson, C. C., Grajewski, B., Petersen, M. R., Pinkerton, L. E., Ward, E. M. & Schnorr, T. M. (2003). Prevalence of respiratory symptoms among female flight attendants and teachers. *Occup Environ Med* **60**(12), 929-934.

WHO (2003a). Consensus document on the epidemiology of severe acute respiratory syndrome (SARS). WHO/CDS/CSR/GAR/2003.11, Geneva: World Health Organization. (Retrieved January, 2008, from, http://www.who.int/csr/sars/en/WHOconsensus.pdf).

WHO (2003b). Summary of SARS and air travel. Geneva: World Health Organization. (Retrieved January, 2008, from, http://www.who.int/csr/sars/travel/airtravel/en/print.html).

WHO (2003c). Update 62-more than 8000 cases reported globally, situation in Taiwan, data on in-flight transmission, report on Henan province, China. Geneva: World Health Organization. (Retrieved January, 2008, from, http://www.who.int/csr/don/2003_05_22/en/print.html).

WHO (2007). Acute respiratory infections in children. (Retrieved January, 2008, from, http://www.who.int/fch/depts/cah/resp_infections/en/).

Wichman, H. A. & Van Den Bussche, R. A. (1992). In search of retrotransposons: exploring the potential of the PCR. *Biotechniques* **13**(2), 258-265.

Wigand, R. & Sabin, A. B. (1961). Properties of ECHO types 22, 23 and 24 viruses. *Arch Gesamte Virusforsch* **11**, 224-247.

Wilder-Smith, A., Leong, H. N. & Villacian, J. S. (2003). In-flight transmission of Severe Acute Respiratory Syndrome (SARS): a case report. *J Travel Med* **10**(5), 299-300.

Wilhelmi, I., Roman, E. & Sanchez-Fauquier, A. (2003). Viruses causing gastroenteritis. *Clin Microbiol Infect* **9**(4), 247-262.

Will, C., Schewe, C., Schluns, K. & Petersen, I. (2006). HPV typing and CGH analysis for the differentiation of primary and metastatic squamous cell carcinomas of the aerodigestive tract. *Cell Oncol* **28**(3), 97-105.

Wilson, C. A., Wong, S., Muller, J., Davidson, C. E., Rose, T. M. & Burd, P. (1998). Type C retrovirus released from porcine primary peripheral blood mononuclear cells infects human cells. *J Virol* **72**(4), 3082-3087.

Wise, A. G., Kiupel, M. & Maes, R. K. (2006). Molecular characterization of a novel coronavirus associated with epizootic catarrhal enteritis (ECE) in ferrets. *Virology* **349**(1), 164-174.

Wolfel, R., Paweska, J. T., Petersen, N., Grobbelaar, A. A., Leman, P. A., Hewson, R., Georges-Courbot, M. C., Papa, A., Gunther, S. & Drosten, C. (2007). Virus detection and monitoring of viral load in Crimean-Congo hemorrhagic fever virus patients. *Emerg Infect Dis* **13**(7), 1097-1100.

Woo, P. C., Lau, S. K., Chu, C. M., Chan, K. H., Tsoi, H. W., Huang, Y., Wong, B. H., Poon, R. W., Cai, J. J., Luk, W. K., Poon, L. L., Wong, S. S., Guan, Y., Peiris, J. S. & Yuen, K. Y. (2005). Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* **79**(2), 884-895.

Woo, P. C., Lau, S. K., Li, K. S., Poon, R. W., Wong, B. H., Tsoi, H. W., Yip, B. C., Huang, Y., Chan, K. H. & Yuen, K. Y. (2006). Molecular diversity of coronaviruses in bats. *Virology* **351**(1), 180-187.

Woo, P. C., Wang, M., Lau, S. K., Xu, H., Poon, R. W., Guo, R., Wong, B. H., Gao, K., Tsoi, H. W., Huang, Y., Li, K. S., Lam, C. S., Chan, K. H., Zheng, B. J. & Yuen, K. Y. (2007). Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *J Virol* **81**(4), 1574-1585.

Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S. G. & Ding, J. (2003). Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res* **31**(24), 7117-7130.

Yang, C. F., Chen, H. Y., Jorba, J., Sun, H. C., Yang, S. J., Lee, H. C., Huang, Y. C., Lin, T. Y., Chen, P. J., Shimizu, H., Nishimura, Y., Utama, A., Pallansch, M., Miyamura, T., Kew, O. & Yang, J. Y. (2005). Intratypic recombination among lineages of type 1 vaccine-derived poliovirus emerging during chronic infection of an immunodeficient patient. *J Virol* **79**(20), 12623-12634.

Yoo, D., Pei, Y., Christie, N. & Cooper, M. (2000). Primary structure of the sialodacryoadenitis virus genome: sequence of the structural-protein region and its application for differential diagnosis. *Clin Diagn Lab Immunol* **7**(4), 568-573.

Yu, G. Y. & Lai, M. M. (2005). The ubiquitin-proteasome system facilitates the transfer of murine coronavirus from endosome to cytoplasm during virus entry. *J Virol* **79**(1), 644-648.

Yu, I. T., Li, Y., Wong, T. W., Tam, W., Chan, A. T., Lee, J. H., Leung, D. Y. & Ho, T. (2004). Evidence of airborne transmission of the severe acute respiratory syndrome virus. *N Engl J Med* **350**(17), 1731-1739.

Zelus, B. D., Schickli, J. H., Blau, D. M., Weiss, S. R. & Holmes, K. V. (2003). Conformational changes in the spike glycoprotein of murine coronavirus are induced at 37 degrees C either by soluble murine CEACAM1 receptors or by pH 8. *J Virol* **77**(2), 830-840.

Zhang, J., Guy, J. S., Snijder, E. J., Denniston, D. A., Timoney, P. J. & Balasuriya, U. B. (2007). Genomic characterization of equine coronavirus. *Virology* **369**(1), 92-104.

Ziebuhr, J. (2004). Molecular biology of severe acute respiratory syndrome coronavirus. *Curr Opin Microbiol* **7**(4), 412-419.

Zitter, J. N., Mazonson, P. D., Miller, D. P., Hulley, S. B. & Balmes, J. R. (2002). Aircraft cabin air recirculation and symptoms of the common cold. *JAMA* **288**(4), 483-486.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**(13), 3406-3415.