

Viral metagenomics of six bat species in close contact with humans in southern China

Xue-yan Zheng^{1,2} · Min Qiu¹ · Wei-jie Guan³ · Jin-ming Li⁴ · Shao-wei Chen¹ · Ming-ji Cheng¹ · Shu-ting Huo¹ · Zhong Chen⁵ · Yi Wu⁶ · Li-na Jiang¹ · Qing Chen¹

Received: 16 January 2017 / Accepted: 14 September 2017 / Published online: 5 October 2017
© Springer-Verlag GmbH Austria 2017

Abstract Accumulating studies have shown that bats could harbor various important pathogenic viruses that could be transmitted to humans and other animals. Extensive metagenomic studies of different organs/tissues from bats have revealed a large number of novel or divergent viruses. To elucidate viral diversity and epidemiological and phylogenetic characteristics, six pooled fecal samples from bats were generated (based on bat species and geographic regions characteristic for virome analysis). These contained 500 fecal samples from six bat species, collected

in four geographic regions. Metagenomic analysis revealed a plethora of divergent viruses originally found in bats. Multiple contigs from influenza A virus and coronaviruses in bats shared high identity with those from humans, suggesting possible cross-species transmission, whereas a number of contigs, whose sequences were taxonomically classifiable within *Alphapapillomavirus*, *Betaretrovirus*, *Alpharetrovirus*, *Varicellovirus*, *Cyprinivirus*, *Chlorovirus* and *Cucumovirus* had low identity to viruses in existing databases, which indicated possible evolution of novel viral species. None of the established caliciviruses and picornaviruses were found in the 500 fecal specimens. Papillomaviruses with high amino acid identity were found in *Scotophilus kuhlii* and *Rhinolophus blythi*, challenging the hypotheses regarding the strict host specificity and co-evolution of papillomaviruses. Phylogenetic analysis showed that four bat rotavirus A strains might be tentative G3 strains, according to the Rotavirus Classification Working Group classification.

Handling Editor: Zhongjie Shi.

Electronic supplementary material The online version of this article (doi:[10.1007/s00705-017-3570-3](https://doi.org/10.1007/s00705-017-3570-3)) contains supplementary material, which is available to authorized users.

✉ Qing Chen
qch.2009@163.com

- ¹ Department of Epidemiology, School of Public Health, Southern Medical University, 1838 Guangzhou Avenue North, Guangzhou, Guangdong, China
- ² Institute of Non-communicable Disease Control and Prevention, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, China
- ³ State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Disease, Guangzhou, Guangdong, China
- ⁴ Department of Bioinformatics, School of Basic Medical Sciences, Southern Medical University, Guangzhou, Guangdong, China
- ⁵ College of Life Science, Hainan Normal University, Haikou, Hainan, China
- ⁶ College of Life Science, Guangzhou University, Guangzhou, Guangdong, China

Introduction

As members of the order *Chiroptera*, the second largest group (~25%) of mammalian species [1], more than 1,240 bat species are currently found worldwide, everywhere except the South and North Poles. Recent studies have successfully identified novel bat viruses on the basis of consensus primers or sequence-independent PCR amplification [2–4]. Over 80 virus species have been detected in bats, including several emerging human pathogens (i.e., severe acute respiratory syndrome coronaviruses, lyssaviruses, henipaviruses, Marburg virus and Ebola virus) [5]. Furthermore, some bat species live in close proximity to human habitation, highlighting the potential role of bats as reservoirs of zoonotic diseases [6].

Using polymerase chain reaction (PCR) screening with consensus primers, closely related viral species in bats have been detected [2–4]. However, the high assay specificity of these methods has limited the detection of divergent or unknown viruses. The advent of next-generation sequencing (NGS) technology has facilitated metagenomic analyses to simultaneously screen for all viral families. NGS has been successfully applied to detect novel viruses in different types of samples (including lungs, liver, brain, pharyngeal swabs and anal swabs) [7] and is expected to provide more in-depth information on previously unidentified viruses and viral taxonomy. Studies from the US [8, 9], France [10], New Zealand [11] and China [12–15] have analyzed the virome profiles of bats with NGS technology and have identified various unknown mammalian viruses. Most of these studies analyzed viral contigs by homogenizing different organs and tissues of bats; however, this approach is restricted by the unclear tissue distribution and route of transmission of bat viruses. The paucity of knowledge on bat-derived viruses warrants more extensive investigation and surveillance of different bat species in various geographic areas. The digestive tract is connected to the outer environment through ingestion of foods and liquids, providing a niche for virus transmission from bats to humans and other animals. Investigation of the virome in fecal samples from bats, and subsequent epidemiological and phylogenetic characterization, of any viruses, would significantly advance our understanding, especially with regards to prevention and management of zoonotic diseases in different geographic regions.

Papillomaviruses (PVs), rotavirus A (RVA), caliciviruses (CVs) and picornaviruses (PiVs) are widespread pathogens causing human diseases and have been detected in fecal samples of some bat species [16–19]. However, it is still unknown whether bats play a role as reservoirs in transmission of these viruses. More extensive studies of the prevalence and characteristics of these viruses isolated from different bat species and in different geographic areas is needed. To conduct this analysis, fecal specimens from six bat species, *Cynopterus sphinx*, *Miniopterus schreibersii*, *Rousettus leschenaultii*, *Hipposideros larvatus*, *Rhinolophus blythi* and *Scotophilus kuhlii*, were collected.

In this study, we report on the virome profile of fecal samples from six bat species in four cities in southern China, using metagenomic analysis, sequence-independent amplification and high-throughput sequencing (Solexa, Illumina).

Materials and methods

Ethics statement

The study protocol was reviewed and approved by the Animal Ethics and Welfare Committee of the School of Public

Health and Tropical Medicine, Southern Medical University, China. All animals were treated in strict accordance with the guidelines for Laboratory Animal Use and Care from Southern Medical University and the Rules for the Implementation of Laboratory Animal Medicine (1998) from the Ministry of Health, China.

Sample collection

Fecal sample collection from bats was performed in Hainan province [Haikou city (latitude: 20.02° N; longitude: 110.35° E)], Guangdong province [Huizhou (latitude: 23.11° N; longitude: 114.42° E), Guangzhou (latitude: 23.13° N; longitude: 113.26° E) and Yunfu (latitude: 22.92° N; longitude: 112.04° E) cities], all in southern China. Briefly, fecal samples from bats were immersed into maintenance medium in a virus sampling tube and temporarily stored at -20 °C. After completion of sampling, fecal samples were transported to the laboratory and stored at -80 °C. Bat species identification was confirmed by amplification and sequencing of the cytochrome B (*cytB*) gene, which has been commonly applied in archaeology [20].

Purification of fecal samples and extraction of viral nucleic acids

Tubes with fecal samples in maintenance medium were vigorously vortexed to homogenize the samples. Samples of each species from the same site were then pooled by adding 1 ml from each maintenance medium sample into a fresh sample tube. Six pooled samples, classified by species and collection site, were then centrifuged twice at 13,000×g for 20 min at 4 °C. The supernatant was then filtered through a 0.22 µm filter (Millipore Inc., USA) twice to remove eukaryotic cell- and bacterium-sized particles. Filtrates were concentrated in a 100-kDa Pellicon II filter (Millipore Inc., USA). To remove the remaining extra-cellular nucleic acids, the filtrates were treated with DNaseI (NEB, USA) and RNaseIf (NEB). 375 µl of the filtrate from each pooled sample was then digested in a mixture of 3 U DNase and 25 U RNase at 37 °C for 90 min in 10× DNase buffer (NEB). The viral DNA and RNA were simultaneously extracted using a TaKaRa MiniBEST Viral RNA/DNA Extraction Kit Ver.5.0 (TaKaRa Inc., Japan).

Reverse transcription (RT) and sequence-independent PCR amplification of viral nucleic acids

Reverse transcription was performed using a Transcriptor First-Strand cDNA Synthesis Kit 6.0 (Roche Inc., USA) and 100 pmol of primer K-8N (GACCATCTAGCGACC TCCACNNNNNNNN) [21]. The first-strand cDNA of each sample was converted into double-stranded cDNA in the

presence of 5 U of Klenow fragment (NEB) in 10× NEB buffer 2, which was incubated at 37 °C for 1 h and inactivated at 75 °C for 10 min.

Sequence-independent PCR amplification was conducted with 5 µl of double-stranded cDNA template in a final reaction volume of 50 µl, which contained 2× Gflex PCR Buffer, 200 µM deoxynucleoside triphosphate (dNTP), 1 µM primer K (GACCATCTAGCGACCTCCAC) and 1.25 U Tks Gflex DNA polymerase (TaKaRa). The PCR cycles were set as follows: 94 °C for 1 min, followed by 40 cycles of 98 °C for 10 s, 55 °C for 30 s and 68 °C for 1 min.

Library construction and sequencing

A total of 1 µg of DNA per sample was used as input material for DNA sample preparation. Sequencing libraries were generated using a NEBNext® Ultra™DNA Library Prep Kit for Illumina (NEB) following the manufacturer's recommendations. Briefly, the DNA sample was fragmented by sonication to a size of ~300 bp, and fragments were end-polished, poly-A tailed and ligated with full-length adaptors for Illumina sequencing by further PCR amplification. Finally, PCR products were purified (AMPure XP system, USA), and libraries were analyzed for size distribution using an Agilent 2100 Bioanalyzer (USA) and quantified using real-time PCR.

Pre-processing sequence filter

A pre-processing workflow was applied to ensure that high-quality reads were obtained. The following reads were removed: (1) reads containing low-quality bases (Sanger quality value ≤ 5 and low-quality bases exceeding 40% of the length of the reads); (2) reads with excessive N bases ($\geq 10\%$ of given length of the reads); (3) reads of less than 50% of a given length; (4) reads containing an overlap of 15 bp with sequences on a user-defined primer/adaptor sequence list; (5) homopolymer-containing reads; and (6) duplicated reads.

The most closely related host genome sequences in this study were scanned and discarded with SOAP2 mapper software [22], by comparison with the available genome sequences of *Myotis brandtii*, *Myotis davidii*, *Myotis lucifugus* and *Eptesicus fuscus* (whole-genome shotgun sequencing project accession numbers: ANKR00000000.1, ANKR00000000.1, AAPE00000000.2 and ALEH00000000.1). Sequences originating from the host genome ($\geq 90\%$ consistency) were deleted.

Analysis of sequence reads and *de novo* metagenomic assembly

Valid sequence reads were aligned with sequences in the NCBI non-redundant nucleotide database (NT), the

non-redundant protein database (NR) and viral Refseq databases (downloaded from the NCBI FTP server in November 2014) using BLASTn, BLASTx and tBlastx, respectively. The BLAST hits were defined as significant when e-value $\leq 10^{-5}$ [14]. A rapid, highly restrictive BLASTn homology search (with word size 40) against a non-redundant nucleotide (nt) database was performed to eliminate additional host reads and eukaryotic contaminants. The virus-like clean data was then subjected to further analysis.

Clean data from the above filter process were assembled using SOAP *de novo* assembling software [23] using different K-mer references (45, 55, 59). Contigs with the largest N50 were selected. Assembled sequence contigs greater than 150 bp were compared with sequences in the GenBank non-redundant nucleotide database using BLASTn. The reference sequence of the source organism with the best contig alignment, with an e-value cut-off of 10^{-5} , was retrieved for further phylogenetic analysis. More than one hit was sometimes obtained for each unassembled read and contig, at different taxonomic levels. To guarantee biological significance, the eligible sequences were parsed and exported with MetaGenome Analyzer (MEGAN 4) using the LCA algorithm to assign each sequence to the appropriate taxon using the NCBI taxonomic database [24]. The first and best hit was regarded as the taxonomic annotation of the contigs or unassembled reads.

Molecular epidemiology of bat PVs, RVA, CVs and PiVs

To delineate the molecular epidemiology of bat PVs, RVA, CVs and PiVs from different bat species in four geographic regions of Guangdong and Hainan provinces in southern China, previously described PCR primers were used to amplify the conserved regions of the Late 1 (L1) gene of PVs (450 bp), the major capsid region (VP7) gene of RVA (847 bp), the RdRp gene of CVs (319-331 bp) and the 3Dpol gene of PiVs (571 bp), respectively.

Phylogenetic analysis

Sequence editing and identity calculations were conducted by using BioEdit, version 7.0.4 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). Nucleotide and amino acid sequence alignments were constructed for different open reading frames (ORFs) and the corresponding amino acid alignment using Clustal W version 2.0 [25]. Phylogenetic trees were constructed using Clustal W version 2.0 and Mega 6 [26] according to the neighbor-joining and maximum-likelihood algorithms. Bootstrap values of the constructed phylogenetic trees were generated by iteration using 1,000 replicates.

Nucleotide sequence accession numbers

All sequences in this study have been submitted to GenBank. The accession numbers for the *pol* gene of retrovirus, IVa2 gene of adenovirus (AdV), Vp7 gene of rotavirus and L1 gene of PV are KY321921, KY321918, KU746889-KU746892 and KU727225-KU727235, respectively.

Results

Bat sampling

Bat samples were collected between August 2011 and October 2014 in four cities (Huizhou, Guangzhou, Yunfu and Haikou) in southern China. We sampled the fecal specimens from 500 bats (6 species). Eighty-six fecal samples, classified by species and collection site, were pooled as six samples (E-Table 2). These samples were processed as follows: nucleic acid purification, RNA/DNA extraction and sequence-independent RT-PCR for NGS. Molecular epidemiological analysis of of RVA, PV, CV and PiV in the 500 bats was also conducted (E-Table 3).

Virome analysis

Viral metagenomic overview

After removing the contaminating reads, 108,443,256 sequence reads with an average length of 150 bp had been generated from the extracted nucleic acids (Table 1). The fecal samples from bats revealed significant microbial diversity. A total of 90,181,623 sequence reads were classified as cellular organisms, including bacteria, archaea, eukaryota and other unassigned reads. There were 18,261,633 reads

that best matched to viral proteins. The sequences from each pool were assembled *de novo* into contigs of variable lengths. A total of 13,466,976 sequences were assembled into viral contigs (average length: 182 bp, e-value ≤ 0.0001), among which 5,496 were viral contigs longer than 500 bp (average length: 776 bp).

The total viral contigs could be assigned into five viral clades, including 54 families and 219 genera: Reverse-Transcribing Viruses, dsDNA Viruses, dsRNA Viruses, ssDNA Viruses and ssRNA Viruses. Vertebrate viruses comprised the largest population and included viruses classifiable as members of the *Alloherpesviridae*, *Herpesviridae*, *Adenoviridae*, *Orthomyxoviridae*, *Papillomaviridae*, *Retroviridae*, *Hepadnaviridae*, *Flaviviridae*, *Polyomaviridae*, *Paramyxoviridae*, *Rhabdoviridae*, *Coronaviridae*, *Picornaviridae*, *Arenaviridae*, *Astroviridae*, *Bunyaviridae*, *Iridoviridae*, *Poxviridae* and *Circoviridae* families (Figure 1, Table 2; please refer to Table 2 for details of other viral contigs, plant, bacterial, fungal, archaeal and algal in nature).

There were 1424 deltaretrovirus-related contigs, accounting for most of the virus-related contigs. These had the highest amino acid identity with the Pol protein of *Pteropus vampyrus* endogenous retrovirus group K (90%). Sequences of contigs from members classifiable within the genera *Phlebovirus*, *Tospovirus*, *Rotavirus A*, *Mastarenavirus*, *Picobirnavirus*, *Simplexvirus*, *Varicellovirus*, *Mastadenovirus*, *Circovirus*, *Mamastrovirus*, *Coronavirus*, *Orthohepadnavirus*, *Reoviridae*, *Iflavirus*, *Bracovirus*, *Orthopoxvirus*, *Vesiculovirus* and *Chrysovirus* (as well as all phage-like contigs) shared an identity of greater than 80% with sequences in the existing database. In addition, the contigs representing influenza A viruses (with identities of 89.6–100% to sequences in the database, average: 92.8%; 13.2% of the sequence identities were greater than 95%) and coronaviruses (identities of 95–100%, average: 99.8%) found in bats shared high

Table 1 Overview of Solexa sequencing

Pool	Data size	Viral Reads	Others	Contigs	A.length	Contigs >500bp	A.length >500bp
GZ.CS	10,337,998	5,442,833	4,895,165	5,081,665	168	386	861
HN.MS	15,584,606	8,015,941	7,568,665	6,742,580	197	1176	783
HN.RL	35,722,372	244,671	35,477,701	80,568	209	805	650
HZ.HL	22,860,172	337,084	22,523,088	107,376	201	1267	847
YF.RB	8,191,142	2,580,806	5,610,336	1,021,724	162	1232	829
YF.SK	15,746,966	1,640,298	14,106,668	433,066	156	630	687
Total	108,443,256	18,261,633	90,181,623	13,466,976	182	5496	776

GZ.CS: *Cynopterus sphinx* collected in Guangzhou; HN.MS: *Miniopterus schreibersi* collected in Hainan; HN.RL: *Rousettus leschenaulti* collected in Hainan; HZ.HL: *Hipposideros larvatus* collected in Huizhou; YF.RB: *Rhinolophus blythi* collected in Yunfu; YF.SK: *Scotophilus kuhlii* collected in Yunfu; Data size: the total reads after removal of contamination by host genome; Reads: the total length of viral reads; Contigs: the total length of contigs after overlapping; A. length: the average length of contigs; A. Length >500bp: the average length of contigs >500bp; Others: the total length of reads from eukaryotic, bacterial and unknown sequences without a hit in the GenBank non-redundant database

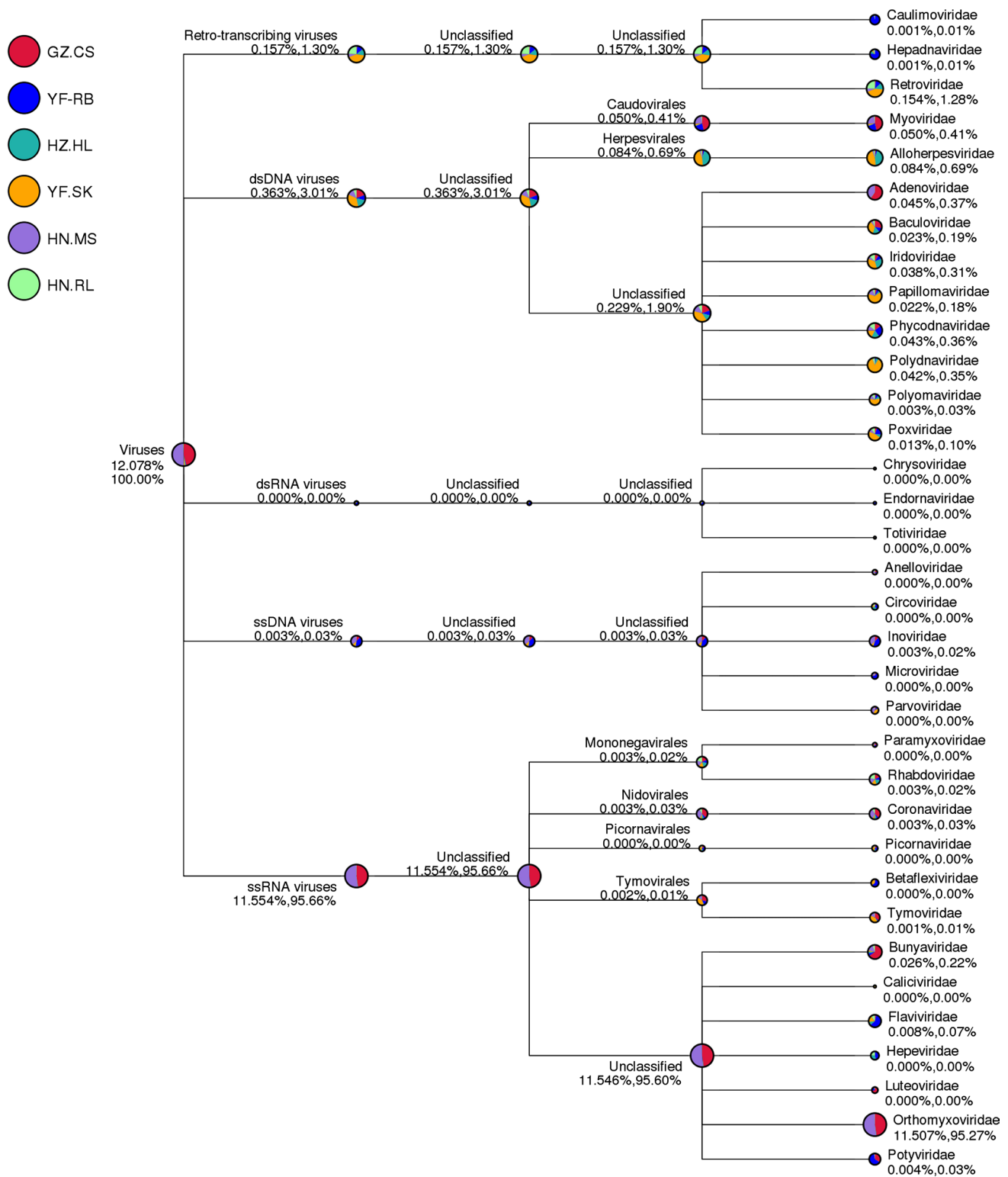


Fig. 1 Taxonomy information for the six bat groups using the MEGAN software. GZ.CS: *Cynopterus sphinx* collected in Guangzhou; HN.MS: *Miniopterus schreibersi* collected in Hainan; HN.RL:

Rousettus leschenaulti collected in Hainan; HZ.HL: *Hipposideros larvatus* collected in Huizhou; YF.RB: *Rhinolophus blythi* collected in Yunfu; YF.SK: *Scotophilus kuhlii* collected in Yunfu

Table 2 Overview of the viral contigs identified in this study

Virus host	Viral family	Viral genus	HN.MS	GZ.CS	YF.RB	YF.SK	HZ.HL	HN.RL	Total	
Vertebrate	<i>Alloherpesviridae</i>	<i>Cyprinivirus</i>	6	0	9	3	16	18	52	
		<i>Ictalurivirus</i>	0	0	0	6	15	63	84	
	<i>Herpesviridae</i>	<i>Cytomegalovirus</i>	15	3	15	12	16	11	69	
		<i>Lymphocryptovirus</i>	12	0	6	9	4	5	36	
		<i>Macavirus</i>	3	0	0	0	2	1	6	
		<i>Mardivirus</i>	0	0	0	0	10	14	24	
		<i>Muromegalovirus</i>	0	0	0	0	0	1	1	
		<i>Percavirus</i>	0	0	0	0	0	2	2	
		<i>Proboscivirus</i>	0	0	0	0	4	8	12	
		<i>Rhadinovirus</i>	0	0	3	9	1	4	17	
		<i>Roseovirus</i>	3	3	0	9	7	25	47	
		<i>Simplexvirus</i>	0	0	3	9	2	13	27	
		<i>Varicellovirus</i>	0	0	0	3	16	26	42	
	<i>Adenoviridae</i>	<i>Mastadenovirus</i>	15	24	0	0	1	1	41	
	<i>Orthomyxoviridae</i>	<i>Influenzavirus A</i>	3	3	78	0	162	254	500	
	<i>Papillomaviridae</i>	<i>Alphapapillomavirus</i>	15	3	15	8	142	118	301	
	<i>Retroviridae</i>	<i>Alpharetrovirus</i>	54	0	9	12	8	10	93	
		<i>Gammaretrovirus</i>	105	5	21	0	23	78	232	
		<i>Lentivirus</i>	15	0	3	6	21	78	123	
		<i>Rotavirus</i>	0	0	12	9	4	18	43	
		<i>Betaretrovirus</i>	0	0	9	15	16	44	84	
		<i>Deltaretrovirus</i>	0	0	3	39	18	1364	1424	
	<i>Hepadnaviridae</i>	<i>Unclassified</i>	0	25	48	120	16	66	275	
		<i>Orthohepadnavirus</i>	0	0	3	0	2	3	8	
	<i>Flaviviridae</i>	<i>Hepevirus</i>	0	0	6	0	5	4	15	
		<i>Hepacivirus</i>	3	0	30	0	8	15	56	
	<i>Polyomaviridae</i>	<i>Pestivirus</i>	0	0	36	0	15	36	87	
		<i>Polyomavirus</i>	3	0	0	3	15	45	66	
	<i>Paramyxoviridae</i>	<i>Respirovirus</i>	0	0	0	0	0	2	2	
	<i>Rhabdoviridae</i>	<i>Bracorhabdovirus</i>	0	0	0	0	0	6	6	
		<i>Novirhabdovirus</i>	0	0	0	0	3	0	3	
	<i>Coronaviridae</i>	<i>Coronavirus</i>	0	0	0	0	0	1	1	
	<i>Picornaviridae</i>	<i>Aphthovirus</i>	0	0	0	0	0	3	3	
		<i>Enterovirus</i>	0	0	0	3	1	1	5	
		<i>Parechovirus</i>	0	0	3	0	0	0	3	
	<i>Arenaviridae</i>	<i>Arenavirus</i>	0	0	0	0	1	5	5	
	<i>Astroviridae</i>	<i>Mamastrovirus</i>	0	0	0	0	0	1	1	
	<i>Bunyaviridae</i>	<i>Hantavirus</i>	0	0	0	0	0	9	9	
		<i>Orthobunyavirus</i>	0	0	3	0	1	12	16	
		<i>Phlebovirus</i>	0	0	0	0	7	0	7	
	<i>Iridoviridae</i>	<i>Iridovirus</i>	0	0	0	0	0	4	4	
	<i>Poxviridae</i>	<i>Avipoxvirus</i>	0	0	0	0	0	1	1	
		<i>Molluscipox virus</i>	0	0	3	0	4	9	16	
		<i>Orthopoxvirus</i>	0	0	9	0	13	31	53	
	<i>Circoviridae</i>	<i>Circovirus</i>	0	0	3	0	0	0	3	
			Subtotal	252	66	330	275	579	2410	3912
	Plant	<i>Luteoviridae</i>	<i>Polerovirus</i>	0	0	0	0	0	0	0
<i>Bromoviridae</i>		<i>Cucumovirus</i>	0	0	0	0	2	4	6	
<i>Bunyaviridae</i>		<i>Tospovirus</i>	0	0	0	3	19	151	173	
<i>Potyviridae</i>		<i>Potyvirus</i>	0	0	36	0	0	5	41	

Table 2 (continued)

Virus host	Viral family	Viral genus	HN.MS	GZ.CS	YF.RB	YF.SK	HZ.HL	HN.RL	Total
		Subtotal	0	0	36	3	21	160	220
Fungi	<i>Chrysoviridae</i>	<i>Chrysovirus</i>	0	0	0	0	2	12	14
	<i>Narnaviridae</i>	<i>Mitovirus</i>	0	0	0	6	1	2	9
		Subtotal	0	0	0	6	3	14	23
Bacteria	<i>Siphoviridae</i>	<i>Lambdalikeyvirus</i>	15	0	24	0	0	0	39
		<i>Tunalikeyvirus</i>	0	0	0	30	0	1	31
		<i>Tm4likevirus</i>	0	0	0	0	1	1	2
	<i>Myoviridae</i>	<i>T4 like virus</i>	0	0	9	0	1	1	11
		<i>T7 like virus</i>	0	0	3	36	0	1	40
	<i>Bicaudaviridae</i>	<i>Bicaudavirus</i>	0	0	0	0	0	1	1
		Subtotal	15	0	36	66	2	5	124
Insect	<i>Parvoviridae</i>	<i>Iteradensovirus</i>	18	0	0	15	0	0	33
	<i>Potyviridae</i>	<i>Potyviridae</i>	0	0	0	0	0	4	4
	<i>Baculoviridae</i>	<i>Alphabaculovirus</i>	0	0	0	0	9	12	21
		<i>Betabaculovirus</i>	0	0	0	0	1	4	5
	<i>Polydnaviridae</i>	<i>Bracovirus</i>	0	0	3	6	53	96	158
		<i>Ichnovirus</i>	0	0	0	0	8	22	30
		Subtotal	18	0	3	21	71	138	251
Archaea	<i>Rudiviridae</i>	<i>Rudiviridae</i>	0	0	0	0	0	2	2
		Subtotal	0	0	0	0	0	2	2
Alga	<i>Phycodnaviridae</i>	<i>Chlorovirus</i>	0	0	3	0	1	12	16
		<i>Coccolithovirus</i>	0	0	0	0	4	0	4
		Subtotal	0	0	3	0	5	14	22
		Total	285	66	408	371	681	2741	4552

sequence identity with those from humans. Other contigs including those classifiable as members of the genera *Alphapapillomavirus* (sequence identities of 82.8–92.5%, average: 83.4%), *Betaretrovirus* (61.9–84.3%, average: 69.7%), *Alpharetrovirus* (46.5–86.1%, average: 67.5%), *Varicellovirus* (28.7–100%, average: 49.7%), *Cyprinivirus* (35.6–100%, average: 58.5%), *Chlorovirus* (33.3–77.8%, average: 53.4%) and *Cucumovirus* (46.9–68.8%, average: 57.1%) had low sequence identity to viruses in the existing databases (data not shown).

Bat ReVs (retroviruses)

A total of 2274 contigs were identified as belonging to the five genera of retroviruses (alpha-, beta-, gamma-, delta- and unclassified retroviruses), with e-values of less than 10^{-5} (Table 2). The retroviral sequences identified herein represented two genes, encoding the protease/polymerase (Pol) and envelope glycoproteins (Env). According to the BLASTx analysis, some of these genes were very closely related to those of *Pteropus vampyrus* endogenous retroviruses. All retroviral contigs that encoded the Env protein also contained stop codons within the translated region, according to BLASTx analysis. One of the longest retrovirus

coding sequences encoding a Pol protein was used to construct a phylogenetic tree (Figure 2). The phylogenetic trees indicated that the longest Pol (HN.RL, accession number: KY321921) sequence identified clustered with those of *Pteropus vampyrus* endogenous retroviruses. HN.RL shared the highest amino acid sequence identity (85.7%) with *Pteropus vampyrus* endogenous retroviruses, when compared to other retroviruses detected in bats (56.7–59.8% sequence identity), duck (57.4%), fowl (56.7%), galidia (62.1%), murine (57.7–60.1%), baboon (58.1%), koala (58.1%) and porcine (57.4%) sources (Table 3).

Bat AdVs (adenoviruses)

We identified 41 contigs representing AdV genomes in *Cynopterus sphinx* (GZ.CS), *Miniopterus schreibersii* (HN.MS), *Hipposideros larvatus* (HZ.HL) and *Rousettus leschenaultii* (HN.RL). No AdV contigs were found in *Scotophilus kuhlii* (YF.SK) or *Rhinolophus blythi* (YF.RB). These AdV contigs represented 25 E4orf1 protein-encoding and 16 IVa2 mature protein-encoding genes. A phylogenetic tree based on the IVa2 gene sequences (Figure 3) revealed that the contig representing AdV in HN.MS fell within the lineage of human adenovirus C, sharing a sequence identity

Fig. 2 Phylogenetic analysis of the pol gene of ReVs detected in bats from southern China, based on a partial 171 amino acid sequence. The tree was generated using the neighbor-joining algorithm and the p-distance model. A bootstrap test of 1000 replicates was used. Numbers above the branches indicate NJ bootstrap values. Bold triangles indicate retroviruses detected in the present study. RL: *Rousettus leschenaultii*; HN: Hainan province

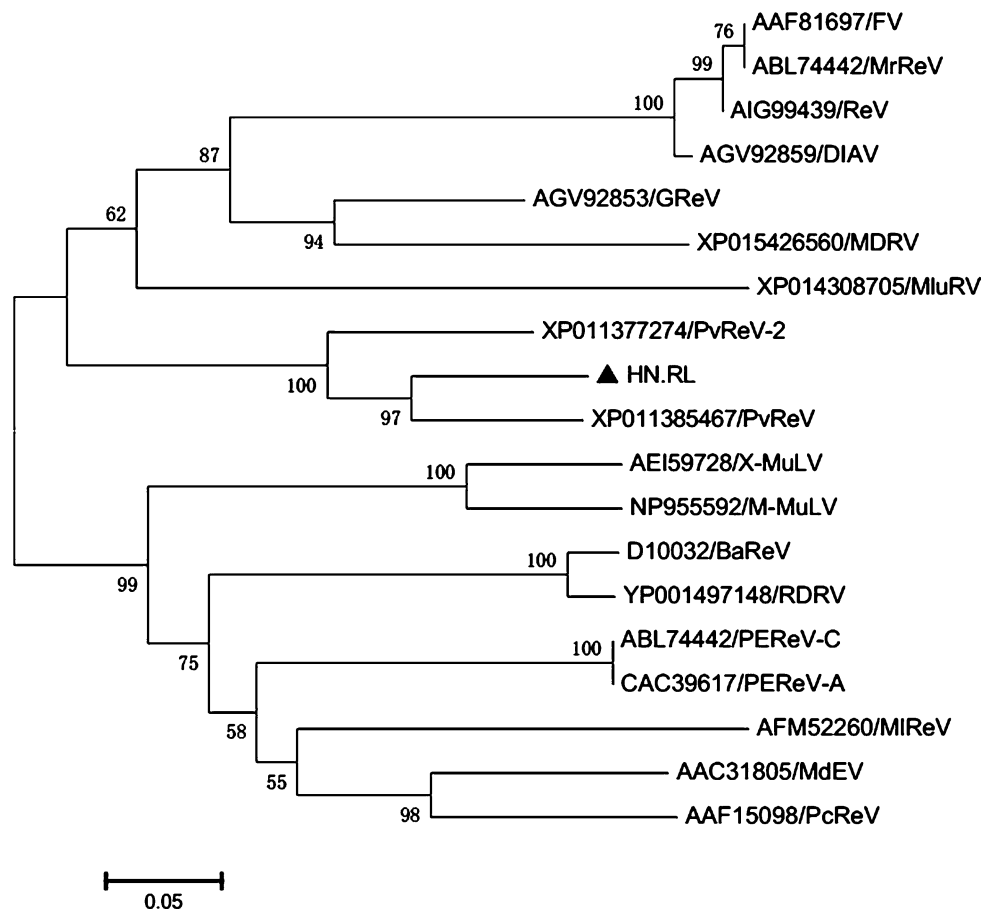
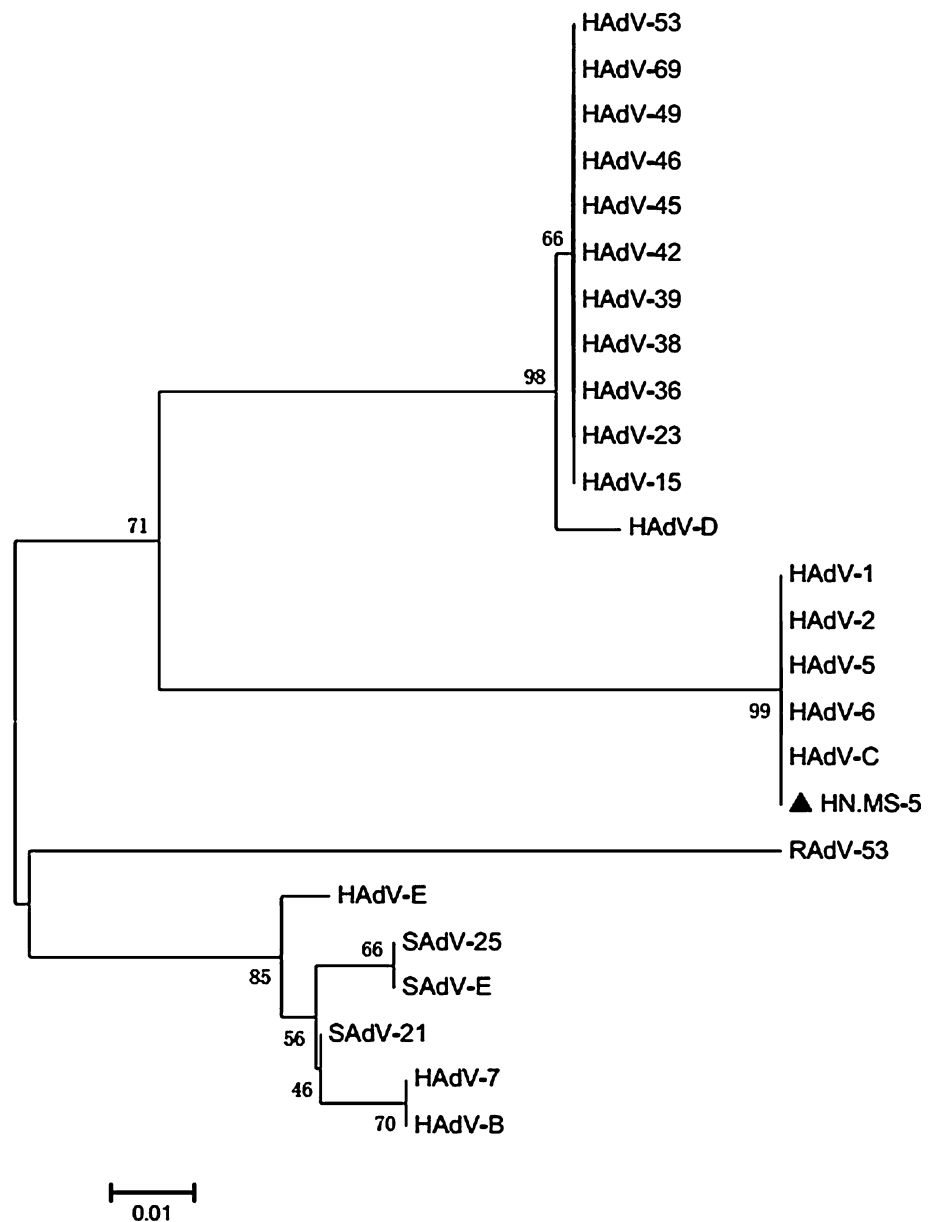


Table 3 Amino acid sequence identities (%) comparing the partial POL gene between bat ReVs and other established ReVs

ReVs in this study	Accession number	Host	Identity
XP_011385467/PvReV	XP_011385467	<i>Pteropus vampyrus</i>	85.7
XP_014308705/MluReV	XP_014308705	<i>Myotis lucifugus</i>	59.8
AFM52260/MlyReV	AFM52260	<i>Megaderma lyra</i>	56.7
XP_015426560/MdReV	XP_015426560	<i>Myotis davidii</i>	59.4
ABL74442/MrReV	ABL74442	<i>Myotis ricketti</i>	56.7
AGV92859/DIAV	AGV92859	Duck	57.4
AAF81697/FV	AAF81697	Fowl	56.7
AGV92853/GReV	AGV92853	Galidia	62.1
NP_955592/M-MuLV	NP_955592	Murine	60.1
AEI59728/X-MuLV	NP_040333	Murine	58.1
D10032/BaReV	D10032	Baboon	58.1
AAF15098/PcReV	AAF15098	<i>Phascolarctos cinereus</i>	58.1
AAC31805/MdEV	AAC31805	<i>Mus dunni</i>	57.4
CAC39617/PEReV-A	CAC39617	Porcine	57.4
ABL74442/PEReV-C	ABL74442	Porcine	57.4

PvRe: *Pteropus vampyrus* endogenous retrovirus; MluReV: *Myotis lucifugus* retrovirus; MlyReV: *Megaderma lyra* retrovirus; MrReV: *Myotis ricketti* retrovirus; MdReV: *Myotis davidii* retrovirus; DIAV: Duck infectious anemia virus; FV: Fowlpox virus; GReV: Galidia retrovirus; X-MuLV: Xenotropic murine leukemia virus; M-MuLV: Moloney murine leukemia virus; BaReV: Baboon endogenous retrovirus; PcReV: *Phascolarctos cinereus* retrovirus; PEReV-A: Porcine endogenous type A; PEReV-C: Porcine endogenous type C retrovirus class C; MdEV: *Mus dunni* endogenous virus

Fig. 3 Phylogenetic analysis of the IVa2 gene of AdVs detected in bats from southern China, based on a 102 amino acid sequence. The tree was generated using the neighbor-joining algorithm and the p-distance model. A bootstrap test of 1000 replicates was used. Numbers above the branches indicate NJ bootstrap values. Bold triangles indicate adenoviruses detected in the present study. MS: *Miniopterus schreibersi*; HN: Hainan province



of 100%. Additionally, HN.MS shared a sequence identity of 87.2–100% with human AdVs, compared with only 60.7–62.7% sequence identity with bat AdVs (Table 4). This indicates the putative evolution of a novel viral strain.

Prevalence and phylogenetic analysis of bat PVs (papillomaviruses)

In total, 2.2% (11/500) of fecal samples tested positive for PVs, of which two were found in *Scotophilus kuhlii* and nine in *Rhinolophus blythi* (E-Table 3).

A phylogenetic tree was constructed based on an L1 gene nucleotide sequence alignment of 47 PV types (representing multiple PV species/genera) in combination with the bat PVs detected herein. The phylogenetic tree constructed

Table 4 Amino acid sequence identities (%) for the partial IVa2 gene between bat AdVs and other established AdVs

AdVs in this study	Accession number	Host	Identity
HAdV-C	AGT75926	Human	100
HAdV-6	ACN88093	Human	100
HAdV-D	AGT77812	Human	88.2
HAdV-39	AFK92646	Human	88.2
HAdV-E	YP_068022	Human	87.2
SAdV-E	AAS10359	Simian	87.2
SAdV-21	AP_000265	Simian	88.2
HAdV-B	AFQ34416	Human	87.2
RAdV-53	YP_009110798	Rhesus	83.3
BAdV-TJM	YP_005271181	Bat	62.7
BAdV-2	YP_004782099	Bat	60.7

Fig. 4 Phylogenetic analysis of the L1 gene of papillomavirus detected in bats from southern China, based on a 392 nucleotide sequence. The phylogenetic tree was generated using the neighbor-joining algorithm with the p-distance or Maximum Composite Likelihood model. A bootstrap test of 1000 replicates was used. The numbers above the branches indicate the NJ bootstrap values. Bold triangles indicate the papillomavirus detected in this study. Bold squares indicated the papillomavirus detected in bats. RB: *Rhinolophus blythi*; SK: *Scotophilus kuhlii*; YF: Yunfu; HZ: Huizhou

by the neighbor-joining algorithm clustered the established PVs into 17 distinct, previously defined genera, ranging from *Alphapapillomavirus* to *Sigmampapillomavirus* [27, 28]. Within this phylogenetic tree, the 11 bat PVs from *Scotophilus kuhlii* and *Rhinolophus blythi* clustered in the same clade, constituting a monophyletic clade with viruses classified within the genus *Betapapillomavirus*. However, the bat PVs reported here actually appear to belong to an unassigned genus because the bootstrap support value for the closest node was less than 80% (Figure 4). Eleven bat PVs shared high sequence similarities in the L1 gene region. Notably, the similarities of the PVs from *Scotophilus kuhlii* and *Rhinolophus blythi* ranged from 94.5% to 100%. Likewise, low sequence similarity was found between the PVs reported here and previously detected bat PVs Ms-PVs-1 (42.1–43.6% similarity), MrPVs1 (45.8–47.3%) and RaPVs1 (47.7–50.0%). Even when compared to the most closely related clade (*Betapapillomavirus*), the bat PVs shared a sequence similarity of only 55.0–58.9%. The similarities, calculated by pairwise alignment of the corresponding amino acids, are provided in Table 5.

Prevalence and phylogenetic analysis of bat RVA (rotavirus A)

In total, 0.8% (4/500) of fecal samples tested positive for RVA strains, all of which were found in *Scotophilus kuhlii* (E-Table 3).

A phylogenetic tree was constructed based on a VP7 gene nucleotide sequence alignment of established genotypes, calculated using the neighbor-joining algorithm with the p-distance or Maximum Composite Likelihood algorithm. The four RVA strains detected herein clustered with mammalian RVAs, constituting a monophyletic clade among the genotype 3 strains; however, the bootstrap value of the node was only 50%, which failed to confirm whether the bat RVAs detected herein belong to G3 (Figure 5).

The putative VP7 genes of the four RVA strains detected in *Scotophilus kuhlii* were 847 bp in length and encoded a 282 amino acid protein. When compared with other mammalian rotaviruses, the VP7 nucleotide sequences of the four *Scotophilus kuhlii* isolates showed low levels of sequence identity to the established G3 genotypes (58.0–80.2%). Compared with human, bat, cat, dog

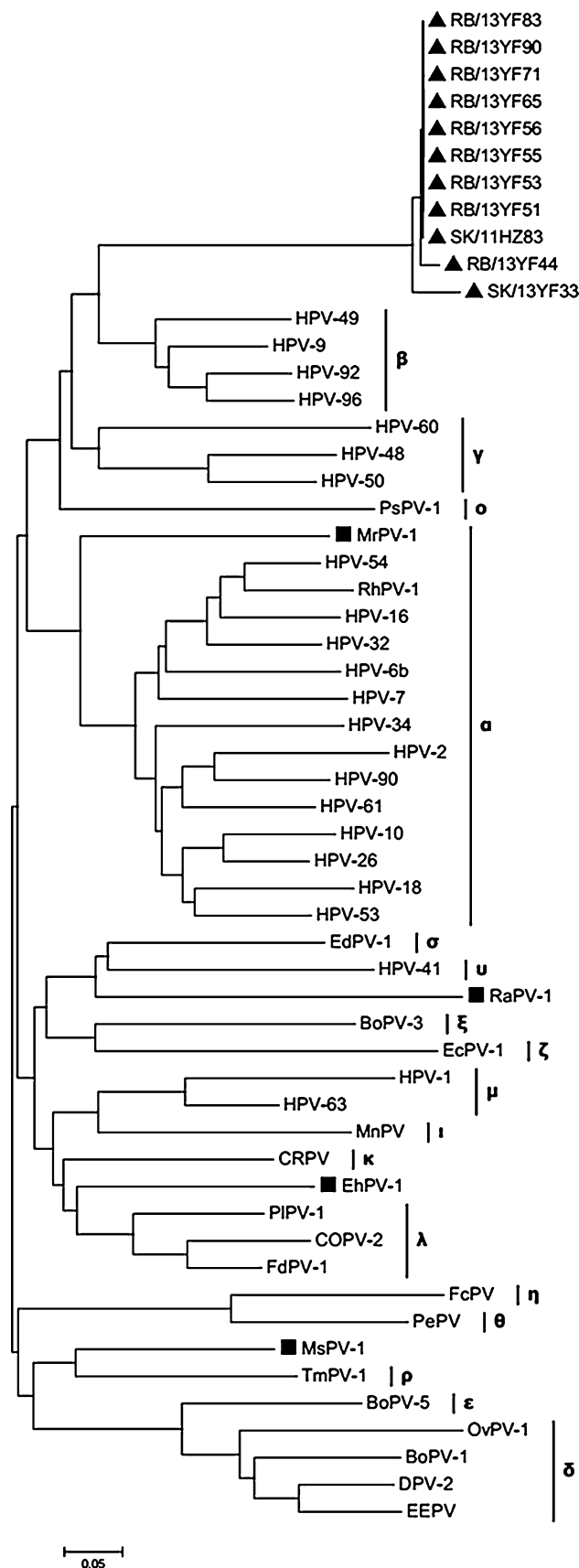


Table 5 Amino acid sequence identities (%) for the partial L1 gene between bat papillomaviruses identified in this study and other previously reported papillomaviruses

PV in this study	Betapapillomavirus				Gammapapillomavirus		
	HPV-9	HPV-49	HPV-92	HPV-96	HPV-60	HPV-48	HPV-50
SK/11HZ83	57.3	57.3	58.9	58.1	52.2	53.8	56.1
RB/13YF51	57.3	57.3	58.9	58.1	52.2	53.8	56.1
RB/13YF44	56.5	56.5	57.3	57.3	51.5	52.3	55.3
RB/13YF53	57.3	57.3	58.9	58.1	52.2	53.8	56.1
RB/13YF55	57.3	57.3	58.9	58.1	52.2	53.8	56.1
RB/13YF56	57.3	57.3	58.9	58.1	52.2	53.8	56.1
RB/13YF65	57.3	57.3	58.9	58.1	52.2	53.8	56.1
RB/13YF71	57.3	57.3	58.9	58.1	52.2	53.8	56.1
RB/13YF83	57.3	57.3	58.9	58.1	52.2	53.8	56.1
RB/13YF90	57.3	57.3	58.9	58.1	52.2	53.8	56.1
SK/13YF33	55.0	55.0	56.5	55.8	50.7	52.3	54.6

and simian G3 RVA strains, the bat RVAs reported herein had sequence identities of 77.1–79.5%, 79.4–79.6%, 79.3–79.4%, 78.3–80.2% and 77.6–80.0%, respectively (Table 6).

The bat VP7 sequences shared a maximum nucleotide identity of 80.0% with that of a simian RVA strain (RVA_simian-tc_USA_RRV_1975_G3P3), which is approaching the cut-off value (80.0%) for VP7 classification proposed by the Rotavirus Classification Working Group (RCWG). The amino acid identities of the VP7 antigenic regions: A (87–100), B (141–150) and C (208–224) within these bat RVAs, when compared with the G3 types was 85.7–92.8%, 80.0–100.0% and 82.3–88.2%, respectively, whereas these sequences exhibited identities of 50.5–85.7%, 50.0–100.0% and 64.7–88.2% to other G types, respectively (Table 6).

Prevalence and phylogenetic analysis of bat CVs and PiVs

No CVs or PiVs were found in the 500 fecal specimens studied.

Discussion

To determine the potential for zoonotic virus transmission from bats to humans, we collected bat samples from habitats mainly in residential areas, city parks, abandoned houses and mine caves, all of which were close to the living environment of humans. This study examined the virome composition and viral abundance in fecal and rectal samples from six bat species collected in four cities in southern China. This represents the first time, to our knowledge, a report on the viral diversity of *Cynopterus sphinx*, *Rousettus leschenaultii*,

Hipposideros larvatus and *Rhinolophus blythi* in southern China.

Virome analysis

Similarly to previous studies [5–8], a considerable number of viral sequence reads could be classified into virus families: plant, fungal, bacterial, insect, archaeal and algal in nature. The identification of insect viruses, fungal viruses and plant viruses presumably corresponds to the habits and habitats of the bats, including insect-eating, cave-dwelling and social structure. In addition, the identification of phages likely reflects the bacterial flora harbored inside the bats. Since we focused on the putative role of bats in zoonotic disease transmission, only mammalian virus contigs were analyzed further. Similarly to previous studies, most mammalian viruses were typical zoonotic viruses, including viruses classifiable as *Influenza virus A*, as well as members of the genera *Mastadenovirus*, *Rotavirus*, *Mamastrovirus*, *Coronavirus*, *Hantavirus*, *Rhabdovirus* and *Parechovirus*, and also members of the family *Papillomaviridae*. Metagenomic analyses in other studies have identified the full-length sequences of adenoviruses [8, 10, 12, 14], papillomaviruses [13, 14, 27, 28], parvoviruses [8, 13, 29], circoviruses [29], herpesviruses [11] and rotaviruses [12]. However, it is challenging to make specific or direct comparisons with our study due to significant methodological variations; for instance, the sample type (e.g., fresh guano, urine and roost guano), virus enrichment procedures (e.g., nuclease treatment, centrifugation and filtration) and amplification methods and/or high-throughput sequencing platforms [30]. The most important finding of our study was that the influenza A virus and coronavirus sequences we identified in bats shared their highest genetic identity with previous isolates from humans, suggesting possible cross-species transmission. However, contig sequences representing viruses classifiable

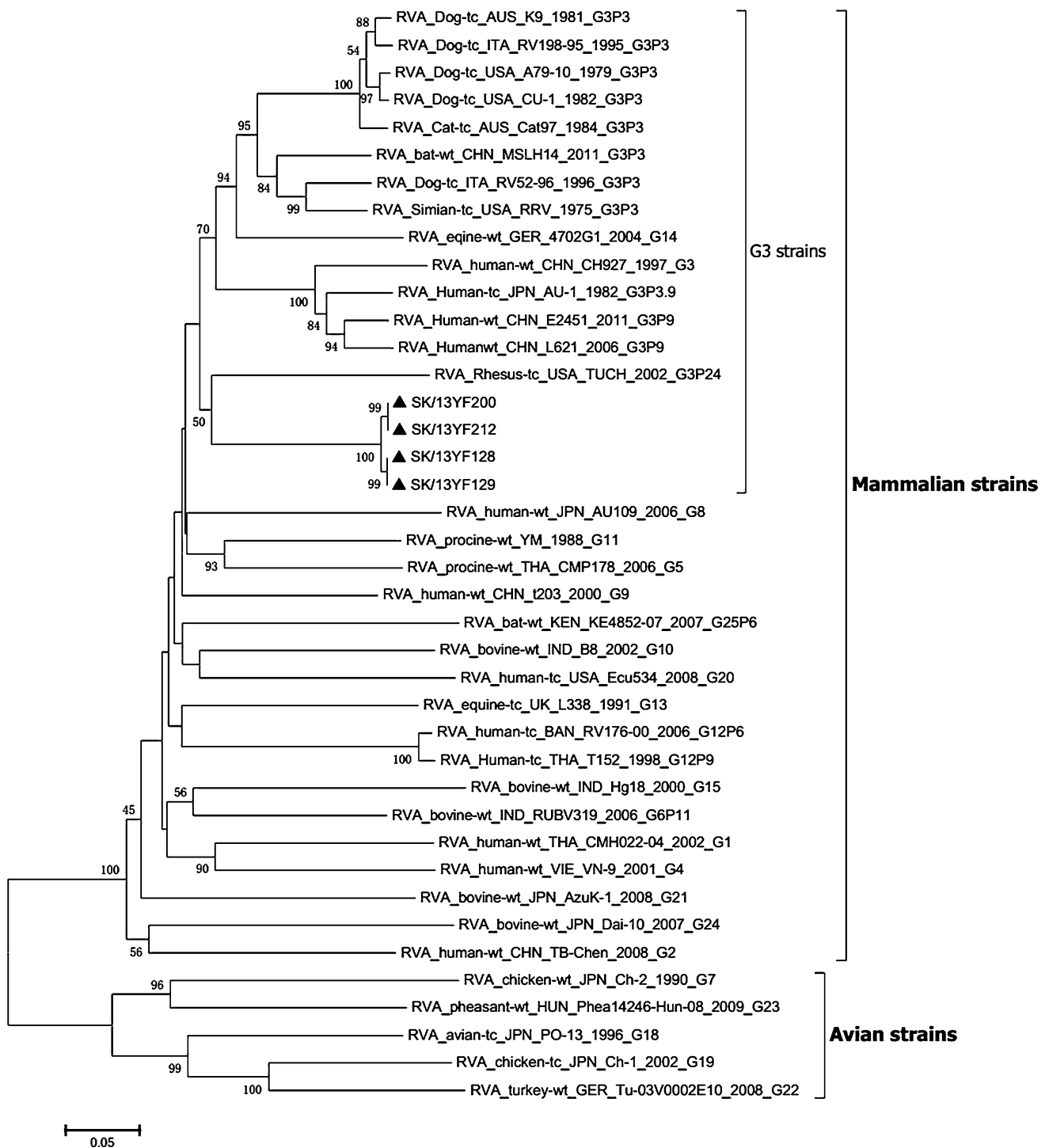


Fig. 5 Phylogenetic analysis of the major capsid region (vp7) of rotaviruses from bats in southern China, based on a 847 nucleotide sequence. The tree was generated using the neighbor-joining method with the p-distance or Maximum Composite Likelihood model. A bootstrap test of 1000 replicates was used. Numbers above the

branches indicate NJ bootstrap values. Bold triangles indicate rotaviruses detected in the present study. Abbreviations of virus names are shown in Figure 2, footnote. YF: Yunfu; SK: *Scotophilus kuhlii*; RB: *Rhinolophus blythi*

within the genera *Alphapapillomavirus*, *Betaretrovirus*, *Alpharetrovirus*, *Varicellovirus*, *Cyprinivirus*, *Chlorovirus* and *Cucumovirus* were also identified that had identities

of less than 70% when compared with viruses in existing databases, indicating the possible evolution of novel viral strains.

Table 6 Nucleotide identities (%) of partial Vp7 gene sequences between bat rotaviruses and other established rotaviruses

RVA strains	Genotype (% nucleotide sequence identity)			
	SK/13YF128	SK/13YF129	SK/13YF200	SK/13YF212
RVA_human-wt_THA_CMH022-04_2002_G1	72.40	72.40	72.30	72.30
RVA_human-wt_CHN_TB-Chen_2008_G2	70.00	70.00	70.30	70.30
RVA_human-wt_CHN_CH927_1997_G3	77.10	77.10	77.20	77.20
RVA_bat-wt_CHN_MSLH14_2011_G3P3	79.60	79.60	79.40	79.40
RVA_cat-tc_AUS_Cat97_1984_G3P3	79.40	79.40	79.30	79.30
RVA_dog-tc_ITA_RV198-95_1995_G3P3	78.50	78.50	78.30	78.30
RVA_dog-tc_AUS_K9_1981_G3P3	78.30	78.30	78.30	78.30
RVA_simian-tc_USA_RRV_1975_G3P3	80.00	80.00	80.00	80.00
RVA_dog-tc_ITA_RV52-96_1996_G3P3	80.20	80.20	80.10	80.10
RVA_dog-tc_USA_A79-10_1979_G3P3	78.30	78.30	78.30	78.30
RVA_human-tc_JPN_AU-1_1982_G3P3.9	78.90	78.90	79.00	79.00
RVA_humanwt_CHN_L621_2006_G3P9	79.30	79.30	79.30	79.30
RVA_rhesus-tc_USA_TUCH_2002_G3P24	77.40	77.40	77.60	77.60
RVA_procine-wt_THA_CMP178_2006_G5	76.70	76.70	76.60	76.60
RVA_bovine-wt_IND_RUBV319_2006_G6P11	75.40	75.40	75.60	75.60
RVA_chicken-wt_JPN_Ch-2_1990_G7	59.60	59.60	59.90	59.90
RVA_human-wt_JPN_AU109_2006_G8	74.40	74.40	74.60	74.60
RVA_bovine-wt_IND_B8_2002_G10	72.30	72.30	72.20	72.20
RVA_procine-wt_YM_1988_G11	77.20	77.20	77.00	77.00
RVA_equine-tc_UK_L338_1991_G13	75.60	75.60	75.60	75.60
RVA_eqine-wt_GER_4702G1_2004_G14	77.20	77.20	77.10	77.10
RVA_bovine-wt_IND_Hg18_2000_G15	70.90	70.90	71.00	71.00
RVA_avian-tc_JPN_PO-13_1996_G18	63.10	63.10	63.10	63.10
RVA_chicken-tc_JPN_Ch-1_2002_G19	58.80	58.80	58.90	58.90
RVA_human-tc_USA_Ecu534_2008_G20	74.10	74.10	74.10	74.10
RVA_turkey-wt_GER_Tu-03V0002E10_2008_G22	58.00	58.00	58.10	58.10
RVA_pheasant-wt_HUN_Phea14246-Hun-08_2009_G23	61.60	61.60	61.50	61.50
RVA_bovine-wt_JPN_Dai-10_2007_G24	68.10	68.10	68.10	68.10
RVA_bat-wt_KEN_KE4852-07_2007_G25P6	73.50	73.50	73.50	73.50

Similar to the findings of Yuan et al [15], retroviruses were one of the most commonly found viruses in bats. *Rousettus leschenaultii* from Hainan was found to harbor phylogenetically distant gammaretroviruses from those found in *Myotis lucifugus*, *Megaderma lyra*, *Myotis davidii* and *Myotis ricketti*, which suggested that bats might be suitable vectors for retrovirus transmission [31, 32]. Interestingly, we found that the AdV contigs from HN.RL (representing the IVa2 genes) were clustered with human adenovirus-C and shared higher amino acid sequence identity with human adenoviruses than previously reported bat AdVs. This might indicate the possible evolution of novel viral strains. Owing to the limited size and number of contigs generated by Illumina sequencing, conventional PCR assays are warranted to generate longer and more abundant sequences for further phylogenetic analysis [33, 34].

Comparable with previous studies [13, 15], except for retrovirus and AdV, analysis of more in-depth virus

sequencing information (including phylogeny and sequence identity) was challenging for the following reasons: 1) many reads could not be assembled into longer scaffolds because the NGS reads were randomly amplified throughout the whole viral genome; 2) incomplete DNase/RNase enzyme digestion resulting in a lack of hits for a considerable number of sequence reads (including those from cellular organisms) during raw data analysis; 3) some viruses may have been present in the samples in relatively small amounts below the sensitivity of NGS detection; 4) host genome and other non-target nucleic acids could have lowered the sensitivity of high-throughput sequencing and introduced background noise into the results.

Papillomaviruses

The L1 ORF is characterized as the most conserved region in the PV genome and is used for PV classification. According

to recent classification criteria: 1) PVs belong to different PV genera when they share less than 60% nucleotide sequence identity across the entire L1 ORF [28]; 2) the taxonomic status of PV types, subtypes and variants is based on the conventional criteria that the sequence of their L1 genes should be at least 10%, 2–10% and maximally 2% dissimilar from one another. Because the bat L1 sequences shared a maximum of 58.9% nucleotide identity with the L1 sequence of HPVs-92, these PVs cannot be placed in one of the existing genera. The bat PVs therefore represent tentative members of a novel, yet innominate, PV genus.

Our findings argue against the mainstream theory that PVs are highly species specific and co-evolve with their host. Compatible with the results of Garcia-Perez et al [35], this hypothesis of virus-host co-evolution is rejected by the existence of 11 closely related bat PVs in *Scotophilus kuhlii* and *Rhinolophus blythi* and by the lack of congruence between bats and bat PV phylogenies. Second, the possible inter-transmission of PVs to infect two different bat species (*Scotophilus kuhlii* and *Rhinolophus blythi*) argues against strict host specificity. However, the L1 sequence detected herein was only 450 bp in length, which does not represent the whole of the L1 gene. Further studies with additional genomic sequences are required to examine the potential pathogenic role of PVs in other bat species, which would provide insights into the ecology and evolution of PVs in bats. These findings might help address whether the prevailing paradigms regarding PV evolution should be reconsidered.

Rotavirus

The RVA VP7 glycoprotein (G genotypes) elicits neutralizing antibodies, and, to date, 27 G and 37 P genotypes from various hosts have been identified [36]. The nearly full-length VP7 sequences (847 nt) of four *Scotophilus kuhlii* rotavirus strains (SK/13YF128/129/200/212), encoding a 282-amino acid protein, were determined. These segments were most closely related to simian and dog RVA G3 strains (RVA_simian-tc_USA_RRV_1975_G3P3, RVA_dog-tc_ITA_RV52-96_1996_G3P3), sharing a similarity of 80.0% and 80.2%, respectively. Even though the similarity of these VP7 sequences did not exceed the appropriate cut-off (80.0%), the VP7 sequences detected herein represented more than 500 nt of the ORF (>50% of the ORF). Hence, the four bat RVA strains reported herein might be tentative G3 strains according to the RCWG classification [37].

G3P [3] and G3P [9] are genotype combinations commonly found in canines (RV198, RV52, A79-10, K9, CU-1, AUS and ITA) [38, 39]. Previous reports have identified canine RVA strains that infect humans, including the strains Ro1845, HCR3A, AU-1, E2451, L621, T152, CMH222, PA260, PAH136, PAI58, 6212, CMH120 and CU-1 [38–41].

In addition, some isolated animal RVA strains are believed to have a distant ancestor in common with canine RVA strains, when taking into account the majority of their gene segments [41]. Two distinct simian RVA strains (RRV and TUCH) have multiple genotypes more typical of canine-feline rotaviruses [40], suggesting potential transmission of canine-like RVA strains to simians. Although the genetic relationship of VP7 sequences between the four bat RVA strains and ITA/RRV is relatively distant, they might share a common ancestor. In light of the fact that the four bat RVA strains are not closely related to any known RVA strain, we speculate that these viruses are true bat RVA strains rather than viruses transmitted from other species.

The positive detection rate of RVA strains was low, which might have arisen for the following reasons: 1) we collected samples from adult bats which were immunocompetent and can clear viruses rapidly; 2) no neonates were involved, whereas RVA strains are known to infect mainly infants and young animals. To better understand the prevalence of RVAs in bats, further molecular and serological investigations with larger sample sizes and more species of adult and neonatal bats in different geographical regions are required.

Calicivirus and picornavirus

Notably, CVs and PiVs were not detected in the present study. The detection rates of CVs and PiVs in bats was also low in previous studies. In 2014, Kemenesi et al [17] found only three novel CVs in 447 bat fecal specimens in Europe, segregating with the *Hipposideros pomona* sapovirus genus found in 2012 [16]. In another study only 12 picornavirus-positive respiratory and alimentary specimens were identified from 1108 bats and 18 species [18]. The lack of CVs and PiVs detected in our study may be accounted for by low carrying rates for these two viruses. Other causes might be variations in duration of the virus shedding in feces, or geographical/seasonal/bat species differences.

Conclusion

In conclusion, this study describes the virome of six bat species in southern China and provides a more comprehensive understanding of virus ecology in common bat species found in human habitats. The detection of PVs in *Scotophilus kuhlii* and *Rhinolophus blythi* argues against the hypothesis that PVs have strict host specificity and co-evolution. The four bat RVA strains we isolated might be tentative G3 strains, according to the RCWG classification.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant no. 30972525). We thank Jian-peng Xiao for his insightful knowledge of the phylogenetic analysis.

Author contributions The contribution of each author is outlined briefly here: XZ, MQ, JL, SC and SH performed experiments and analyzed data; XZ and JL contributed bioinformatics analysis; ZC and YW identified bat species; XZ, QC and WG conceived and designed the experiments and wrote the paper.

Compliance with ethical standards

Conflict of interest No competing financial interests exist.

References

- Wilson DE, Reeder DM (2005) Mammal species of the world. A taxonomic and geographic reference, 3rd edn. Johns Hopkins University Press, Baltimore, pp 154–196
- Lau SK, Woo PC, Lai KK, Huang Y, Yip CC, Shek CT et al (2011) Complete genome analysis of three novel picornaviruses from diverse bat species. *J Virol* 85:8819–8828
- Li Y, Ge X, Zhang H, Zhou P, Zhu Y, Zhang Y et al (2010) Host range, prevalence, and genetic diversity of adenoviruses in bats. *J Virol* 84:3889–3897
- Wibbelt G, Kurth A, Yasmum N, Bannert M, Nagel S, Nitsche A et al (2007) Discovery of herpesviruses in bats. *J Gen Virol* 88:2651–2655
- Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH et al (2005) Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci USA* 102:14040–14045
- Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T (2006) Bats: important reservoir hosts of emerging viruses. *Clin Microbiol Rev* 19:531–545
- Victoria JG, Kapoor A, Dupuis K, Schnurr DP, Delwart EL (2008) Rapid identification of known and new RNA viruses from animal tissues. *PLoS Pathog* 4(9):e1000163
- Li L, Victoria JG, Wang C, Jones M, Fellers GM, Kunz TH et al (2010) Bat guano virome: predominance of dietary viruses from insects and plants plus novel mammalian viruses. *J Virol* 84(14):6955–6965
- Donaldson EF, Haskew AN, Gates JE, Huynh J, Moore CJ, Frieman MB (2010) Metagenomic analysis of the viromes of three North American bat species: viral diversity among different bat species that share a common habitat. *J Virol* 84(24):13004–13018
- Dacheux L, Cervantes-Gonzalez M, Guigon G, Thiberge JM, Vandebogaert M, Maufrais C et al (2014) Preliminary study of viral metagenomics of French bat species in contact with humans: identification of new mammalian viruses. *PLoS One* 9(1):e87194
- Wang J, Moore NE, Murray ZL, McInnes K, White DJ, Tompkins DM, Hall RJ (2015) Discovery of novel virus sequences in an isolated and threatened bat species, the New Zealand lesser short-tailed bat (*Mystacina tuberculata*). *J Gen Virol* 96(8):2442–2452
- He B, Li Z, Yang F et al (2013) Virome profiling of bats from Myanmar by metagenomic analysis of tissue samples reveals more novel mammalian viruses. *PLoS One* 8(4):e61950
- Ge X, Li Y, Yang X, Zhang H, Zhou P, Zhang Y, Shi Z (2012) Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* 86(8):4620–4630
- Wu Z, Ren X, Yang L, Hu Y, Yang J, He G et al (2012) Virome analysis for identification of novel mammalian viruses in bat species from Chinese provinces. *J Virol* 86(20):10999–11012
- Yuan L, Li M, Li L, Monagin C, Chmura AA, Schneider BS et al (2014) Evidence for retrovirus and paramyxovirus infection of multiple bat species in China. *Viruses* 6(5):2138–2154
- Tse H, Tsang AK, Tsoi HW, Leung AS, Ho CC, Lau SK et al (2012) Identification of a novel bat papillomavirus by metagenomics. *PLoS One* 7(8):e43986
- Kemenesi G, Dallos B, Görfö T et al (2014) Molecular survey of RNA viruses in Hungarian bats: discovering novel astroviruses, coronaviruses, and caliciviruses. *Vector Borne Zoonotic Dis* 14(12):846–855
- Lau SK, Woo PC, Lai KK et al (2011) Complete genome analysis of three novel picornaviruses from diverse bat species. *J Virol* 85(17):8819–8828
- He B, Yang F, Yang W, Zhang Y, Feng Y, Zhou J et al (2013) Characterization of a novel G3P[3] rotavirus isolated from a lesser horseshoe bat: a distant relative of feline/canine rotaviruses. *J Virol* 87(22):12357–12366
- Linacre A, Lee JC (2005) Species determination: the role and use of the cytochrome b gene. *Methods Mol Biol* 297:45–52
- Djikeng A, Halpin R, Kuzmickas R, Depasse J, Feldblyum J, Sengamalay N et al (2008) Viral genome sequencing by random priming methods. *BMC Genom* 7(9):5
- Li R, Yu C, Li Y, Lam TW, Yiu SM et al (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25:1966–1967
- Luo R, Liu B, Xie Y et al (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1(1):18
- Huson DH, Auch AF, Qi J, Schuster SC (2008) MEGAN analysis of metagenomic data. *Genome Res* 17:377–386
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H et al (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596–1599
- Rector A, Bossart GD, Ghim SJ, Sundberg JP, Jenson AB, Van Ranst M (2004) Characterization of a novel close-to-root papillomavirus from a Florida manatee by using multiply primed rolling-circle amplification: *Trichechus manatus* latirostris papillomavirus type 1. *J Virol* 78:12698–12702
- de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H (2004) Classification of papillomaviruses. *Virology* 324:17–27
- Lima FE, Cibulski SP, Dall Bello AG, Mayer FQ, Witt AA, Roehe PM et al (2015) A novel chiropteran circovirus genome recovered from a Brazilian insectivorous bat species. *Genome Announc* 25(6):e01393–15
- Daly GM, Bexfield N, Heaney J et al (2011) A viral discovery methodology for clinical biopsy samples utilising massively parallel next generation sequencing. *PLoS One* 6:e28879
- Chua KB, Cramer G, Hyatt A et al (2007) A previously unknown retrovirus of bat origin is associated with an acute respiratory disease in humans. *Proc Natl Acad Sci USA* 104:11424–11429
- Gifford RJ, Katzourakis A, Tristem M et al (2008) A transitional endogenous lentivirus from the genome of a basal primate and implications for lentivirus evolution. *Proc Natl Acad Sci USA* 105:20362–20367
- Drexler JF, Corman VM, Muller MA et al (2012) Bats host major mammalian paramyxoviruses. *Nat Commun* 24(3):796
- Baker KS, Leggett RM, Bexfield NH et al (2013) Metagenomic study of the viruses of African straw-colored fruit bats: detection of a chiropteran poxvirus and isolation of a novel adenovirus. *Virology* 441:95–106
- Garcia-Perez R, Ibáñez C, Godínez JM et al (2014) Novel papillomaviruses in free-ranging Iberian bats: no virus-host co-evolution,

- no strict host specificity, and hints for recombination. *Genome Biol Evol* 6(1):94–104
36. Matthijssens J, Ciarlet M, McDonald SM et al (2011) Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). *Arch Virol* 156:1397–1413
 37. Matthijssens J, Ciarlet M, Rahman M et al (2008) Recommendations for the classification of group A rotaviruses using all 11 genomic RNA segments. *Arch Virol* 153(8):1621–1629
 38. Matthijssens J, De Grazia S, Piessens J et al (2011) Multiple reassortment and interspecies transmission events contribute to the diversity of feline, canine and feline/canine-like human group A rotavirus strains. *Infect Genet Evol* 11:1396–1406
 39. Tsugawa T, Hoshino Y (2008) Whole genome sequence and phylogenetic analyses reveal human rotavirus G3P[3] strains Ro1845 and HCR3A are examples of direct virion transmission of canine/feline rotaviruses to humans. *Virology* 380:344–353
 40. Wang YH, Pang BB, Zhou X et al (2013) Complex evolutionary patterns of two rare human G3P[9] rotavirus strains possessing a feline/canine-like H6 genotype on an AU-1-like genotype constellation. *Infect Genet Evol* 16:103–112
 41. Khamrin P, Maneekarn N, Peerakome S et al (2006) Molecular characterization of a rare G3P[3] human rotavirus reassortant strain reveals evidence for multiple human-animal interspecies transmissions. *J Med Virol* 78:986–994