Discovery of a Highly Divergent Coronavirus in the Asian House Shrew from China Illuminates the Origin of the Alphacoronaviruses

Wen Wang,^a Xian-Dan Lin,^b Yong Liao,^c Xiao-Qing Guan,^a Wen-Ping Guo,^a Jian-Guang Xing,^d Edward C. Holmes,^e Yong-Zhen Zhang^a

State Key Laboratory for Infectious Disease Prevention and Control, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Department of Zoonoses, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, Beijing, China^a; Wenzhou Center for Disease Control and Prevention, Wenzhou, Zhejiang Province, China^b; Ganzhou Center for Disease Control and Prevention, Ganzhou, Jiangxi Province, China^c; Wencheng Center for Disease Control and Prevention, Wencheng, Zhejiang Province, China^d; Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, University of Sydney, Sydney, New South Wales, Australia^e

ABSTRACT Although shrews are one of the largest groups of mammals, little is known about their role in the evolution and transmission of viral pathogens, including coronaviruses (CoVs). We captured 266 Asian house shrews (*Suncus murinus*) in Jiangxi and Zhejiang Provinces, China, during 2013 to 2015. CoV RNA was detected in 24 Asian house shrews, with an overall prevalence of 9.02%. Complete viral genome sequences were successfully recovered from the RNA-positive samples. The newly discovered shrew CoV fell into four lineages reflecting their geographic origins, indicative of largely allopatric evolution. Notably, these viruses were most closely related to alphacoronaviruses but sufficiently divergent that they should be considered a novel member of the genus *Alphacoronavirus*, which we denote Wénchéng shrew virus (WESV). Phylogenetic analysis revealed that WESV was a highly divergent member of the alphacoronaviruses and, more dramatically, that the S gene of WESV fell in a cluster that was genetically distinct from that of known coronaviruses. The divergent position of WESV suggests that coronaviruses have a long association with Asian house shrews. In addition, the genome of WESV contains a distinct NS7 gene that exhibits no sequence similarity to genes of any known viruses. Together, these data suggest that shrews are natural reservoirs for coronaviruses and may have played an important and long-term role in CoV evolution.

IMPORTANCE The subfamily *Coronavirinae* contains several notorious human and animal pathogens, including severe acute respiratory syndrome coronavirus, Middle East respiratory syndrome coronavirus, and porcine epidemic diarrhea virus. Because of their genetic diversity and phylogenetic relationships, it has been proposed that the alphacoronaviruses likely have their ultimate ancestry in the viruses residing in bats. Here, we describe a novel alphacoronavirus (Wénchéng shrew virus [WESV]) that was sampled from Asian house shrews in China. Notably, WESV is a highly divergent member of the alphacoronaviruses and possesses an S gene that is genetically distinct from those of all known coronaviruses. In addition, the genome of WESV contains a distinct NS7 gene that exhibits no sequence similarity to those of any known viruses. Together, these data suggest that shrews are important and longstanding hosts for coronaviruses that merit additional research and surveillance.

KEYWORDS coronavirus, alphacoronavirus, Asian house shrew, evolution, phylogeny, recombination

Received 5 May 2017 Accepted 13 June 2017

Accepted manuscript posted online 21 June 2017

Citation Wang W, Lin X-D, Liao Y, Guan X-Q, Guo W-P, Xing J-G, Holmes EC, Zhang Y-Z. 2017. Discovery of a highly divergent coronavirus in the Asian house shrew from China illuminates the origin of the alphacoronaviruses. *J Virol* 91:e00764-17. <https://doi.org/10.1128/JVI.00764-17>.

Editor Julie K. Pfeiffer, University of Texas Southwestern Medical Center

Copyright © 2017 American Society for Microbiology. All Rights Reserved.

Address correspondence to Yong-Zhen Zhang, zhangyongzhen@icdc.cn.

W.W., X.-D.L., and Y.L. contributed equally to this work.

Most emerging infectious diseases described recently are due to previously unknown zoonotic pathogens (1, 2), particularly rapidly evolving RNA viruses that frequently jump species boundaries (3–7). In addition to their rapid evolution, ongoing changes in the natural environment and in the behavior of their hosts have facilitated the emergence of viral diseases by providing new ecological niches (8–11). Such a process of disease emergence is predicted to occur with increased frequency as humans continually change their interactions with the animal world.

Coronaviruses (CoVs) (subfamily *Coronavirinae*, family *Coronaviridae*, order *Nidovirales*) are single-stranded positive-sense RNA viruses and produce enveloped virions (12). Their genomes (26 to 32 kb) contain six open reading frames (ORFs) that are conserved across the subfamily and arranged in the order 5′-replicase ORF1ab-spike (S)-envelope (E)-membrane (M)-nucleocapsid (N)-3′ (12). The replicase gene ORF1ab encodes 16 nonstructural proteins (termed nsp1 to nsp16). On the basis of phylogeny and pairwise evolutionary distances in the conserved domains of the replicase polyprotein, the currently known coronaviruses are classified into 30 species within four genera: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* (13, <http://ictv.global/report>). These viruses can infect humans, other mammals, and birds, causing respiratory, enteric, hepatic, and neurological diseases of varying severity (12). More importantly, the pandemic of severe acute respiratory syndrome (SARS) that occurred during 2002 to 2003 (5) and the subsequent emergence of the Middle East respiratory syndrome (MERS) in 2012 (14), both of which were caused by previously unknown coronaviruses, remind us that these viruses will likely remain a considerable challenge to public health for the foreseeable future. In addition, the discovery of SARS-like CoV in Himalayan palm civets (15) and bats (16, 17) highlights the essential role that mammalian species play in coronavirus evolution and has heightened interest in documenting novel coronaviruses in animals and humans on a global scale.

All known alphacoronaviruses form a monophyletic group within the subfamily *Coronavirinae* (13). Two genetic features set them apart from other coronaviruses: (i) a unique type of nsp1, distinct in size and sequence from the betacoronavirus nsp1 and with no apparent counterpart in gammacoronaviruses and deltacoronaviruses, and (ii) the presence of a commonly shared accessory gene for a dispensable multispanning alphacoronavirus membrane protein (α mp) (13). At present, the genus *Alphacoronavirus* includes 11 species (<http://ictv.global/report>) and some tentative species (13, 18–20). These virus species have been sampled from bats, as well as a variety of other mammals, including humans. On the basis of their diversity and phylogeny, it has been proposed that the alphacoronaviruses likely have their ultimate ancestry in bats (21, 22). However, the recent discovery of Lucheng Rn rat coronavirus (LRNV) in a brown rat (*Rattus norvegicus*) sampled from China suggests that the evolutionary history of these viruses is more complex than previously thought (18). Indeed, as RNA viruses likely exist in every species of cellular life (23, 24), our current knowledge of the origins and evolutionary history of alphacoronaviruses from such sparse sampling is likely to be biased.

Shrews (Mammalia: Eulipotyphla: Soricidae) are small mole-like mammals that are broadly distributed globally. The shrew family is the fourth largest among mammals, comprising approximately 376 species (25). As the former name of the Eulipotyphla (i.e., Insectivora) implies, insects make up a large portion of the typical shrew diet. Our recent studies have revealed a remarkable diversity of viruses in invertebrates, especially in arthropods (24, 26). Additionally, the discovery of distinct nidoviruses in insects suggests that coronaviruses may have an invertebrate origin (27, 28). Importantly, multiple viruses (e.g., arenavirus, hantaviruses, and rotavirus) have also been identified in insect-eating shrews over the past decade (29–31). Hence, like bats, shrews may play an important role in the evolution and transmission of viruses, including coronaviruses, among animals, or from animals to humans. In this study, we tested shrew samples collected in the Jiangxi and Zhejiang Provinces of China for the presence of coronaviruses. Based on the discovery of a distinct shrew virus, we explore the origin and evolution of alphacoronaviruses as a whole.

FIG 1 Locations (red circles) within China of trap sites in which shrews were captured.

RESULTS

CoV identification in Asian house shrews. During 2013 to 2015, a total of 266 Asian house shrews were captured in Zhejiang (214) and Jiangxi (52) Provinces, China (Fig. 1). Species identification was based on morphological identification and amplification and subsequent sequencing of the mitochondrial cytochrome *b* (*mt-cyt b*) gene (3). Reverse transcription (RT)-PCR targeting a 440-bp fragment of the viral RdRp (RNA-dependent RNA polymerase) gene was performed to detect CoV RNA, as described previously (18, 19). Viral RNA was identified in a total of 24 shrews, with an overall detection rate of 9.02%. The detection rate was 8.7% (2/23) in Ruian, 12.4% (12/97) in Wencheng, 10% (4/40) in Yudu, and 50% (6/12) in Xingguo. However, no CoV was detected in 94 Asian house shrews from Longwan. Genetic analysis revealed that these viruses were closely related to each other, with 87.8 to 100% nucleotide similarity in the RdRp gene, and were generally most closely related to members of the genus *Alphacoronavirus* in the RdRp gene (65.6 to 72.8% nucleotide similarity). However, they exhibited more than 35.3% nucleotide difference from known alphacoronaviruses, suggesting that a novel CoV circulates in Asian house shrews. Finally, although rodents were also captured from the same geographic regions, no similar CoV was identified in those animals (data not shown).

Genomic features of the newly discovered shrew virus. Since the newly discovered shrew CoV might represent a novel member of the genus *Alphacoronavirus*, seven complete genome sequences were recovered from the viral-RNA-positive samples

TABLE 1 Key features of WESV strains with complete or nearly complete genome sequences

Strain	Genome size	Gender of host ^a	Sampling yr	Sampling location
Wénchéng-554	26,028 nt	M	2014	Wencheng
Wénchéng-562	26,028 nt	F	2014	Wencheng
Wénchéng-578	26,028 nt	F	2014	Wencheng
Ruǎn-90	26,042 nt	M	2014	Ruian
Ruǎn-133	26,041 nt	F	2014	Ruian
Yúdū-76	26,002 nt	M	2014	Yudu
Yúdū-19	26,031 nt	M	2015	Yudu
Xīngguó-101 ^b	25,995 nt	M	2015	Xingguo
Xīngguó-74 ^b	25,984 bp	M	2015	Xingguo

^aM, male; F, female.^bStrain with nearly complete genome sequence.

collected in Wencheng (strains Wénchéng-554, Wénchéng-562, and Wénchéng-578), Ruian (Ruǎn-90 and Ruǎn-133), and Yudu (Yúdū-76 and Yúdū-19), as well as two nearly complete genome sequences (Xīngguó-74 and Xīngguó-101) from Xingguo. The key features of these CoV sequences are described in Tables 1 and 2 and Fig. 2. Genetic analysis revealed that the nucleotide similarities among these viruses were 88.2 to 99.9%. Generally, they shared 48.7 to 55.1% nucleotide similarity with known alphacoronaviruses and less than 57.1% nucleotide similarity with other coronaviruses. Further comparison of the replicase domains (i.e., ADP-ribose 1"-phosphatase [ADRP], chymotrypsin-like protease [3CLpro], RdRp, helicase [Hel], 3'-to-5' exonuclease [ExoN], nidoviral endoribonuclease specific for uridylate [NendoU], and ribose-2'-O-methyltransferase [O-MT]) revealed more than 29.2% amino acid differences between the newly discovered shrew viruses and known alphacoronaviruses (see Table S1 in the supplemental material). In addition, all phylogenetic analyses were consistent in showing that the newly discovered shrew viruses were distinct from the known alphacoronaviruses (see below). Therefore, these shrew viruses represent a novel member of the genus *Alphacoronavirus*: we have termed this Wénchéng shrew virus (WESV) based on its host and the location of its first identification.

TABLE 2 Coding potential and putative transcription-regulatory sequences of the Wénchéng-562, Ruǎn-90, and Yúdū-76 viruses

Coronavirus	ORF	Location (nt)	Length (nt)	Length (aa)	TRS location	TRS ^a
Wénchéng-562	ORF1ab	266–19233 (shift at 11239)	18,968	6,322	72–77	CUAAAC(188)AUG
	S	19240–22644	3,405	1,134	19233–19238	AACUAA(1)AUG
	NS3	22644–23357	714	237		
	E	23338–23565	228	75	23313–23318	CUAAAC(19)AUG
	M	23578–24267	690	229	23569–23574	CUAAAC(3)AUG
	N	24271–25368	1,098	365	24264–24269	CUAAAC(1)AUG
	NS7	25355–25762	408	135		
Ruǎn-90	ORF1ab	265–19241 (shift at 11247)	18,977	6,325	71–76	CUAAAC(188)AUG
	S	19248–22652	3,405	1,134	19241–19246	AACUAA(1) AUG
	NS3	22652–23365	714	237		
	E	23346–23573	228	75	23321–23326	CUAAAC(19)AUG
	M	23586–24275	690	229	23577–23582	CUAAAC(3)AUG
	N	24279–25379	1,101	366	24272–24277	CUAAAC(1)AUG
	NS7	25366–25773	408	135		
Yúdū-76	ORF1ab	266–19200 (shift at 11206)	18,935	6,311	72–77	CUAAAC(188)AUG
	S	19207–22614	3,408	1,135	19200–19205	AACUAA(1) AUG
	NS3	22614–23327	714	237		
	E	23308–23535	228	75	23283–23288	CUAAAC(19)AUG
	M	23548–24237	690	229	23539–23544	CUAAAC(3)AUG
	N	24241–25341	1,101	366	24234–24239	CUAAAC(1)AUG
	NS7	25328–25735	408	135		

^aNumbers in parentheses represent the number of nucleotides to the putative start codon.

FIG 2 Schematic of the annotated WESV genome in comparison to those of representative alphacoronaviruses.

Excluding the polyadenylated tail at the 3' terminus, the genomes of this novel virus varied from 25,986 to 26,026 nucleotides (nt), with a lower G+C content (31.53 to 31.97%) than that of known alphacoronaviruses (34.46 to 42.02%). The genome organization of WESV was similar to that of other alphacoronaviruses (Fig. 2), showing the characteristic gene order: 5'-replicase ORF1ab-spike (S)-envelope (E)-membrane (M)-nucleocapsid (N)-3'. Remarkably, two additional ORFs coding for nonstructural proteins (NSPs) NS3 and NS7 were identified (Fig. 2). In addition, a putative transcription regulatory sequence (TRS) motif (5'-CUAAAC-3'), similar to that in other alphacoronaviruses, was documented at the 3' end of the leader sequence and preceded each ORF except the S, NS3, and NS7 genes. An alternative TRS motif (5'-AACUAA-3') was discovered preceding the S gene in the shrew CoV genomes (Table 2). Finally, the putative mature NSPs within ORF1ab encoding the replicase were calculated based on the cleavage and recognition pattern of the chymotrypsin-like protease (3CLpro) and papain-like proteinase (PLpro).

Like other alphacoronaviruses, the S protein of WESV was predicted to be a type I membrane glycoprotein, with most of the protein (residues 16 to 1080 or 16 to 1081) exposed on the outside of the virus. A transmembrane domain was located at residues 1081 to 1103 or 1082 to 1104 at the C terminus. However, WESV shared only 20.1 to 37.7% amino acid identity in the S protein with other members of the genus *Alphacoronavirus* and 20.0 to 25.0% amino acid identity with coronaviruses of the remaining genera but 34% amino acid identity with LRNV, which was sampled in rats collected from Lucheng district (a geographic neighbor of Wencheng and Ruian) of Wenzhou City (18), and two bat viruses (Rhinolophus bat coronavirus HKU2 and BtRf-AlphaCoV/YN2012) also sampled in China (32) (GenBank accession no. [NC_028824](#)).

ORF NS3 encodes a putative 237-amino-acid (aa) nonstructural protein that is located between the S and E genes of WESV. Although the NS3 genes within the same geographic region were closely related to each other (96.2 to 100%, 100%, 97.9%, and 98.7% amino acid identities for the Wencheng, Ruian, Yudu, and Xingguo strains, respectively), the difference among the WESVs from different regions reached 23.5% (Table 3). TMHMM

TABLE 3 Comparison of NS3 genes between WESV and alphacoronaviruses

No.	Virus	Size (bp)	% identity with virus no.ª:													
			1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Xingguó-101	714		99.6	89.9	90.2	91.5	91.5	80.7	80.3	80.8	43.9	43.9	46.4	39.2	39.5
2	Xingguó-74	714	98.7		89.8	90.1	91.3	91.3	80.5	80.3	80.7	43.8	44.1	46.4	38.9	39.8
3	Yúdu-76	714	89.9	89.9		98.3	93.4	93.4	80.3	79.4	80.4	43.9	42.9	46.3	40.2	39.8
4	Yúdu-19	714	90.3	90.3	97.9		93.7	93.7	80.5	80.0	80.7	43.8	42.9	46.4	40.2	39.7
5	Ruian-133	714	90.8	90.8	94.5	94.1		100	81.0	79.8	81.1	43.8	43.7	47.2	40.6	41.0
6	Ruian-90	714	90.8	90.8	94.5	94.1	100.0		81.0	79.8	81.1	43.8	43.7	47.2	40.6	41.0
7	Wénchéng-554	714	79.0	78.6	76.5	76.9	79.0	79.0		96.6	99.9	44.7	42.8	45.2	40.2	40.1
8	Wénchéng-578	714	77.7	77.3	75.2	76.5	77.3	77.3	96.2		96.5	44.7	42.6	45.1	40.6	39.4
9	Wénchéng-562	714	79.0	78.6	76.5	76.9	79.0	79.0	100.0	96.2		44.5	42.6	45.2	40.2	40.1
10	BatCoV HKU2	690	20.3	20.3	19.4	18.9	21.1	21.1	19.8	19.4	19.8		53.0	53.6	50.7	36.5
11	Lucheng-19	645	23.3	23.3	21.4	21.4	23.3	23.3	21.9	21.4	21.9	31.6		49.3	46.9	36.8
12	HCoV-NL63	678	22.7	22.7	21.8	21.3	23.6	23.6	22.2	22.2	22.2	41.8	33.2		47.3	44.3
13	PEDV	675	19.3	19.3	19.7	19.3	21.1	21.1	20.2	21.1	20.2	35.1	29.4	34.8		33.9
14	BatCoV HKU9	663	13.6	13.6	14.1	13.2	13.6	13.6	13.2	12.3	13.2	11.6	10.6	8.5	9.5	

ªPercent identities for nucleotide (above the diagonal) and amino acid (below the diagonal) sequences.

analysis (see below) revealed there were two putative transmembrane domains in WESV NS3, at residues 53 to 70 and 90 to 112 of the Wencheng strains, at residues 49 to 71 and 91 to 113 in the Ruian and Yudu strains, and at residues 53 to 70 and 91 to 113 for the Xingguo strains. In addition, the NS3 gene of the WESV strains was longer than that of other alphacoronaviruses and distinct from those of known alphacoronaviruses and betacoronaviruses.

One of the most striking genomic features was the presence of an NS7 gene encoding a putative nonstructural protein of 136 amino acid residues located downstream of the N protein (Fig. 2). Notably, at the amino acid level, the NS7 gene did not show homology to any known genes in GenBank. Additionally, although an ORF (or ORFs) downstream of the N gene was also reported in the genomes of some alphacoronaviruses, including BtKYNL63-9a, HKU8, transmissible gastroenteritis virus (TGEV), porcine respiratory coronavirus (PRCV), HKU2, and BtCoV/512/2005, there was no sequence similarity in NS7 between WESV and these CoVs, indicative of markedly different origins.

Phylogenetic relationship between WESV and known coronaviruses. To better understand the evolutionary relationship between WESV and other members of the genus *Alphacoronavirus*, we estimated phylogenetic trees based on the amino acid sequences of the nonstructural and structural genes (Fig. 3 to 5). In the RdRp tree (Fig. 3A and B), WESV formed a distinct cluster that was separated from the other alphacoronaviruses by a relatively long branch. The WESV strains clearly clustered according to their geographic origins, indicative of the *in situ* evolution of WESVs in shrews (Fig. 3C). However, although the Ruian and Wencheng strains were both sampled in Wenzhou, the Ruian strains were more closely related to those sampled from Ganzhou city (Jiangxi Province) than to those from Wencheng.

A similar clustering pattern was observed in the trees estimated using the amino acid sequences of the nonstructural genes (Fig. 4) and the structural gene N (Fig. 4). Even more striking was the phylogenetic tree of the S gene (Fig. 5), in which WESV formed a divergent cluster with LRNV, HKU2, and BtRf-AlphaCoV/YN2012 that was genetically distinct not only from the genus *Alphacoronavirus*, but also from the other genera of coronaviruses, so these are clearly genetically distinct members of the subfamily *Coronavirinae*. Within this cluster, the rat virus and two bat viruses shared a common ancestry, with the WESVs again forming a distinct cluster.

Coronavirus recombination. We performed recombination analyses of the genomes of the Wencheng, Ruian, Yudu, and Xingguo strains using the Recombination Detection Program, version 4 (RDP4) (33). Multiple methods statistically supported a significant recombination event in Wénchéng-578. From the similarity plot, two recombination breakpoints at bp 5248 and 7663 of the sequence alignment (with reference

FIG 3 Maximum-likelihood phylogenetic trees of the amino acid sequences of the putative RdRp protein. (A) WESV and other coronaviruses. (B) WESV and other alphacoronaviruses. (C) WESV only. The asterisks indicate well-supported nodes (>70% bootstrap support). The scale bars indicate the numbers of amino acid substitutions per site. The virus genomes used in this study and their GenBank accession numbers are as follows: AlpacaCoV, alpaca respiratory coronavirus isolate CA08-1/2008 (JQ410000); BatCoV CDPHE15, bat coronavirus CDPHE15/USA/2006 (KF430219); BatCoV FJ2012, BtMf-AlphaCoV/FJ2012 (KJ473799); BatCoV YN2012, BtRf-AlphaCoV/YN2012 (KJ473808); BatCoV HuB2013, BtRf-AlphaCoV/HuB2013 (KJ473807); CamelCoV, camel alphacoronavirus isolate camel/Riyadh/Ry141/2015 (KT368907); CCoV K378, canine coronavirus strain K378 (KC175340); FCoV C1Je, feline coronavirus strain FCoV C1Je (DQ848678); BatCoV HKU2, bat coronavirus HKU2 strain HKU2/GD/430/2006 (EF203064); BatCoV HKU8, bat coronavirus HKU8 strain AFCD77 (EU420139); HCoV-229E, human coronavirus 229E (AF304460); HCoV-NL63, human coronavirus NL63 (AY567487); BatCoV JTAC2, bat coronavirus JTAC2 (KU182966); LrNV, Lucheng Rn rat coronavirus isolate Lucheng-19 (KF294380); BatCoV 1A, bat coronavirus 1A strain AFCD62 (EU420138); MCoV, mink coronavirus strain WD1127 (HM245925); PEDV, porcine epidemic diarrhea virus isolate ZJU/G1/2013 (KU664503); BatCoV HKU10, Roussetus bat coronavirus HKU10 isolate 183A (JQ989270); BatCoV SAX2011, BtMr-AlphaCoV/SAX2011 (KJ473806); BtCoV/512/2005, Scotophilus bat coronavirus 512 (DQ648858); TGEV, transmissible gastroenteritis virus virulent Purdue (DQ811789); BatCoV Zhejiang2013, bat Hp-betacoronavirus/Zhejiang2013 (KF636752); MERS-CoV, human betacoronavirus 2c EMC/2012 (JX869059); HCoV-HKU1, human coronavirus HKU1 (AY597011); BatCoV HKU9, bat coronavirus HKU9 (EF065513); SARS-CoV, SARS coronavirus WH20 (AY772062); BuCoV HKU11, bulbul coronavirus HKU11-934 (FJ376619); PorCoV HKU15, porcine coronavirus HKU15 strain HKU15-44 (JQ065042); MRCov HKU18, magpie-robin coronavirus HKU18 strain HKU18-chu3 (JQ065046); WiCoV HKU20, wigeon coronavirus HKU20 strain HKU20-9243 (JQ065048); AIBV-Beaudette, avian infectious bronchitis virus Beaudette (NC_001451); DKCoV, duck coronavirus isolate DK/CH/HN/ZZ2004 (JF705860); BWCoV SW1, beluga whale coronavirus SW1 (EU111742); TCoV, turkey coronavirus isolate TCoV-ATCC (EU022526).

jvi.asm.org 8

FIG 4 (Continued)

similarity to Ruiān-90 and -133 as opposed to 88.0% sequence similarity to Wénchéng-554 and -562; in contrast, in parental region B, they are more closely related to Wénchéng-554 and -562 (97.7 to 97.8% similarity) than to Ruiān-90 and -133 (89.1%). This recombination event was confirmed by phylogenetic analyses of the different parental regions and with high bootstrap values (Fig. 6B).

FIG 5 Maximum-likelihood phylogenetic tree of the amino acid sequences of the putative S proteins of WESV and other coronaviruses. The asterisks indicate well-supported nodes (>70% bootstrap support). The scale bars indicate the numbers of amino acid substitutions per site. The virus genomes used are the same as those shown in Fig. 3.

Although readily apparent in the amino acid phylogenies, the recombination event between WESV and other (and/or unknown) coronaviruses did not receive significant statistical support in the RDP analysis and similarity plot analysis (Fig. 6C), likely because these nucleotide sequences are highly divergent (for example, the S gene of WESVs differs from those of alphacoronaviruses by 26.6 to 62.6% at the nucleotide level). Similar suggestions have been made with respect to the recombination involving *Rhinolophus* bat coronavirus HKU2 and Lucheng Rn rat coronavirus (18, 32).

Numbers of synonymous and nonsynonymous substitutions across the WESV genome. An analysis of the numbers of synonymous (dS) and nonsynonymous (dN) substitutions per site (dN/dS ratio) in the genome sequences of WESV and other alphacoronaviruses revealed relatively low dN/dS values, reflecting the predominance of purifying selection (Table 4). The exception was NS7, in which the far higher supported ratio for WESV (0.514) was indicative of a markedly different selection pressure.

DISCUSSION

We describe a novel coronavirus, denoted Wénchéng shrew coronavirus (WESV), in shrews in four counties of Jiangxi and Zhejiang Provinces, China. WESV was highly divergent from other alphacoronaviruses, exhibiting $\leq 71.1\%$ amino acid similarity to any known members of the genus *Alphacoronavirus* in the coronavirus-wide conserved domains in the replicase polyprotein pp1ab and less than 61.3% amino acid similarity to the other three coronavirus genera. The Coronaviridae Study Group of the Interna-

FIG 6 Recombination analysis of the WESV genome. (A) Sequence similarity plot revealing two recombination breakpoints with their locations shown by the red numbers on the x axis. The plot shows genome scale similarity comparisons of the Wénchéng-578 sequence (query) against Wénchéng-554 and -562 (parental group 1; red) and Ruiān-90 and -133 (parental group 2; blue). The (Continued on next page)

TABLE 4 Comparison of the mean numbers of nonsynonymous and synonymous substitutions per site, and their ratios in the coding regions of WESV, BatCoV HKU2, PEDV, and HCoV-NL63

Gene	WESV (n = 9)			BatCoV HKU2 (n = 5)			PEDV (n = 7)			HCoV-NL63 (n = 6)		
	dN	dS	dN/dS	dN	dS	dN/dS	dN	dS	dN/dS	dN	dS	dN/dS
nsp1	0.090	0.418	0.215	0.014	0.085	0.165	0.012	0.026	0.462	0.006	0.031	0.194
nsp2	0.075	0.365	0.205	0.022	0.154	0.143	0.010	0.051	0.196	0.006	0.023	0.261
nsp3	0.058	0.245	0.237	0.038	0.233	0.163	0.009	0.040	0.225	0.006	0.017	0.353
nsp4	0.043	0.297	0.145	0.009	0.101	0.089	0.005	0.048	0.104	0.002	0.020	0.100
nsp5	0.034	0.317	0.107	0.005	0.061	0.082	0.007	0.038	0.184	0.001	0.013	0.077
nsp6	0.073	0.280	0.261	0.005	0.136	0.037	0.004	0.046	0.087	0.002	0.009	0.222
nsp7	0.033	0.254	0.130	0.000	0.166		0.002	0.042	0.048	0.002	0.006	0.333
nsp8	0.018	0.248	0.073	0.009	0.153	0.059	0.001	0.036	0.028	0.001	0.012	0.083
nsp9	0.039	0.369	0.106	0.005	0.204	0.025	0.000	0.044		0.000	0.013	
nsp10	0.016	0.275	0.058	0.010	0.099	0.101	0.001	0.029	0.034	0.000	0.043	
nsp11	0.040	0.124	0.323	0.000	0.000		0.000	0.029		0.000	0.040	
nsp12	0.018	0.240	0.075	0.002	0.097	0.021	0.007	0.043	0.163	0.001	0.008	0.125
nsp13	0.021	0.243	0.086	0.001	0.097	0.010	0.002	0.053	0.038	0.000	0.007	
nsp14	0.032	0.305	0.105	0.003	0.041	0.073	0.002	0.066	0.030	0.001	0.012	0.083
nsp15	0.032	0.225	0.142	0.003	0.065	0.046	0.006	0.062	0.097	0.001	0.005	0.200
nsp16	0.029	0.207	0.140	0.002	0.075	0.027	0.005	0.043	0.116	0.000	0.014	
S	0.039	0.093	0.419	0.067	0.407	0.165	0.023	0.089	0.258	0.007	0.041	0.171
NS3	0.085	0.383	0.222	0.022	0.267	0.082	0.009	0.032	0.281	0.001	0.020	0.050
E	0.045	0.342	0.132	0.009	0.088	0.102	0.011	0.059	0.186	0.000	0.029	
M	0.032	0.318	0.101	0.007	0.137	0.051	0.008	0.032	0.250	0.006	0.016	0.375
N	0.056	0.338	0.166	0.036	0.260	0.138	0.011	0.068	0.162	0.004	0.016	0.250
NS7	0.242	0.471	0.514									
NS7a				0.050	0.190	0.263						

tional Committee on Taxonomy of Viruses (ICTV) has established the following genus and species demarcation criteria in the family *Coronaviridae*: coronaviruses that do not cluster together and share less than 46% sequence identity in the conserved replicase domains with any other established member are considered a new genus, while viruses that share more than 90% amino acid sequence identity in the conserved replicase domains are considered to belong to the same species (13). Hence, the virus harbored by the Asian house shrew is sufficiently divergent that it should be considered a distinct member of the genus *Alphacoronavirus*, although not a new genus under the current ICTV criteria.

Our analysis also revealed that WESV has had a complex evolutionary history. Although WESVs exhibited distinct geographic clustering, indicative of *in situ* evolution, the evolutionary relationships among viruses sampled from four counties were not consistent with their geographic locations. Such a phylogeographic pattern might reflect the influence of geographic barriers, such as mountains, rather than simple isolation by distance. In addition, the fact that the S gene of WESV was divergent from those of all known coronaviruses suggests that an intergenus recombination event may have occurred, and strong evidence for intraspecies recombination was obtained. It is also striking that the WESVs possess a distinct NS7 gene. Although a gene named ORF7 has been observed in the bat virus HKU8 (34), the NS7 gene of WESV exhibited no sequence similarity with genes of HKU8 or any other known viruses, so it has an unknown origin. In addition, the NS3 gene of WESV was genetically distinct from those of known alphacoronaviruses and betacoronaviruses.

Diverse alphacoronaviruses and betacoronaviruses have now been identified in a variety of bats globally (16, 17, 20, 34–40), based on which it has been proposed that alphacoronaviruses and betacoronaviruses in other animals have their ultimate ancestry in bats (21, 22). However, we observed that the WESVs harbored by shrews were

FIG 6 Legend (Continued)

background color of parental region A is gray, while that of parental region B is white. (B) Phylogenies of parental region A (nt 5248 to 7663) and region B (nt 1 to 5247 and 7664 to the end of the sequence). The numbers (>70) above or below branches indicate percent bootstrap values. (C) Recombination analyses of Wénchéng-554 and other known alphacoronaviruses.

phylogenetically distinct within the genus *Alphacoronavirus*, suggesting that they may have emerged early in Asian house shrews, and it is striking that WESV possesses an especially divergent S gene. Together, these results suggest that alphacoronaviruses have a far more complex evolutionary history than previously realized, with insectivores likely playing a more important role. Hence, greater effort is needed to infer the evolutionary history of alphacoronaviruses in a wider sample of mammalian species.

Shrews, classified in the order Eulipotyphla, have a broad geographic distribution and exhibit substantial diversity, rivaled only by members of the muroid families, Muridae and Cricetidae, and the bat family, Vespertilionidae (25). Asian house shrews (*Suncus murinus*) have a wide distribution throughout the Old World tropics. However, unlike bats and rodents, these mammals have not attracted attention with respect to virus evolution, emergence, and transmission. The recent discovery of erinaceus coronavirus (EriCoV) in West European hedgehogs (*Erinaceus europaeus*) indicates that insectivores are the natural reservoir of CoV (41). Over the past decade, additional novel viruses have been identified in shrews (29–31), indicating that these animals may play an important role in the evolution and transmission of viruses, including coronaviruses. WESV was identified in 24 of 266 shrews sampled from four counties of two provinces, with an overall detection rate of 9.02%, but not in rodents captured from the same areas. Therefore, shrews appear to be a natural reservoir of coronaviruses, and thus, their role in coronavirus evolution clearly merits further investigation.

MATERIALS AND METHODS

Trapping of small animals and sample collection. During 2013 to 2015, shrews were trapped in mountainous regions of Xingguo and Yudu Counties in Ganzhou City, Jiangxi Province, and in the Longwan district and Ruian and Wencheng Counties of Wenzhou City, Zhejiang Province, China (Fig. 1), as described previously (3, 42). All the animals were initially identified by morphological examination and were further confirmed by sequence analysis of the mt-cyt *b* gene (3). Euthanasia was performed before necropsy. Every effort was made to minimize suffering. Rectal samples were collected from shrews for CoV detection.

The study was reviewed and approved by the ethics committee of the National Institute for Communicable Disease Control and Prevention of the Chinese Center for Disease Control and Prevention (CDC). All animals were treated in strict accordance with the Guidelines for Laboratory Animal Use and Care of the Chinese CDC and the Rules for the Implementation of Laboratory Animal Medicine (1998) from the Ministry of Health, China, under the protocols approved by the National Institute for Communicable Disease Control and Prevention.

DNA and RNA extraction and virus detection. Total RNA was extracted from fecal samples using TRIzol reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. The RNA was eluted in 50 μ l of diethyl pyrocarbonate (DEPC) water and was used as the template for reverse transcription-PCR. Total DNA was extracted from rectal samples using a DNeasy blood and tissue kit (Qiagen, Valencia, CA, USA) according to protocols suggested by the manufacturer.

CoV RNA was detected by RT-PCR as described previously (18, 19). Complete genomes of coronaviruses were amplified using primers based on the conserved regions of known genome sequences (18, 19). The 5' and 3' ends of the genomes of the newly discovered shrew coronaviruses were obtained by 5' and 3' rapid amplification of cDNA ends (RACE) using a RACE kit (TaKaRa, Dalian, China). Sequences were assembled and manually edited to produce the final viral genomes. Amplification of the mt-cyt *b* gene was performed as described previously (3).

RT-PCR amplicons of <700 bp were purified using the QIAquick gel extraction kit (Qiagen, Valencia, CA, USA) according to the manufacturer's recommendations and subjected to direct sequencing. Purified DNA of >700 bp was cloned into the pMD18-T vector (TaKaRa, Dalian, China) and subsequently transformed into JM109-143 competent cells.

Phylogenetic analysis. Analysis of protein families was performed using the PFAM and InterProScan programs (43, 44). Prediction of transmembrane domains was performed using the TMHMM program, version 2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>).

Because of extensive sequence divergence between the nucleotide sequences of different CoV genera, all phylogenetic analyses were based on amino acid sequences. Accordingly, amino acid sequence alignments were conducted using the MAFFT program employing the G-INS-i algorithm (45). After alignment, gaps and ambiguously aligned regions were removed using Gblocks (v0.91b) (46). Phylogenetic analyses were then performed using the sequences of eight complete CoV proteins: (i) nsp5 (3CLpro), (ii) RdRp (nsp12), (iii) nsp13 (Hel), (iv) nsp14 (ExoN), (v) nsp15 (NendoU), (vi) nsp16 (O-MT), (vii) spike protein (S), and (viii) nucleocapsid protein (N) (12). Phylogenetic trees from these data were estimated using the maximum-likelihood (ML) method implemented in PhyML v3.0 (47), with bootstrap support values calculated from 1,000 replicate trees. The best-fit amino acid substitution models were determined using MEGA version 5 (48).

Recombination detection. The full genome alignment of all WESV sequences was screened for recombination using the RDP, GENECONV, and BootScan methods available within RDP4 (33). Only

sequences with significant evidence ($P < 0.05$) of recombination detected by at least two methods and confirmed by phylogenetic analysis were taken to represent strong evidence for recombination. In addition, we visualized the recombinant and the parental strains determined as described above using similarity plot analysis as implemented in Simplot version 3.5.1 (49), with a window size of 400 nt and a step size of 40 nt.

Estimation of the numbers of synonymous and nonsynonymous substitutions. The numbers of synonymous substitutions per synonymous site (dS) and nonsynonymous substitutions per nonsynonymous site (dN) for each coding region between each pair of WESV, BatCoV HKU2, porcine epidemic diarrhoea virus (PEDV), and HCoV-NL63 strains were calculated using the Kimura 2-parameter method applied to synonymous and nonsynonymous sites as implemented in MEGA (v5) (48).

Accession number(s). All the viral sequences obtained in this study have been deposited in GenBank under accession numbers [KY967715](#) to [KY967735](#) and [KF294384](#) to [KF294386](#).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <https://doi.org/10.1128/JVI.00764-17>.

SUPPLEMENTAL FILE 1, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

This study was supported by the Special National Project on Research and Development of Key Biosafety Technologies (2016YFC1201900), the 12th Five-Year Major National Science and Technology Projects of China (2014ZX10004001-005), and the National Natural Science Foundation of China (grants 81290343, 81273014). E.C.H. is funded by an NHMRC Australia Fellowship (GNT1037231). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

- Lloyd-Smith JO, George D, Pepin KM, Pitzer VE, Pulliam JR, Dobson AP, Hudson PJ, Grenfell BT. 2009. Epidemic dynamics at the human-animal interface. *Science* 326:1362–1367. <https://doi.org/10.1126/science.1177345>.
- Wolfe ND, Dunavan CP, Diamond J. 2007. Origins of major human infectious diseases. *Nature* 447:279–283. <https://doi.org/10.1038/nature05775>.
- Guo WP, Lin XD, Wang W, Tian JH, Cong ML, Zhang HL, Wang MR, Zhou RH, Wang JB, Li MH, Xu J, Holmes EC, Zhang YZ. 2013. Phylogeny and origins of hantaviruses harbored by bats, insectivores, and rodents. *PLoS Pathog* 9:e1003159. <https://doi.org/10.1371/journal.ppat.1003159>.
- Holmes EC, Zhang YZ. 2015. The evolution and emergence of hantaviruses. *Curr Opin Virol* 10:27–33. <https://doi.org/10.1016/j.coviro.2014.12.007>.
- Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, SARS Working Group. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348:1953–1966. <https://doi.org/10.1056/NEJMoa030781>.
- Morens DM, Folkers GK, Fauci AS. 2004. The challenge of emerging and re-emerging infectious diseases. *Nature* 430:242–249. <https://doi.org/10.1038/nature02759>.
- Zhang YZ, Zhou DJ, Xiong Y, Chen XP, He YW, Sun Q, Yu B, Li J, Dai YA, Tian JH, Qin XC, Jin D, Cui Z, Luo XL, Li W, Lu S, Wang W, Peng JS, Guo WP, Li MH, Li ZJ, Zhang S, Chen C, Wang Y, de Jong MD, Xu J. 2011. Hemorrhagic fever caused by a novel tick-borne Bunyavirus in Huaiyangshan, China. *Zhonghua Liu Xing Bing Xue Za Zhi* 32:209–220.
- Daszak P, Cunningham AA, Hyatt AD. 2000. Emerging infectious diseases of wildlife: threats to biodiversity and human health. *Science* 287:443–449. <https://doi.org/10.1126/science.287.5452.443>.
- Hoberg EP, Brooks DR. 2015. Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. *Philos Trans R Soc Lond B Biol Sci* 370:20130553. <https://doi.org/10.1098/rstb.2013.0553>.
- Keesing F, Belden LK, Daszak P, Dobson A, Harvell CD, Holt RD, Hudson P, Jolles A, Jones KE, Mitchell CE, Myers SS, Bogich T, Ostfeld RS. 2010. Impacts of biodiversity on the emergence and transmission of infectious diseases. *Nature* 468:647–652. <https://doi.org/10.1038/nature09575>.
- Zhang YZ, Xu J. 2016. The emergence and cross species transmission of newly discovered tick-borne Bunyavirus in China. *Curr Opin Virol* 16:126–131. <https://doi.org/10.1016/j.coviro.2016.02.006>.
- Masters P, Perlman S. 2013. *Coronaviridae*, p 825–858. In Knipe DM, Cohen JL, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), *Fields virology*, vol 1. Lippincott Williams & Wilkins, Philadelphia, PA.
- de Groot RJ, Baker SC, Baric R, Enjuanes L, Gorbalenya A, Holmes KV, Perlman S, Poon L, Rottier PJ, Talbot PJ, Woo PC, Ziebuhr J. 2011. *Coronaviridae*, p 806–828. In King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), *Virus taxonomy: classification and nomenclature of viruses*. Ninth report of the International Committee on Taxonomy of Viruses. Academic Press, London, United Kingdom.
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367:1814–1820. <https://doi.org/10.1056/NEJMoa1211721>.
- Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL. 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302:276–278. <https://doi.org/10.1126/science.1087139>.
- Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY. 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* 102:14040–14045. <https://doi.org/10.1073/pnas.0506735102>.
- Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF. 2005. Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310:676–679. <https://doi.org/10.1126/science.1118391>.
- Wang W, Lin XD, Guo WP, Zhou RH, Wang MR, Wang CQ, Ge S, Mei SH, Li MH, Shi M, Holmes EC, Zhang YZ. 2015. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology* 474:19–27. <https://doi.org/10.1016/j.virol.2014.10.017>.
- Lin XD, Wang W, Hao ZY, Wang ZX, Guo WP, Guan XQ, Wang MR, Wang HW, Zhou RH, Li MH, Tang GP, Wu J, Holmes EC, Zhang YZ. 2017. Extensive diversity of coronaviruses in bats from China. *Virology* 507:1–10. <https://doi.org/10.1016/j.virol.2017.03.019>.
- Smith CS, de Jong CE, Meers J, Henning J, Wang L, Field HE. 2016.

- Coronavirus infection and diversity in bats in the Australasian region. *Ecohealth* 13:72–82. <https://doi.org/10.1007/s10393-016-1116-x>.
21. Vijaykrishna D, Smith GJ, Zhang JX, Peiris JS, Chen H, Guan Y. 2007. Evolutionary insights into the ecology of coronaviruses. *J Virol* 81: 4012–4020. <https://doi.org/10.1128/JVI.02605-06>.
 22. Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Bai R, Teng JL, Tsang CC, Wang M, Zheng BJ, Chan KH, Yuen KY. 2012. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J Virol* 86:3995–4008. <https://doi.org/10.1128/JVI.06540-11>.
 23. Koonin EV, Senkevich TG, Dolja VV. 2006. The ancient virus world and evolution of cells. *Biol Direct* 1:29. <https://doi.org/10.1186/1745-6150-1-29>.
 24. Shi M, Lin XD, Tian JH, Chen LJ, Chen X, Li CX, Qin XC, Li J, Cao JP, Eden JS, Buchmann J, Wang W, Xu J, Holmes EC, Zhang YZ. 2016. Redefining the invertebrate RNA virosphere. *Nature* 540:539–543. <https://doi.org/10.1038/nature20167>.
 25. Wilson DE, Reeder DM. 2005. Mammal species of the world. A taxonomic and geographic reference, 3rd ed. Johns Hopkins University Press, Baltimore, MD.
 26. Li CX, Shi M, Tian JH, Lin XD, Kang YJ, Chen LJ, Qin XC, Xu J, Holmes EC, Zhang YZ. 2015. Unprecedented genomic diversity of RNA viruses in arthropods reveals the ancestry of negative-sense RNA viruses. *eLife* 4:e05378. <https://doi.org/10.7554/eLife.05378>.
 27. Nga PT, Parquet Mdel C, Lauber C, Parida M, Nabeshima T, Yu F, Thuy NT, Inoue S, Ito T, Okamoto K, Ichinose A, Snijder EJ, Morita K, Gorbalenya AE. 2011. Discovery of the first insect nidovirus, a missing evolutionary link in the emergence of the largest RNA virus genomes. *PLoS Pathog* 7:e1002215. <https://doi.org/10.1371/journal.ppat.1002215>.
 28. Zirkel F, Kurth A, Quan PL, Briese T, Ellerbrok H, Pauli G, Leendertz FH, Lipkin WI, Ziebuhr J, Drosten C, Junglen S. 2011. An insect nidovirus emerging from a primary tropical rainforest. *mBio* 2:e00077-11. <https://doi.org/10.1128/mBio.00077-11>.
 29. Li K, Lin XD, Huang KY, Zhang B, Shi M, Guo WP, Wang MR, Wang W, Xing JG, Li MH, Hong WS, Holmes EC, Zhang YZ. 2016. Identification of novel and diverse rotaviruses in rodents and insectivores, and evidence of cross-species transmission into humans. *Virology* 494:168–177. <https://doi.org/10.1016/j.virol.2016.04.017>.
 30. Li K, Lin XD, Wang W, Shi M, Guo WP, Zhang XH, Xing JG, He JR, Wang K, Li MH, Cao JH, Jiang ML, Holmes EC, Zhang YZ. 2015. Isolation and characterization of a novel arenavirus harbored by rodents and shrews in Zhejiang Province, China. *Virology* 476:37–42. <https://doi.org/10.1016/j.virol.2014.11.026>.
 31. Zhang YZ. 2014. Discovery of hantaviruses in bats and insectivores and the evolution of the genus *Hantavirus*. *Virus Res* 187:15–21. <https://doi.org/10.1016/j.virusres.2013.12.035>.
 32. Lau SK, Woo PC, Li KS, Huang Y, Wang M, Lam CS, Xu H, Guo R, Chan KH, Zheng BJ, Yuen KY. 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* 367:428–439. <https://doi.org/10.1016/j.virol.2007.06.009>.
 33. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefevre P. 2010. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* 26:2462–2463. <https://doi.org/10.1093/bioinformatics/btq467>.
 34. Chu DK, Peiris JS, Chen H, Guan Y, Poon LL. 2008. Genomic characterizations of bat coronaviruses (1A, 1B and HKU8) and evidence for co-infections in *Miniopterus* bats. *J Gen Virol* 89:1282–1287. <https://doi.org/10.1099/vir.0.83605-0>.
 35. Corman VM, Rasche A, Diallo TD, Cottontail VM, Stocker A, Souza BF, Correa JI, Carneiro AJ, Franke CR, Nagy M, Metz M, Knornschild M, Kalko EK, Ghanem SJ, Morales KD, Salsamendi E, Spinola M, Herrler G, Voigt CC, Tschapka M, Drosten C, Drexler JF. 2013. Highly diversified coronaviruses in neotropical bats. *J Gen Virol* 94:1984–1994. <https://doi.org/10.1099/vir.0.054841-0>.
 36. Corman VM, Ithete NL, Richards LR, Schoeman MC, Preiser W, Drosten C, Drexler JF. 2014. Rooting the phylogenetic tree of Middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. *J Virol* 88:11297–11303. <https://doi.org/10.1128/JVI.01498-14>.
 37. Corman VM, Baldwin HJ, Tateno AF, Zerbini RM, Annan A, Owusu M, Nkrumah EE, Maganga GD, Oppong S, Adu-Sarkodie Y, Vallo P, da Silva Filho LV, Leroy EM, Thiel V, van der Hoek L, Poon LL, Tschapka M, Drosten C, Drexler JF. 2015. Evidence for an ancestral association of human coronavirus 229E with bats. *J Virol* 89:11858–11870. <https://doi.org/10.1128/JVI.01755-15>.
 38. Drexler JF, Corman VM, Drosten C. 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. *Antiviral Res* 101:45–56. <https://doi.org/10.1016/j.antiviral.2013.10.013>.
 39. He B, Zhang Y, Xu L, Yang W, Yang F, Feng Y, Xia L, Zhou J, Zhen W, Feng Y, Guo H, Zhang H, Tu C. 2014. Identification of diverse alphacoronaviruses and genomic characterization of a novel severe acute respiratory syndrome-like coronavirus from bats in China. *J Virol* 88:7070–7082. <https://doi.org/10.1128/JVI.00631-14>.
 40. Huang C, Liu WJ, Xu W, Jin T, Zhao Y, Song J, Shi Y, Ji W, Jia H, Zhou Y, Wen H, Zhao H, Liu H, Li H, Wang Q, Wu Y, Wang L, Liu D, Liu G, Yu H, Holmes EC, Lu L, Gao GF. 2016. A bat-derived putative cross-family recombinant coronavirus with a reovirus gene. *PLoS Pathog* 12:e1005883. <https://doi.org/10.1371/journal.ppat.1005883>.
 41. Corman VM, Kallies R, Philipps H, Gopner G, Muller MA, Eckerle I, Brunink S, Drosten C, Drexler JF. 2014. Characterization of a novel betacoronavirus related to Middle East respiratory syndrome coronavirus in European hedgehogs. *J Virol* 88:717–724. <https://doi.org/10.1128/JVI.01600-13>.
 42. Mills JN, Childs JE, Ksiazek TG, Peters CJ, Velleca WM. 1995. Methods for trapping and sampling small mammals for virologic testing. Centers for Disease Control and Prevention, Atlanta, GA.
 43. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res* 29:37–40. <https://doi.org/10.1093/nar/29.1.37>.
 44. Bateman A, Birney E, Cerruti L, Durbin R, Ewiler L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL. 2002. The Pfam protein families database. *Nucleic Acids Res* 30:276–280. <https://doi.org/10.1093/nar/30.1.276>.
 45. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
 46. Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 56:564–577. <https://doi.org/10.1080/10635150701472164>.
 47. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59: 307–321. <https://doi.org/10.1093/sysbio/syq010>.
 48. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28:2731–2739. <https://doi.org/10.1093/molbev/msr121>.
 49. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73:152–160.