








Article

# Shared Common Ancestry of Rodent Alphacoronaviruses Sampled Globally

Theocharis Tsoleridis <sup>1</sup>, Joseph G. Chappell <sup>1</sup>, Okechukwu Onianwa <sup>1</sup>, Denise A. Marston <sup>3</sup>, Anthony R. Fooks <sup>3</sup>, Elodie Monchatre-Leroy <sup>4</sup>, Gérald Umhang <sup>4</sup>, Marcel A. Müller <sup>5</sup>, Jan F. Drexler <sup>5</sup>, Christian Drosten <sup>5</sup>, Rachael E. Tarlinton <sup>2</sup>, Charles P. McClure <sup>1</sup>, Edward C. Holmes <sup>6</sup> and Jonathan K. Ball <sup>1,\*</sup>

<sup>1</sup> School of Life Sciences, University of Nottingham, Nottingham NG7 2UH, UK;

t.tsoleridis@nottingham.ac.uk (T.T.); nixjc1@exmail.nottingham.ac.uk (J.G.C.);

okechukwu.onianwa@phe.gov.uk (O.O.); patrick.mcclure@nottingham.ac.uk (C.P.M.)

<sup>2</sup> School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington Campus,

Loughborough LE12 5RD, UK; rachael.tarlinton@nottingham.ac.uk

<sup>3</sup> Wildlife Zoonoses and Vector-borne Diseases Research Group, Animal and Plant Health Agency (APHA),

Weybridge-London KT15 3NB, UK; Denise.Marston@apha.gsi.gov.uk (D.A.M.);

Tony.Fooks@apha.gsi.gov.uk (A.R.F.)

<sup>4</sup> Anses, Laboratoire de la rage et de la faune sauvage, 54220 Malzéville, France;

elodie.monchatre-leroy@anses.fr (E.M.-L.); gerald.umhang@anses.fr (G.U.)

<sup>5</sup> Charité-Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu

Berlin, and Berlin Institute of Health, Institute of Virology, 10117 Berlin, Germany;

marcel.mueller@charite.de (M.A.M.); felix.drexler@charite.de (J.F.D.); christian.drosten@charite.de (C.D.)

<sup>6</sup> Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and

Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006,

Australia; edward.holmes@sydney.edu.au

\* Correspondence: jonathan.ball@nottingham.ac.uk

Received: 14 January 2019; Accepted: 28 January 2019; Published: 30 January 2019



**Abstract:** The recent discovery of novel alphacoronaviruses (alpha-CoVs) in European and Asian rodents revealed that rodent coronaviruses (CoVs) sampled worldwide formed a discrete phylogenetic group within this genus. To determine the evolutionary history of rodent CoVs in more detail, particularly the relative frequencies of virus-host co-divergence and cross-species transmission, we recovered longer fragments of CoV genomes from previously discovered European rodent alpha-CoVs using a combination of PCR and high-throughput sequencing. Accordingly, the full genome sequence was retrieved from the UK rat coronavirus, along with partial genome sequences from the UK field vole and Poland-resident bank vole CoVs, and a short conserved ORF1b fragment from the French rabbit CoV. Genome and phylogenetic analysis showed that despite their diverse geographic origins, all rodent alpha-CoVs formed a single monophyletic group and shared similar features, such as the same gene constellations, a recombinant beta-CoV spike gene, and similar core transcriptional regulatory sequences (TRS). These data suggest that all rodent alpha CoVs sampled so far originate from a single common ancestor, and that there has likely been a long-term association between alpha CoVs and rodents. Despite this likely antiquity, the phylogenetic pattern of the alpha-CoVs was also suggestive of relatively frequent host-jumping among the different rodent species.

**Keywords:** coronavirus; alphacoronavirus; rodents; ancestry; recombination; evolution

## 1. Introduction

Among the viruses that infect both humans and animals, coronaviruses are common and important pathogens. The first coronavirus (CoV) was discovered during the 1930s [1], and the first human coronavirus was characterized in the 1960s [2]. Since then, five novel human coronaviruses have been discovered. The coronavirus family gained particular notoriety with the emergence of the severe acute respiratory syndrome (SARS) in 2002/2003 and the Middle East respiratory syndrome (MERS) in 2013. In humans, coronaviruses are associated with respiratory disease, causing a range of symptoms from mild common colds to more severe lower respiratory tract infections that can be lethal. CoV infection in a range of non-human mammals and birds is associated with enteric and respiratory diseases as well as hepatitis and neurological disorders [3].

Rodents are an important source of emerging virus infections [4]. Rodentia is the single largest mammalian order and comprises ~40% of all mammals [5], including ~2200 species such as rats, mice, and voles [6]. Despite the key role played by rodents in virus emergence, until recently, only two rodent CoVs had been characterized—murine hepatitis virus (MHV) [7] and rat sialodacryoadenitis coronavirus (SADV) [8], both of which are from the same viral species and are betacoronaviruses.

We previously reported the discovery of novel alpha-CoVs in European rodents and shrews, encompassing several species such as field voles (*Microtus agrestis*), bank voles (*Myodes glareolus*), and rats (*Rattus norvegicus*) [9]. Since then, additional alpha-CoVs have been described infecting a variety of Chinese rodents, including field mice (*Apodemus chevrieri*), grey red-backed voles (*Myodes rufocanus*), and lesser rice field rats (*Rattus losea*) [10,11]. Interestingly, a recently discovered rabbit alphacoronavirus also clustered with rodent alpha-CoVs [12].

To gain a better understanding of the evolution of alphacoronaviruses in rodents, particularly how long these virus-host associations have been established, we recovered and characterized complete or partial genomes for European Norway rat, field vole, and bank vole CoVs, and compared these sequences to those available for other rodent CoVs obtained in other parts of the world. Our new phylogenetic analyses provide compelling evidence that the alpha-CoVs sampled from rodents to date have arisen from a single common ancestor, suggestive of a long-term virus-host association.

## 2. Materials and Methods

### 2.1. Samples

Samples previously reported to be positive for alpha-CoVs were used in this study [9,12]. These samples included a UK *Rattus norvegicus* (UKRn3), UK *Microtus agrestis* (UKMa1), Poland-resident *Myodes glareolus* (PLMg1), and the France-resident *Oryctolagus cuniculus* (L232) (Table 1).

**Table 1.** List of sample name, species, country of origin, organ, and original publication.

Name	Species	Country	Organ	Reference
UKRn3	<i>Rattus norvegicus</i>	United Kingdom	Liver	[9]
UKMa1	<i>Microtus agrestis</i>	United Kingdom	Gut	[9]
PLMg1	<i>Myodes glareolus</i>	Poland	Liver	[9]
L232	<i>Oryctolagus cuniculus</i>	France	Intestinal fluid	[12]

### 2.2. Nucleic Acid Preparation

Total RNA was extracted from approximately one cubic millimetre sections of liver (UKRn3, PLMg1) or intestinal tissue samples (UKMa1) using the GenElute™ Mammalian Total RNA Miniprep Kit (Sigma Aldrich, Steinheim, Germany). Total RNA was extracted from intestinal fluid for the France resident L232 rabbit CoV using the QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany). The RNA to cDNA EcoDry™ Premix - Random Hexamers kit (Clontech, Saint-Germain-en-Laye, France) was used for cDNA synthesis.

### 2.3. High-Throughput Sequencing

High-throughput sequencing was performed on total RNA extracted from UKMa1 and PLMg1 using Illumina MiSeq. Genomic DNA (gDNA) was depleted using DNaseI, and ribosomal RNA (rRNA) was removed using NEBNext® rRNA Depletion Kit (Human/Mouse/Rat) with RNA Sample Purification Beads (New England Biolabs, Ipswich, Massachusetts, USA). The samples were run on an Illumina MiSeq platform, generating a total of 8,045,758 paired reads for PLMg1 and 14,140,770 paired reads for UKMa1. Each read length was  $2 \times 150$  bp, and the insert size was 400 bp on average. All the Next Generation Sequencing (NGS) data were analysed using Geneious 11.1.2 software. The reads were assembled into contigs with de novo assembly using Geneious. The assembled contigs were compared against the RefSeq sequence database of all the virus proteins downloaded from GenBank using blastX. The minimum *e*-value was  $1 \times 10^{-5}$  to maintain high sensitivity and low rates of false hits [13]. The contigs that matched known viruses were verified by PCR.

### 2.4. Genome Retrieval with PCR

The primers used for primer walking were either designed based on the Lucheng Rn coronavirus (Genbank Acc. No. KF294380) reference sequence or based on the known genome sequences recovered from UKRn3. The primers for contig confirmation and gap-filling between contigs for UKMa1 and PLMg1 were designed based on the sequences of the contigs. Geneious 11.1.2 software was used for genome mapping, annotating, and analysis. All the primers were evaluated in silico with the Primer3 online tool. For the retrieval of a short *ORF1b* fragment for L232, a previously described primer combination was used [9].

PCR reactions were carried out in a PCT-200 Peltier Thermal Cycler (MJ Research). The PCR reactions were performed with HotStarTaq Polymerase (Qiagen) according to manufacturer's instructions.

The coronavirus sequences generated in this study were deposited in GenBank under the accession numbers MK163627 and MK249067-MK249069.

### 2.5. Phylogenetic Analysis

Reference sequences of  $\alpha$ ,  $\beta$ , and one  $\gamma$ -CoV were downloaded from GenBank for the *ORF1b*, *S*, and *N* genes and aligned using ClustalW [14] recognising codon positions. To provide evolutionary context, phylogenetic analyses were performed using the maximum likelihood (ML) method within the Molecular Evolutionary Genetics Analysis version 7 (MEGA7) [15] package. These analyses utilized a General Time Reversible (GTR) nucleotide substitution model with a gamma distribution of rate variation, a class of invariant sites ( $\Gamma+I$ ) model of evolution, and complete deletion of gaps, with statistical robustness assessed using bootstrap resampling (1,000 pseudoreplicates). Equivalent ML phylogenies were also inferred using partial *ORF1b*, *S*, and *N* amino acid sequences and employing the Jones-Taylor-Thornton (JTT) substitution model with uniform rates, complete deletion of gaps, and statistical robustness of 1000 pseudoreplicates. Details regarding the number of sequences used for the phylogenetic analysis of each gene along with their length are given in Table 2, and the coronavirus reference sequences used in the phylogenetic analysis are given in Tables S1 and S2.

**Table 2.** List of the different genes used for phylogenetic analyses, number of sequences, sequence length, and corresponding positions on the Lucheng CoV genome.

Gene	Number of Sequences	Length (nt)	Corresponding Positions on Lucheng CoV Genome
<i>ORF1b</i>	46	576	14,140–14,715
<i>S</i>	31	1398	23,242–24,639
<i>N</i>	32	1170	26,977–28,146

To determine the phylogenetic relationships among the host species analysed, the mitochondrial *cytochrome b* reference sequences of *Rattus*, *Apodemus*, *Myodes*, *Microtus* spp., and *Oryctolagus cuniculus* were downloaded from Genbank. In total, 44 full length (1136 nt) *cytochrome b* sequences were aligned as described above, and an ML tree was estimated using the same procedure as described above.

### 2.6. Analysis of the Virus-Host Co-Divergence

To compare the evolutionary history of rodents and their coronaviruses, particularly the relative frequencies of co-divergence versus cross-species transmission, a tanglegram was inferred by comparing the ML phylogenies of the rodent *cytochrome b* and the coronavirus *ORF1b* sequences, connecting hosts to their corresponding coronaviruses. The reconciliation between the alpha-CoV *ORF1b* and the host mtDNA *cytochrome b* phylogenies was performed using the Jane software with default parameters [16].

## 3. Results

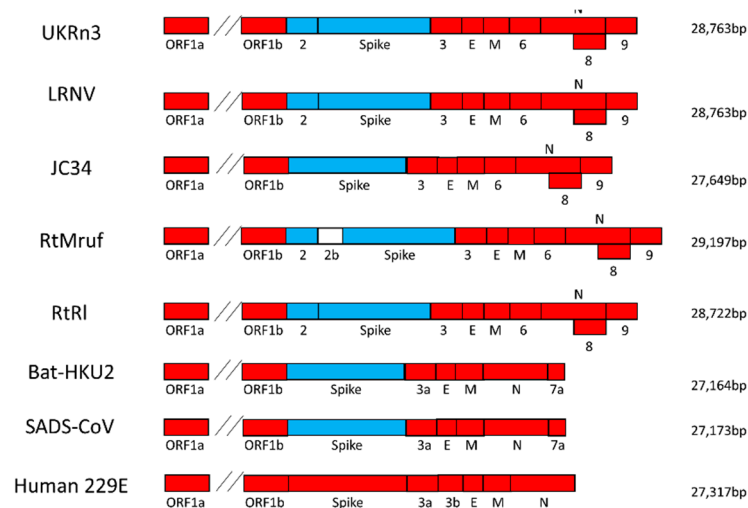
### 3.1. Genome Retrieval of Rodent and Rabbit Alphacoronaviruses

The near-full genome sequence of the UK rat CoV (UKRn3) was retrieved with primer-walking PCR. The 5' and 3' ends corresponded to the rat Lucheng CoV (KF294380)-derived primer sequences used for amplification of the genome termini and were not confirmed by the rapid amplification of cDNA ends (RACE) method. The genome of UKRn3 had a length of 28,763 nt and 40.2% GC content. Its pairwise distance to Lucheng CoV at the nucleotide level was 95.5%, indicating that they belong in the same species. Additionally, the rat UKRn3 was 90.2% similar to RtMruf (KY370045), 95.4% to RtRI-CoV (KY370050), and 75% to AcCoV-JC34 (NC034972) in the nucleotide level.

For the genome retrieval of the UK field vole CoV (UKMa1) and the Poland-resident bank vole (PLMg1), primer-walking PCR was not applicable due to the lack of a very close reference sequence. Thus, NGS performed on total RNA extracted from the *Microtus agrestis* gut and *Myodes glareolus* liver was used as an alternative method. Genome fragments representing 17,483 and 8603 nucleotides spanning various genes were retrieved for UKMa1 and PLMg1, respectively. Full-length *N* gene and partial *ORF1ab*, *ORF2*, *S*, *ORF6*, *ORF8*, and *ORF9* genes were retrieved for UKMa1, whereas partial *ORF1b* and *N* genes were retrieved for PLMg1. For the L232 coronavirus, which was obtained from a rabbit living in France [12], additional sequences up to 621 bp of the *ORF1b* gene were retrieved by targeting conserved regions with PCR. Unfortunately, presumably due to the relatively poor quality of the recovered nucleic acid, we were unable to recover additional sequences.

### 3.2. Virus Genome Characterization

Genome analysis of all the rodent alpha-CoVs revealed a number of common features (Figure 1). All viruses, with the exception of the field mouse AcCoV-JC34 (NC034972), possessed a predicted *ORF2* gene that was approximately 274 amino acids (aa) in length and was located between the *ORF1b* and *S* gene. Blast search revealed that this gene was not present in any other characterized alpha-CoVs and its closest match, at 44% amino acid similarity, was the NS2a gene of the Betacoronavirus1 Human CoV OC43. Interestingly, all the rodent  $\alpha$ -CoVs possessed a  $\beta$ -CoV-like spike, which indicates recombination between  $\alpha$  and  $\beta$ -CoV. All the rodent  $\alpha$ -CoVs possessed a predicted 214 aa long *ORF3* gene located between the *S* and the *E* gene. The closest match for *ORF3* was the 3c-like protein of ferret alphacoronavirus with 47% similarity at the amino acid level. Another striking feature of rodent alphacoronavirus genomes was the presence of the predicted *ORF6* (165 aa) gene between the *M* and *N* genes, which is observed in no other members of the genus *Alphacoronavirus*. Other common features of all the rodent  $\alpha$ -CoV were the presence of a predicted *ORF8* (~181 aa) within the *N* gene and *ORF9* (~111 aa) downstream of the *N* gene. Finally, the grey red-backed vole RtMruf-CoV (KY370045) appeared to possess an additional 154 aa long predicted gene called *ORF2b*, located between *ORF2* and *S*.



**Figure 1.** Genome organisation of rodent alphacoronaviruses. Schematic representation of the UKRn3 genome organisation along with those of other rodent alphacoronaviruses ( $\alpha$ -CoVs): Lucheng Rn CoV (LRNV), AcCoV-JC34, RtRI-CoV, RtMruf-CoV, and the related Bat-HKU2 and Swine acute diarrhoea syndrome coronaviruses (CoV) (SADS-CoV). Its genome organisation was similar to those of members of the genus *Alphacoronavirus*, with the characteristic 5' - ORF1ab - spike (S) - envelope (E) - membrane (M) - nucleocapsid (N) - 3' gene order. The red colour represents genes that belong to the genus *Alphacoronavirus*, whereas light blue denotes genes that belong to the genus *Betacoronavirus*, indicative of recombination. The Human 229E genome represents a typical alpha-CoV genome organisation.

Further analysis revealed similarities in the core part of the transcription regulatory sequence (TRS) located upstream of the 5' end of each ORF (Table 3). All the rodent  $\alpha$ -CoVs possessed the same 5'-AACUAA-3' core TRS sequence in the majority of their genes. However, some variations of the core TRS were observed; specifically, a 5'-AACUUUAA-3' for ORF2, 5'-UACUAA-3' for ORF3, 5'-CACUAA-3', and 5'-GACUAA-3' for ORF8.

### 3.3. Phylogenetic Analysis of Rodent Alphacoronaviruses Suggests a Shared Common Ancestry

Having retrieved the whole genome for UKRn3 and partial genomes for UKMa1, PLMg1, and L232, we next evaluated the evolutionary relationships between the novel  $\alpha$ -CoVs and the coronavirus sequences publicly available on GenBank. Accordingly, an analysis of a conserved 576 bp region of the ORF1b gene, which encodes the RNA-dependent-RNA-polymerase (RdRp), revealed that all the rodent alpha-CoVs sampled from geographic localities as diverse as Western Europe and East Asia, as well as the rabbit L232 CoV, formed a distinct clade within the *Alphacoronavirus* genus with 100% bootstrap support (Figure 2). Within that cluster, viruses obtained from UK and Chinese rat species formed a single sub-clade with 100% bootstrap support. Other well supported subclades (i.e., all supported by >80% bootstrap replicates) contained (i) the Chinese grey red-backed vole (RtMruf), the German bank vole (RCoV/RMU10 3212/Myo gla/GER/2010), the Chinese field mouse (AcCoV-JC34), and (ii) the UK field vole (UKMa1) and French rabbit CoV (L232). The virus sampled from a Poland-resident bank vole (PLMg1) formed a distinct lineage.

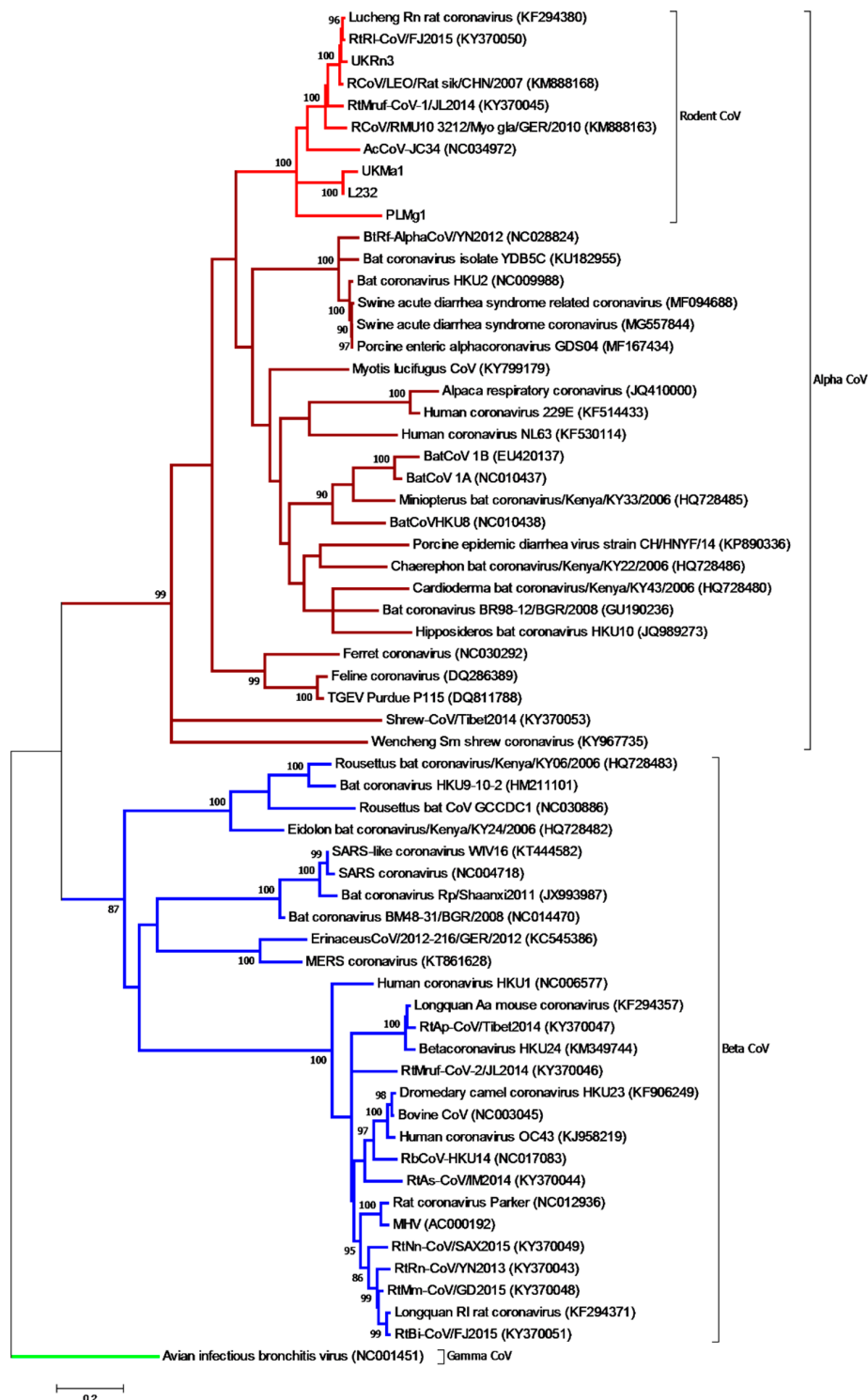
Phylogenetic analysis of the N gene (Figure 3) revealed similar results to the ORF1b phylogeny, with all the available rodent coronaviruses forming a distinct clade within the genus *Alphacoronavirus*. The rat viruses contained within a single cluster supported by 100% of bootstrap trees, while PLMg1, UKMa1, and AcCoV-JC34 fell in strongly supported more divergent positions. The phylogenetic relationships within the rodent clade are therefore the same as those observed in the ORF1b phylogeny (rabbit L232 is not shown because it was not possible to retrieve the corresponding N gene sequence).

**Table 3.** List of all the predicted core transcription regulatory sequences (TRS) for all the genes of the rodent alpha-CoVs. Core TRS could not be predicted for ORF1ab, S, ORF3, E, and M for UKMa1 as there was no available sequence for those genes. AcCoV-JC34 did not possess an ORF2 such that there was no core TRS for that gene.

Virus	Genes										
	<i>ORF1ab</i>	<i>ORF2</i>	<i>S</i>	<i>ORF3</i>	<i>E</i>	<i>M</i>	<i>ORF6</i>	<i>N</i>	<i>ORF8</i>	<i>ORF9</i>	
<b>UKRn3</b>	AACUAA	AACUUUAA	AACUAA	UACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA
<b>LRNV</b>	AACUAA	AACUAA	AACUAA	UACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA
<b>UKMa1</b>	N/A	AACUUUAA	N/A	N/A	N/A	N/A	AACUAA	AACUAA	CACUAA	AACUAA	AACUAA
<b>RtRI</b>	AACUAA	AACUUUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA	AACUAA
<b>RtMruf</b>	AACUAA	AACUUUAA	AACUAA	UACUAA	AACUAA	AACUAA	AACUAA	AACUAA	GACUAA	AACUAA	AACUAA
<b>JC34</b>	AACUAA	N/A	AACUUA	UACUAAA	AACUAA	AACUAA	AACUAA	AACUAA	CACUAA	AACUAA	AACUAA

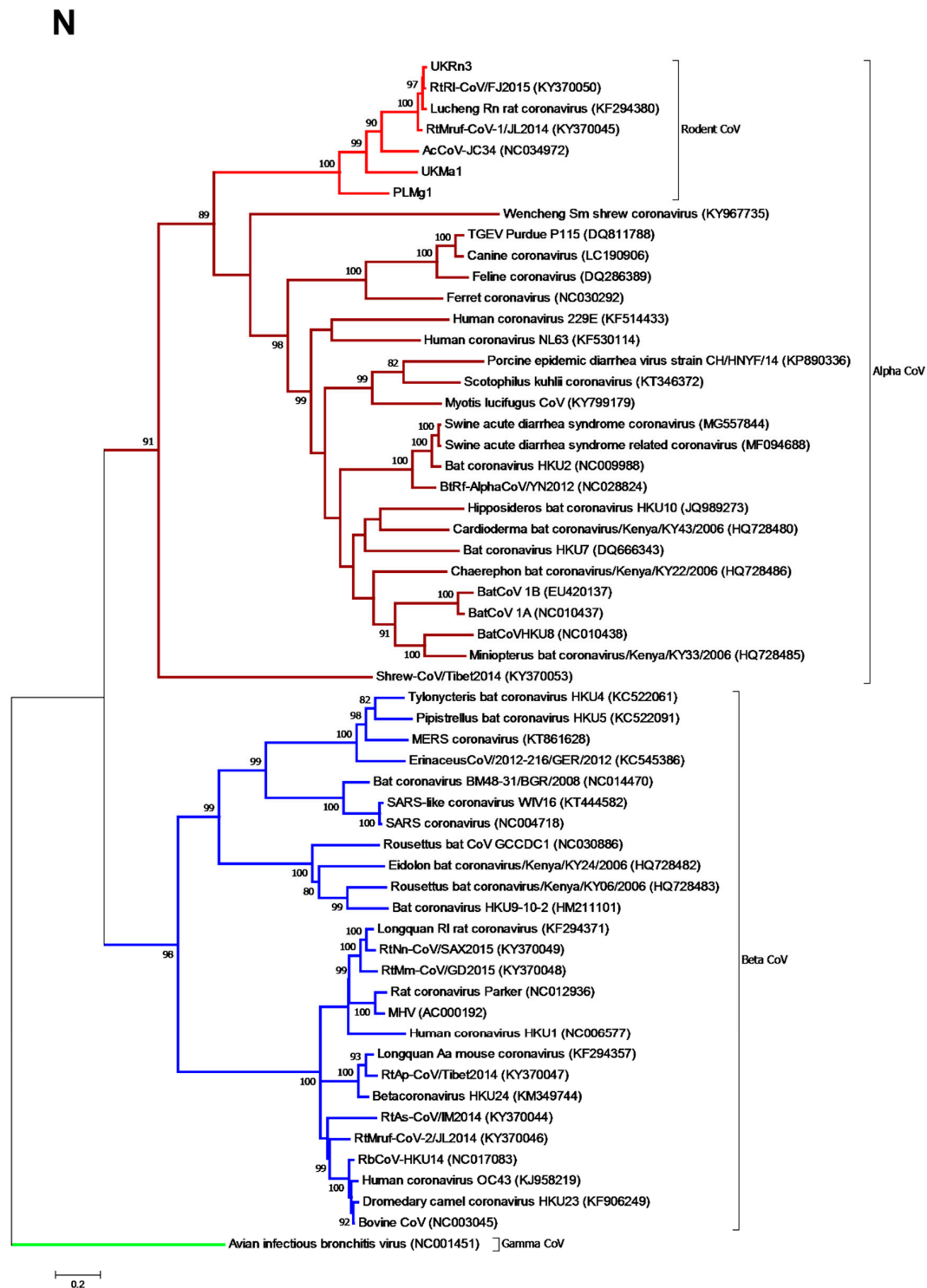


# ORF1b



**Figure 2.** Phylogenetic relationships of the rodent alphacoronaviruses based on the *ORF1b* gene. Maximum likelihood phylogenetic analysis of coronavirus partial 576 nt *ORF1b* gene sequences, revealing that the rodent alphaviruses formed a single clade with the novel rabbit L232 CoV within the genus alphacoronavirus. The novel coronavirus sequence obtained from a French rabbit L232 CoV was analysed alongside reference sequences representing three different coronavirus genera (*Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus* as an outgroup). Reference sequences are

indicated by their GenBank accession numbers. Branch lengths are drawn to a scale of nucleotide substitutions per site. Numbers above individual branches indicate the percentage that that branch was found in 1000 bootstrap replicates; only values >80% are shown. Alpha-CoVs are highlighted in brown, betacoronaviruses in blue, gammacoronaviruses in green, and the rodent coronaviruses in light red.



**Figure 3.** Phylogenetic relationships of the rodent alphacoronaviruses based on the *N* gene. Maximum likelihood phylogenetic analysis of coronavirus partial *N* gene sequences corresponding to positions 10–1173 of the Lucheng Rn CoV (KF294380). Novel coronavirus sequences obtained from PL-resident



*Myodes glareolus* (PLMg1) and the UK *Microtus agrestis* (UKMa1) and the UK *Rattus norvegicus* (UKRn3) were analysed alongside reference sequences representing three different coronavirus genera (*Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus* as an outgroup). Reference sequences are indicated by their GenBank accession numbers. Branch lengths are drawn to a scale of nucleotide substitutions per site. Numbers above individual branches indicate the percentage that that branch was found in 1000 bootstrap replicates; only values >80% are shown. Alpha-CoVs are highlighted in brown, betacoronaviruses in blue, gammacoronaviruses in green, and the rodent coronaviruses in light red.

A markedly different evolutionary relationship was observed in the case of the partial *S* gene phylogeny, in which all the rodent-borne CoVs previously assigned as alpha-CoVs based on their *ORF1* and *N* sequences clustered within the genus *Betacoronavirus* (Figure 4). Within this genome region, these viruses clustered with other known recombinant coronaviruses, including Wencheng Sm shrew CoV, Bat CoV HKU2, BtRf-AlphaCoV, Swine acute diarrhoea syndrome CoV, and swine acute diarrhoea syndrome related (bat) CoV. The rodent coronaviruses again formed a single branch with 100% bootstrap support, indicative of an ancient recombination event between the  $\alpha$ - and  $\beta$  CoVs.

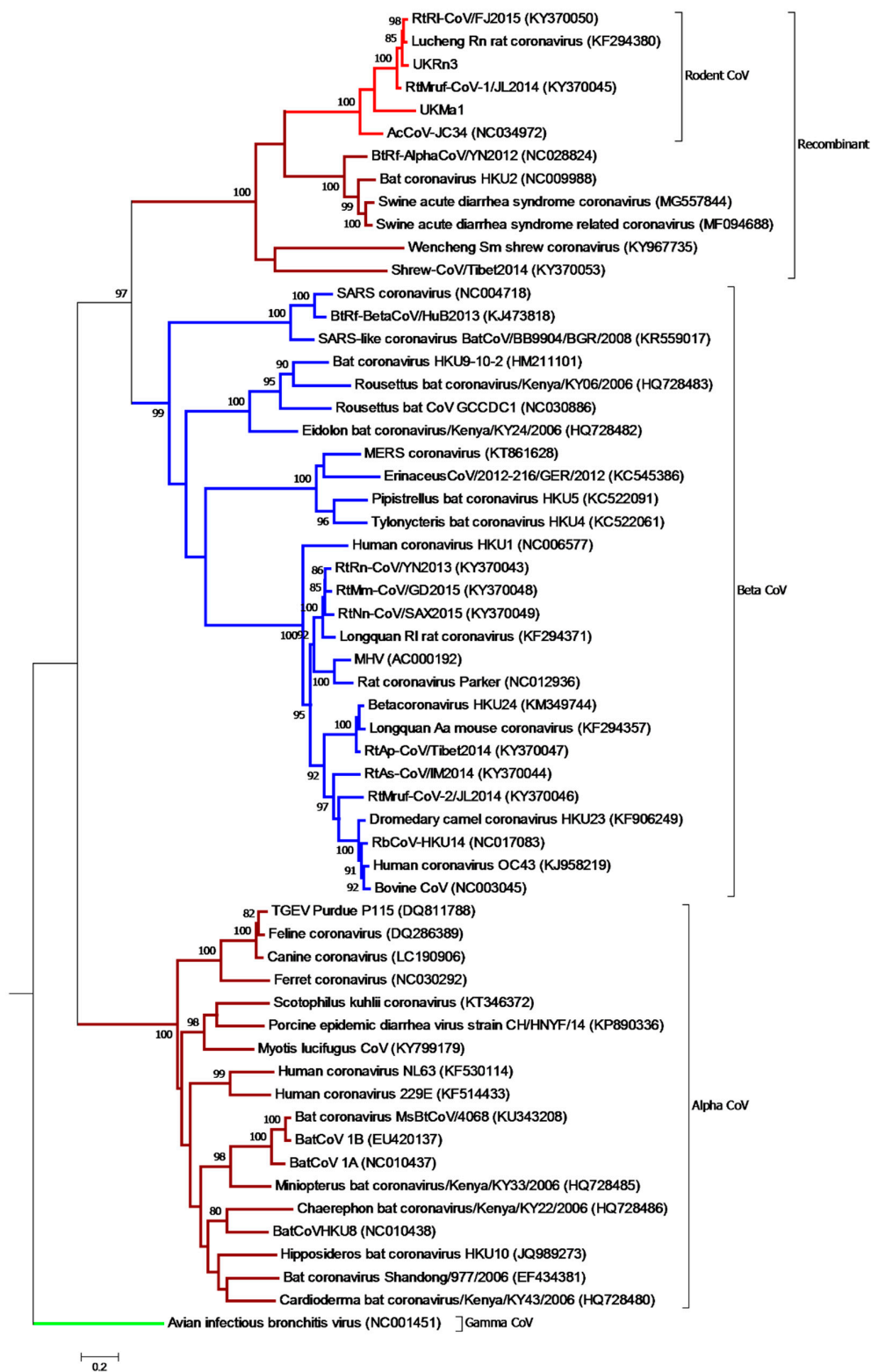
Importantly, amino acid-based ML trees were also inferred for the *ORF1b*, *N*, and *S*, and exhibited the same clustering of rodent alpha-CoVs with high bootstrap support (not shown, available on request).

#### 3.4. Comparison of Virus and Host Phylogenies

To determine the extent to which the rodent CoVs may have co-diverged with their rodent hosts, indicative of a very long-term virus-host association, we examined the extent of congruence between the virus *ORF1b* (with the same topology observed in the *N* gene) and host mitochondrial (mt) *cytochrome b* phylogenies by inferring tanglegrams (Figure 5A). The tanglegram was generated by linking the phylogenies of rodent alpha-CoV *ORF1b* to their equivalent rodent species and assessing whether there were changes in the branching order in the two phylogenies, which would indicate cross-species transmission. The rabbit (*Oryctolagus cuniculus*; order Lagomorpha) sequence was used as outgroup as it has previously been shown to be closely related to the *Rodentia* order [5]. Notably, the topology of the *cytochrome b* tree did not fully match that observed for the rodent alpha-CoVs, in which PLMg1 (*Myodes* spp.) fell basal to all the other viruses, and RtMruf and RCoV/RMU10 3212/Myo gla/GER/2010 (*Myodes* spp.) were more closely related to *Rattus* spp. coronaviruses (UKRn3, Lucheng Rn, RtRl). Moreover, the rodent UKMa1 (*Microtus* spp.) clustered with L232 (*Oryctolagus cuniculus*) in the *ORF1b* tree, again indicative of some cross-species virus transmission.

In addition to the tanglegram, we conducted a co-phylogenetic analysis of the different types of evolutionary events that likely shaped the evolution of the rodent CoVs with their mammalian hosts (mtDNA *cytochrome b*) (Figure 5B). The co-phylogenetic JANE analysis between the rodent alpha-CoV *ORF1b* and the rodent *cytochrome b* phylogenies was performed to assess whether the two phylogenies matched, which would be indicative of co-evolution of virus and host. Under the most likely co-phylogenetic scenario, virus-host co-divergence explained 22% of all evolutionary events, with host-switching (i.e., cross-species transmission) more frequent at 45% (and duplications accounting for 33%), although the relative frequency of these events is likely to change with more extensive sampling. Hence, these data suggest that although the rodent alpha-CoV formed a monophyletic group indicative of a long virus-host association, the evolution of this group of viruses was likely characterized by relatively frequent cross-species transmission.

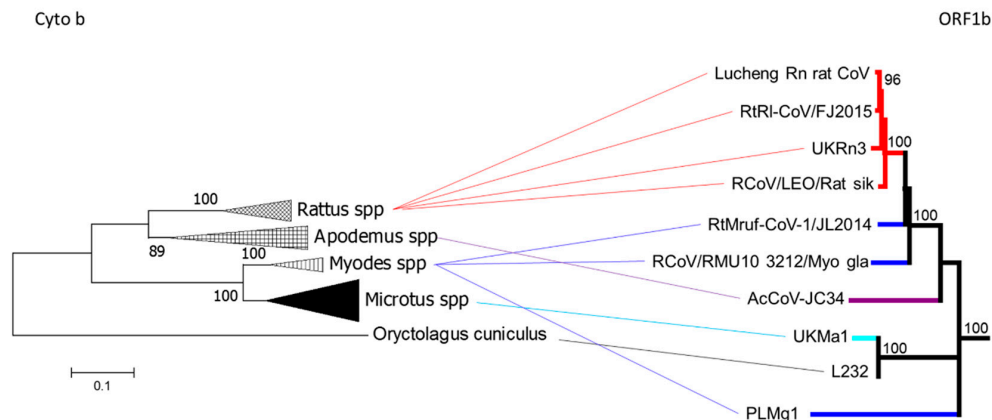
S



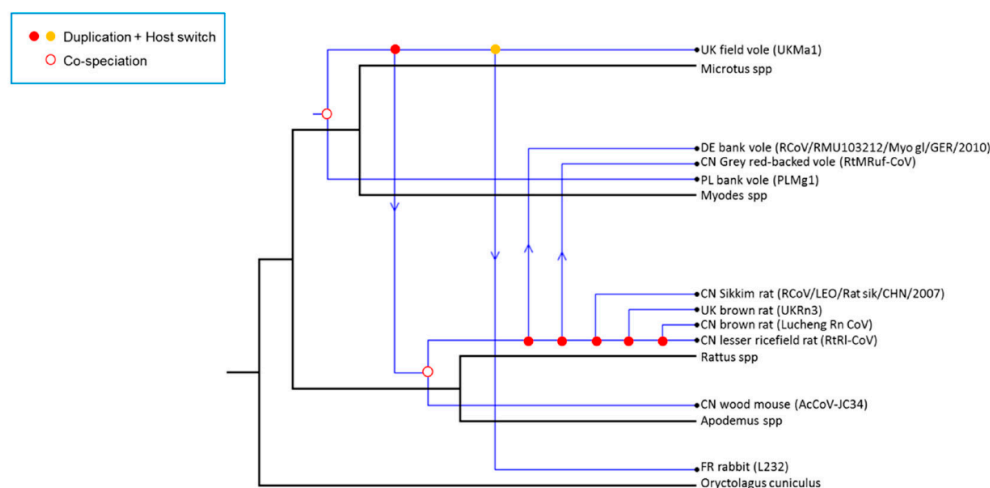
**Figure 4.** Phylogenetic relationships of the rodent alphacoronaviruses based on the S gene. Maximum likelihood phylogenetic analysis of coronavirus partial S gene sequences corresponding to positions 1831–3228 of the Lucheng Rn CoV (KF294380), indicating that the rodent alpha-CoVs formed a distinct clade within the genus *Betacoronavirus*. The novel coronavirus sequence obtained from a UK-resident

*Microtus agrestis* was analysed alongside reference sequences representing three different coronavirus genera (*Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus* as an outgroup). Reference sequences are indicated by their GenBank accession numbers. Branch lengths are drawn to a scale of nucleotide substitutions per site. Numbers above individual branches indicate the percentage that the branch was found in 1000 bootstrap replicates; only values >80% are shown. Alpha-CoVs are highlighted in brown, betacoronaviruses in blue, gammacoronaviruses in green, and the rodent coronaviruses in light red.

A.



B.



**Figure 5.** (A) Tanglegram of rodent *cytochrome b* and CoV *ORF1b* evolution. Maximum likelihood phylogenetic analysis of full-length rodent *cytochrome b* nucleotide sequences compared to rodent alpha-CoV *ORF1b* phylogeny. *Cytochrome b* sequences were obtained from GenBank for the rodent species *Microtus*, *Myodes*, *Rattus*, and *Apodemus*, and the rabbit *Oryctolagus cuniculus*, which was used as an outgroup. Branch lengths are drawn to scale: the bar indicates 0.1 nucleotide substitutions per site. Numbers above individual branches indicate the percentage that that branch was found in 1000 bootstrap replicates; only percentages >80 are shown. Blue indicates the *Myodes* spp., cyan the *Microtus* spp., purple the *Apodemus* spp., red *Rattus* spp., and black indicates *Oryctolagus cuniculus*. (B) Reconciliation of the sampled rodent alpha-CoV phylogeny with that of their mammalian hosts, utilizing the co-phylogenetic method implemented in the Jane package [16]. This figure shows the maximum possible co-divergence, duplication, and host switching events.

#### 4. Discussion

We [9] and others [10–12,17,18] have previously shown the presence of novel CoVs in a range of rodent species. Here, we show that all rodent alpha-CoVs sampled to date—and from West Europe and East Asia—form a monophyletic group and have similar genome structure and common gene constellations and a recombinant spike gene. Hence, these data suggest that the rodent alpha-CoVs sampled to date originate from a single common ancestor and that there is a long association between rodents and CoVs. Moreover, our co-phylogenetic analysis revealed that the phylogenetic diversity in the rodent group was likely shaped by a combination of virus-host divergence and cross-species transmission, which appears to be a common mode of macroevolution in RNA viruses [19].

Also of note was the presence of a recombinant *Betacoronavirus* spike gene that clustered with a number of previously reported recombinant viruses that infect a diverse array of mammalian species, one of which was associated with a recent fatal outbreak in pigs in China [20–23]. Hence, the recombination event involving the *S* gene clearly occurred early on in the evolutionary history of the alpha- and betacoronaviruses.

In addition to the presence of a *Betacoronavirus* spike, the rodent CoVs also had similar, but not identical, gene constellations and genome features. Again, the most parsimonious explanation of this would be that all the rodent alpha-CoVs sampled to date originate from a single common ancestor, and the lack of an *ORF2* (complete deletion) in AcCoV-JC34 likely reflects a secondary evolutionary loss.

Further analysis showed that the rodent alpha-CoVs shared the same predicted core TRS sequences upstream to each gene with a few exceptions. The predominant core TRS was 5'-AACUAA-3'. The same core sequence was also observed in Human NL63 (5'-AACUAAA-3') [24] and Bat HKU2 CoV (5'-AACUAAA-3') [20]. Although it was mentioned that no TRS was detected for *ORF3* in Lucheng Rn CoV [17], we used the core TRS of AcCoV-JC34 (5'-UACUAAA-3') [10] for *ORF3* as a reference, and we were able to find the same sequence in Lucheng Rn CoV, UKRn3, and RtMruf. Other variations such as 5'-CACUAA-3' and 5'-GACUAA-3' were observed in *ORF8* of UKMa1 and RtMruf, respectively. The variation 5'-CUAAAC-3' has been observed in Feline infectious peritonitis virus (FIPV) and Transmissible gastroenteritis virus (TGEV) [25,26]. The conserved core TRS among rodent alpha-CoVs is another indicator that they originate from a common ancestor.

Comparison of rodent species' *cytochrome b* and coronavirus *ORF1b* evolution showed incomplete congruence, such that despite forming a monophyletic group, these rodent CoVs also experienced cross-species transmission. In particular, according to the *cytochrome b* analysis, *Rattus* spp. formed a distinct clade with *Apodemus* spp., whereas *Microtus* spp. formed a separate clade with *Myodes* spp. However, according to *ORF1b* phylogeny, *Rattus* spp. CoVs clustered with two *Myodes* spp. CoV (RtMruf and RCoV/RMU10 3212/Myo gla/GER/2010), suggesting cross-species transmission (and this was confirmed in the co-phylogenetic analysis). Cross-species transmission is frequently observed in coronaviruses and representative examples are SARS-CoV, MERS-CoV, and SADS/SADSR-CoV [21,27].

That the rodent alpha-CoVs form a monophyletic group despite being sampled from multiple localities suggests that they have been associated with rodents for an extended period, although providing an exact chronological context is complex. *Rattus norvegicus* originated in Southeast Asia and spread to Northeast Asia approximately 200,000 years ago. It then spread from Southeast Asia to the Middle East ~3100 years ago, then to Africa ~2000 years ago, and finally to Europe ~1800 years ago [28]. Therefore, it is possible that even though the European UKRn3 coronavirus appears to be approximately as old as Lucheng Rn CoV, the global spread of the brown rat implies that the virus introduction into Europe occurred with that host and was a relatively recent event.

In sum, we show that all the rodent alpha-CoVs described to date share a single common ancestor and have likely evolved through a combination of virus-host co-divergence and cross-species transmission. Not only is this common ancestry evident in the phylogenetic relationships inferred from multiple genes, but also genomic signatures are the recombinant spike gene, same gene constellations (*ORF2*, *ORF3*, *ORF6*, *ORF8*, and *ORF9*), and similar core TRS sequences. Finally, although the current

sequence data set of rodent alphacoronaviruses is relatively small, it is representative of a number of different rodent species living across a wide geographical area.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1999-4915/11/2/125/s1>, Table S1 and Table S2.

**Author Contributions:** T.T., C.P.M. and J.K.B. conceived and designed the experiments; T.T., J.G.C. and O.O. performed the experiments; D.A.M. and A.R.F. performed NGS on our rodent alpha-CoV samples; E.M.-L., G.U. and R.E.T. provided samples; M.A.M., J.F.D. and C.D. provided training and contributed in genome retrieval of the viruses; T.T., E.C.H. and J.K.B. analysed the data and wrote the paper; all authors edited the final draft.

**Funding:** This study was supported by the Biotechnology and Biological Sciences Research Council [grant number BB/J014508/1]. ECH is supported by an ARC Australian Laureate Fellowship (FL170100022). JFD is supported by the European Union's Horizon 2020 research and innovation program through the COMPARE project (grant agreement no. 643476).

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Hudson, C.; Beaudette, F.R. Infection of the cloaca with the virus of infectious bronchitis. *Science* **1932**, *76*, 34. [[CrossRef](#)]
- Hamre, D.; Procknow, J.J. A new virus isolated from the human respiratory tract. *Proc. Soc. Exp. Biol. Med.* **1966**, *121*, 190–193. [[CrossRef](#)]
- Weiss, S.; Navas-Martin, S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 635–664. [[CrossRef](#)]
- Luis, A.D.; Hayman, D.T.; O'Shea, T.J.; Cryan, P.M.; Gilbert, A.T.; Pulliam, J.R.; Mills, J.N.; Timonin, M.E.; Willis, C.K.; Cunningham, A.A.; et al. A comparison of bats and rodents as reservoirs of zoonotic viruses: Are bats special? *Proc. Biol. Sci.* **2013**, *280*. [[CrossRef](#)]
- Huchon, D.; Madsen, O.; Sibbald, M.J.; Ament, K.; Stanhope, M.J.; Catzeflis, F.; de Jong, W.W.; Douzery, E.J. Rodent phylogeny and a timescale for the evolution of Glires: Evidence from an extensive taxon sampling using three nuclear genes. *Mol. Biol. Evol.* **2002**, *19*, 1053–1065. [[CrossRef](#)]
- Phan, T.G.; Kapusinszky, B.; Wang, C.; Rose, R.K.; Lipton, H.L.; Delwart, E.L. The fecal viral flora of wild rodents. *PLoS Pathog.* **2011**, *7*, e1002218. [[CrossRef](#)]
- Cheever, F.S.; Daniels, J.B.; Pappenheimer, A.M.; Bailey, O.T. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin. *J. Exp. Med.* **1949**, *90*, 181–194. [[CrossRef](#)]
- Parker, J.C.; Cross, S.S.; Rowe, W.P. Rat coronavirus (RCV): A prevalent, naturally occurring pneumotropic virus of rats. *Arch. Gesamte Virusforsch.* **1970**, *31*, 293–302. [[CrossRef](#)] [[PubMed](#)]
- Tsoleridis, T.; Onianwa, O.; Horncastle, E.; Dayman, E.; Zhu, M.; Danjitrang, T.; Wachtl, M.; Behnke, J.M.; Chapman, S.; Strong, V.; et al. Discovery of Novel Alphacoronaviruses in European Rodents and Shrews. *Viruses* **2016**, *8*, 84. [[CrossRef](#)] [[PubMed](#)]
- Ge, X.Y.; Yang, W.H.; Zhou, J.H.; Li, B.; Zhang, W.; Shi, Z.L.; Zhang, Y.Z. Detection of alpha- and betacoronaviruses in rodents from Yunnan, China. *Virol. J.* **2017**, *14*, 98. [[CrossRef](#)]
- Hu, D.; Zhu, C.; Wang, Y.; Ai, L.; Yang, L.; Ye, F.; Ding, C.; Chen, J.; He, B.; Zhu, J.; et al. Virome analysis for identification of novel mammalian viruses in bats from Southeast China. *Sci. Rep.* **2017**, *7*, 10917. [[CrossRef](#)]
- Monchatre-Leroy, E.; Boué, F.; Boucher, J.-M.; Renault, C.; Moutou, F.; Ar Gouilh, M.; Umhang, G. Identification of Alpha and Beta Coronavirus in Wildlife Species in France: Bats, Rodents, Rabbits, and Hedgehogs. *Viruses* **2017**, *9*, 364. [[CrossRef](#)]
- Shi, M.; Lin, X.D.; Tian, J.H.; Chen, L.J.; Chen, X.; Li, C.X.; Qin, X.C.; Li, J.; Cao, J.P.; Eden, J.S.; et al. Redefining the invertebrate RNA virosphere. *Nature* **2016**. [[CrossRef](#)]
- Thompson, J.D.; Higgins, D.G.; Gibson, T.J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680. [[CrossRef](#)]
- Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874. [[CrossRef](#)]
- Conow, C.; Fielder, D.; Ovadia, Y.; Libeskind-Hadas, R. Jane: A new tool for the cophylogeny reconstruction problem. *Algorithms for Mol. Biol.* **2010**, *5*, 16. [[CrossRef](#)]



17. Wang, W.; Lin, X.D.; Guo, W.P.; Zhou, R.H.; Wang, M.R.; Wang, C.Q.; Ge, S.; Mei, S.H.; Li, M.H.; Shi, M.; et al. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology* **2015**, *474*, 19–27. [[CrossRef](#)]
18. Lau, S.K.; Woo, P.C.; Li, K.S.; Tsang, A.K.; Fan, R.Y.; Luk, H.K.; Cai, J.P.; Chan, K.H.; Zheng, B.J.; Wang, M.; et al. Discovery of a novel coronavirus, China Rattus coronavirus HKU24, from Norway rats supports the murine origin of Betacoronavirus 1 and has implications for the ancestor of Betacoronavirus lineage A. *J. Virol.* **2015**, *89*, 3076–3092. [[CrossRef](#)]
19. Geoghegan, J.L.; Duchene, S.; Holmes, E.C. Comparative analysis estimates the relative frequencies of co-divergence and cross-species transmission within viral families. *PLoS Pathog.* **2017**, *13*, e1006215. [[CrossRef](#)]
20. Lau, S.K.; Woo, P.C.; Li, K.S.; Huang, Y.; Wang, M.; Lam, C.S.; Xu, H.; Guo, R.; Chan, K.H.; Zheng, B.J.; et al. 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* **2007**, *367*, 428–439. [[CrossRef](#)]
21. Zhou, P.; Fan, H.; Lan, T.; Yang, X.-L.; Shi, W.-F.; Zhang, W.; Zhu, Y.; Zhang, Y.-W.; Xie, Q.-M.; Mani, S.; et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* **2018**, *556*, 255–258. [[CrossRef](#)]
22. Wang, W.; Lin, X.D.; Liao, Y.; Guan, X.Q.; Guo, W.P.; Xing, J.G.; Holmes, E.C.; Zhang, Y.Z. Discovery of a highly divergent coronavirus in the Asian house shrew from China illuminates the origin of the alphacoronaviruses. *J. Virol.* **2017**. [[CrossRef](#)]
23. Wu, Z.; Yang, L.; Ren, X.; He, G.; Zhang, J.; Yang, J.; Qian, Z.; Dong, J.; Sun, L.; Zhu, Y.; et al. Deciphering the bat virome catalog to better understand the ecological diversity of bat viruses and the bat origin of emerging infectious diseases. *ISME J.* **2016**, *10*, 609–620. [[CrossRef](#)]
24. Pyrc, K.; Jebbink, M.F.; Berkhout, B.; van der Hoek, L. Genome structure and transcriptional regulation of human coronavirus NL63. *Virol. J.* **2004**, *1*, 7. [[CrossRef](#)]
25. Dye, C.; Siddell, S.G. Genomic RNA sequence of Feline coronavirus strain FIPV WSU-79/1146. *J. Gen. Virol.* **2005**, *86*, 2249–2253. [[CrossRef](#)]
26. Hiscox, J.A.; Mawditt, K.L.; Cavanagh, D.; Britton, P. Investigation of the control of coronavirus subgenomic mRNA transcription by using T7-generated negative-sense RNA transcripts. *J. Virol.* **1995**, *69*, 6219–6227.
27. De Wit, E.; van Doremalen, N.; Falzarano, D.; Munster, V.J. SARS and MERS: Recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **2016**, *14*, 523–534. [[CrossRef](#)]
28. Zeng, L.; Ming, C.; Li, Y.; Su, L.-Y.; Su, Y.-H.; Otecko, N.O.; Dalecky, A.; Donnellan, S.; Aplin, K.; Liu, X.-H.; et al. Out of Southern East Asia of the Brown Rat Revealed by Large-Scale Genome Sequencing. *Mol. Biol. Evol.* **2018**, *35*, 149–158. [[CrossRef](#)]

