JOURNAL OF GENERAL VIROLOGY

RESEARCH ARTICLE

Apaa *et al.*, *Journal of General Virology* 2023;104:001859
DOI 10.1099/jgv.0.001859

MICROBIOLOGY SOCIETY

OPEN ACCESS

Sarbecoviruses of British horseshoe bats; sequence variation and epidemiology

Ternenge Apaa[1,2], Amy J. Withers[1,2], Ceri Staley[1], Adam Blanchard[1], Malcolm Bennett[1], Samantha Bremner-Harrison[3,4], Elizabeth A. Chadwick[3,5], Frank Hailer[5], Stephen W. R. Harrison[3], Matthew Loose[6], Fiona Mathews[7] and Rachael Tarlinton[1,*]

Abstract

Horseshoe bats are the natural hosts of the *Sarbecovirus* subgenus that includes SARS-CoV and SARS-CoV- 2. Despite the devastating impact of the COVID-19 pandemic, there is still little known about the underlying epidemiology and virology of sarbecoviruses in their natural hosts, leaving large gaps in our pandemic preparedness. Here we describe the results of PCR testing for sarbecoviruses in the two horseshoe bat species (*Rhinolophus hipposideros* and *R. ferrumequinum*) present in Great Britain, collected in 2021–22 during the peak of COVID-19 pandemic. One hundred and ninety seven *R. hipposideros* samples from 33 roost sites and 277 *R. ferrumequinum* samples from 20 roost sites were tested. No coronaviruses were detected in any samples from *R. ferrumequinum* whereas 44 and 56% of individual and pooled (respectively) faecal samples from *R. hipposideros* across multiple roost sites tested positive in a sarbecovirus-specific qPCR. Full genome sequences were generated from three of the positive samples (and partial genomes from two more) using Illumina RNAseq on unenriched samples. Phylogenetic analyses showed that the obtained sequences belong to the same monophyletic clade, with >95% similarity to previously-reported European isolates from *R. hipposideros*. The sequences differed in the presence or absence of accessory genes ORF 7b, 9b and 10. All lacked the furin cleavage site of SARS-CoV-2 spike gene and are therefore unlikely to be infective for humans. These results demonstrate a lack, or at least low incidence, of SARS-CoV-2 spill over from humans to susceptible GB bats, and confirm that sarbecovirus infection is widespread in *R. hipposideros*. Despite frequently sharing roost sites with *R. ferrumequinum*, no evidence of cross-species transmission was found.

BACKGROUND

The most widely accepted explanation for the origin of the SARS-CoV-2 pandemic is that it arose from animals held in the Wuhan market in late 2019 [1]. It has been proposed that masked palm civets (*Paguma larvatai*) acted a bridging host for SARS-CoV transmission from an unknown ancestral reservoir in a species of horseshoe bats to humans [2, 3], and it is likely that intermediate hosts were also involved in the emergence of SARS-CoV-2. Widespread onward transmission of SARS-CoV-2 from humans to other mammals occurred during the human covid pandemic, including large outbreaks in farmed mink (*Neovison vison*) [4–7], the establishment of a new reservoir in wild white-tailed deer (*Odocoileus virginianus*) in the USA [8, 9], and repeated infection (though without the establishment of endemic transmission) in domestic cats (*Felis catus*) [10–13]. Reports of transmission

Received 20 February 2023; Accepted 11 May 2023; Published 14 June 2023
Author affiliations: [1]School of Veterinary Medicine and Science, University of Nottingham, Sutton Bonington, UK; [2]Animal and Plant Health Agency (APHA), Addlestone, UK; [3]School of Animal, Rural and Environmental Sciences, Nottingham Trent University, Southwell, UK; [4]Vincent Wildlife Trust, Herefordshire, UK; [5]Organisms and Environment, School of Biosciences, Cardiff University, UK; [6]School of Life Sciences, University of Nottingham, Nottingham, UK; [7]School of Life sciences, University of Sussex, Brighton, UK.
***Correspondence:** Rachael Tarlinton, rachael.tarlinton@nottingham.ac.uk
Keywords: coronavirus; horseshoe bat; *Rhinolophus hipposideros*; *Rhinolophus ferrumequinum*; sarbecovirus; UK.
Abbreviations: E, envelope; hACE2, human angiotensin converting enzyme 2; RBD, receptor binding domain; RBM, receptor binding motif; RDRP, RNA-dependent RNA polymerase; S, spike; SARs-CoV-2, severe acute respiratory syndrome coronavirus 2; TPMRSS2, transmembrane serine protease 2.
All GB lesser horseshoe Sarbecovirus genomes assembled from this study have been deposited in the NCBI GenBank database under accession numbers OP776338 — OP776340 and OP837780 — OP837781. In addition, Illumina read datasets generated have been submitted under Bioproject accession number PRJNA895830, SRA accession numbers SRR22103527 — SRR22103531 and Biosample accession numbers SAMN31525260 — SAMN31525264.
Eight supplementary figures and six supplementary tables are available with the online version of this article.

back to humans from these species have also occurred [6, 14, 15] highlighting the risk of new animal reservoirs developing, potentially as sources of new viral variants that might evade vaccine-induced immunity. Sporadic cases have also been reported in domestic dogs (*Canis familiaris*), hamsters (*Mesocricetus auratus*), cattle (*Bos Taurus*) and camels (*Camelus dromedarius*), and a large range of primates, non-domestic felids, mustelids, hippos (*Hippopotamus amphibius*) and manatees (*Trichechus manatus*) (reviewed in [16]).

The natural hosts of sarbecoviruses (SARS-like betacoronaviruses) are insectivorous horseshoe bats (represented by a single extant genus *Rhinolophus* within the family Rhinolophidae; superfamily Rhinolophoidea; sub-order Yinpterochiroptera) and the related old world roundleaf bats of the family Hipposideridae which share the same superfamily. There are around 180 known species of Rhinolophidae and Hipposideridae with ranges across Eurasia and Africa, and sarbecoviruses have been detected in at least 30 of them [17]. For species and sites where more extensive studies have been performed (generally SE Asia), it is apparent that species- and site-specific clustering of virus sequences occurs [18, 19]. Geographical hotspots of both bat and viral species are apparent in SE Asia [17, 20].

There are, in contrast, very few reports of sarbecoviruses in the sub-order Yangochiroptera: a small number of isolated reports in individual animals or pooled samples of Molossidae and Vespertilionidae families have been recorded in multispecies studies [17, 21, 22].

There is also evidence from multiple studies for considerable variation in both *Sarbecovirus* carriage and sequences among host and viral lineages in bats, with implications for potential cross species transmission [22]. While the exact determinants of what makes sarbecoviruses more likely to cross the species barrier into humans are not known, one clear limiting factor is the ability of the virus spike protein (the protein that binds to the host cell facilitating virus entry) to bind to the human ACE-2 protein (the receptor for SARS-CoV and SARS-CoV-2 in humans) [23]. Considerable mutation and antigenic variation is present in this spike protein in SARS-CoV-2 isolates and bat sarbecoviruses [18, 24]. In SARs-CoV-2 this protein also displays a distinctive cleavage site for the furin protease that is missing in many bat isolates and appears to be a critical factor in transmission and pathogenesis in humans [25]. Complicating matters further, it is apparent that there is frequent viral recombination between different sarbecovirus lineages [18, 19].

Two species of horseshoe bats are present in Great Britain, the greater horseshoe bat (*Rhinolophus ferrumequinum*) and lesser horseshoe bat (*Rhinolophus hipposideros*), both with distributions largely restricted to southwest England and Wales. Neither are common at a national scale and both are regarded as rare in Europe, and hence are protected under the Habitats Directive and Eurobats Agreement [26, 27]. In Britain, populations comprise approximately 13000 greater horseshoe bats and 50000 lesser horseshoe bats [28, 29]. Britain and Ireland form the extreme north-western end of their geographic ranges (only *R. hipposideros* is found in Ireland), which extends through central and southern Europe to Central Asia for lesser horseshoe bats and across Asia to Japan for greater horseshoe bats. The two species can share roost sites and both use a variety of roosts across the year, including maternity and hibernation sites [30].

Sarbecoviruses have been reported previously from both bat species. Partial RDRP (Rnase dependent RNA polymerase) sequences have been described from Chinese, Italian and Bulgarian greater horseshoe bats [31–35] at a prevalence of 2.2–42% (6/45 13%, 2/53 3.8%, 19/45 42%, 1/45 2.2%) in European samples. Full sequences are also available for a number of Chinese, one Korean and one Russian greater horseshoe bat sarbecoviruses; the Asian sequences are phylogenetically very distinct from the European isolates [36–39].

Similarly, partial RDRP sequences have been reported in Slovenian lesser horseshoe bats [40], at a prevalence of 31% (14/36), and Slovenian and Spanish lesser horseshoe bats at a prevalence of 14% (5/36) and 7.1% (21/285) [35]. More recently, full genome sequences from Russian and British lesser horseshoe bats [36, 41] have also been reported.

Resequencing of partial RDRP genes from Spanish, Italian and Slovenian isolates (and the full genome sequences for the UK and Russian isolates) from both species has demonstrated a lack of the furin cleavage site present in SARS-CoV-2 [38]. The European greater and lesser horseshoe bat sequences are also phylogenetically distinct from each other [38] with the two Russian isolates (one from each species) displaying 59–95% amino acid similarity to each other depending on the gene [36].

In addition to their being hosts of various sarbecoviruses [20], horseshoe bats have been identified as potentially susceptible to infection with SARS-CoV-2 [41–44]. This led us to screen British horseshoe bat populations, at the height of the SARS-CoV-2 pandemic in western Europe, for the presence of sarbecoviruses with the aim of confirming or ruling out the establishment of SARS-CoV-2 circulation in these animals.

METHODS

Sample collection

A total of 517 bat samples (oronasal swabs, external swabs of the rectal region, and faecal samples), including 474 from horseshoe bats, (Table 1) were collected from roost sites or animals in care in the UK between April 2021 and February 2022, during which

Table 1. Summary of bat species, sample type and positivity rates for Sarbecovirus E gene qPCR

Bat species	Type of sample	no. of samples	Sarbecovirus E gene qPCR
Lesser horseshoe	Oronasal swabs	71	0 (0 %)
	Rectal swabs	56	3 (5.4 %)
	Individual faeces samples	39	17 (43.5 %)
	Pooled faeces samples	31	17 (54.8 %)
Greater horseshoe	Oronasal swabs	90	0
	Rectal swabs	69	0
	Individual faeces samples	91	0
	Pooled faeces samples	28	0
Common pipistrelle	Oronasal swabs	3	0
	Rectal swabs	–	–
	Individual faeces samples	7	0
	Pooled faeces samples	–	–
Natterer's	Oronasal swabs	1	0
	Rectal swabs	–	–
	Individual faeces samples	–	–
	Pooled faeces samples	–	–
Daubenton's	Oronasal swabs	5	0
	Rectal swabs	8	0
	Individual faeces samples	16	0
	Pooled faeces samples	–	–

time human infection in the UK underwent successive alpha, delta and omicron SARS-CoV-2 epidemic waves [45]. Faecal samples included those collected from individuals in capture bags during routine population monitoring (during which sex and age were also recorded where possible) and pooled samples of fresh droppings collected from the floor of bat roosts. All samples were collected secondarily to regular licenced population monitoring efforts. Handling of animals was conducted under Natural England Licence 2022–61108-Sci-Sci (Mathews), and followed best practice guidelines for minimising the risk of human to bat SARS-CoV-2 transmission. Ethical approval was granted by the University of Nottingham School of Veterinary Medicine and Science Committee for Animal Research and Ethics (CARE), and the University of Sussex Animal Welfare and Ethical Review Board. Samples were preserved in RNAlater at room temperature and sent to the University of Nottingham where they were stored at −20 °C until RNA extraction. Samples were collected from five bat species from 54 sites: 197 from lesser horseshoe bats (*Rhinolophus hipposideros*) at 33 sites, 277 for greater horseshoe bats (*Rhinolophus ferrumequinum*) from 20 sites, 10 from common pipistrelles (*Pipistrellus pipistrellus*), 32 from Daubentons bats (*Myotis daubentonii*) and one from a Natterer's bat (*Myotis nattereri*) (Table 1 and Supplementary information, available in the online version of this article).

RNA extraction, reverse transcriptase (RT) and RNA-dependent RNA polymerase (RDRP) gene coronaviruses generic conventional PCR and envelope gene sarbecovirus-specific real-time PCR

RNA extraction from bat faecal and oronasal swabs, and cell culture supernatant as positive control, was carried out using the Macherey-Nagel RNA tissue extraction kit as per manufacturer's instructions. The Wuhan SARS-CoV-2 strain positive control sample used throughout this study was kindly donated by Dr Christopher Coleman (Division of Infection, Immunity and Microbes, School of Life Sciences, University of Nottingham, UK). RT was performed in two steps, using M-MLV-RT and random hexamer primers (Promega) as per manufacturer's instructions. All cDNA products were stored at −20 °C for conventional PCR.

A generic pan-coronavirus PCR assay published by [46] was used to amplify a 440 bp fragment of the coronavirus RDRP gene using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs cat no: M0493S). PCR products were purified using the Nucleospin extract II kit (Macherey-Nagel) according to manufacturer's instructions and were Sanger sequenced (Eurofins UK).

Real-time PCR was carried out using the Promega GoTaq Probe 1-Step RT-qPCR System (Promega) with Sarbecovirus-specific envelope gene primers from bat RNA samples as published by [47].

RNA and cDNA quality control was assessed via partial amplification of 148 bp of the bat mitochondrial cytochrome b gene using a published conventional PCR protocol [48].

Next generation sequencing and genome analyses

RNA sequencing was performed by Novogene UK, using the Illumina NovaSeq 6000 platform. Quality filtering and trimming to remove adapters, duplicates and low quality reads was achieved using fastp v0.23.1 [49]. Kraken2 v2.1.2 was used for taxonomic classification reads against the Kraken2 viral Refseq database [50] (retrieved on 9 June 2022). Reads were assembled using the coronaSPAdes option in SPAdes genome assembler v3.15.4 [51] using default parameters. While CheckV v1.0.1, a fully automated command-line pipeline, was used for identification and quality assessment of contigs, contigs were also queried against the NCBI custom BLASTN (v2.12.0) viral database [52] (retrieved on 3 July 2022) to confirm results from CheckV. Assembled contigs were indexed and contigs that were classified and assessed as complete bat Sarbecovirus genomes were extracted for downstream analysis using samtools v1.16.1 faidx option [53]. Individual reads for each sample were mapped back to identified contigs using Minimap2 [54], read coverage and depth were generated using samtools [53]. Assembled genomes were annotated in Geneious Prime (v.2022.2.2) using NCBI coronavirus reference sequences and visualized for the presence of the structural, non-structural protein, accessory genes and to generate linear genome maps and data on individual gene location, composition, and nucleotide length. Reads were also mapped to the prototype European horseshoe bat (*Rhinolophus blasii,* Blasius horseshoe bat Bulgarian isolate) sarbecovirus reference genome (NC014470) using BBMap for variant calling, SNPs were viewed, and data exported in Geneious Prime (v.2022.2.2).

Phylogenetic analysis

Complete coronavirus genomes, extracted RDRP, spike, envelope and nucleocapsid nucleotide sequences from *Sarbecovirus* genomes assembled in this study, and a total of 198 reference *Sarbecovirus* genomes (including non-human *Sarbecovirus* isolates, SARS-CoV and SARS-CoV-2) downloaded from NCBI, were aligned using mafft v7.490 [55]. Maximum likelihood phylogenetic trees were reconstructed based on complete coronavirus genomes, and four different genes using IQ-TREE v2.0.7 [56], using 1000 bootstrap approximations following implementation of UFBoot2 within IQ-TREE v2.0.7 to evaluate branch support [57]. The ModelFinder option was included in command-line to select the best fitting nucleotide substitution model for phylogenetic reconstruction [58]. Phylogenetic trees were visualized and annotated in FigTree v1.4.4 (https://github.com/rambaut/figtree/).

Spike glycoprotein comparison, identification of furin cleavage site (FCS) and transmembrane protease serine 2 (TMPRSS2), receptor binding domain (RBD) homology modelling and structural analysis

To identify and compare the presence of FCS and TMPRSS2 between UK Coronavirus genomes and related CoVs, RBD spike proteins from the six UK Coronaviruses (five from this study and one reported by [41]), SARS-CoV, SARS-CoV-2 and related Beta coronaviruses were aligned, viewed and pictures generated in Jalview [59]. The model for the receptor binding domain (RBD) and hACE2 protein complex were constructed using SWISS-MODEL to assess the amino acid residues and structural differences between UK bat CoV, SARS-CoV, and SARS-CoV-2 using an RBD-hACE2 complex. The SWISS-MODEL template library was search for evolutionary related structures matching the target UK Coronavirus RBD (residues 321–515) and hACE2 (PDB: Q9BYF1) protein sequences using BLAST HHblits database. The most suitable template (PDB: 6vw1) with the highest selected Global Model Quality Estimate (GMQE) of 0.75 (range: 0–1) and the lowest average Quaternary Structure Quality Mean Estimate (QMEAN) of – 0.64 was selected to build models for the five UK Coronavirus RBD proteins generated from this study. The structural analysis and verification server (SAVES) was used to validate 3D structures, create Ramachandran plots, and classify amino acid residue torsion angles of each structure via implementation of PROCHECK AND ERRAT2 [60], respectively. In addition, receptor binding residues detection and binding energies of 3D model complex structures generated were calculated using PRODIGY [61, 62]. Visualization, superimposition, alignment, and generation of figures for presentation was done using UCSF Chimaera [63].

Recombination analysis

To assess for the presence of potential recombination events, a total of 24 complete coronavirus genomes, including 19 reference and the five bat *Sarbecovirus* assembled genomes from this study (Supplementary information), were aligned using mafft v7.490 [55]. Eight recombination detection methods were executed in RDP5 to detect combination events including RDP (R), GENECONV (G), Bootscan (B), Maxchi (M), Chimaera (C), SiSscan (S), 3Seq (T), and LARD (L) [64]. Only a series of recombination events detected by at least six of these methods were considered as indicative of recombination.

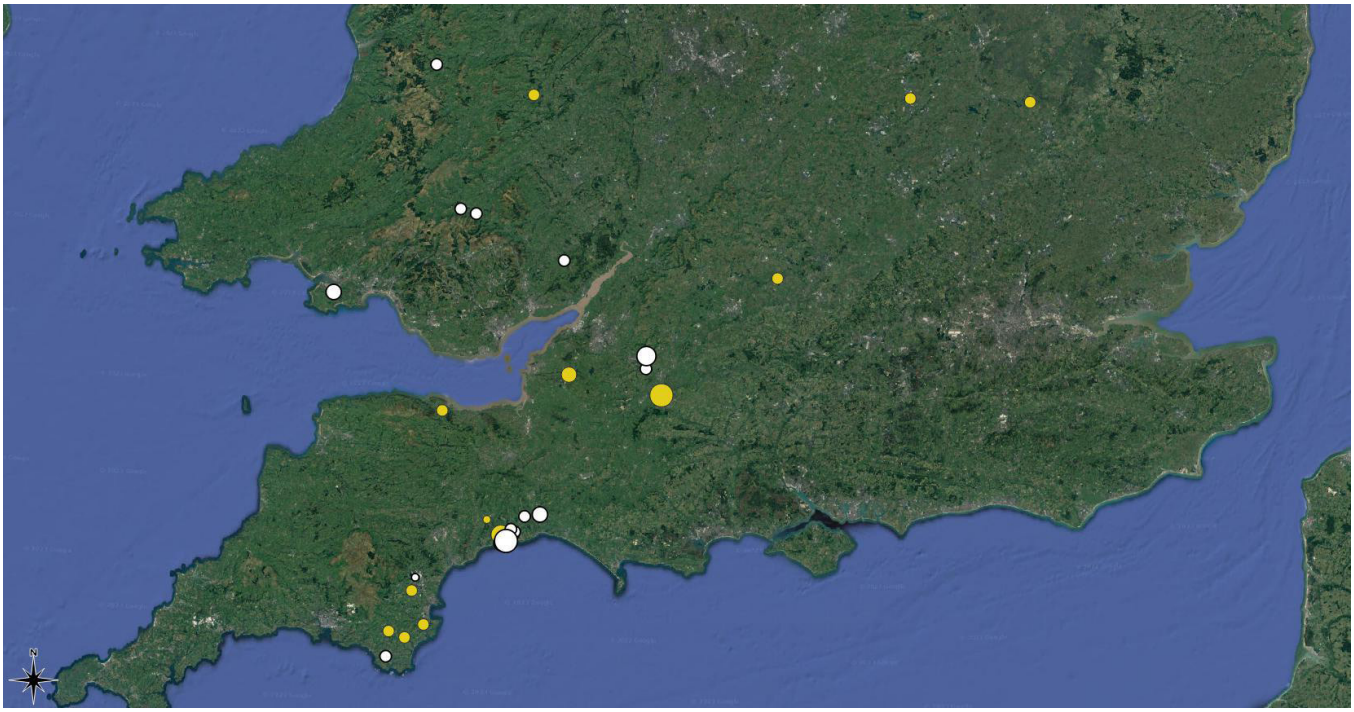


Fig. 1. Roost sites of *Sarbecovirus* detected by PCR in faeces from lesser horseshoe bats (*Rhinolophus hipposideros*), yellow=negative, white=positive. Circle sizes are proportional to number of samples from the site.

Geospatial mapping and data analysis

Maps of roost sites positive and negative for coronavirus PCR were created in QGIS (version 3.28.2). To account for pooled and individual samples, map circle sizes were set to five proportional sizes as follows: <3 bats, 3–5 or one pooled sample, 5–10 (pools count as 3), 10–120 (pools count as 3) and >20 (pools count as 3). Chi-square test was used to evaluate potential differences in prevalence with sex and age-class (SPSS, Amos 28.0.0).

RESULTS

No samples tested positive in the pan-coronavirus assay, and no oronasal swabs tested positive in the *Sarbecovirus*-specific PCR assay for either species, however, numerous lesser horseshoe bat faecal samples tested positive with the *Sarbecovirus* qPCR assay (Table 1). Within the five roosts from which >5 individual faecal samples and/or rectal swabs were collected, all yielded at least one positive sample, and the percentage positivity ranged from 5–25% (mean at individual level across those sites, 16/100 individuals, 16%, 95% CI [Wilson's] 10.0–24.6%). Twenty-two sites yielded only single, pooled samples, of which 10 (45%; 95% CI [Wilson's] 27.0–65.3%) were positive (Supplementary information). The locations and infection status of lesser horseshoe bat roost sites yielding either >3 individual faecal samples or at least one pooled faecal sample are shown in Fig. 1; of these, 21/31 (68%; 95% CI 50.0–81.4%) roosts yielded at least one positive sample.

Age and sex were recorded for 41 and 37 animals, respectively, when individual faecal samples were collected and examined; among this subset of animals tested, there were no significant (chi-square) differences in the frequency of positive samples by age (19.5% adult, 8% juvenile) or sex (12.5% female, 21.9% male).

Taxonomic classification, genome assembly

Taxonomic classification using Kraken2 identified reads assigned to other viral operational taxonomic units, however, only reads classified to the Coronaviridae viral family are reported in this study. Paired end reads were generated from RNA extracted from six lesser horseshoe bat faecal samples. Between 98–99% of reads were unclassified. The percentage of classified viral reads from the datasets examined varied from 0.49–1.18%, while the total percentage of classified viral reads assigned to the family Coronaviridae varied from 0.01–63.58%. Three samples including RhGB04, RhGB05 and RhGB06 recorded approximately 50% of classified viral reads assigned to Coronaviridae (Supplementary information).

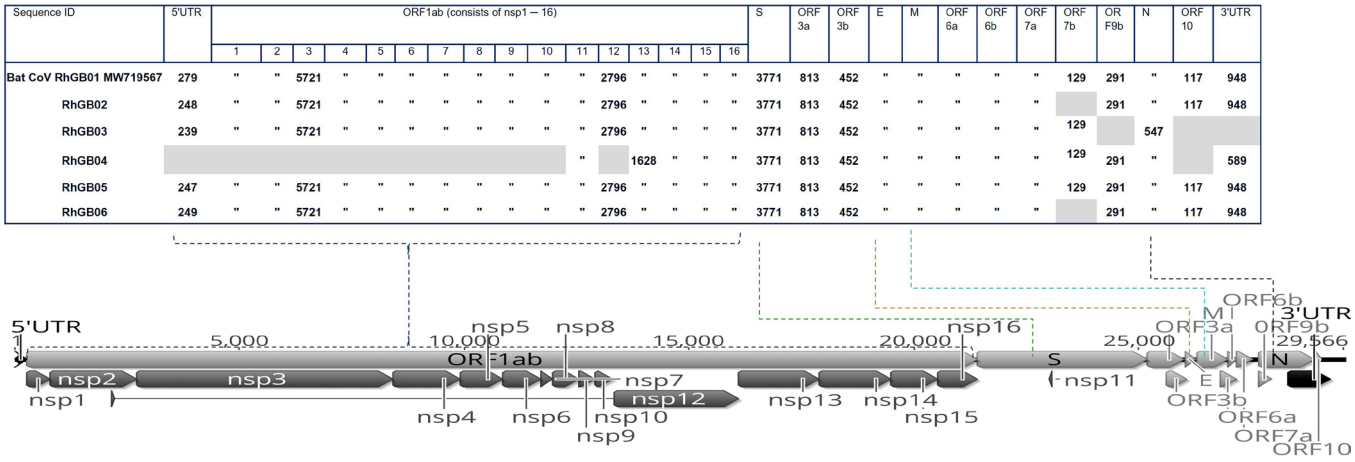


Fig. 2. Genome organisation of lesser horseshoe bat *Sarbecovirus* sequences derived from this study and the reference UK genome RhGB001. Missing genes are shown in grey, lengths of nucleic acid segments are listed for each gene, LTR and isolate. RhGB04 is a partial sequence missing the 5′ end of the genome. RhGB03 is likewise an incomplete genome missing ORF10 and the 3′UTR.

De novo assembly of datasets from lesser horseshoe bats yielded five genomes made up of single contigs. The length of the assembled genomes varied from 28.2–30.6 kb with one short contig of 12.6 k. CheckV analysis demonstrated that all the assembled contigs were closely related to beta coronaviruses, specifically the subgenus Sarbecovirus. Sarbecovirus genomes identified were assessed to have 94–100% (4/5) and 42% (1/%) quality/completeness, all genomes had 0% contamination except for RhGB05 with 3.17% contamination. Overall, all the assembled genomes shared 94.5–97.7% average amino acid identity with the reference genome (RhGB01 MW1957) identified by CheckV (Supplementary information).

Assessment of reads mapped to the assembled genomes demonstrated over 7000, 4000, 302000, 420000 and 500000 reads mapped to RhGB02, RhGB03, RhGB04, RhGB05 and RhGB06, respectively. The genome assembled from RhGB03 had the least average mean depth of 22X and was missing ORF10 and the 3′ UTR, while RhGB04 (partial genome of 12.6 kb missing the 5′ end of ORF 1ab) recorded the highest mean depth coverage of 3446X (Supplementary information). RhGB03 sequence was not of sufficient quality for submission to Genbank due to multiple stop codons. The other sequences were submitted with accession numbers OP776338–OP776340 and OP837780–OP837781. RhGB03 analysis has been included for the sake of completeness except for phylogenetic analysis, where its sequence quality was too low (likely due to very low sequence coverage).

Genome annotation and organization of British bat CoVs

Genome annotation in Geneious Prime using NCBI SARS-CoV-2 and SARS-related reference sequences, confirmed a similar genome organization to the UK lesser horseshoe bat reference genome available from the NCBI database (Fig. 2). *Sarbecovirus* genomes consist of a leader sequence (5′UTR), followed by ORF1ab gene with sixteen non-structural proteins (nsp1–16) making up about 2/3 of the viral genome. The assembled sarbecoviruses were made up of four major structural proteins including the spike (S), membrane (M), envelope (E) and nucleocapsid (N) proteins. While accessory proteins including ORF3b, 6a, 6b, 7a and variable ORFs recognisable for 7b, 9b and 10 were reported within the 3′ region interspaced between the major structural proteins (Fig. 2).

Variant calling

Variant calling following mapping of raw reads to the bat Bulgarian Sarbecovirus reference genome (Bat CoV BM48-31) available in Europe, demonstrated the presence of a series of single nucleotide polymorphism (SNPs) characterized by amino acid substitution distributed throughout the entire length of the genome. However, only a total of 22 unique SNPs were identified within the spike glycoprotein (Supplementary information) with minimum read coverage ≥10X, mapped quality ≥10% and variant frequency >95% from the four nearly full genomes assembled (Supplementary information for details). Overall, 15/22 SNPs resulting in amino acid substitution were found in RhGB02, 12/22 SNPs identified in RhGB03 (12/22) though sequencing depth was low for this sample and 8/22 SNPs were reported from both RhGB05 and RhGB06 with the highest read-depth coverage. One unique SNP at coding sequence position 3641 (V to A) was found in all four assembled genomes, and 5/22 unique SNPs, including changes at position 2399 (SR to KR), 2766 (D to E), 2781 (TT to TA), 3208 (E to K), and 3268 (I to V), in at least 3/4 of the assembled genomes. The position of these SNPs coincided with location of Sarbecovirus SD-1 and SD-2 subdomains, S1/S2 cleavage region and the S2 fusion subunit (Supplementary information).

Phylogenetic analysis

Results from the maximum likelihood phylogenetic trees drawn using complete coronavirus genomes (Supplementary information), spike glycoprotein (Fig. 3), RDRP, envelope and nucleocapsid (Supplementary information) nucleotide sequences showed that the *Sarbecovirus* genomes assembled from UK lesser horseshoe bats belong to the same monophyletic clade as the published European horseshoe bat *Sarbecovirus* genomes (RhGB01, Khosta1 and 2 and BM48-31), and are more similar to the small number of African horseshoe bat sarbecoviruses than to any Asian horseshoe bat isolates, even those from the same host species (*R. ferrumequinum*).

Spike glycoprotein comparison, identification of furin cleavage site (FCS) and Transmembrane protease serine 2 (TMPRSS2), receptor binding domain (RBD) homology modelling and structural analysis

All the UK bat coronaviruses reported in this study shared the same spike glycoprotein amino acid length of 1256 aa with the only available complete bat SARS-like coronavirus genome reported from UK lesser horseshoe bats [41]. The S1 RBD of UK bat coronaviruses reported from this study and the previously reported RhGB01 [41] consisted of 220 amino acids, the RBM (receptor binding motifs) of UK lesser horseshoe bat coronaviruses reported from this study have the same amino acid length (71 aa) as RhGB01.

The RBD (receptor binding domain) pairwise alignment percentage identity estimation (Fig. 4b) between the reported UK bat coronaviruses and selected Betacoronavirus reference sequences, revealed an estimated percentage identity of 68% with SARS-CoV and 65–67% with SARS-CoV-2 related viruses. The RBD of all five UK bat coronaviruses reported from this study shared ≥95% amino acid homology with RhGB01 reported by Crook, Murphy *et al.* [41], and 76% amino acid homology with the Bulgarian horseshoe bat *Sarbecovirus* (BM48-31) reported by Drexler, Gloza-Rausch *et al.* [33]. Further assessment of the S1 and S2 regions of UK bat coronaviruses RBDs showed the absence of a furin or S1/S2 cleavage site (Fig. 4a), while demonstrating the presence of host transmembrane serine protease 2 (TMPRSS2) or S2 cleavage site from reported UK bat CoVs (Fig. 4a).

Comparative homology modelling of this study's sarbecoviruses to examine the interaction between RBD and the human ACE2, yielded five UK bat CoV 3D models sharing ≥92% identity and 3.0 Å root-mean-square deviation (RMSD) of C-atoms with the SARS-CoV-2 (PDB: 6vw1) X-ray crystal template structure (Fig. 5a, b). The overall quality factor of 3D models was calculated through validation with ERRAT2 and Ramachandran plots (Supplementary information) as ≥96 for all RBD and hACE2 3D structures constructed. Amino acid residues rejected at 95 and 99% confidence intervals and those known to be characteristic for the interaction between SARS-CoV-2 and hACE2 are highlighted along the RBD proteins in Fig. 5. Ramachandran plots generated following ERRAT2 validation demonstrated that 91% of residues were in the most favoured regions, 8.7% in additional allowed regions and 0.3% were found to be in the generously allowed regions (Supplementary information).

Prediction of amino acid residues using Prodigy showed that amino acid residues at interface with hACE2 receptor in British bat Sarbecovirus spike glycoproteins can be located at positions V409, R431, Q437, Y441, L443, F444, Y461, S462, P463, S464, G465, E471, F473, Y482, A483, F484, S486, P487, G488, I489, G490, T491, and Y493; as reported previously [41]. RBM of these sarbecoviruses is located between amino acid position 426 aa to 496 aa of the spike protein, coinciding with the location of hACE2 amino acid residues. In addition, RBD comparative analysis revealed the presence of critical amino acid residues in British bat sarbecoviruses (L443, L443, F473, A480, H481 and I489) previously reported to play a key role in cross-species transmission of SARS-like CoVs (Fig. 5).

Recombination

The only potential recombination events between the sarbecoviruses detected in this study all involved the partial or low quality sequences RhGB03 and RhGB05, and were with other sequences from this study (Supplementary information). These can and must be treated with caution due to both sequence quality and the lack of potentially intermediate sequences from more animals.

DISCUSSION

This study found no detectable SARS-CoV-2 in British bats, in particular none in horseshoe bats, which might be expected to be most at risk of infection. This is similar to the findings of other studies of European bats during the COVID-19 pandemic, none of which demonstrated SARS-CoV-2 infection including from, three common pipistrelles in the UK [65], 503 samples from 20 bat species including 58 lesser horseshoe bats in Poland [66], 197 samples from five bat species including 82 samples from lesser horseshoe bats, 104 from greater horseshoe bats and five from Mediterranean horseshoe bats (*Rhinolophus euyale*) from Sochi in Russia [36] and 53 lesser horseshoe bats in the UK [41].

The pan-coronavirus screening assay used in this study [46] is known to be relatively insensitive and may miss some coronavirus strains due to sequence mismatch. For example, it is not a good match for the known UK isolate MW719567, RhGB01. The *Sarbecovirus* E gene qPCR [47] was designed to detect all known sarbecoviruses and is in common use in human SARS-CoV-2 diagnostics due to the relative stability of the E gene in SARS-CoV-2 isolates and sarbecoviruses in

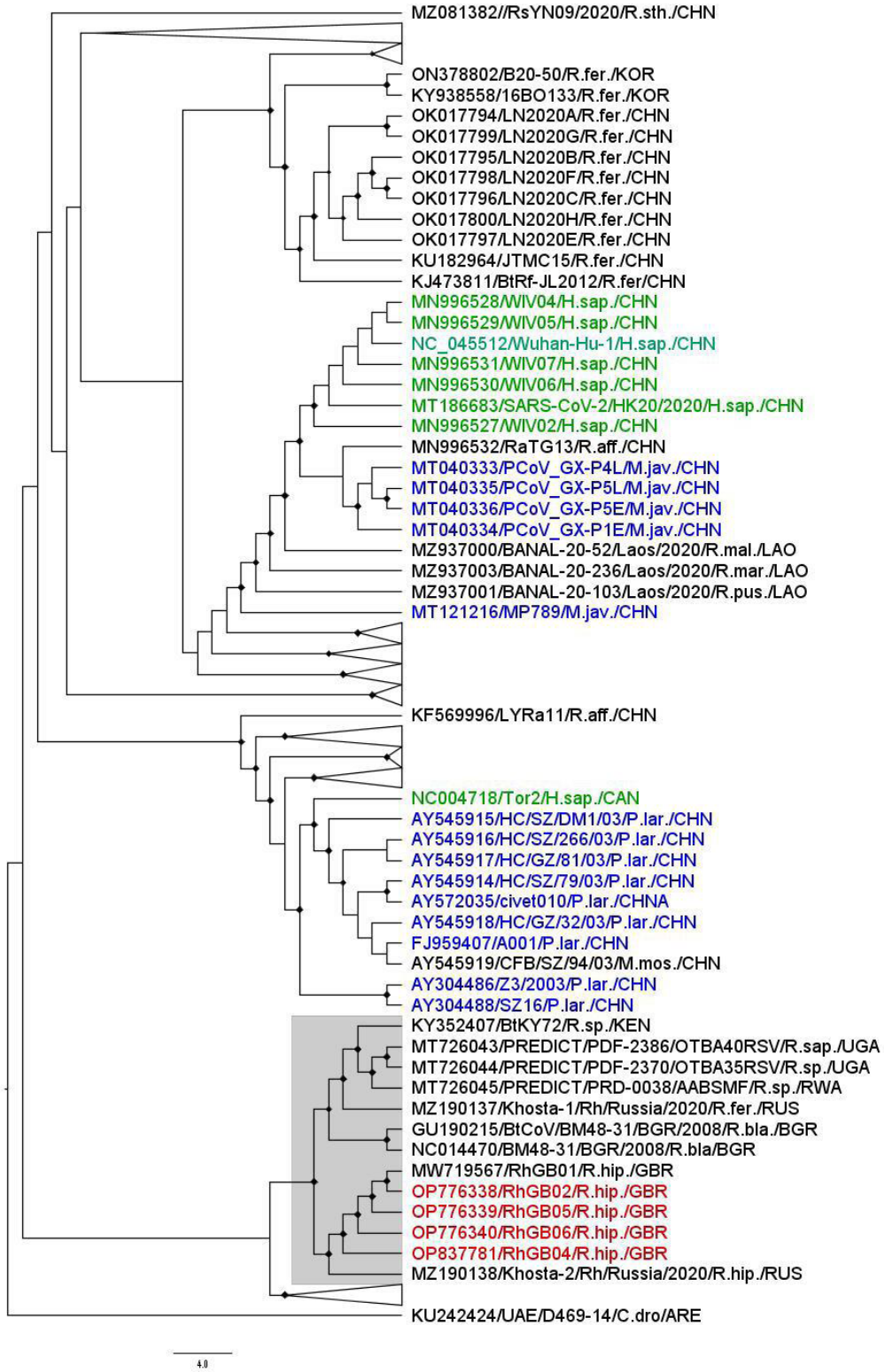
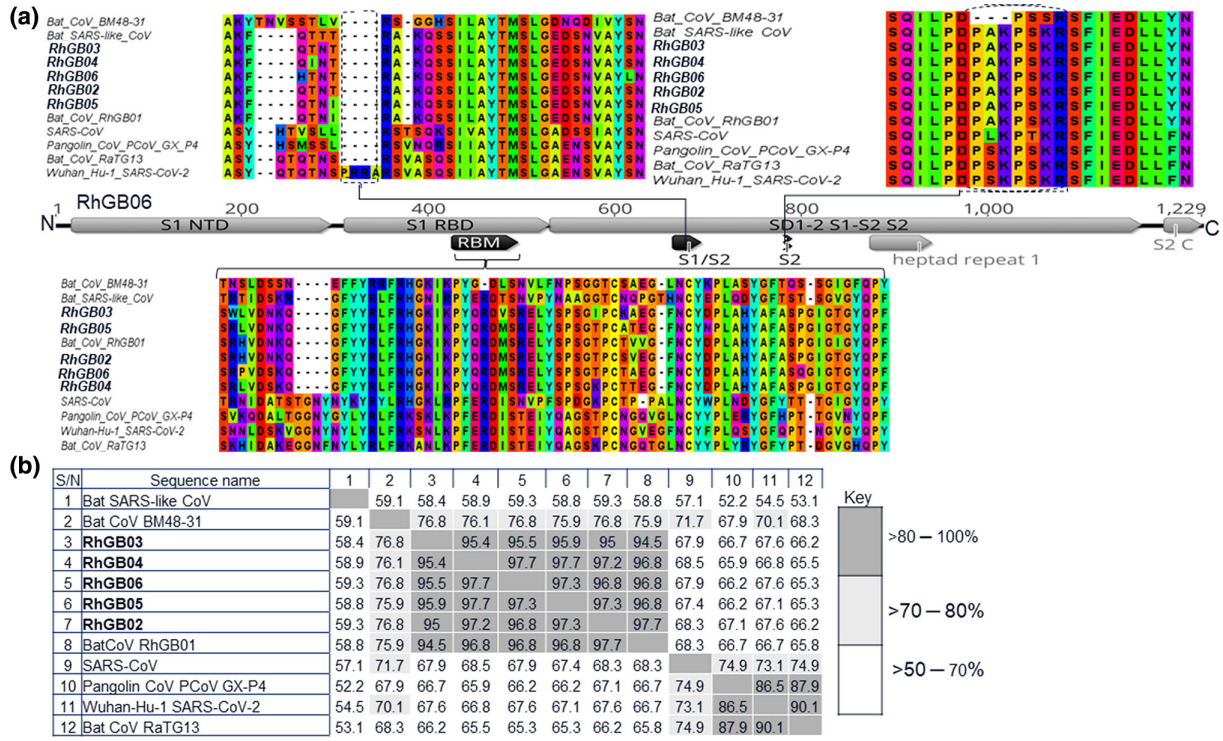


Fig. 3. Maximum likelihood phylogenetic tree of S gene nucleic acid constructed with 1000 bootstrap approximation, rooted on the MERs coronavirus reference sequence. One hundred and ninety eight non-human *Sarbecovirus* genomes and reference sequences for major variants of SARS-CoV and SARS-CoV-2 were included (non-human SARS-CoV-2 isolates were not included). Clades of Asian bat coronavirus sequences (apart from greater horseshoe bat sequences) have been collapsed for clarity (represented as broad triangles). Red=isolates from this study, green=human isolates, blue=isolates from other mammals. Sequences are named with Genbank ID, name from original study species of origin (e.g. R.hip=*Rhinolophus hipposideros*) and country of origin (e.g. GBR=Great Britain). Diamond symbols on nodes represent bootstrap support values ≥95%.



S/N	Sequence name	1	2	3	4	5	6	7	8	9	10	11	12
1	Bat SARS-like CoV		59.1	58.4	58.9	59.3	58.8	59.3	58.8	57.1	52.2	54.5	53.1
2	Bat CoV BM48-31	59.1		76.8	76.1	76.8	75.9	76.8	75.9	71.7	67.9	70.1	68.3
3	**RhGB03**	58.4	76.8		95.4	95.5	95.9	95	94.5	67.9	66.7	67.6	66.2
4	**RhGB04**	58.9	76.1	95.4		97.7	97.7	97.2	96.8	68.5	65.9	66.8	65.5
5	**RhGB06**	59.3	76.8	95.5	97.7		97.3	96.8	96.8	67.9	66.2	67.6	65.3
6	**RhGB05**	58.8	75.9	95.9	97.7	97.3		97.3	96.8	67.4	66.2	67.1	65.3
7	**RhGB02**	59.3	76.8	95	97.2	96.8	97.3		97.7	68.3	67.1	67.6	66.2
8	BatCoV RhGB01	58.8	75.9	94.5	96.8	96.8	96.8	97.7		68.3	66.7	66.7	65.8
9	SARS-CoV	57.1	71.7	67.9	68.5	67.9	67.4	68.3	68.3		74.9	73.1	74.9
10	Pangolin CoV PCoV GX-P4	52.2	67.9	66.7	65.9	66.2	66.2	67.1	66.7	74.9		86.5	87.9
11	Wuhan-Hu-1 SARS-CoV-2	54.5	70.1	67.6	66.8	67.6	67.1	67.6	66.7	73.1	86.5		90.1
12	Bat CoV RaTG13	53.1	68.3	66.2	65.5	65.3	65.3	66.2	65.8	74.9	87.9	90.1	

Fig. 4. British bat Sarbecovirus (in bold text) spike glycoprotein organization and sites of interest. (a) The presence of conserved S2 or TMPRSS2 cleavage site (top right) and the absence of furin cleavage site (S1/S2 top left). The FCS consisted of four amino acids (PRRA) identified to be present in only SARS-CoV-2. Below these two is the British bat CoVs spike glycoprotein linear map. (b) Estimated percentage pairwise alignment identity heatmap following alignment of RBD proteins from sarbecovirus reference genomes and those reported in this study.

general. Serial dilution of the SARS-CoV-2 positive control in this study (data not shown) found that the E gene qPCR was approximately a hundred times more sensitive than the pan coronavirus screening assay, though SARS-CoV-2 is routinely detected by both assays. This difference in sensitivity of the two assays may explain the discrepancies in results between the two screens (CT values for the E gene qPCR ranged between 22 and 40). It is therefore possible that other coronaviruses, in particular non-sarbecoviruses are present in these sample but were not detected.

British lesser horseshoe bats were, however, found to be frequently infected with a *Sarbecovirus* similar to that described previously in this species [36, 38, 40, 41, 66], and distinct from sarbecoviruses previously described in greater horseshoe bats in Bulgaria [35] and Russia [36]. Although the sampling strategy in this study, based on opportunistic sampling linked to bat survey and conservation studies, did not allow a prevalence to be calculated, the frequency at which the virus was detected within and between populations suggests a prevalence not dissimilar to that found in Slovenia of around 30% [66]. That study, unlike this, detected virus in oronasal swabs as well as faeces, and this may reflect the smaller amount of material collected on swabs in this study (rectal swabs were also less frequently positive in the qPCR assay than faecal samples). The sites sampled reflected the distribution of horseshoe bats in Great Britain, and infection was clearly common within and between populations of lesser horseshoe bats in this study, with no clear sex or age differences in likelihood of shedding. This high frequency of shedding and widespread distribution geographically and demographically, likely indicates either persistent infection and excretion, or frequent reinfection, of the gastrointestinal tract as has been reported for many other coronaviruses, such as those of cats, chickens and pigs [67–69].

Many of the roosts of lesser horseshoes bats in this study were shared with greater horseshoe bats, yet none of the latter were shedding detectable virus. This suggests lack of cross-species transmission and that the virus is relatively host-species restricted. This contrasts with several studies of SE Asian horseshoe bat species and their sarbecoviruses [70–72], which found strong evidence of cross-species transmission of coronaviruses. Lesser horseshoe bats were primarily sampled in the summer due to the greater sensitivity of this species to disturbance, while greater horseshoe bats were sampled across both summer and winter (53/277). The numbers of greater horseshoe bats in this study should have been adequate to detect, if it were present, the virus previously detected in Italian, Bulgarian and Russian greater horseshoe bats at the prevalence rates of 2.2–42% recorded in those studies [31–33, 35, 36] and the assay should be able to detect the Russian isolate (the only

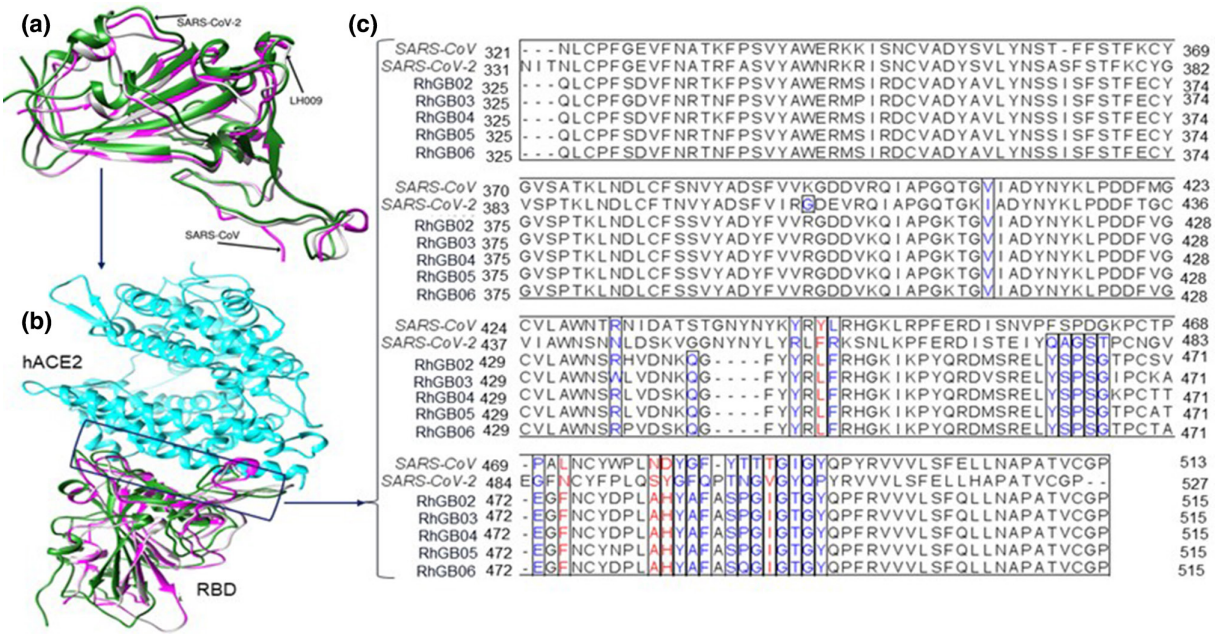


Fig. 5. Protein-protein homology complex model of British bat and related sarbecoviruses. (a) Superimposed RBD 3D complexes from UK bat CoVs (grey LH009), SARS-CoV (magenta), and SARS-CoV-2 (green). (b) Superimposed model between RBDs of UK bat CoVs (grey), SARS-CoV (magenta) and SARS-CoV-2 (green) and hACE2 (cyan). (c) Sequence alignment and comparison of S1 RBD and RBM of UK bat CoVs, SARS-CoV, and SARS-CoV-2. RBM for UK bat CoVs is positioned between 426 aa to 496 aa. Amino acid residues predicted to interact with hACE2 are shown in blue text, amino acid residues in red text are the critical residues within the RBM previously reported to play key role in cross-species transmission.

one for which an E gene is available). It is not clear why no sarbecoviruses were detected in these animals, though annual variation in viral loads is possible, as is a 'founder effect' in an island population on the edge of the host's geographic range.

The apparent host specificity of the European lesser horseshoe bat Sarbecovirus, combined with its lack of the furin cleavage site thought to be critical for human spread of SARS-Cov-2, indicates that these viruses are likely of low potential for zoonotic transmission, although modelling studies [38] have indicated that these viruses could acquire such features with minimal mutation and that these viruses could potentially bind to the human ACE-2 receptor [41].

The sarbecoviruses found in European lesser horseshoe bats cluster monophyletically and probably represent a distinct genus of *Sarbecovirus*. Recombination analysis with RDP5 (data not shown) indicates that these viruses are not recombinants.

Some variation in the presence of the accessory genes 7b, 9b and 10 was found amongst the isolates studied. These genes are not essential for viral replication, but are modulators of the host's innate immune system response and as such can affect strain pathogenicity [73–75]. All three of these genes interfere with MAVS (Mitochondrial Antiviral Signalling Protein) and IFN-β signalling and probably act synergistically in IFN inhibition in human SARS-CoV-2 infection [73, 75–77]. Plasticity in accessory gene complement has also been evident in different SARs-CoV-2 isolates and appears to be a feature of *Sarbecovirus* isolate variability [74, 75].

We do not have any indication of whether or not infection with these viruses has any adverse effects on their hosts or whether excretion patterns vary with age or reproductive status. This type of work is mature in relation to other bat-borne viruses, for example Hendra virus (a member of the Paramyxoviridae family) in bats of the genus *Pteropus* (flying foxes), for which it is clear that virus excretion is associated with maternity roosts with large numbers of birthing and juvenile animals and that virus excretion peaks in times of nutritional stress [78]. Such work is, however, in its infancy with horseshoe bat sarbecoviruses, although some longitudinal studies in SE Asia hint at a summer/maternity roost excretion pattern [70].

A better understanding of the ecology of horseshoe bat viruses requires further and longer term studies of sarbecoviruses in their natural hosts. The samples with the highest detection rates in this study were pooled faecal samples from bat roosts, so this sample type may provide the most reliable method of detection of virus in a roost site as well as being the most convenient, causing least disruption to bat colonies and presenting least threat of cross-species transmission between bats or between bats and humans. Serological studies to estimate lifetime exposure to SARS-CoV-2 (or this sarbecovirus) might also be useful, although existing SARS-CoV-2 serological assays have not been assessed for either these bats or their sarbecoviruses. There are also practical and regulatory issues with blood sampling very small locally endangered bat species.

Overall, this study provides several critical pieces of data in the overall picture of sarbecoviruses in their natural hosts. The growing picture is one of relative stability in terms of viral diversity and cross-species transmission potential in European horseshoe bat species, in contrast to higher diversity and cross species sharing of isolates in SE Asia, potentially related to the greater diversity and number of horseshoe bat species in the latter geographic region. This likely partly explains the repeated cross-species spill over of sarbecoviruses with an ancestral origin in bats, into other mammals including humans in SE Asia, though human behaviour and wildlife/human interactions, habitat disruption and farming of wildlife for meat are also likely to be contributing factors [1]. Ongoing work with these European sarbecoviruses remains to determine the epidemiology, temporal changes in viral prevalence and sequence, and effects of the virus in its natural host, with a real need for longitudinal studies of roost sites across the host range and spatial comparisons of sequences and prevalence.

Funding information
This work was funded by the BBSRC Grant number: BB/W009501/1

Acknowledgements
We are grateful for the help of many licensed bat workers who assisted in the collection of samples for this project, particularly Devon Bat Research and Conservation Group, Wiltshire Bat Group, and the Vincent Wildlife Trust. We also appreciate constructive discussions with the Bat Conservation Trust that informed this project.

Conflicts of interest
The authors declare that there are no conflicts of interest

Ethical statement
Ethical approval was granted by the University of Nottingham School of Veterinary Medicine and Science Committee for Animal Research and Ethics (CARE) and the Animal Welfare and Ethical Review Board of the University of Sussex.

References

1. Holmes EC, Goldstein SA, Rasmussen AL, Robertson DL, Crits-Christoph A, *et al*. The origins of SARS-CoV-2: A critical review. *Cell* 2021;184:4848–4856.
2. Lau SKP, Woo PCY, Li KSM, Huang Y, Tsoi H-W, *et al*. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci* 2005;102:14040–14045.
3. Wang M, Yan M, Xu H, Liang W, Kan B, *et al*. SARS-CoV infection in a restaurant from palm civet. *Emerg Infect Dis* 2005;11:1860–1865.
4. Eckstrand CD, Baldwin TJ, Rood KA, Clayton MJ, Lott JK, *et al*. An outbreak of SARS-CoV-2 with high mortality in mink (Neovison vison) on multiple Utah farms. *PLoS Pathog* 2021;17:e1009952.
5. Larsen HD, Fonager J, Lomholt FK, Dalby T, Benedetti G, *et al*. Preliminary report of an outbreak of SARS-CoV-2 in mink and mink farmers associated with community spread, Denmark, June to November 2020. *Euro Surveill* 2021;26:2100009.
6. Lu L, Sikkema RS, Velkers FC, Nieuwenhuijse DF, Fischer EAJ, *et al*. Adaptation, spread and transmission of SARS-CoV-2 in farmed minks and associated humans in the Netherlands. *Nat Commun* 2021;12:6802.
7. Rabalski L, Kosinski M, Smura T, Aaltonen K, Kant R, *et al*. Severe acute respiratory syndrome coronavirus 2 in farmed mink (Neovison vison), Poland. *Emerg Infect Dis* 2021;27:2333–2339.
8. Hale VL, Dennis PM, McBride DS, Nolting JM, Madden C, *et al*. SARS-CoV-2 infection in free-ranging white-tailed deer. *Nature* 2022;602:481–486.
9. Willgert K, Didelot X, Surendran-Nair M, Kuchipudi SV, Ruden RM, *et al*. Transmission history of SARS-CoV-2 in humans and white-tailed deer. *Sci Rep* 2022;12:12094.
10. Bienzle D, Rousseau J, Marom D, MacNicol J, Jacobson L, *et al*. Risk factors for SARS-CoV-2 infection and illness in cats and dogs(1). *Emerg Infect Dis* 2022;28:1154–1162.
11. Goletic S, Goletic T, Softic A, Zahirovic A, Rukavina D, *et al*. The evidence of SARS-CoV-2 human-to-pets transmission in household settings in Bosnia and Herzegovina. *Front Genet* 2022;13:839205.
12. Hosie MJ, Hofmann-Lehmann R, Hartmann K, Egberink H, Truyen U, *et al*. Anthropogenic infection of cats during the 2020 COVID-19 pandemic. *Viruses* 2021;13:185.
13. Jairak W, Chamsai E, Udom K, Charoenkul K, Chaiyawong S, *et al*. SARS-CoV-2 delta variant infection in domestic dogs and cats, Thailand. *Sci Rep* 2022;12:8403.
14. Pickering B, Lung O, Maguire F, Kruczkiewicz P, Kotwa JD, *et al*. Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission. *Nat Microbiol* 2022;7:2011–2024.
15. Sila T, Sunghan J, Laochareonsuk W, Surasombatpattana S, Kongkamol C, *et al*. Suspected cat-to-human transmission of SARS-CoV-2, Thailand, July-september 2021. *Emerg Infect Dis* 2022;28:1485–1488.
16. Pappas G, Vokou D, Sainis I, Halley JM. SARS-CoV-2 as a zooanthroponotic infection: spillbacks, secondary spillovers, and their importance. *Microorganisms* 2022;10:11.
17. Muylaert RL, Kingston T, Luo J, Vancine MH, Galli N, *et al*. Present and future distribution of bat hosts of sarbecoviruses: implications for conservation and public health. *Proc Biol Sci* 2022;289:20220397.
18. Boni MF, Lemey P, Jiang X, Lam TT-Y, Perry BW, *et al*. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* 2020;5:1408–1417.
19. Lytras S, Hughes J, Martin D, Swanepoel P, de Klerk A, *et al*. Exploring the natural origins of SARS-CoV-2 in the light of recombination. *Genome Biol Evol* 2022;14:evac018.
20. Wong ACP, Li X, Lau SKP, Woo PCY. Global epidemiology of bat coronaviruses. *Viruses* 2019;11:174.
21. Seifert SN, Bai S, Fawcett S, Norton EB, Zwezdaryk KJ, *et al*. An ACE2-dependent Sarbecovirus in Russian bats is resistant to SARS-CoV-2 vaccines. *PLoS Pathog* 2022;18:e1010828.
22. Wang J, Pan Y-F, Yang L-F, Yang W-H, Luo C-M, *et al*. Individual bat viromes reveal the co-infection, spillover and emergence risk of potential zoonotic viruses. *bioRxiv* 2022:2022.11.23.517609. DOI: 10.1101/2022.11.23.517609.
23. Conceicao C, Thakur N, Human S, Kelly JT, Logan L, *et al*. The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. *PLoS Biol* 2020;18:e3001016.
24. Li J, Lai S, Gao GF, Shi W. The emergence, genomic diversity and global spread of SARS-CoV-2. *Nature* 2021;600:408–418.
25. Chan YA, Zhan SH. The emergence of the spike furin cleavage site in SARS-CoV-2. *Mol Biol Evol* 2022;39:msab327.

26. JNCC. 1303 Lesser horseshoe bat Rhinolophus hipposideros. Species list; 2022. https://sac.jncc.gov.uk/species/S1303/
27. JNCC. 1304 Greater horseshoe bat Rhinolophus ferrumequinum. Species list; 2022. https://sac.jncc.gov.uk/species/S1304/
28. Mathews F, Kubasiewicz LM, Gurnell J, Harrower CA, McDonald RA, *et al*. A review of the population and conservation status of British mammals. A report by the Mammal Society under contract to Natural England, Natural Resources Wales and Scottish Natural Heritage. *Natural England Joint Report* 2018;JP025. pp701.
29. Temple HJ, Terry A. *The Status and Distribution of European Mammals*. Luxembourg: Office for Official Publications of the European Communities; 2007.
30. Russo D. *Chiroptera*. Springer, 2022.
31. Balboni A, Gallina L, Palladini A, Prosperi S, Battilani M. A real-time PCR assay for bat SARS-like coronavirus detection and its application to Italian greater horseshoe bat faecal sample surveys. *Sci World J* 2012;2012:989514.
32. Balboni A, Palladini A, Bogliani G, Battilani M. Detection of a virus related to betacoronaviruses in Italian greater horseshoe bats. *Epidemiol Infect* 2011;139:216–219.
33. Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, *et al*. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. *J Virol* 2010;84:11336–11349.
34. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, *et al*. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog* 2017;13:e1006698.
35. Muth D, Corman VM, Roth H, Binger T, Dijkman R, *et al*. Attenuation of replication by a 29 nucleotide deletion in SARS-coronavirus acquired during the early stages of human-to-human transmission. *Sci Rep* 2018;8:15177.
36. Alkhovsky S, Lenshin S, Romashin A, Vishnevskaya T, Vyshemirsky O, *et al*. SARS-like coronaviruses in horseshoe bats (*Rhinolophus* spp.) in Russia, 2020. *Viruses* 2022;14:113.
37. Lin X-D, Wang W, Hao Z-Y, Wang Z-X, Guo W-P, *et al*. Extensive diversity of coronaviruses in bats from China. *Virology* 2017;507:1–10.
38. Sander AL, Moreira-Soto A, Yordanov S, Toplak I, Balboni A, *et al*. Genomic determinants of Furin cleavage in diverse European SARS-related bat coronaviruses. *Commun Biol* 2022;5:491.
39. Wu Z, Han Y, Wang Y, Liu B, Zhao L, *et al*. A comprehensive survey of bat sarbecoviruses across China in relation to the origins of SARS-CoV and SARS-CoV-2. *Natl Sci Rev* 2022.
40. Rihtaric D, Hostnik P, Steyer A, Grom J, Toplak I. Identification of SARS-like coronaviruses in horseshoe bats (Rhinolophus hipposideros) in Slovenia. *Arch Virol* 2010;155:507–514.
41. Crook JM, Murphy I, Carter DP, Pullan ST, Carroll M, *et al*. Metagenomic identification of a new sarbecovirus from horseshoe bats in Europe. *Sci Rep* 2021;11:14723.
42. Logeot M, Mauroy A, Thiry E, De Regge N, Vervaeke M, *et al*. Risk assessment of SARS-CoV-2 infection in free-ranging wild animals in Belgium. *Transbound Emerg Dis* 2022;69:986–996.
43. Common SM, Shadbolt T, Walsh K, Sainsbury AW. The risk from SARS-CoV-2 to bat species in England and mitigation options for conservation field workers. *Transbound Emerg Dis* 2022;69:694–705.
44. Cook JD, Grant EHC, Coleman JTH, Sleeman JM, Runge MC. Risks posed by SARS-CoV-2 to North American bats during winter fieldwork. *Conserv Sci Pract* 2021;3:e410.
45. UKHSA. Sarscov2 variants of concern and variants under investigation in England technical briefing 8 April 2022 report 40; (n.d.)
46. Woo PCY, Lau SKP, Chu C, Chan K, Tsoi H, *et al*. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* 2005;79:884–895.
47. Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, *et al*. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020;25:2000045.
48. Lopez-Oceja A, Gamarra D, Borragan S, Jiménez-Moreno S, de Pancorbo MM. New cyt b gene universal primer set for forensic analysis. *Forensic Sci Int Genet* 2016;23:159–165.
49. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
50. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257.
51. Meleshko D, Hajirasouliha I, Korobeynikov A. coronaSPAdes: from biosynthetic gene clusters to RNA viral assemblies. *Bioinformatics* 2021;38:1–8.
52. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990;215:403–410.
53. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, *et al*. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10.
54. Li H, Birol I. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094–3100.
55. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–780.
56. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, *et al*. Corrigendum to: IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 2020;37:1530–1534.
57. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;35:518–522.
58. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;14:587–589.
59. Procter JB, Carstairs GM, Soares B, Mourão K, Ofoegbu TC, *et al*. Correction to: alignment of Biological Sequences with Jalview. *Methods Mol Biol* 2021;2231:C1.
60. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr* 1993;26:283–291.
61. Honorato RV, Koukos PI, Jiménez-García B, Tsaregorodtsev A, Verlato M, *et al*. Structural biology in the clouds: the WeNMR-EOSC ecosystem. *Front Mol Biosci* 2021;8:729513.
62. Vangone A, Bonvin AM. Contacts-based prediction of binding affinity in protein-protein complexes. *Elife* 2015;4:e07454.
63. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, *et al*. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–1612.
64. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, *et al*. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol* 2021;7:veaa087.
65. Jones S, Bell T, Coleman CM, Harris D, Woodward G, *et al*. Testing bats in rehabilitation for SARS-CoV-2 before release into the wild. *Conserv Sci Pract* 2022;4:e12707.
66. Orłowska A, Smreczak M, Thor K, Niedbalska M, Pawelec D, *et al*. The genetic characterization of the first detected bat coronaviruses in Poland revealed SARS-related types and alphacoronaviruses. *Viruses* 2022;14:1914.
67. Quinteros JA, Noormohammadi AH, Lee SW, Browning GF, Diaz-Méndez A. Genomics and pathogenesis of the avian coronavirus infectious bronchitis virus. *Aust Vet J* 2022;100:496–512.
68. Thayer V, Gogolski S, Felten S, Hartmann K, Kennedy M, *et al*. 2022 AAFP/everycat feline infectious peritonitis diagnosis guidelines. *J Feline Med Surg* 2022;24:905–933.
69. Zhang Y, Chen Y, Zhou J, Wang X, Ma L, *et al*. Porcine epidemic diarrhea virus: an updated overview of virus epidemiology, virulence variation patterns and virus-host interactions. *Viruses* 2022;14:2434.

70. Cappelle J, Furey N, Hoem T, Ou TP, Lim T, *et al*. Longitudinal monitoring in Cambodia suggests higher circulation of alpha and betacoronaviruses in juvenile and immature bats of three species. *Sci Rep* 2021;11:24145.
71. Latinne A, Hu B, Olival KJ, Zhu G, Zhang L, *et al*. Origin and cross-species transmission of bat coronaviruses in China. *Nat Commun* 2020;11:4235.
72. Zhou H, Ji J, Chen X, Bi Y, Li J, *et al*. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. *Cell* 2021;184:4380–4391.
73. Fang P, Fang L, Zhang H, Xia S, Xiao S. Functions of coronavirus accessory proteins: overview of the state of the art. *Viruses* 2021;13:1139.
74. Forni D, Cagliani R, Molteni C, Arrigoni F, Mozzi A, *et al*. Homology-based classification of accessory proteins in coronavirus genomes uncovers extremely dynamic evolution of gene content. *Mol Ecol* 2022;31:3672–3692.
75. McGrath ME, Xue Y, Dillen C, Oldfield L, Assad-Garcia N, *et al*. SARS-CoV-2 variant spike and accessory gene mutations alter pathogenesis. *Proc Natl Acad Sci U S A* 2022;119:e2204717119.
76. Li X, Hou P, Ma W, Wang X, Wang H, *et al*. SARS-CoV-2 ORF10 suppresses the antiviral innate immune response by degrading MAVS through mitophagy. *Cell Mol Immunol* 2022;19:67–78.
77. Shemesh M, Aktepe TE, Deerain JM, McAuley JL, Audsley MD, *et al*. SARS-CoV-2 suppresses IFNβ production mediated by NSP1, 5, 6, 15, ORF6 and ORF7b but does not suppress the effects of added interferon. *PLoS Pathog* 2021;17:e1010146.
78. Eby P, Peel AJ, Hoegh A, Madden W, Giles JR, *et al*. Pathogen spillover driven by rapid changes in bat ecology. *Nature* 2023;613:340–344.