Surveillance of 16 UK native bat species through conservationist networks uncovers coronaviruses with zoonotic potential

Cedric C.S. Tan[1,2,*], Jahcub Trew[3,*], Thomas P. Peacock[4,*], Kai Yi Mok[4], Charlie Hart[3], Kelvin Lau[5], Dongchun Ni[6], David Orme[3], Emma Ransome[3], William D. Pearse[3], Christopher M. Coleman[7], Dalan Bailey[8], Nazia Thakur[8], Jessica L. Quantrill[4], Ksenia Sukhova[4], Damien Richard[1], Guy Woodward[3], Thomas Bell[3], Lisa Worledge[9], Joe Nunez-Mino[9], Wendy Barclay[4], Lucy van Dorp[1], Francois Balloux[1,+], Vincent Savolainen[3,+]

[1]UCL Genetics Institute, University College London, Gower St, London WC1E 6BT, UK.

[2]The Francis Crick Institute, 1 Midland Rd, London NW1 1AT, UK.

[3]Georgina Mace Centre for the Living Planet, Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK.

[4]Department of Infectious Disease, Imperial College London, Exhibition Rd, South Kensington, London SW7 2BX, UK.

[5]Protein Production and Structure Core Facility (PTPSP), School of Life Sciences, École polytechnique fédérale de Lausanne, Rte Cantonale, 1015 Lausanne, Switzerland.

[6]Laboratory of Biological Electron Microscopy (LBEM), School of Basic Science, École polytechnique fédérale de Lausanne, Rte Cantonale, 1015 Lausanne, Switzerland.

[7]Queen's Medical Centre, University of Nottingham, Derby Rd, Lenton, Nottingham NG7 2UH, UK.

[8]The Pirbright Institute, Surrey GU24 0NF, UK.

[9]The Bat Conservation Trust, Studio 15 Cloisters House, Cloisters Business Centre, 8 Battersea Park Road, London SW8 4BG, UK.

* These authors contributed equally

[+]These authors contributed equally

Correspondence: V.S. (v.savolainen@imperial.ac.uk)

Abstract

While the COVID-19 pandemic, caused by SARS-CoV-2, has renewed genomic surveillance efforts in wildlife, there has been limited characterisation of bat-borne coronaviruses in Europe. We collected 48 faecal samples from all but one of the 17 bat species breeding in the UK, through an extensive network of bat rehabilitators and conservationists, and screened them for coronaviruses using deep RNA sequencing. We assembled nine novel, high-quality coronaviral genomes, comprising four alphacoronaviruses from *Myotis daubentonii* and *Pipistrellus pipistrellus*, a Middle East respiratory syndrome (MERS)-related coronavirus from *Plecotus auritus*, and four closely-related sarbecoviruses isolated from both horseshoe bat species *Rhinolophus hipposideros* and *R. ferrumequinum*. We further used *in vitro* assays to demonstrate that at least one of these sarbecoviruses can bind ACE2, the receptor used by SARS-CoV-2 to infect human cells, which was also supported using *in silico* structural and sequence analyses. Although this sarbecovirus can enter human cells *in vitro* when ACE2 is over-expressed, our analyses indicate that it is unlikely to infect humans and would require adaptations to do so. Our findings highlight the importance of working collaboratively with conservation networks to enable larger, coordinated viral surveillance efforts and prevent the emergence of zoonoses from wildlife.

Introduction

The majority of emerging infectious diseases in humans are zoonotic - arising from animal-to-human transmission of a pathogen[1] - and more than 70% originate in wildlife[2]. Bats (*Chiroptera*) are an ancient and diverse order of mammals, with 1,447 extant species[3]; due to this diversity, bats, as an order, represent a major wild reservoir for viruses, some of which have a high potential for being zoonotic[4]. These include viruses that are capable of subsequent human-to-human transmission, such as the Marburg and Nipah viruses[5,6]. More prominently, some bat species host coronaviruses closely related to those responsible for recent human epidemics, including Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV)[7], Middle East Respiratory Syndrome Coronavirus (MERS-CoV)[8,9], and SARS-CoV-2[10–12], the agent of the COVID-19 pandemic, suggesting an evolutionary origin in bats for all three viruses.

Multiple factors have to align for a successful zoonotic spillover to occur, including the frequency of exposure, the ability of the pathogen to infect humans and its capacity for onward human-to-human transmission[13]. The high degree of habitat overlap places many bats close to humans and domestic, farmed or hunted animals that are potential bridging hosts for the transmission of bat-borne viruses to humans[14]. Coronaviruses can infect a broad range of animals and are prone to zoonotic spillovers from their animal hosts. The seven major coronavirus infecting humans include SARS-CoV-2, the agent of the COVID-19 pandemic, its relative SARS-CoV-1, which caused a major international outbreak in 2002-2004 with around 8000 recorded cases and at least 774 deaths[15], and MERS-CoV which fuels recurrent disease outbreaks in humans through repeated host jumps from its reservoir host in dromedary camels[16]. Four coronaviruses (HCoV-229, -NL63, -OC43 and -HKU1) circulate endemically in humans; the ancestral reservoir of these alphacoronaviruses are believed to be from species of bat[17]. Additionally, multiple cases of host jumps into humans leading to isolated cases or small clusters have been documented for multiple species of coronavirus[17]. Given the current health burden exerted by coronaviruses and the risk they pose as possible agents of future epidemics and pandemics, global, robust surveillance of bat-borne coronaviruses should be a public health priority.

Several studies over the last decade have screened bats across Asia, Africa, the Middle East and Europe for coronaviruses (summarised in Supplementary Table 1)[12,18–28], finding anywhere from 1.5-23% coronavirus prevalence of animals tested. These estimates were primarily obtained via a reverse transcription real-time PCR (RT-PCR) using degenerate primers designed to target most coronaviruses species (i.e., pan-coronavirus primers; Supplementary Table 1). Given the vast diversity of coronaviruses, including those yet to be discovered, it is difficult to design primers that can amplify and capture the full diversity of coronaviruses. Our own comparative analysis of published primer sets show that existing RT-PCR assays[29–33] underestimate coronavirus prevalence (Supplementary Information; Supplementary Figure 1 and 2). Difficulties with primer design is exacerbated by low RNA concentrations in field samples and RNA degradation, so the large variability in prevalence estimates in these studies may be due to the sensitivity of the primer set used rather than the epidemiology of bat-coronaviruses. While sample RNA quality remains mainly dependent on sample collection and laboratory practices, because untargeted RNA sequencing does not require a priori knowledge of sequence information, it provides a more accurate estimate of viral diversity and prevalence. Hence, we chose this approach over RT-PCR to survey coronaviruses in UK bats.

Sequencing-based surveillance data can be used to assess the zoonotic potential of novel viruses. This includes *in silico* assessments that determine the degree of sequence and structural homology to other known and closely-related human-infecting viruses[18,20]. Even more compelling evidence can be obtained *in vitro* by measuring the binding efficiency of viral entry proteins to host receptors[12]. One of the most direct assessments is to test the efficiency of viral entry into human cell lines via a pseudovirus assay[18]. Despite the importance of functional validation, many studies to date fall short of providing *in vitro* or even *in silico* assessments of zoonotic risk (Supplementary Table 1).

There are 17 bat species that breed in the United Kingdom (UK), most of which roost in domestic buildings, churches, barns and other artificial structures. This proximity to humans may be a potential risk factor for zoonotic spillover of novel bat-borne viruses[5]. Additionally, the high habitat overlap with humans also places bats in close proximity to domesticated and farmed animals, which can serve as potential bridging hosts for transmitting bat-borne viruses to humans[14]. All UK bat species are protected by law across the UK with licences required for work related to bats, so although direct contact is rare among the general public, it is far more common for the small proportion of the population comprised of bat scientists, ecologists, conservationists and bat rehabilitators that undertake regular research, monitoring, surveillance, and bat rehabilitation work.

Only two coronavirus surveillance studies have been conducted in UK bats to date[20,28]. The first, published a decade ago, screened seven bat species and detected alphacoronaviruses in Daubenton's bat and Natterer's bat (*Myotis daubentonii and M. nattereri, respectively*)[28]. The other, from 2021, screened faecal samples from lesser horseshoe bats (*Rhinolophus hipposideros*) and recovered the whole genome sequence of a sarbecovirus, RhGB01 (MW719567)[20]. However, neither study provided direct *in vitro* assessments of zoonotic risk. Accordingly, the viral diversity and zoonotic potential of UK bat viruses remains largely unknown. This is equally true of most other UK mammals. However, given that the evolutionary origins of many coronaviruses of human health concern can be traced back to bats, assessing their zoonotic potential in UK bats is a top priority, before moving on to other animal groups.

To address this knowledge gap, we used an existing UK network of bat rehabilitators and conservationists to collect faecal samples from UK bats. Faeces from all but one bat species breeding in the UK (the grey long-eared bat, *Plecotus austriacus*, the rarest species in the UK) were collected and subsequently screened using deep RNA sequencing to characterise the genomic diversity of bat-borne coronaviruses in the UK. Going beyond surveillance, we performed pseudovirus assays and *in silico* analyses to assess the zoonotic potential of some of these viruses. Overall, our findings

demonstrate the effectiveness of decentralised surveillance of bat-borne viruses through bat conservationists and highlight the possible zoonotic potential of coronaviruses in the UK.

Results

Untargeted RNA sequencing recovers nine complete coronavirus genomes with novel genes and reveals cross-species transmission of coronaviruses

We performed deep RNA sequencing on 48 faecal samples from 16 of the 17 UK breeding bat species, with wide geographic coverage and over two years (Supplementary Figure 3). We recovered nine complete (97-100%) and five partial (up to 3%) coronavirus genomes across six UK bat species (*M. daubentonii*, *P. pipistrellus*, *P. pygmaeus*, *P. auritus*, *R. ferrumequinum*, and *R. hipposideros*), detecting coronaviruses amongst 29.2% of the samples. The nine complete genomes were of high-quality (as assessed by CheckV[34]; see Methods) and 28-30 kilobases in length (Table 1).

Table 1: Summary statistics for novel coronavirus genomes assembled in this study

Sample	Species	Common name	Closest hit			Novel genomes (this study)				
			Accession	Name	Subgenus	Name	Length (bp)	BLASTn identity (%)	Prop. aligned (%)	CheckV completeness
2-GH106	*Rhinolophus ferrumequinum*	Greater horseshoe bat	MW719567	RhGB01	*Sarbecovirus*	RfGB02	29375	98.1	99.4	97
1-GH087	*Rhinolophus ferrumequinum*		MW719567	RhGB01	*Sarbecovirus*	RfGB01	29308	98.1	99.7	97
2-30B	*Rhinolophus hipposideros*	Lesser horseshoe bat	MW719567	RhGB01	*Sarbecovirus*	RhGB02	29240	98.2	83.9	97
Sample-18	*Rhinolophus hipposideros*		MW719567	RhGB01	*Sarbecovirus*	RhGB03	29217	98.3	69.4	99
5-129B	*Plecotus auritus*	Brown long-eared bat	MG596803	*P. kuhlii* MERS-related CoV	*Merbecovirus*	PaGB01	30084	82.9	76.4	99
Sample-37	*Myotis daubentonii*	Daubenton's bat	MN535733	*M. daubentonii* pedacovirus	*Pedacovirus*	MdGB03	28227	95.4	99.8	100
Sample-30	*Myotis daubentonii*		MN535733	*M. daubentonii* pedacovirus	*Pedacovirus*	MdGB02	28010	95.4	99.8	100
4-126A	*Myotis daubentonii*		MN535731	*M. daubentonii* pedacovirus	*Pedacovirus*	MdGB01	28224	95.8	99.6	100
Sample-25	*Pipistrellus pipistrellus*	Common pipistrelle	MN535731	*M. daubentonii* pedacovirus	*Pedacovirus*	PpiGB02	28247	80.8	82.9	100

A global tree based on genetic distances[35] revealed genus and subgenus membership of these new coronaviruses (Fig. 1a; see Methods). We then followed with local maximum-likelihood phylogenetic analyses to determine their precise placement within each subgenus (Fig. 1b-d). These phylogenetic analyses reveal that the nine novel

genomes we recovered comprise four alphacoronaviruses from the *Pedacovirus* subgenus and five betacoronaviruses including one merbecovirus, and four sarbecoviruses (Fig. 1). Three of the coronaviruses recovered from *M. daubentonii* (which we call MdGB01-03) form a well-supported clade with other pedacoviruses isolated from the same bat species in Denmark (Fig. 1b). One coronavirus sequenced from *P. pipistrellus* (PpiGB01) falls as a sister lineage to the above clade. Another coronavirus from *P. auritus* (PaGB01) is related to MERS-CoV-like merbecoviruses isolated from *Hypsugo*, *Pipistrellus*, and *Vespertilio* spp. from Western Europe and China (Fig. 1c). Four coronaviruses isolated from *R. ferrumequinum and R. hipposideros* (RfGB01-02 and RhGB02-03, respectively) are closely related to the previously described UK bat sarbecovirus, RhGB01[36] (Fig. 1d).

Of these nine coronaviral genomes recovered here, two represent new species. Indeed, pedacovirus PpiGB01 from *P. pipistrellus* was relatively divergent from its closest match, a pedacovirus previously isolated from *M. daubentonii* (less than 81% nucleotide sequence identity; Table 1). Similarly, merbecovirus PaGB01 shares less than 82% sequence identity to its closest match, a merbecovirus isolated from *P. kuhlii* in Italy (Table 1). Overall, our survelliance efforts have extended our knowledge of the existing diversity of coronaviruses. Looking at their genomic structures, we identified one new gene in each of these new species (Supplementary Information; Supplementary Figure 4).

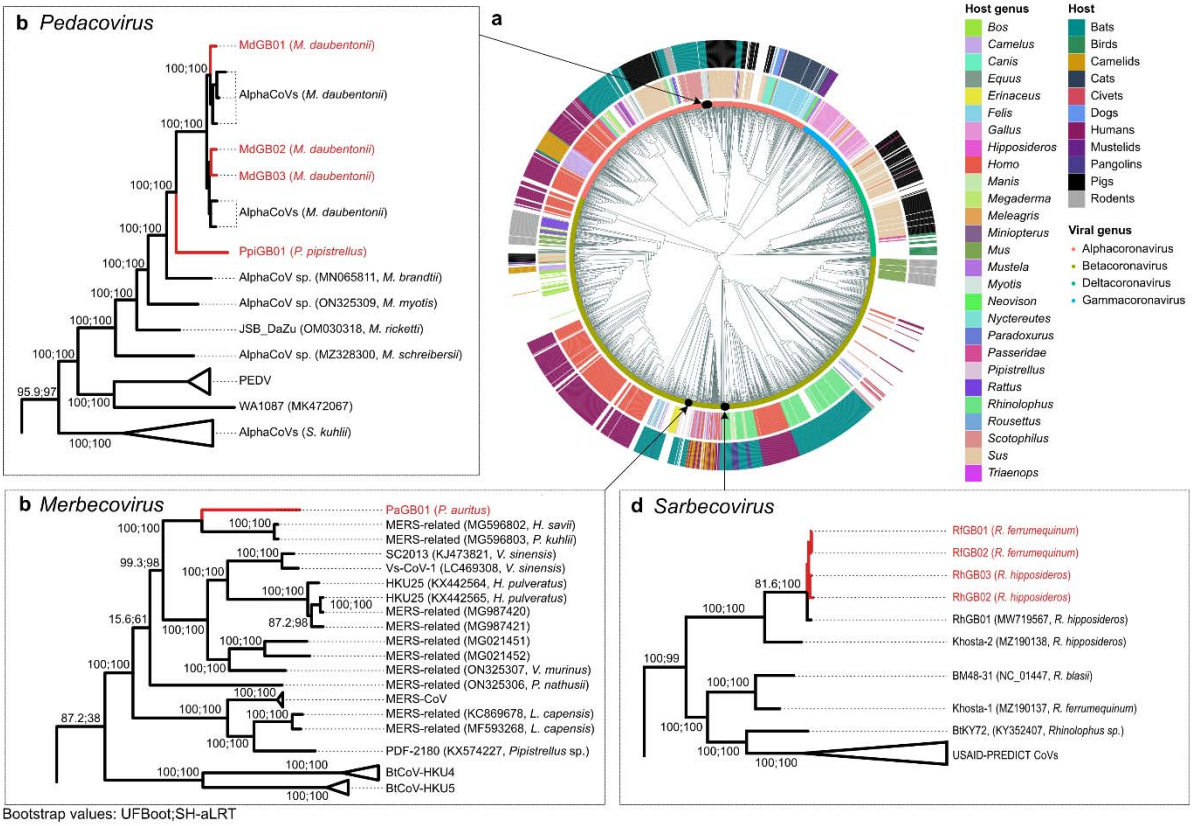

Figure 1. Phylogenetic placement of novel coronaviruses. (a) Alignment-free phylogeny of the global diversity of coronavirus genomes (*n* = 2118) and our nine novel genomes. Host genus (inner ring) and their broader host groups (outer ring) are annotated. Local maximum likelihood trees of (b) pedacoviruses (*n* = 106), (c) merbecoviruses (*n* = 113) and (d) sarbecoviruses (*n* = 534).

Viruses that are able to infect a broad range of hosts have been associated with a higher risk of emerging as infectious diseases that can transmit between humans[37,38]. Here, the four sarbecovirus genomes, representing one viral species, were recovered from two distinct horseshoe bat species, *R. ferrumequinum* and *R. hipposideros*. RhGB02, RhGB03, RfGB01, RfGB02 share 97-100% identity with RhGB01 previously described in *R. hipposideros*[20]. To better understand how these viruses might be shared among the two hosts, we looked at the habitat distribution of each horseshoe bat species. The two horseshoe bat species share a large proportion of their habitats, with 33% of their occurrence records reported at the same geographical coordinates. Furthermore, species distribution modelling predicted that 45% of the total land area

occupied by the two species is shared (Supplementary Figure 5a). Since the two *Rhinolophus* species can share roosts[39], these results indicate a potentially high frequency of direct contact, which may facilitate viral sharing and thus account for the isolation of RhGB01-like sarbecoviruses that are closely related from these two species.

To extend this analysis, we examined both observed and predicted distributions of all 17 UK bat species to identify potential viral sharing hotspots for future surveillance work. By analysing 42,953 occurrence records, we identified three regions near Bristol, Birmingham and Brighton with particularly high species diversity (up to 16 species in a single 5x5 km grid; Supplementary Figure 5b). Additionally, we identified regions within the UK, especially in Wales and the south coast of England where the habitats of the greatest number of different bat species are predicted to coincide (Supplementary Figure 5c). Alongside an understanding of the ecology of native species, including co-roosting and foraging behaviours, such information is a useful resource for future surveillance studies, and for prioritising focal areas of potential high risk.

Sarbecoviruses recovered from UK bats can bind the human ACE2 receptor for cellular entry

Because sarbecoviruses are of public health concern, in particular following the SARS epidemic and COVID-19 pandemic, we assessed the potential zoonotic risk of the ones we recovered from UK bats. All sarbecoviruses recovered from UK horseshoe bats (RhGB01, RhGB02, RhGB03, RfGB01, RfGB02) share 98-100% amino acid sequence identity in their spike proteins, so we chose RhGB02 for further analyses as follows. We first synthesised pseudovirus constructs expressing RhGB02 spike proteins and tested their ability to infect human cells expressing the human ACE2 receptor (hACE2), that is, the receptor used by SARS-CoV-2 to infect human cells. We then measured the binding affinity between spike proteins and hACE2 using bio-layer interferometry (BLI). Demonstrating human cellular entry and detectable spike-hACE2 binding would indicate zoonotic potential. If that was the case, we may then expect RhGB02 spike proteins to have evolved to bind ACE2 receptors from a variety of bat species, which we tested using these same assays.

RhGB02 spike-expressing pseudoviruses showed significantly higher entry into cells overexpressing hACE2 compared to those not expressing hACE2 (Fig. 2a; $p < 0.0001$). For comparison, we performed the same experiment using the spike proteins from other sarbecoviruses, namely BANAL-20-52, RaTG13 and SARS-CoV-2 (wild-type Wuhan-Hu-1), which are all known to effectively use hACE2 for cellular entry[40,41]. As expected, all three spike pseudoviruses showed significantly higher entry than into cells not expressing hACE2, confirming that they can also use hACE2 for cellular entry(Fig. 2a). We also used VSV-G pseudoviruses as a control because it can enter cells regardless of their receptor expression (Fig. 2a). BLI confirmed that RhGB02 spike is able to bind hACE2 with a dissociation constant, K_d = 253nM (Fig. 2b). However, the binding affinity of RhGB02 spike to hACE2 is approximately 17-fold lower than that for SARS-CoV-2 spike (K_d = 15nM) (Fig. 2b).

Given the lower binding affinity of RhGB02 spike compared to SARS-CoV-2, we then investigated if, like SARS-CoV-2, RhGB02 can infect human cells expressing lower (HEK293T-hACE2 – HEK293Ts stably transduced with hACE2) or physiological levels of hACE2 (Calu-3 lung, and Caco-2 colorectal cell lines). We could not detect entry of RhGB02 spike pseudovirus into any of these human cell lines (Fig. 2c). Meanwhile, we detected significant levels of entry for BANAL-20-52 and SARS-CoV-2 spike pseudoviruses (Fig. 2c; $p < 0.001$ see Methods), which were our positive controls. Additionally, we were not able to detect significant entry of RaTG13 spike pseudoviruses into these human cell lines, which has been demonstrated previously[41].

We next investigated if RhGB02 spike proteins can use ACE2 receptors from four bat species (*R. ferrumequinum*, *R. pusillus*, *Myotis lucifugus*, and *Rousettus leschenaultia*) for cell entry. We detected significant cell entry only through *M. lucifugus* and *R. leschenaultii* ($p < 0.01$; Fig. 2a). RhGB02 could not use *R. ferrumequinum* ACE2 receptors, although BLI measurements indicate detectable binding of RhGB02 spike to both *R. ferrumequinum* and *M. lucifugus* ACE2 (Fig. 2d). Unfortunately, the ACE2 sequence for *R. hipposideros* (from which RhGB02 was recovered) was currently unavailable for our assays. These results indicate that RhGB02 can bind the ACE2 receptor of *R. ferrumequinum* but not enter cells, suggesting that binding of host ACE2 alone may not be sufficient for efficient viral entry, and that other host cell-virus interactions (e.g. presence of suitable co-receptors) may be required.

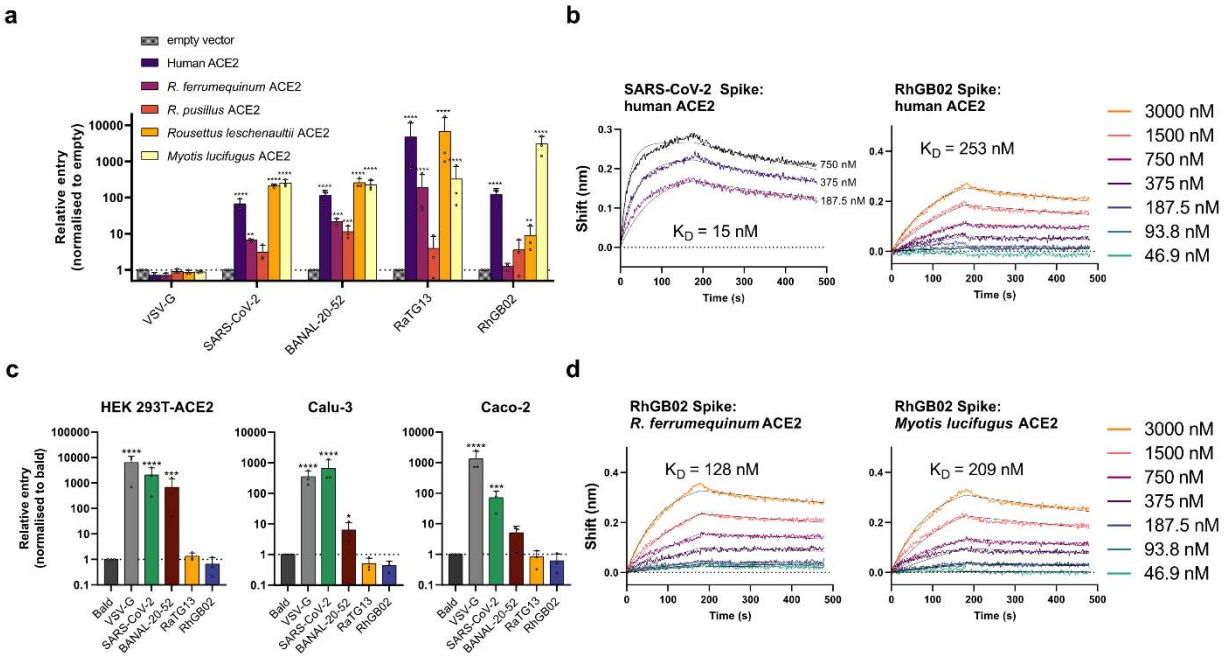

Figure 2. RhGB02 can bind human ACE2 and use it for cell entry *in vitro*. (a) Entry of different spike pseudoviruses expressing viral glycoproteins into HEK293T cells transfected with ACE2 homologues from different species or with an 'empty' vector. All entry measurements are normalised to that for the 'empty' vector. (b) Bio-layer interferometry binding curves showing the association and dissociation of SARS-CoV-2 and RhGB02 spike proteins with hACE2. (c) Entry of pseudoviruses into different 'normal' human cell lines that stably express lower or physiological levels of hACE2. All entry measurements are normalised to those for the 'bald' pseudovirus not expressing any spike protein. Data from panels (a) and (c) are compiled from *n*=3 completely independent repeats and plotted as mean + s.d. Significance was determined by (a) two-way ANOVA or (c) one-way ANOVA on log-transformed data (after determining log normality by the Shapiro–Wilk test and QQ plot) with multiple comparisons against 'empty' vector or 'bald' pseudovirus, respectively. *0.05 ≥ P > 0.01; **0.01 ≥ P > 0.001; ***0.001 ≥ P > 0.0001; ****P ≤ 0.0001. (d) Bio-layer interferometry binding curves showing the association and dissociation of RhGB02 spike proteins with *R. ferrumequinum* or *Myotis lucifugus* ACE2.

Structural and sequence features of RhGB02 spike explain detectable but inefficient usage of hACE2

To better understand the results of the assays described above, we used the AlphaFold artificial intelligence program[42] to predict the 3D structure of the receptor-binding domain (RBD) of the RhGB02 spike protein. We then compared it to the resolved RBD structures of SARS-CoV-2[43], BANAL-236 (a close relative of BANAL-20-52)[44], and RaTG13[41] bound to hACE2. Superposition of the RBD structures showed high structural conservation across all four sarbecoviruses (Fig. 3a). Additionally, the 3D structure of the RhGB02 RBD near the RBD-hACE2 binding interface was highly similar to that for SARS-CoV-2 (Fig. 3b), which was confirmed by comparing the area of contact surface (894 $Å^2$ and 850 $Å^2$, respectively; Supplementary Figure 6). These findings account for the ability of the RhGB02 spike protein to bind hACE2 (Fig. 2a,b). To understand why pseudoviruses exhibiting RhGB02 could not enter cells expressing the ACE2 receptor at physiological levels (Fig. 2c), we compared the contact residues in the SARS-CoV-2[43] and SARS-CoV[45] spike proteins that are crucial for spike-hACE2 binding. The novel RhGB01-like sarbecoviruses share only 9/17 SARS-CoV-2 (Fig. 3c) and 7/14 SARS-CoV contact residues (Fig. 3d), while RaTG13 only shares 11/17 SARS-CoV-2 and 8/14 SARS-CoV contact residues. In contrast, BANAL-20-52 shares 15/17 contact residues with SARS-CoV-2. These results indicate poorer conservation of these key contact residues, which would explain the relatively lower hACE2 usage efficiency, and hence the ability to infect human cells, of RhGB02 and RaTG13 compared to BANAL-20-52.

Notably, the RhGB01-like sarbecoviruses already possess a R-A-K-Q sequence (spike residues 657-660; Supplementary Figure 7), which is one nucleotide away (Gln/C**A**A to Arg/C**G**A) from the canonical R-X-K/R-R motif, a furin cleavage site that enhances the ability of many coronaviruses to infect human cells and hence their transmissibility[46,47]. Additionally, a recombination analysis of the RhGB01-like and other representative sarbecoviruses indicates a high prevalence of recombination (Supplementary Information; Supplementary Figure 8), which may accelerate adaptation for infecting novel hosts. Given these findings, the current zoonotic risk of sarbecoviruses in UK bats cannot be ignored and warrants more extensive surveillance of bats in the region.

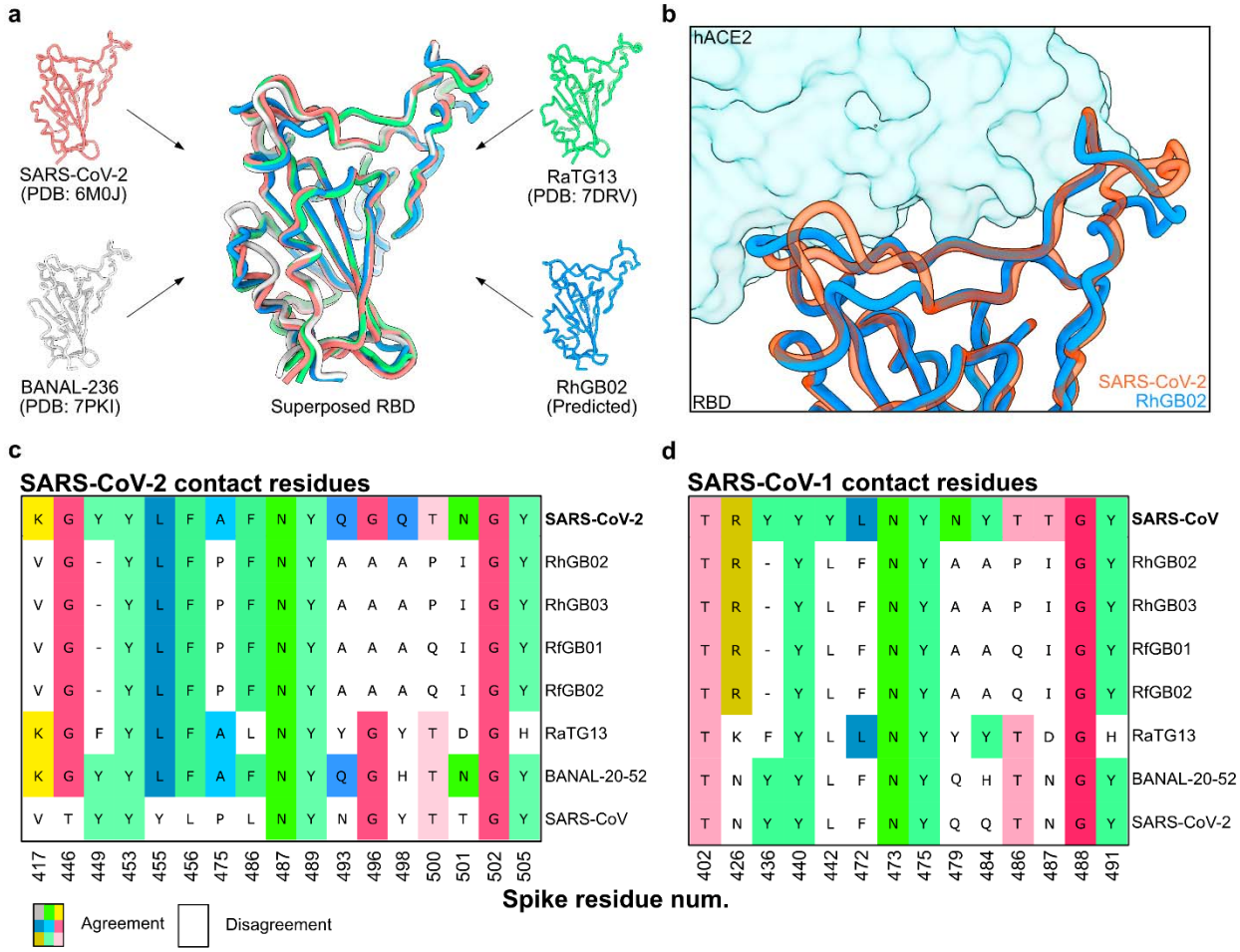

Figure 3. Structural and sequence features of RhGB01-like sarbecoviruses. (a) The solved RBD structures of SARS-CoV-2, RaTG13, BANAL-236 (close relative of BANAL-20-52[44]) and the AlphaFold-predicted structure of RhGB02 were superposed. (b) The 3D surfaces of the RBD-hACE2 binding interface for SARS-CoV-2 and RhGB02. Alignment of coronavirus spike proteins showing the conservation of key contact residues involved interactions between (c) SARS-CoV-1 spike and (d) SARS-CoV-2 spike with hACE2.

Discussion

The emergence of the COVID-19 pandemic in 2019 is a sobering reminder of the massive impact of zoonotic viruses on global health and economy. Despite this, sustained genomic surveillance in wildlife has remained limited. In this study, we used an existing network of bat rehabilitators to obtain geographically and temporally diverse samples from almost all bat species in the UK. We argue that this can be a sustainable and effective surveillance model to identify and characterise novel bat-borne viruses at risk of potential zoonotic emergence.

We provided evidence that at least one sarbecovirus isolated from UK horseshoe bats can bind hACE2 *in vitro*. Given the moderate level of conservation at the key residues of the RhGB02 spike protein that directly interact with hACE2, the RhGB01-like viruses likely require further adaptations, particularly in their spike proteins, before they can make the zoonotic jump. Notably, single mutations in sarbecoviral spike proteins have been shown to enable binding of ACE2 from novel host species[48]. Additionally, a single T403R mutation in the RaTG13 spike has been shown to allow the virus to infect human cells[49], we speculate that the genetic barrier precluding effective hACE2 usage for cellular entry into human cells may be small. This may also be the case for our novel sarbecoviruses, which already share more than half of the SARS-CoV-2 contact residues, and is reflected by the ability of RhGB02 to infect hACE2-overexpressing cells.

Further, we found a high prevalence of genetic recombination amongst sarbecoviruses, particularly in the spike gene (Supplementary Figure 8), which may facilitate viral adaptations to overcome this genetic barrier. This observation is corroborated by other studies that have also suggested an enrichment of recombination signals in or surrounding the sarbecovirus spike gene[50,51]. Co-infections and subsequent recombination of RhGB01-like viruses with other coronaviruses that already effectively use hACE2 may therefore facilitate zoonotic transmission. As such, the possibility of a future host-jump into humans cannot be ruled out, even if the risk is small. This reiterates the need for individuals that are in frequent contact with bats, such as bat

rehabilitators, to adhere to current biosafety practices to reduce their exposure to bat coronaviruses and likewise to reduce the risk of the exposure of bats to human-borne coronaviruses[52], such as SARS-CoV-2 or the endemic HCoVs. Luckily in the UK, the risk of zoonotic exposure is minimised for most people through a lack of direct contact (roosting spaces are often well away from human inhabitants) along with the provision of science-based information to roost owners by organisations such as the Bat Conservation Trust (https://www.bats.org.uk).

Our *in vitro* assays indicate that RhGB01-like sarbecoviruses do not use *R. ferrumequinum* ACE2 as their primary receptor, which is in line with other studies of bat coronaviruses[48,53]. Importantly, this raises the question as to what evolutionary mechanisms drive the acquisition of the ability to use hACE2 in bat sarbecoviruses. Given previous associations between pathogen host breadth and their capacity to emerge as zoonotic diseases[37,38], we speculate that multi-host viruses tend to have 'generalist' cell entry receptors that possess a low genetic barrier in the evolution of zoonotic transmission. More extensive surveillance of the viral sharing dynamics in mammalian hosts, including bats, may provide key insights into the molecular and ecological determinants of zoonotic events. Such studies can leverage both species occurrence data and niche modelling to prioritise regions where a high number of species are likely to be found combined with an understanding of species ecology for quantification of risk.

The initial spread of SARS-CoV-2 in China, its widely publicised evolutionary origin in *Rhinolophus* bats[54], and the subsequent identification of other bat-borne sarbecoviruses in Southeast Asia[12,18], has focused attentions about the zoonotic risk of coronaviruses in those geographical regions. However, our findings highlight the zoonotic risk of sarbecoviruses may extend beyond Asia, stressing the importance of more extensive surveillance globally.

Finally, while it is imperative to better quantify the risk of zoonotic events from bats and design approaches to mitigate risk, bats serve important roles in ecosystems globally,

including services such as arthropod suppression, pollination and seed dispersal[55]. Some bat species have rapidly declining populations – for example, one third of the most threatened mammalian species in the UK are bats[56,57]. Recent studies have shown that human-associated stressors such as habitat loss and changes in land-use can be important drivers of zoonotic spillover from wildlife[58,59]. As such, it is vitally important that an integrated ecological conservation approach is taken that includes maintaining legal protection, rather than destruction of wildlife and its habitat, in future approaches to mitigate zoonotic risk.

Methods

Sample collection

Sampling kits were sent out to various bat rehabilitators in the UK as described previously[60] for the collection of faeces from bats. These faecal samples (0.02-1g) were immediately stored in 5 ml of RNAlater solution to prevent degradation of RNA. The geographical locations and collection dates for all samples are provided in Supplementary Table 2.

RNA extraction

RNA was extracted from faecal samples using the QIAamp Viral RNA Mini Kit (Qiagen) following the protocol for extracting RNA from stool samples. We used up to 0.5 g of faeces, which was vortexed in 2ml of 0.9% NaCl solution, at 6000rpm for 2 minutes. The supernatant was filtered using a 0.2µm syringe filter, 280 µl of which was used for RNA extraction. Total RNA was eluted in 80 µl of AVE buffer and stored at -80°C. RNA was quantified using Qubit 2.0 fluorometer (Invitrogen).

Coronavirus database

To create a database representing the extant global genomic diversity of coronaviruses, we downloaded all complete *Coronaviridae* (taxid:11118) genomes from *NCBI Virus*, excluding provirus sequences (accessed 4th July 2022). Additionally, we downloaded all non-human-associated and non-SARS-CoV-2 betacoronaviruses from GISAID[61] (n = 29). To minimise the overrepresentation of certain viral species, we randomly retained 50 isolates for each of the following species: porcine epidemic diarrhoea virus, avian infectious bronchitis virus, MERS-CoV, SARS-CoV and SARS-CoV-2 sequences. This yielded a final dataset comprising 2118 genomes.

Metagenomic sequencing and assembly

All samples were prepared for sequencing using the NEBNext® Ultra™ Directional RNA Library Prep Kit, with a QIAseq FastSelect rRNA depletion step. Sequencing was carried out using Illumina NovaSeq, paired end 150 bp. Quality control of reads was performed using *bbduk.sh v39.01* from the *BBTools Suite* (https://sourceforge.net/projects/bbmap/). Briefly, we trimmed adapter sequences and read ends below Q10, and discarded trimmed reads with average quality below Q10. *De novo* metagenomic assembly was performed on quality-controlled or raw reads for each sample using coronaSPAdes v3.15.4[62]. Assembled scaffolds were then queried using BLASTn against all 2118 genomes in our coronavirus database to determine their most related reference. Scaffolds that could be aligned using BLASTn to coronaviruses in our database and that were already longer than 28kb were considered as complete genomes.

In some cases, *de novo* assembly yielded multiple scaffolds that were shorter than 28kb but shared the same closest reference. We 'stitched' these scaffolds together using the BLASTn alignment coordinates to the closest coronavirus reference and replaced any gaps with Ns. *De novo* assembly using raw reads produced better results, producing longer and more complete scaffolds, yielding six >28kb scaffolds (MdGB01, MdGB02, MdGB03, PpiGB01, RfGB01, RfGB02), compared to quality-controlled read assembly which yielded only two (RfGB01, PpiGB01). Further, the two >28kb scaffolds, RfGB01 and PpiGB01, generated using either raw or quality-controlled assemblies were identical, suggesting that *de novo* assembly using raw reads were reliable. We hence chose the assemblies generated using raw reads for our downstream analyses. We named the novel complete genomes following the naming convention for the Sarbecovirus previously described in a UK bat, RhGB01 – species: 'Rh' (*R. hipposideros*), region the coronavirus was found in: 'GB' (Great Britain) and the frequency of description: '01' (the first described in that species and country).

Genome annotation and characterisation of novel genes

We performed gene annotations using Prokka v1.14.6[63] to determine if these genomes carry any novel genes. We subsequently used PSI-BLAST on the online webserver (https://blast.ncbi.nlm.nih.gov/), an iterative search program that is more sensitive than the conventional protein BLAST[64], to identify distant homologues of protein sequences. We additionally used InterProScan[65,66] to make functional predictions for potentially novel proteins.

Species niche modelling

Bat occurrence records data were gathered from the online databases NBN Atlas (https://nbnatlas.org/) and GBIF (www.gbif.org). Records from year 2000-present were included, removing replicate records and those with high coordinate uncertainty. The number of occurrence points used for modelling ranged from 32 (*Myotis alcathoe*) to 16,403 (*Pipistrellus pipistrellus*). An initial 17 environmental variables were identified *a priori* to be important for predicting bat distributions. Nine were climatic variables averaged across 1980-2010 as described by Karger et al.[67], and were reduced to five variables using Variance Inflation Factor (VIF), retaining only those with a VIF < 0.5. These were mean annual air temperature, mean diurnal air temperature range, mean daily mean air temperature of the wettest quarter, precipitation seasonality and mean monthly precipitation amount of the warmest quarter. Four variables were derived from the UKCEH Land Cover Map 2019[68]. After merging similar land use classes, distance to woodland, distance to grassland, distance to arable and horticulture, and distance to urban were measured using Euclidean distance tools in ArcMap version 10.8. Two further distance variables were derived from Ordnance Survey polygons (2019, 2021): distance to the nearest road and distance to the nearest river. Elevation and slope were included to describe the topography of Great Britain, and were taken from the LiDAR Composite Digital Terrain Model data at 10m resolution. All spatial data were subsequently reduced to 1000m resolution and projected to British National Grid.

An ensemble of five supervised binary classifiers was trained to predict the suitability of a land area for each of the 17 UK bat species using the *R* package *sdm*[69]: random forest (RF), maximum entropy (MaxEnt), multivariate adaptive regression splines (MARS), boosted regression trees (BRT), and support vector machines (SVM). Classifiers were trained to predict whether a particular species was present or 'absent' based on the 13 ecological variables described above, using the occurrence data for each species and an equal number of randomly generated pseudo-absence data points across the study area. Training and evaluation was performed using a 5-fold cross-validation protocol, where a random subset comprising 80% of the dataset is used for training and the remaining 20% use for the final evaluation. A final ensemble of all five classifiers that were trained was used to generate the species distribution maps, with the contribution of each individual classifier weighted based on its area under the receiver operating characteristic curve (AUROC) score obtained during training. The resultant species distribution maps indicate habitat suitability as a probability score for each 1 km square grid on the study area, which ranges from 0 (unsuitable habitat) to 1 (suitable habitat). All models across all species performed well, with a median AUROC, sensitivity and specificity of 0.827, 0.854, and 0.78, respectively. The individual species distribution maps and model performance metrics are provided in Supplementary Figure 9.

Phylogenetic analyses

To place the novel sequences within the global diversity of coronaviruses sequenced to date, we computed alignment-free pairwise Mash distances using Mash v2.3[35] with a *k*-mer length of 12, and reconstructed neighbour-joining trees[70] using the *nj* function from the Ape v5.6.2 package in *R* (Fig. 1a). This alignment-free phylogenetic reconstruction approach circumvents the challenge of aligning highly diverse sequences at the family level, where high frequency of viral recombination may obscure true evolutionary histories[71] and prevent dataset wide alignments. In accordance with previous work[72], we rooted the neighbour-joining tree to a monophyletic *Deltacoronavirus* clade comprising all 10 representative *Deltacoronavirus* genomes downloaded from NCBI RefSeq.

From this global phylogeny, we retrieved the pedacovirus (*n* = 106), merbecovirus (*n* = 113) and sarbecovirus genomes (*n* = 534) most proximal to the novel assembled genomes. We then aligned genomes from these subgenera separately using the Augur v14.0.0[73] wrapper for MAFFT v7.490[74]. Genome positions where more than 20% of sequences were assigned gaps were removed from the alignment. We subsequently reconstructed finer-scale maximum-likelihood trees with IQTree v2.1.4-beta under a GTR+G model, using ultrafast bootstrapping (UFBoot)[75] and approximate likelihood-ratio tests (SH-aLRT)[76] with 1000 replicates. All phylogenetic trees were visualised either using FigTree v1.4.4 or *ggtree* v3.2.1[77].

Recombination analysis

We selected 218 sarbecovirus genomes from the local sarbecovirus tree (*n* = 534) by retaining only one representative each for SARS-CoV (NC_004718) and SARS-CoV-2 (MW206198). We subsequently aligned these genomes via the same approach described above but masked all positions with >20% of gaps by replacing the positions with Ns, and removed gaps in the alignment relative to the genome used to root the local sarbecovirus tree, NC_025217. This masked alignment was then analysed using RDP v4.101[78]. Gene annotations for NC_025217 were obtained from GenBank and used to annotate predicted recombinant positions.

Spike protein homology and conservation of contact residues

We extracted the Prokka-annotated spike protein sequences from our novel isolates for further analysis. We calculated pairwise amino acid sequence similarities (Figure 5a) by first performing pairwise global alignments[79] of the Spike protein sequences using the *pairwiseAlignment* function as part of the *Biostrings* v2.62.0 package[80] in *R*. The BLOSUM62 scoring matrix was used for pairwise alignment. Pairwise sequence similarities, including gapped positions, were then calculated using the *pid* function in the *Biostrings* package. Separately, we performed multiple sequence alignments of Spike sequences from our novel isolates and other human-infecting *Betacoronaviruses* (BANAL-236, MZ937003.2; SARS-CoV-2, NC_045512.2; SARS-CoV-1, NC_004718.3;

MERS, NC_019843.3) using Mafft v7.490[74]. Subsequently, we visualised and annotated the Spike alignments using *UGENE* v42.0[81]. The accessions of all genome records used in these analyses are provided in Supplementary Table 3.

Pseudovirus assays

To further test the capability of the coronaviruses we identified to infect human cells, we synthesised human codon-optimised, Δ19-truncated spike contructs in pcDNA.3.1. Gene synthesis and codon optimisation was performed by GeneArt (Thermo Fisher). Plasmids (*Rhinolophus pusillus*), Leschenault's rousette fruit bat (*Rousettus leschenaultii*), and little brown bat (*Myotis lucifugus*) in pDisplay were used as previously described[82]. Additionally, Greater horseshoe bat (*Rhinolophus ferrumequinum;* BAH02663.1) ACE2 was synthesised and cloned into pDISPLAY for this study.

We maintained human embryonic kidney cells (HEK 293T; ATCC CRL-11268) in complete media (DMEM, 10% FBS, 1% non-essential amino acids (NEAA) and 1% penicillin-streptomycin (P/S)). Human lung cancer cells (Calu-3; ATCC HTB-55) and Human epithelial colorectal adenocarcinoma cells (Caco-2; ATCC HTB-37) were maintained in DMEM, 20% FBS, 1% NEAA and 1% P/S. All cells were kept at 5% CO_2, 37°C. 293T-hACE2 cells were generated by transducing HEK 293T cells with an ACE2-expressing lentiviral vector, MT126[83] and selecting with 2 μg ml^{-1} puromycin; after selection, cells were subsequently maintained with 1 μg ml^{-1} puromycin.

Lentiviral based pseudotyped viruses were generated as previously described[47]. Briefly, 100 mm dishes of 293T cells were transfected using lipofectamine 3000 (Invitrogen) with a mixture of 1 µg of the HIV packaging plasmid pCAGGs-GAG-POL, 1.5 µg of the luciferase reporter construct (pCSFLW), and 1 µg of the plasmid encoding the spike or glycoprotein of interest in pcDNA3.1. After 24 h supernatant was discarded and replaced. PV-containing supernatants were collected at 48 and 72 h post-transfection, passed through a 0.45 µM filter, and aliquoted and frozen at -80°C.

Pseudovirus entry assays were performed as previously described[47]. Briefly, 100 mm dishes of 293T cells were transfected using lipofectamine 3000 (Invitrogen) with 2 µg of the ACE2 encoding plasmid or empty vector. After 24 h, cells were resuspended by scraping and plated into 96 well plates. Cells were overlayed with pseudovirus for 48 h before lysis with reporter lysis buffer (Promega). Caco-2, Calu-3, and 293T-hACE2 cells were seeded into 96 well plates. Cells were overlayed with pseudovirus for 48 h before lysis with cell culture lysis buffer (Promega). We determined luciferase luminescence on a FLUOstar Omega plate reader (BMF Labtech) using the Luciferase Assay System (Promega).

We assessed expression of transfected receptors using Western blot assays. Cell suspensions were pelleted by centrifugation at 1000 revolutions per minute (RPM) for 7 min at 4°C, then supernatant was removed. Cells were resuspended in 150 µl of cold radioimmunoprecipitation assay (RIPA) buffer (Thermo Fisher) and incubated on ice for 30 min. Then, they were spun down at 3750 RPM for 30 min at 4°C. The protein-containing supernatants were transferred to sterile Eppendorfs and frozen down at -20°C. Before running a gel, 50 µl of 2-Mercaptoethanol (BME; Sigma) diluted 1:10 in 4X Laemmli Sample Buffer (Bio-Rad, USA) was added to lysates and incubated at 80°C for 10 min.

To analyse incorporation of spike into the different sarbecovirus pseudoviruses, we concentrated pseudovirus by ultracentrifugation at 100,000 x g for 2 h over a 20% sucrose cushion.

In all experiments, we confirmed the successful expression of host receptors and spike pseudoviruses using Western blot analyses (Supplementary Figure 10). For Western blotting, membranes were probed with mouse anti-tubulin (diluted 1/5,000; abcam; ab7291), mouse anti-p24 (diluted 1/2,000; abcam; ab9071), rabbit anti-SARS spike protein (diluted 1/2,000; NOVUS; NB100-56578) or rabbit anti-HA tag (diluted 1/2000; abcam; ab9110). Near infra-red secondary antibodies, IRDye® 680RD Goat anti-mouse (diluted 1/10,000; abcam; ab216776), IRDye® 680RD Goat anti-rabbit (diluted 1/10,000;

abcam; ab216777), were subsequently used. Western blots were visualized using an Odyssey DLx Imaging System (LI-COR Biosciences).

Alphafold2 (ColabFold) structural analysis

The protein structure model of the RhGB02 RBD was predicted using Alphafold2 as implemented in ColabFold[84]. Default settings were used. The top ranked model was used for all analyses. Structural representations and calculations were done within ChimeraX[85,86]. RMSD values for structural superpositions were calculated using the matchmaker command. Reported values represent the RMSD of all Calpha carbons. Buried surface area calculations were performed using the measure buriedarea command**.**

Biolayer Interferometry (BLI)

The RhGB02 spike trimer was designed to mimic the native trimeric conformation of the protein. It consists of a gene synthesized by Genscript of CHO codon-optimized sequence of RhGB02, residues 1-1191, preceded by a u-phosphatase signal peptide[87], residues 969 and 970 mutated to proline (2P) to stabilize the prefusion state of the spike trimer, a putative basic site that may be the site of protelysis (RAKQ, residues 669-672, was mutated to GASQ), a C-terminal T4 foldon fusion domain to stabilize the trimer complex, followed by C-terminal 8x His and 2x Strep tags for affinity purification. This gene was cloned with the pcDNA3.1(+) vector. The trimeric RhGB02 spike protein was expressed as previously reported as for the SARS-CoV-2 spike transiently expressed in suspension-adapted ExpiCHO cells (Thermo Fisher) in ProCHO5 medium (Lonza) at 5 x10^6 cells/mL using PEI MAX (Polysciences) for DNA delivery[88]. At 1 h post-transfection, dimethyl sulfoxide (DMSO; AppliChem) was added to 2% (v/v). Following a 7-day incubation with agitation at 31 °C and 4.5% CO_2, the cell culture medium was harvested and clarified using a 0.22 µm filter. The conditioned medium was loaded onto Streptactin XT columns (IBA) washed with PBS and eluted with 50 mM biotin in 150 mM NaCl, 100 mM HEPES 7.5. Eluted protein was then dialyzed overnight into PBS. The purity of spike trimers was determined to be >99% pure by SDS-PAGE analysis.

Human (residues 19-615), little brown bat (19-629) and greater horseshoe bat (19-615) ACE2 genes were synthesized by Genscript and cloned in after the human pregnancy specific glycoprotein 1 signal peptide and is followed by a 3C protease cleavage site, a mouse IgG2a Fc fragment and a 10x His tag (only for the hACE2 construct). Protein production was produced exactly as for the RhGB02 spike. The filtered conditioned media was then subjected to Protein A purification. Eluted protein was dialyzed into PBS.

Experiments were performed on a Gator BLI system. Running buffer was 1X PBS. Dimeric mFc-hACE2 and bat ACE2 were diluted to 10 µg/mL and captured with MFc tips (GatorBio). Loaded tips were dipped into 2-fold serial dilution series (highest concentration 3000 nM) of the RhGB02 spike protein. Curves were processed using the Gator software with a 1:1 fit after background subtraction. Plots were generated in Prism v9.

Data analysis and visualisation

All data analyses were performed using *R* v4.1.0 or *Python* v3.9.12. Visualisations were performed using *ggplot v3.3.5*[89].

References

1. Taylor, L. H., Latham, S. M. & Woolhouse, M. E. J. Risk factors for human disease emergence. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**, 983–989 (2001).
2. Jones, K. E. *et al.* Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008).
3. ASM Biodiversity Committee. Mammal Diversity Database (1.10) [Data set]. *Zenodo* (2022) doi:https://doi.org/10.5281/zenodo.7394529.
4. Olival, K. J. *et al.* Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017).
5. Letko, M., Seifert, S. N., Olival, K. J., Plowright, R. K. & Munster, V. J. Bat-borne virus diversity, spillover and emergence. *Nat. Rev. Microbiol. 2020 188* **18**, 461–471 (2020).
6. Mollentze, N. & Streicker, D. G. Viral zoonotic risk is homogenous among taxonomic orders of mammalian and avian reservoir hosts. *Proc. Natl. Acad. Sci.* **117**, 9423 LP – 9430 (2020).
7. Li, W. *et al.* Bats are natural reservoirs of SARS-like coronaviruses. *Science* **310**, 676–679 (2005).
8. Mohd, H. A., Al-Tawfiq, J. A. & Memish, Z. A. Middle East respiratory syndrome coronavirus (MERS-CoV) origin and animal reservoir. *Virol. J.* **13**, 87 (2016).
9. Anthony, S. J. *et al.* Further evidence for bats as the evolutionary source of Middle East respiratory syndrome coronavirus. *MBio* **8**, (2017).

10. Zhou, H. *et al.* A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr. Biol.* **30**, 2196–2203 (2020).
11. Hul, V. *et al.* A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *bioRxiv* 2021.01.26.428212 (2021) doi:10.1101/2021.01.26.428212.
12. Wacharapluesadee, S. *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat. Commun.* **12**, 972 (2021).
13. Plowright, R. K. *et al.* Pathways to zoonotic spillover. *Nat. Rev. Microbiol.* **15**, 502–510 (2017).
14. Hassell, J. M., Begon, M., Ward, M. J. & Fèvre, E. M. Urbanization and disease emergence: dynamics at the wildlife–livestock–human interface. *Trends Ecol. Evol.* **32**, 55–67 (2017).
15. Peiris, J. S., Guan, Y. & Yuen, K. Severe acute respiratory syndrome. *Nat. Med.* **10**, S88–S97 (2004).
16. Memish, Z. A. *et al.* Human infection with MERS coronavirus after exposure to infected camels, Saudi Arabia, 2013. *Emerg. Infect. Dis.* **20**, 1012–1015 (2014).
17. Balloux, F. *et al.* The past, current and future epidemiological dynamic of SARS-CoV-2. *Oxf. Open Immunol.* **3**, iqac003 (2022).
18. Temmam, S. *et al.* Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. (2021).

19. Kumakamba, C. *et al.* Coronavirus surveillance in wildlife from two Congo basin countries detects RNA of multiple species circulating in bats and rodents. *PloS One* **16**, e0236971 (2021).
20. Crook, J. M. *et al.* Metagenomic identification of a new sarbecovirus from horseshoe bats in Europe. *Sci. Rep.* **11**, 1–9 (2021).
21. Yadav, P. D. *et al.* Detection of coronaviruses in Pteropus & Rousettus species of bats from different States of India. *Indian J. Med. Res.* **151**, 226 (2020).
22. Valitutto, M. T. *et al.* Detection of novel coronaviruses in bats in Myanmar. *PloS One* **15**, e0230802 (2020).
23. Rizzo, F. *et al.* Coronavirus and paramyxovirus in bats from Northwest Italy. *BMC Vet. Res.* **13**, 1–11 (2017).
24. Lin, X.-D. *et al.* Extensive diversity of coronaviruses in bats from China. *Virology* **507**, 1–10 (2017).
25. Waruhiu, C. *et al.* Molecular detection of viruses in Kenyan bats and discovery of novel astroviruses, caliciviruses and rotaviruses. *Virol. Sin.* **32**, 101–114 (2017).
26. Shehata, M. M. *et al.* Surveillance for coronaviruses in bats, Lebanon and Egypt, 2013–2015. *Emerg. Infect. Dis.* **22**, 148 (2016).
27. Kemenesi, G. *et al.* Molecular survey of RNA viruses in Hungarian bats: discovering novel astroviruses, coronaviruses, and caliciviruses. *Vector-Borne Zoonotic Dis.* **14**, 846–855 (2014).
28. August, T. A., Mathews, F. & Nunn, M. A. Alphacoronavirus detected in bats in the United Kingdom. *Vector-Borne Zoonotic Dis.* **12**, 530–533 (2012).

29. De Souza Luna, L. K. *et al.* Generic Detection of Coronaviruses and Differentiation at the Prototype Strain Level by Reverse Transcription-PCR and Nonfluorescent Low-Density Microarray. *J. Clin. Microbiol.* **45**, 1049 (2007).
30. Xiu, L. *et al.* A RT-PCR assay for the detection of coronaviruses from four genera. *J. Clin. Virol.* **128**, 104391 (2020).
31. Watanabe, S. *et al.* Bat coronaviruses and experimental infection of bats, the Philippines. *Emerg. Infect. Dis.* **16**, 1217 (2010).
32. Vijgen, L., Moës, E., Keyaerts, E., Li, S. & Ranst, M. V. A pancoronavirus RT-PCR assay for detection of all known coronaviruses. in *SARS-and Other Coronaviruses* 3–12 (Springer, 2008).
33. Holbrook, M. G. *et al.* Updated and validated pan-coronavirus PCR assay to detect all coronavirus genera. *Viruses* **13**, 599 (2021).
34. Nayfach, S. *et al.* CheckV assesses the quality and completeness of metagenome-assembled viral genomes. *Nat. Biotechnol.* **39**, 578–585 (2021).
35. Ondov, B. D. *et al.* Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* **17**, 132 (2016).
36. Crook, J. M. *et al.* Metagenomic identification of a new sarbecovirus from horseshoe bats in Europe. *Sci. Rep.* **11**, 1–9 (2021).
37. Kreuder Johnson, C. *et al.* Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci. Rep.* **5**, 14830 (2015).
38. Woolhouse, M. E. J. & Gowtage-Sequeria, S. Host range and emerging and reemerging pathogens. *Emerg. Infect. Dis.* **11**, 1842 (2005).

39. Winter, R., Mantilla-Contreras, J. & Schmidt, S. Usage of buildings in the life cycle of two endangered Rhinolophus species in the Mediterranean region: implications for roost protection. *Eur. J. Wildl. Res.* **66**, 1–13 (2020).
40. Temmam, S. *et al.* Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. (2021).
41. Liu, K. *et al.* Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. *Cell* **184**, 3438–3451 (2021).
42. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
43. Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215–220 (2020).
44. Temmam, S. *et al.* Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. (2021).
45. Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864–1868 (2005).
46. Sander, A.-L. *et al.* Genomic determinants of Furin cleavage in diverse European SARS-related bat coronaviruses. *Commun. Biol.* **5**, 1–8 (2022).
47. Peacock, T. P. *et al.* The furin cleavage site in the SARS-CoV-2 spike protein is required for transmission in ferrets. *Nat. Microbiol. 2021 67* **6**, 899–909 (2021).
48. Starr, T. N. *et al.* ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature* **603**, 913–918 (2022).

49. Zech, F. *et al.* Spike mutation T403R allows bat coronavirus RaTG13 to use human ACE2. *bioRxiv* 2021.05.31.446386 (2021) doi:10.1101/2021.05.31.446386.

50. Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* **5**, 1408–1417 (2020).

51. Bobay, L.-M., O'Donnell, A. C. & Ochman, H. Recombination events are concentrated in the spike protein region of Betacoronaviruses. *PLoS Genet.* **16**, e1009272 (2020).

52. Jones, S. *et al.* Testing bats in rehabilitation for SARS-CoV-2 before release into the wild. *Conserv. Sci. Pract.* e12707 (2022) doi:10.1111/CSP2.12707.

53. Menachery, V. D. *et al.* Trypsin treatment unlocks barrier for zoonotic bat coronavirus infection. *J. Virol.* **94**, e01774-19 (2020).

54. Boni, M. F. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* **5**, 1408–1417 (2020).

55. Kunz, T. H., Braun de Torrez, E., Bauer, D., Lobova, T. & Fleming, T. H. Ecosystem services provided by bats. *Ann. N. Y. Acad. Sci.* **1223**, 1–38 (2011).

56. Bat Conservation Trust. *The National Bat Monitoring Programme Annual Report 2021.* (2021).

57. Mathews, F. & Harrower, C. IUCN–compliant Red List for Britain's Terrestrial Mammals. Assessment by the Mammal Society under contract to Natural England, Natural Resources Wales and Scottish Natural Heritage. (2020).

58. Gibb, R. *et al.* Zoonotic host diversity increases in human-dominated ecosystems. *Nature* **584**, 398–402 (2020).

59. Eby, P. *et al.* Pathogen spillover driven by rapid changes in bat ecology. *Nature* 1–3 (2022).

60. Jones, S. *et al.* Testing bats in rehabilitation for SARS-CoV-2 before release into the wild. *Conserv. Sci. Pract.* e12707 (2022) doi:10.1111/CSP2.12707.

61. Khare, S. *et al.* GISAID's Role in Pandemic Response. *China CDC Wkly.* **3**, 1049–1051 (2021).

62. Meleshko, D., Hajirasouliha, I. & Korobeynikov, A. coronaSPAdes: from biosynthetic gene clusters to RNA viral assemblies. *Bioinformatics* **38**, 1–8 (2021).

63. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

64. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

65. Blum, M. *et al.* The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **49**, D344–D354 (2021).

66. Zdobnov, E. M. & Apweiler, R. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).

67. Karger, D. N. *et al.* Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **4**, 1–20 (2017).

68. Morton, R. D., Marston, C. G., O'Neil, A. W. & Rowland, C. S. Land Cover Map 2019 (land parcels, GB). (2020).

69. Naimi, B. & Araújo, M. B. sdm: a reproducible and extensible R platform for species distribution modelling. *Ecography* **39**, 368–375 (2016).

70. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).

71. Zielezinski, A., Vinga, S., Almeida, J. & Karlowski, W. M. Alignment-free sequence comparison: benefits, applications, and tools. *Genome Biol.* **18**, 186 (2017).

72. Tan, C. C. S. *et al.* Pre-existing T cell-mediated cross-reactivity to SARS-CoV-2 cannot solely be explained by prior exposure to endemic human coronaviruses. *Infect. Genet. Evol.* **95**, 105075 (2021).

73. Huddleston, J. *et al.* Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *J. Open Source Softw.* **6**, 2906 (2021).

74. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

75. Minh, B. Q., Nguyen, M. A. T. & von Haeseler, A. Ultrafast Approximation for Phylogenetic Bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).

76. Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).

77. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).

78. Martin, D. P., Murrell, B., Golden, M., Khoosal, A. & Muhire, B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* **1**, (2015).

79. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).

80. Pagès H, P, A., R, G. & S, D. Biostrings: Efficient manipulation of biological strings. R package version 2.64.0, Https://bioconductor.org/packages/Biostrings. Preprint at (2022).

81. Okonechnikov, K. *et al.* Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).

82. Conceicao, C. *et al.* The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. (2020) doi:10.1371/journal.pbio.3001016.

83. Rebendenne, A. *et al.* SARS-CoV-2 triggers an MDA-5-dependent interferon response which is unable to control replication in lung epithelial cells. *J. Virol.* **95**, e02415-20 (2021).

84. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

85. Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).

86. Goddard, T. D. *et al.* UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).

87. Wrobel, A. G. *et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat. Struct. Mol. Biol.* **27**, 763–767 (2020).

88. Fenwick, C. *et al.* Patient-derived monoclonal antibody neutralizes SARS-CoV-2 Omicron variants and confers full protection in monkeys. *Nat. Microbiol.* **7**, 1376–1389 (2022).

89. Wickham, H. ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).

Author contributions

VS, CC, ER, GW and TB wrote the grant application that supported this research. VS and FB supervised the research. Primary analysis on the project was carried out by CCST, TPP, KYM and CH with contributions from FB, LvD, WDP, DO, and WB. CCST wrote the initial draft of manuscript, with subsequent rounds of editing from VS, FB and LvD. All authors provided intellectual contributions to the manuscript.

Declaration of competing interest

Authors declare that to current knowledge, there are no legal, financial or personal competing interests.

Acknowledgments

We would like to acknowledge the following bat rehabilitators, conservationists and researchers for the collection of bat samples for this study: Margaret Grimsey, Hazel Ryan, Gareth Harris, Danielle Linton, Sam Smith, Ross Baker, Lynn Whitfield, Colleen Hope, Josh Sowden, Emily Dickens, Tricia Scott, Jonah Tosney, Eilish Rothney, Dale Irvine, Joe Salkeld, Alice Samuel, Amanda Miller, Sheila Wright, Laura Kahane, Stewart Rowden, Joy Hall, Jessica Dangerfield, Catherine Wood, Rachael Tarlinton, Ternenge Apaa, and Fiona Mathews. We also thank Rachael Tarlinton, Ternenge Apaa, and Fiona Mathews for comments on the manuscript, Scott Jones, Faye Hobbs and Danielle Harris for assisting with sample handling and processing, Xavier Didelot for advice on genetic recombination analyses, all authors and labs that have produced and deposited the genomes used in this study on NCBI GenBank and GISAID (Supplementary Table 2 and 3), and the NERC Omics facility for sequencing. Finally, we thank NERC (UKRI

Covid Urgency grant) and the European Commission (Horizon 2021-2024, END-VOC Project) for funding.

Data and code availability

All novel genomes were uploaded to NCBI GenBank under the BioProject accession XXX. All custom code used to perform the analyses reported here are hosted on GitHub (https://github.com/cednotsed/bat-CoVs.git).