Metagenomic Identification of a New Sarbecovirus from Horseshoe Bats in Europe

Jack Crook
  Public Health England, National Infection Service, Porton Down, Salisbury, UK

Ivana Murphy
  Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich

Daniel Carter
  Public Health England, National Infection Service, Porton Down, Salisbury, UK

Steven Pullan
  Public Health England, National Infection Service, Porton Down, Salisbury, UK

Miles Carroll
  Public Health England, National Infection Service, Porton Down, Salisbury, UK

Richard Vipond
  Public Health England, National Infection Service, Porton Down, Salisbury, UK

Andrew Cunningham (✉ a.cunningham@ioz.ac.uk)
  Institute of Zoology, Zoological Society of London, London

Diana Bell
  Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich

METAGENOMIC IDENTIFICATION OF A NEW SARBECOVIRUS FROM HORSESHOE BATS IN EUROPE

Jack Crook[1,*], Ivana Murphy[2,*], Daniel Carter[1], Steven Pullan[1], Miles Carroll[1,3], Richard Vipond[1], Andrew A. Cunningham[4,¶], Diana Bell[2,¶]

[1] Public Health England, National Infection Service, Porton Down, Salisbury, UK

[2] Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

[3] Wellcome Trust Centre for Human Genetics, Nuffield Department of Medicine, Oxford University, Oxford, UK.

[4] Institute of Zoology, Zoological Society of London, London NW1 4RY, UK

[*]These authors contributed equally to the research presented in this manuscript.

[¶]These authors contributed equally to the research presented in this manuscript.

Corresponding author: Andrew A. Cunningham; a.cunningham@ioz.ac.uk

Abstract

The source of the COVID-19 pandemic is unknown, but the natural host of the progenitor sarbecovirus is thought to be Asian horseshoe (rhinolophid) bats. We identified and sequenced a novel sarbecovirus (RhGB01) from a British horseshoe bat, at the western extreme of the rhinolophid range. Our results extend both the geographic and species ranges of sarbecoviruses and suggest their presence throughout the horseshoe bat distribution. Within the receptor binding domain, but excluding the receptor binding motif, RhGB01 has a 77% (SARS-CoV-2) and 81% (SARS-CoV) amino acid homology. While apparently lacking hACE2 binding ability, and hence unlikely to be zoonotic without mutation, RhGB01 presents opportunity for SARS-CoV-2 and other sarbecovirus homologous recombination. Our findings highlight that the natural distribution of sarbecoviruses and opportunities for recombination through intermediate host co-infection are underestimated. Preventing transmission of SARS-CoV-2 to bats is critical with the current global mass vaccination campaign against this virus.

Introduction

The sources of the current COVID-19 pandemic and of the 2003 Severe Acute Respiratory Syndrome (SARS) epidemic are unknown[1]. Currently, the natural hosts of both SARS-CoV and SARS-CoV-2 (family Coronaviridae; subgenus *Sarbecovirus*)[2], the causative agents of SARS and COVID-19 respectively, are thought to be horseshoe bats (Rhinolophidae), with the zoonotic spillover process involving one or more intermediate hosts, during which time viral mutation, recombination and/or amplification could have occurred[3-7]. Phylogenetic analyses of novel horseshoe bat sarbecoviruses in China have shown these to be most closely related to both SARS-CoV and to SARS-CoV-2[3,4,7]. Recently, a SARS-CoV-2-like virus was also reported from a species of horseshoe bat in Thailand[8].

The range of horseshoe bats extends across much of the Old World, but most sampling for coronaviruses has been conducted in East and South East Asia, where around 50 SARS-related coronaviruses (SARSr-CoVs) have been detected across ten species of bat, with 48 of these being from nine species of horseshoe bat[8-11]. Here we expand the investigation of SARSr-CoVs to the lesser horseshoe bat (*Rhinolophus hipposideros*) in the UK, which is at the western extreme of the range of the Rhinolophidae.

Methods

Sample Collection

Samples were collected directly from 53 individual lesser horseshoe bats *R. hipposideros* caught during routine annual population monitoring surveys at three sites in August and September 2020. These were in Somerset (n=20 bats captured) and Gloucestershire (n = 26) in England and in Monmouthshire, Wales (n=7). Bats were captured using harp traps or mist nets placed near roosts or in woodland under government license following approval by the University of East Anglia Ethics Committee and adhering to UK Government COVID-19 safety regulations in place at the time. A faecal pellet was collected from each of 49 bats held individually in sterile holding bags and the other four samples were collected as anal swabs using rayon-tipped dry swabs (MW100; Medical Wire & Equipment), which were taken from bats that did not defaecate when captured. All bats were released at the site of capture immediately after sample collection. Each sample was transferred directly into an individual sterile tube containing 2 ml RNAlater, refrigerated overnight and stored frozen prior to analysis.

Genomic sequencing

For metagenomic analysis, samples were homogenised by vortexing and spiked with 10^6 genome copies per ml of Hazara virus as an internal control. A 140 μl aliquot of each sample was extracted using the QIAamp Viral RNA extraction kit. Extracts were DNase treated, reverse transcribed and randomly amplified using a Sequence-Independent Single-Primer Amplification (SISPA) based method described in detail previously[12]. Illumina sequencing used the Nextera XT protocol with 2 × 150-bp paired-end sequencing on a MiSeq. Nanopore library preparation was as described previously[12] and sequencing was performed on an Oxford nanopore GridION with base calling via Guppy. Nanopore reads were trimmed using NanoFilt[13] to remove 25bp SISPA primer sequences from the start and end of each read. Raw data and the assembly sequence are deposited at NCBI under BioProject PRNJA706167.

Genomic analyses

Read-level taxonomic classification for each sample was performed using Kraken2 against the RefSeq database (2.0.8-beta)[14]. *De novo* genome assembly was performed using SPAdes (3.15.1) for both Illumina and Illumina/nanopore hybrid assemblies[15]. Contigs of interest were identified using BLASTn[16]. Illumina reads were mapped to the assembled contigs of interest using BWA-MEM[17] and nanopore reads using Minimap2[18]. Read depth values were generated using SAMtools (1.10)[19].

The assembled genome was aligned with selected reference genomes (NC_014470.1, KJ473814.1, NC_045512.2) in MEGA X (10.2.4) using MUSCLE alignment[20,21]. Nucleotide and codon alignments were generated for each gene with visual depictions of alignment and pairwise alignment scores generated in JALVIEW (2.11.1.3)[22].

Aligned nucleotide sequences from 21 sarbecoviruses obtained from GenBank were used to generate maximum likelihood trees using IQTREE (2.0.3)[23]. Using the best fitting nucleotide substitution model as selected by ModelFinder[24]. For the entire coding sequence and S protein, the nucleotide substitution mode is GTR+F+R10, for the ORF1ab sequence the model selected was GTR+F+G4. Models were selected according to the lowest Bayesian information criterion (BIC) score. Nodes were evaluated using UFBoot with 1,000 bootstrap approximations[25]. Further optimisation of each bootstrap was provided using a hill-climbing nearest neighbour interchange search on each bootstrap repeat. Phylogenetic tree visualisation was carried out in iTOL (5.7)[26].

Results

Metagenomic analysis revealed an unclassified betacoronavirus in a single sample with genome organisation consistent with *Sarbecovirus*

An initial screening of the 53 samples identified one sample from Gloucestershire with >650 reads classified to the *Coronavirinae* family, with the positive control spike of Hazara virus detected in 49/53 samples. The percentage of reads classified in total in each sample ranged from 28% to 96% and in both the extraction and library preparation negative controls, no other significant level of reads classified to RNA viruses. Hazara virus was detected in the extraction negative, but not in the library preparation negative.

Classification of the reads in the positive sample identified 0.41% of reads (2550/614996) as being viral in origin. Of these viral reads, 68% of reads (1668) were classified at species level to bat betacoronavirus BM48-31/BGR/2008 (GenBank reference NC_014470.1)[27]. *De novo* assembly of the Illumina reads generated multiple contigs with homology to members of the subgenus *Sarbecovirus* as assessed via BLASTn; the largest single contig being ~7kb. To

investigate, further additional sequence data were generated using Oxford nanopore technology. Using the 562,461 nanopore reads with an average length of ~600kb as scaffold, a hybrid assembly generated a single 29 kb contig similar to sarbecovirus genomes in both size and gene organisation (Figure 1a). To increase confidence in the assembly we performed further Illumina sequencing. With the increased depth, Illumina data alone were assembled into a contig of 21kb, which was again further assembled to a 29kb contig with the inclusion of the nanopore data in a hybrid assembly. Mapping all raw reads to this contig shows that 0.97% and 2.37% map to the contig for Illumina and Nanopore respectively. In total, mean read depth along the assembly is ~25x for Illumina data and 20x for nanopore data (Figure 1b). Combined depth coverage across the assembly is ~50x, confidently supporting the presence of this virus, in the positive sample.

BLASTn analysis of the GenBank nr/nt database shows the assembly shares the highest nucleotide identity with a bat betacoronavirus, BtRs-betaCoV/Hub2013 (GenBank reference KJ473814.1), with 81.01% identity across 85% of the assembly. By comparison it shares 79.78% nucleotide identity across 85% of the assembly with SARS-CoV-2 (isolate SARS-CoV-2/human/USA/FL-CDC-STM-000005640/2021, MW586221.1). We named the virus identified as RhGB01 (*Rhinolophus hipposideros*, Great Britain 01) representing the first detection of a sarbecovirus from *R. hipposideros* in Great Britain.

Genomic organisation of RhGB01 mostly mirrors that of other sarbecoviruses with 10 coding genes, with start codons at sites identical or one codon separated from defined sarbecoviruses. In comparison, SARS-CoV-2, SARS and viruses related to SARS-CoV-2 harbour 11 coding genes. In RhGB01, ORF8 and 20 bases in the 5' region of the ORF7b transcript are absent,

comparable to BM48-31/BGR/2008, the closest related virus as determined by phylogenetic analysis (Figure 2; Supplementary Figure 1).

Phylogenetic analysis

Maximum likelihood phylogenies of betacoronavirus ORF1ab, S protein and entire coding regions (CDS) demonstrate that in all three phylogenies, RhGB01 clusters with BM48-31/BGR/2008, a sarbecovirus isolated from a Blasius's horseshoe bat (*Rhinolophus blasii*) in 2008 in Bulgaria (Figure 2; Supplementary Figure 1). RhGB01 is in a distinct clade from the human pathogenic betacoronaviruses SARS-CoV and SARS-CoV-2 but, of these, is more closely related to SARSr-CoVs.

Spike protein comparison

The major human cellular entry receptor for both SARS and SARS-CoV-2 is Angiotensin-converting enzyme 2 (hACE2). This binding ability is conferred by a receptor binding motif (RBM) within the receptor binding domain (RBD) of the spike protein. RhGB01 shares amino acid identity of 68% and 67% across the RBD with SARS and SARS-COV-2 respectively and just 43% and 48% within the RBM (Figure 3a). By comparison across the RBD, the closest SARSr-CoV-2 viruses from bat and pangolin host shares 89% and 86% amino acid homology to SARS-CoV-2 and 75.77% to SARS (Supplementary Figure 2). Within the RBM SARSr-CoV-2 viruses from bat and pangolin hosts both share 75% amino acid homology to SARS-CoV-2 and 50% and 49% to SARS (Figure 3b). RhGB01 also shows little homology to the RBM of Middle East Respiratory Syndrome (MERS) virus (Figure 3a). The RhGB01 spike amino acid sequence contains motifs comparable to host transmembrane serine protease 2 cleavage site (TMPRSS2) seen in both SARS-CoV-2 and SARS in the S2' target site but lacks

the additional furin cleavage site specific to SARS-CoV-2 at the S1/S2 intersection (Figure 3a).

Discussion

Here we discovered a novel sarbecovirus (RhGB01), the first to be described in the UK, after sampling just 53 lesser horseshoe bats. While other sarbecoviruses have been identified in rhinolophid bats in other European countries by polymerase chain reaction and partial gene sequence analyses, RhGB01 is only the second from Europe to be fully sequenced[28], and the first from a lesser horseshoe bat. The only other full sequence betacoronavirus from a European horseshoe bat is BM48-31/BGR/2008 from *R. blasii*. Our results, therefore, extend the geographic and species ranges of SARSr-CoVs and suggest that sarbecoviruses are likely to be present throughout the range of the Rhinolophidae, which are distributed from Australia and Japan to Europe and Africa.

The range of the lesser horseshoe bat extends from Western Europe to Central Asia, overlapping with those of other rhinolophid species, including the greater horseshoe bat (*R. ferrumequinum*), which ranges from Western Europe to Japan[29,30]. Where they co-exist, the species can be syntopic allowing opportunity for cross-species virus transfer. Prior to our results, the observed and predicted (cut off >=0.9821) number of coronaviruses in the greater horseshoe bat were 13 and 19 respectively, and in the lesser horseshoe bat these figures were zero and three respectively[31]. This suggests that the complement of *Sarbecovirus* species in horseshoe bats is greater than predicted so far, with the possibility of virus sharing across species and large geographic areas.

Genomic alignments between RhGB01 and related sarbecoviruses highlight key genomic differences between RhGB01 and known zoonotic sarbecoviruses. Host specificity is dependent on the ability of a virus to attach to host receptors and enter host cells; a binding process facilitated by contact residues contained within the receptor binding motif[32]. RhGB01 demonstrates low amino acid homology to SARS and SARS-CoV-2 in the receptor binding motif compared to that between SARS, SARS-CoV-2 and the closely related bat (RaTG13) and pangolin (PCoV_GX_P4L) sequences identified in Asia[33,34]. The low level of homology most likely indicates a lack of ability to bind hACE2 and, hence, RhGB01 is unlikely to be zoonotic without mutation. To confirm the absence of hACE2 or other human cell receptor binding, *in silico* structural modelling or *in vitro* binding assays are required.

Aside from the variation observed in amino acid homology within the RBM, RhGB01 also exhibits variation within the furin cleavage site and ORF8 when compared to zoonotic sarbecoviruses. The absence of the furin cleavage site indicates the absence of enhanced efficiency of host cell entry observed with SARS-CoV-2[35]. However, RhGB01 does retain a similar motif responsible for cleavage in the S2' region by host transmembrane serine protease 2; also required for spike protein proteolytic priming for hACE2 attachment[32]. SARS-CoV-2 variants with a functional ORF8 are associated with greater pathogenicity, thought to be due to downregulation of major histocompatibility complex class 1 (MHC 1), and thus a reduction in antigen presentation to CD8+ T lymphocytes which facilitates prolonged infection[36,37]. The absence of ORF8 from the genome of RhGB01 suggests that this virus lacks these immune evasion properties.

It has been postulated that both SARS-CoV and SARS-CoV-2 evolved through mutation, possibly involving homologous recombination, during passage through at least one

intermediate host; probably civets (viverrids) or mustelids for SARS-CoV[38,39] and possibly a species of pangolin for SARS-CoV-2[34]. In this way, the progenitor virus from the natural host (a species of horseshoe bat) gained genetic adaptations to allow successful infection of, and transmission between, human beings. Where there is opportunity for homologous recombination of sarbecoviruses through co-infection, there is the possibility of novel zoonotic emergence. Thus, co-infection of horseshoe bats with their natural suites of coronaviruses and with SARS-CoV-2 could lead to the development of novel zoonotic emergence. While there is a need to increase surveillance for coronaviruses in horseshoe bats across their range, and also in other bat species, especially those syntopic with, or closely related to, horseshoe bats (e.g. the Old World leaf-nosed bats, family Hipposideridae), it is also important that steps are taken to minimise opportunities of virus transmission between novel hosts.

In Europe, unlike in Asia, direct contact between people and bats most commonly occurs when the animals are captured by bat researchers or when sick animals are taken in by bat rescuers and wildlife rehabilitation centres. While the risk of reverse spill over of SARS-CoV-2 from researchers to bats and onward spread within bat populations has been shown to be medium to high[40], it is the caring of sick or injured bats, in particular, that provides most opportunity for long-term close contact and virus transfer in either direction. Although the IUCN Bat Specialist Group has produced guidelines to minimise this risk[41], the degree to which these are known or followed is unclear. Our findings highlight that the natural distribution of sarbecoviruses and opportunities for recombination through intermediate host co-infection are underestimated. Preventing transmission of SARS-CoV-2 to horseshoe bats, with the risk this presents of further mutation, is of particular significance with the current roll out of a global mass vaccination campaign against this virus.

References

1. WHO, 2020. *Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19).* [online] Available at: <https://www.who.int/publications/i/item/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)> [Accessed 24 February 2021].

2. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5**, 536–544 (2020).

3. Boni MF, Lemey P, Jiang X, Lam TT, Perry BW, Castoe TA, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol 2020;5:1408-17. https://doi.org/10.1038/s41564-020-0771-4

4. Li, W. Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science* **310**, 676–679 (2005).

5. Lau, S. K. P. *et al.* Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc Natl Acad Sci U S A* **102**, 14040–14045 (2005).

6. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. & Garry, R. F. The proximal origin of SARS-CoV-2. *Nat Med* 1–3 (2020) doi:10.1038/s41591-020-0820-9.

7. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).

8. Wacharapluesadee, S. *et al.* Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nat Commun* **12**, 972 (2021).

9. Luk, H. K. H., Li, X., Fung, J., Lau, S. K. P. & Woo, P. C. Y. Molecular epidemiology, evolution and phylogeny of SARS coronavirus. *Infect Genet Evol* **71**, 21–30 (2019).

10. Murakami, S., Kitamura, T., Suzuki, J., Sato, R., Aoi, T., Fujii, M., Matsugo, H., Kamiki, H., Ishida, H., Takenaka-Uema, A., Shimojima, M. & Horimoto, T. (2020) Detection and Characterization of Bat Sarbecovirus Phylogenetically Related to SARS-CoV-2, Japan. *Emerg Infect Dis* **26**, 3025-3029. DOI: https://doi.org/10.3201/eid2612.203386

11. Hul, V., Delaune, D., Karlsson, E. A., Hassanin, A., Tey, P.O., Baidaliuk, A., Gámbaro, F., Tu, V.T., Keatts, L., Mazet, J., Johnson, C., Buchy, P., Dussart, P., Goldstein, T., Simon-Lorière, E. & Duong, V. (2021) A novel SARS-CoV-2 related coronavirus in bats from Cambodia. BioRxiv.

12. Lewandowski, K. *et al.* Metagenomic Nanopore Sequencing of Influenza Virus Direct from Clinical Respiratory Samples. *J Clin Microbiol* **58**, (2019).

13. De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M. & Van Broeckhoven, C. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* **34**, 2666–2669 (2018).

14. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**, 257 (2019).

15. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **19**, 455–477 (2012).

16. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).

17. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997 [q-bio]* (2013).

18. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

19. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

20. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol* **35**, 1547–1549 (2018).

21. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).

22. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

23. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).

24. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast Model Selection for Accurate Phylogenetic Estimates. *Nat Methods* **14**, 587–589 (2017).

25. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution* **35**, 518–522 (2018).

26. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**, W256–W259 (2019).

27. Drexler, J. F. *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* **84**, 11336–11349 (2010).

28. Ar Gouilh, M. *et al.* SARS-CoV related Betacoronavirus and diverse Alphacoronavirus members found in western old-world. *Virology* **517**, 88–97 (2018).

29. Taylor, P. 2016. Rhinolophus hipposideros. The IUCN Red List of Threatened Species 2016: e.T19518A21972794. https://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T19518A21972794.en. Downloaded on 24 February 2021.

30. Piraccini, R. 2016. *Rhinolophus ferrumequinum. The IUCN Red List of Threatened Species* 2016: e.T19517A21973253. https://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T19517A21973253.en. Downloaded on 24 February 2021.

31. Wardeh, M., Baylis, M. & Blagrove, M. S. C. Predicting mammalian hosts in which novel coronaviruses can be generated. *Nat Commun* **12**, 780 (2021).

32. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).

33. 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.e3 (2020).

34. Lam, T. T.-Y. *et al.* Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* **583**, 282–285 (2020).

35. Papa, G. *et al.* Furin cleavage of SARS-CoV-2 Spike promotes but is not essential for infection and cell-cell fusion. *PLOS Pathogens* **17**, e1009246 (2021).

36. Young, B. E. *et al.* Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study. *Lancet* **396**, 603–611 (2020).

37. Zhang, Y. *et al. The ORF8 Protein of SARS-CoV-2 Mediates Immune Evasion through Potently Downregulating MHC-I.*

http://biorxiv.org/lookup/doi/10.1101/2020.05.24.111823 (2020) doi:10.1101/2020.05.24.111823.

38. Guan, Y. Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. *Science* **302**, 276–278 (2003).

39. Bell, D., Roberton, S. & Hunter, P. Animal origins of SARS Coronavirus: Possible links with the international trade in small carnivores. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **359**, 1107–14 (2004).

40. Common, S. M., Shadbolt, T., Walsh, K. & Sainsbury, A. W. The risk from SARS-CoV-2 to bat species in England and mitigation options for conservation field workers. *Transbound Emerg Dis* (2021) doi:10.1111/tbed.14035.

41. Jolliffe, T. *et al. IUCN SSC Bat Specialist Group (BSG) recommendations to reduce the risk of transmission of SARS-CoV-2 from humans to bats in bat rescue and rehabilitation centers - MAP: Minimize, Assess, Protect.* (2020).

Acknowledgments

Author Contributions

IM, DB & AAC devised the study. IM, JC, and DC conducted the study, JC & SP analysed the sequence data. MC, RV, DB & AAC obtained the funding. AAC, JC, SP, IM & DB wrote the manuscript, and all authors reviewed the manuscript.

Competing Interests

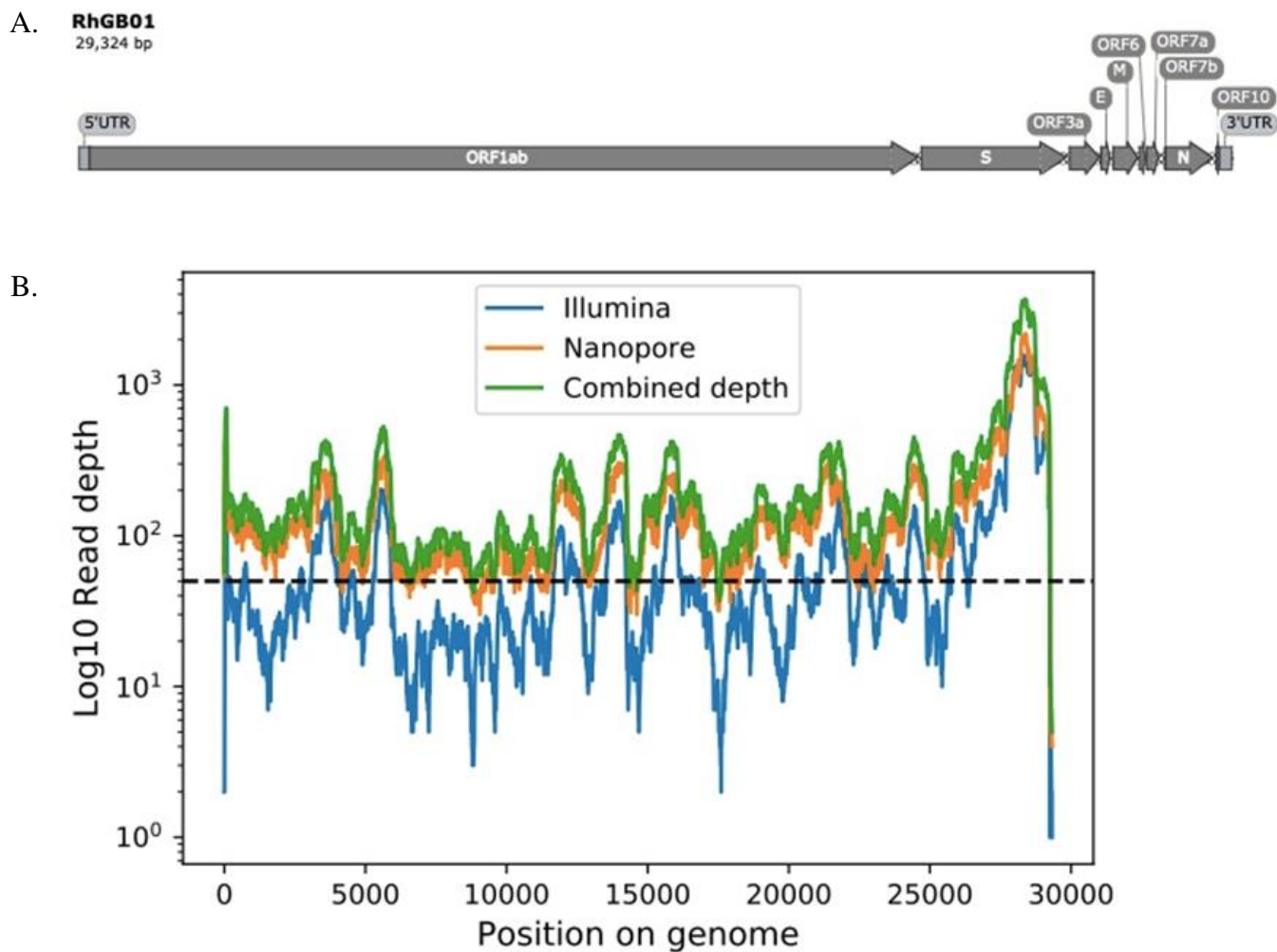None of the authors have any competing interests.

A.

B.

Figure 1. The genomic organisation of and the read depth across the assembled genome.
A) gene organisation of RhGB01 with 10 predicted coding genes. In RhGB01, lengths of the
genomic features are 5'UTR (279bp), ORF1ab (21kb), S (3.7kb), ORF3a (813bp), E (231bp),
M (669bp), ORF6 (189bp), ORF7ab (465), N (1254bp), ORF10 (78bp), 3'UTR (266bp). B)
Per base depth of coverage plotted across the genome from alignment of Illumina (blue),
Nanopore (orange) and combined reads (green). The horizontal line represents 50x read depth.
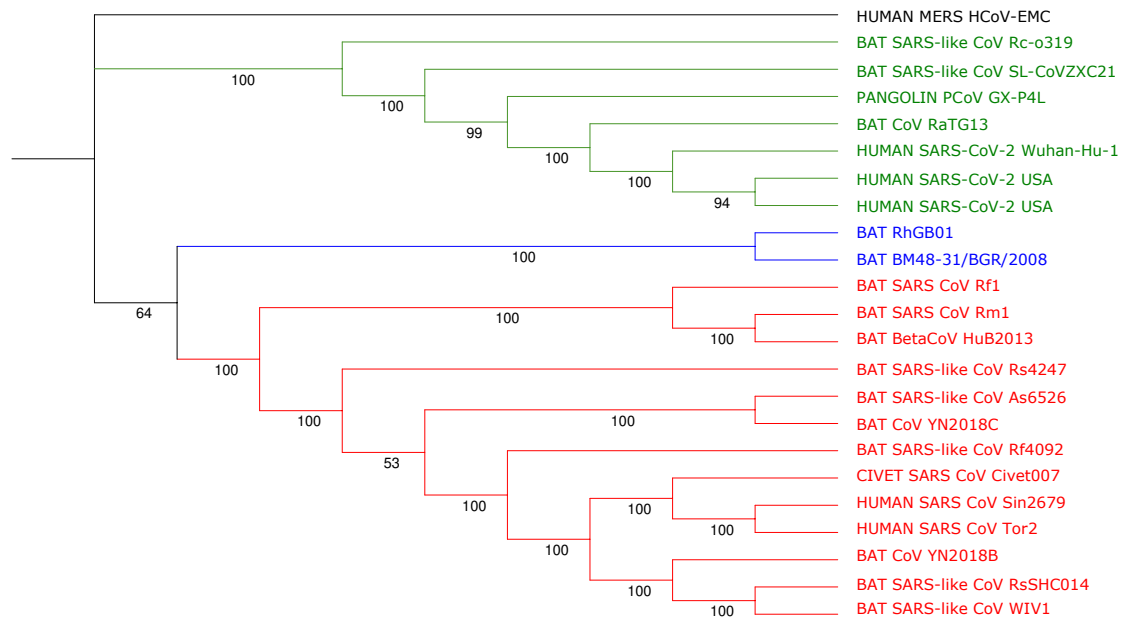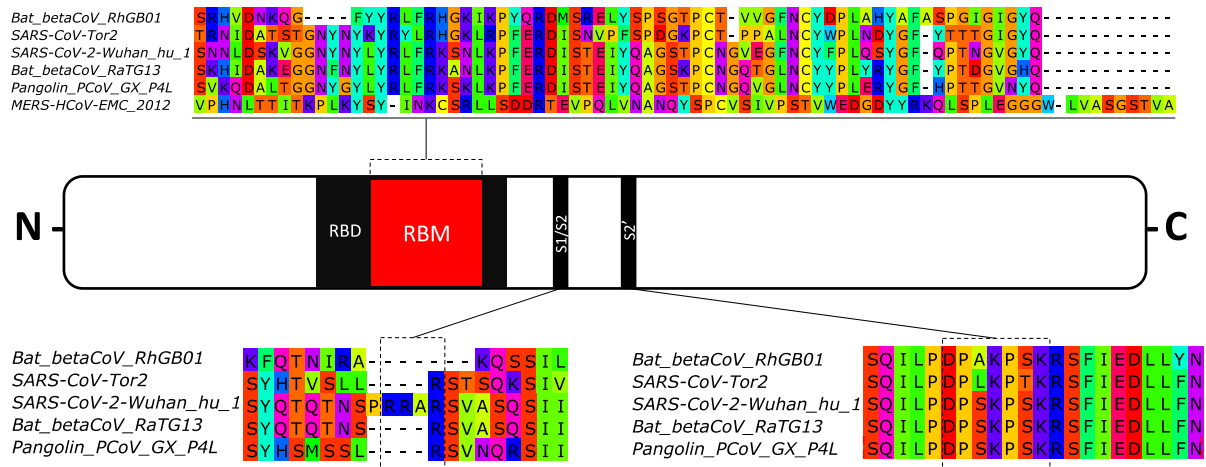
Figure 2.
Maximum likelihood phylogenies of RbGB01 and other sarbecoviruses of interest. Phylogeny inferred from the S protein nucleotide sequence. MERS (a *Merbecovirus*) is included as outgroup. SARS or nominally SARS-like betacoronaviruses are highlighted in red, SARS-CoV-2 or closely related viruses in green, with additional clade containing RhGB01, highlighted in blue. Bootstrap support values indicate confidence of divergence at each divergence event.

A.

B.

	RBM (%)		RBD (%)	
	SARS	SARS-CoV-2	SARS	SARS-CoV-2
RhGB01	43.48	48.57	68.21	67.35
RaTG13	50.72	75.36	75.77	89.18
PCoV	49.28	75.36	75.77	86.80

Figure 3.
A schematic representation of the entire S protein with the receptor binding motif and cleavage sites highlighted and compared between RhBG01, Pangolin and Bat derived related virus, SARS and SARS-CoV-2 and percentage identity values for the RBD and RBM regions. A) Amino acid residues are coloured according to the Taylor colour scheme. Receptor binding motif comparison highlights higher amino acid conservation within the SARS-CoV-2 and SARSr-CoV-2 viruses. RhGB01 exhibits low amino acid conservation to SARS and SARS-CoV-2. The receptor binding motif of MERS-HCoV-EMC-2012, subgenus *Merbecovirus*, demonstrates little amino acid homology to receptor binding motifs from the *Sarbecovirus* subgenus. The furin cleavage site (S1/S2) is only present in SARS-CoV-2, distinct from the TMPRSS2 cleavage motif (S2') which is more conserved. B) A table with percentage identity scores for the RBM and RBD to SARS and SARS-CoV-2. RaTG13 and PCoV represent the most closely SARSr-CoV-2 virus from zoonotic hosts.
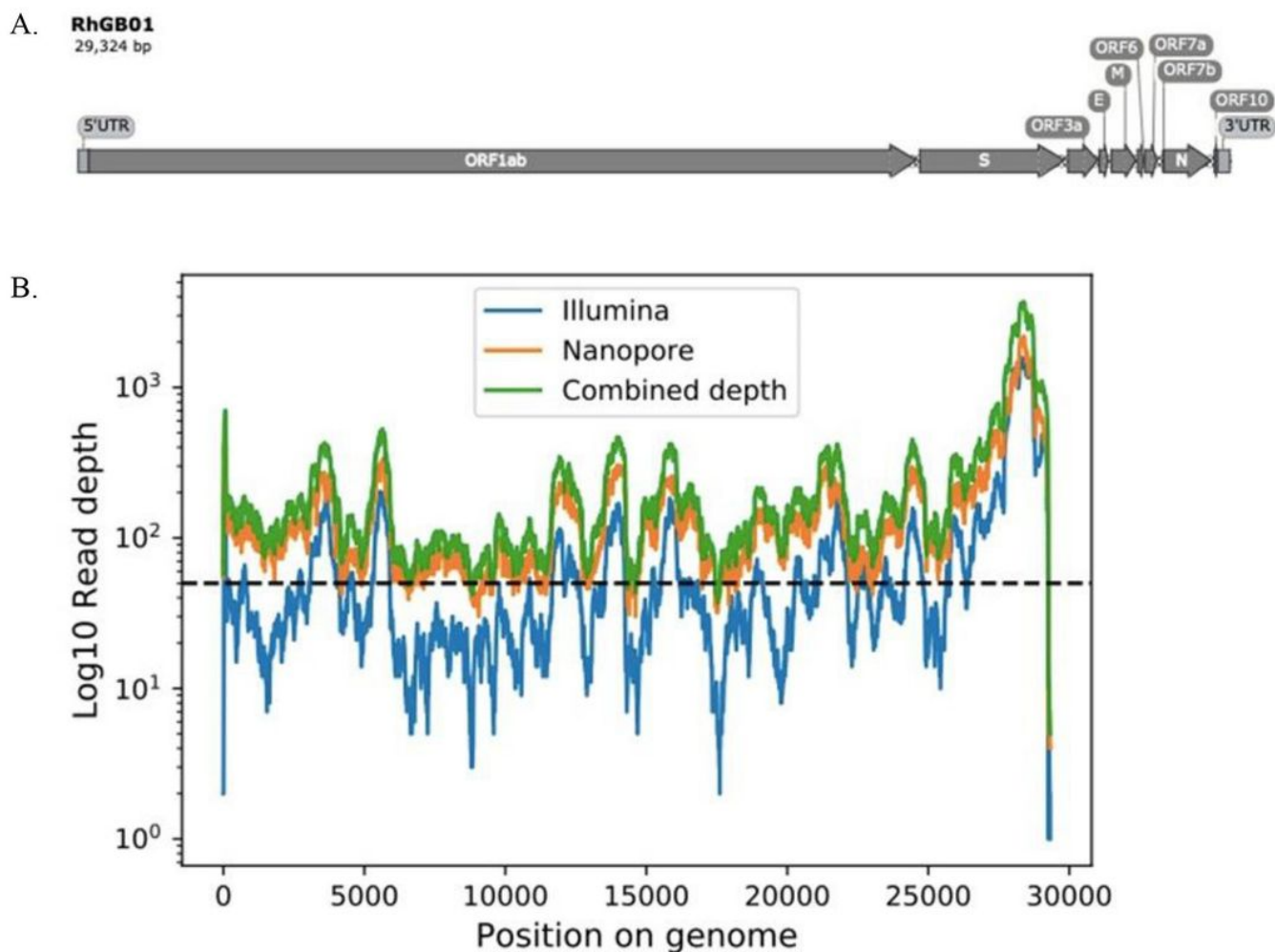
Figures

Figure 1

The genomic organisation of and the read depth across the assembled genome. A) gene organisation of RhGB01 with 10 predicted coding genes. In RhGB01, lengths of the genomic features are 5'UTR (279bp), ORF1ab (21kb), S (3.7kb), ORF3a (813bp), E (231bp), M (669bp), ORF6 (189bp), ORF7ab (465), N (1254bp), ORF10 (78bp), 3'UTR (266bp). B) Per base depth of coverage plotted across the genome from alignment of Illumina (blue), Nanopore (orange) and combined reads (green). The horizontal line represents 50x read depth.
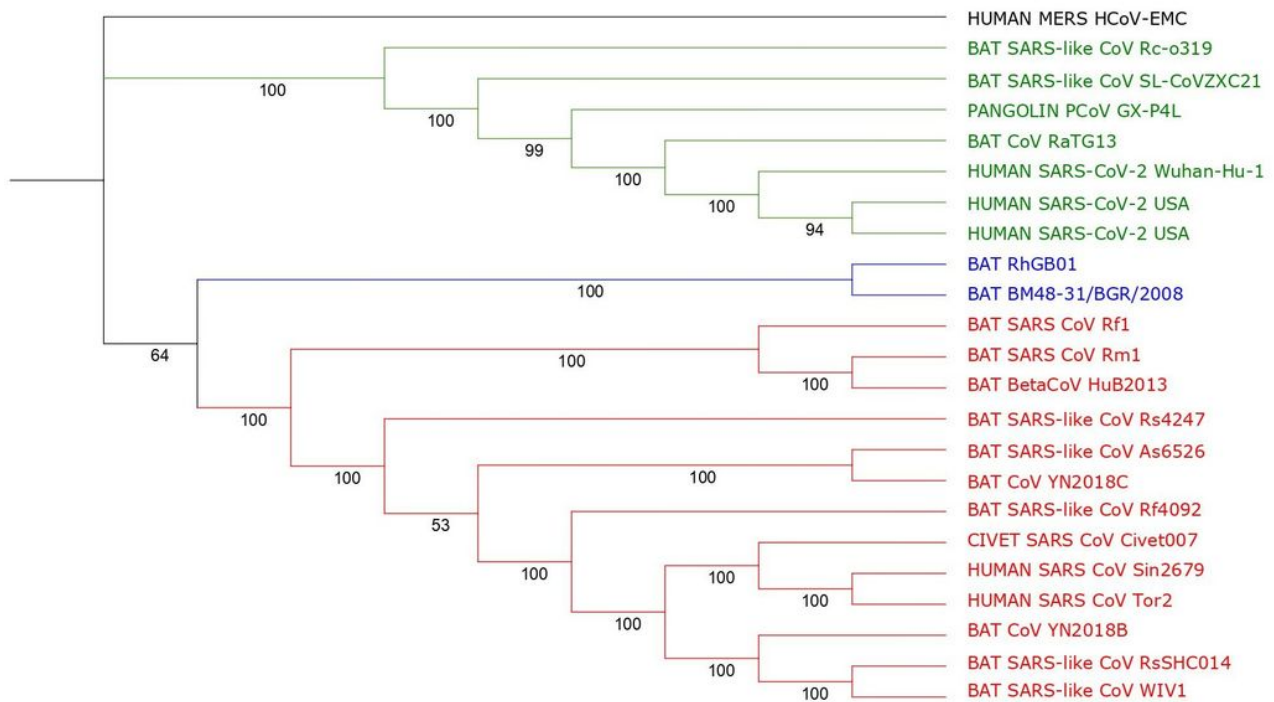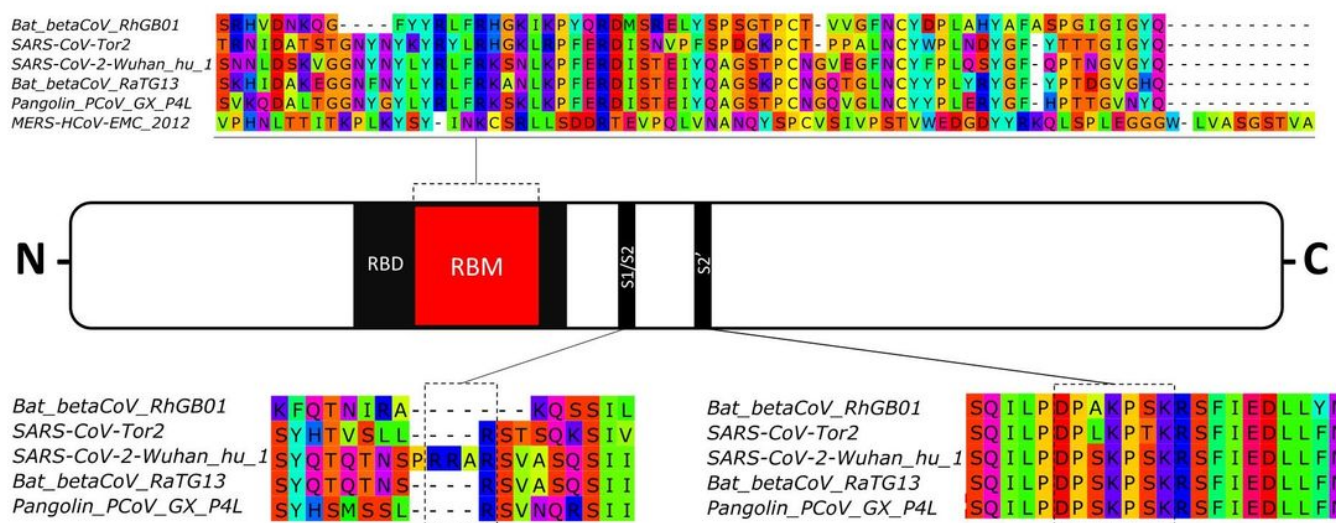
Figure 2

Maximum likelihood phylogenies of RbGB01 and other sarbecoviruses of interest. Phylogeny inferred from the S protein nucleotide sequence. MERS (a Merbecovirus) is included as outgroup. SARS or nominally SARS-like betacoronaviruses are highlighted in red, SARS‑CoV-2 or closely related viruses in green, with additional clade containing RhGB01, highlighted in blue. Bootstrap support values indicate confidence of divergence at each divergence event.

A.

B.

	RBM (%)		RBD (%)	
	SARS	SARS-CoV-2	SARS	SARS-CoV-2
RhGB01	43.48	48.57	68.21	67.35
RaTG13	50.72	75.36	75.77	89.18
PCoV	49.28	75.36	75.77	86.80

Figure 3

A schematic representation of the entire S protein with the receptor binding motif and cleavage sites highlighted and compared between RhBG01, Pangolin and Bat derived related virus, SARS and SARS-CoV-2 and percentage identity values for the RBD and RBM regions. A) Amino acid residues are coloured according to the Taylor colour scheme. Receptor binding motif comparison highlights higher amino acid conservation within the SARS-CoV-2 and SARSr-CoV-2 viruses. RhGB01 exhibits low amino acid conservation to SARS and SARS⬚CoV-2. The receptor binding motif of MERS-HCoV-EMC-2012, subgenus Merbecovirus, demonstrates little amino acid homology to receptor binding motifs from the Sarbecovirus subgenus. The furin cleavage site (S1/S2) is only present in SARS-CoV-2, distinct from the TMPRSS2 cleavage motif (S2') which is more conserved. B) A table with percentage identity scores for the RBM and RBD to SARS and SARS-CoV-2. RaTG13 and PCoV represent the most closely SARSr-CoV-2 virus from zoonotic hosts

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- CrookUKbatsarbecovirusSuppInfo.pdf