Sequence determinants of human-cell entry identified in ACE2-independent bat sarbecoviruses: A combined laboratory and computational network science approach

Ehdieh Khaledian,[a] Sinem Ulusan,[b] Jeffery Erickson,[c] Stephen Fawcett,[b] Michael C. Letko,[b,1]* and Shira L. Broschat [a,b,d,1]*

[a]School of Electrical Engineering and Computer Science, Washington State University, PO Box 640125, Pullman, WA 99164-2752, USA
[b]Paul G. Allen School for Global Health, Washington State University, PO Box 647090, Pullman, WA 99164-7090, USA
[c]School of Molecular Biosciences, Washington State University, Pullman, WA, USA
[d]Department of Veterinary Microbiology Pathology, Washington State University, Pullman, WA, USA

Summary

Background The sarbecovirus subgenus of betacoronaviruses is widely distributed throughout bats and other mammals globally and includes human pathogens, SARS-CoV and SARS-CoV-2. The most studied sarbecoviruses use the host protein, ACE2, to infect cells. Curiously, the majority of sarbecoviruses identified to date do not use ACE2 and cannot readily acquire ACE2 binding through point mutations. We previously screened a broad panel of sarbecovirus spikes for cell entry and observed bat-derived viruses that could infect human cells, independent of ACE2. Here we further investigate the sequence determinants of cell entry for ACE2-independent bat sarbecoviruses.

Methods We employed a network science-based approach to visualize sequence and entry phenotype similarities across the diversity of sarbecovirus spike protein sequences. We then verified these computational results and mapped determinants of viral entry into human cells using recombinant chimeric spike proteins within an established viral pseudotype assay.

Findings We show ACE2-independent viruses that can infect human and bat cells in culture have a similar putative receptor binding motif, which can impart human cell entry into other bat sarbecovirus spikes that cannot otherwise infect human cells. These sequence determinants of human cell entry map to a surface-exposed protrusion from the predicted bat sarbecovirus spike receptor binding domain structure.

Interpretation Our findings provide further evidence of a group of bat-derived sarbecoviruses with zoonotic potential and demonstrate the utility in applying network science to phenotypic mapping and prediction.

Funding statement This work was supported by Washington State University and the Paul G. Allen School for Global Health.

Keywords: Coronavirus; Zoonosis; Sequence similarity network; Entry assay; Pseudotype; Mathematical model

Introduction

Over the past 20 years, cross-species transmission of betacoronaviruses from animals to humans has resulted in three novel, pathogenic human viruses: SARS-CoV, MERS-CoV, and SARS-CoV-2. Virus discovery efforts have revealed a trove of closely related coronaviruses, largely in bats. Merbecoviruses (formerly lineage C betacoronaviruses) include MERS-CoV and thousands of related animal viruses, while sarbecoviruses (formerly lineage B betacoronaviruses) include SARS-CoV, SARS-CoV-2, and at least several hundred related animal viruses. Unfortunately, most of these animal-derived viruses have never been isolated and only exist as

*Corresponding authors at: Paul G. Allen School for Global Health, Washington State University, PO Box 647090, Pullman, WA 99164-7090, USA.
E-mail addresses: michael.letko@wsu.edu (M.C. Letko), shira@wsu.edu (S.L. Broschat).
[1] co-senior authors

Research in context

Evidence before this study

Cross-species transmission of coronaviruses poses a serious threat to global health. While numerous coronavirus have been discovered in wildlife, our ability to predict which pose the greatest threat to humans remains limited. In a previous study, we developed tools to study novel coronavirus entry and characterized the majority of sarbecoviruses for their ability to infect human cells.

Added value of this study

Here, we take a mathematical modelling approach based on viral sequences to better understand what distinguishes animal viruses that can infect human cells from those that cannot. Viruses that infect human cells using a pathway distinct from other common human coronaviruses varied within a region of the viral genome responsible for binding host cell receptors. Importantly, these results were verified through experimental laboratory approaches using both human and bat cells, providing further evidence for viral-host compatibility.

Implications of all the available evidence

Our findings provide additional evidence that a group of bat coronaviruses possess the ability to infect human and bat cells, independent of known coronavirus receptors. Our sequence-based modelling approach accurately distinguished this group of viruses, which may serve as a foundation for databases that integrate viral sequences with functional laboratory data.

sequences, severely hindering our understanding of their zoonotic risk.[1]

Host cell entry is the first step in the viral life cycle and a critical species barrier for coronaviruses.[2−6] Entry occurs following a physical interaction between the receptor binding domain (RBD) of the viral spike protein and a host cell receptor. We and others have previously defined distinct "clades" of sarbecovirus RBDs based on the presence or absence of conserved deletions in regions that presumably engage with the host receptor. Clade 1 sarbecovirus RBDs utilize host Angiotensin-converting enzyme 2 (ACE2) to infect cells and contain no deletions, clade 2 RBDs contain two deletions and do not bind to ACE2, and clade 3 RBDs contain one deletion and also lack ACE2 binding capacity for any species tested to date.[2,7,8]

To better understand the zoonotic risk posed by novel coronaviruses, we recently developed a laboratory platform to functionally characterize cell tropism for the sarbecoviruses.[7] In this BSL2-based, viral pseudotype approach, we replaced the RBD of the SARS-CoV spike gene with the RBD from other animal-derived sarbecoviruses, resulting in functional, chimeric spike proteins possessing the receptor tropism of the animal sarbecovirus. We generated and screened a large panel of chimeric spike proteins, representing the majority of the natural variation published for sarbecoviruses. While most animal viruses exhibited little to no tropism for human cells, we observed a subgroup of clade 2 RBD viruses that can infect human cells independent of known coronavirus receptors, including ACE2. Similar to some animal coronaviruses related to MERS-CoV, clade 2 RBD sarbecovirus cell entry was only observed when exogenous trypsin was added during the infection suggesting that protease processing may be a species barrier for these particular viruses rather than host-receptor incompatibility.[7,9]

The protein interface between both the SARS-CoV and SARS-CoV-2 spikes and ACE2 encompasses more than 15 contact points, and variation in many of these residues can severely disrupt viral entry.[10−14] Thus, it is possible that the interaction between clade 2 RBDs and "receptor X" also encompasses multiple contact points, confounding our early efforts to modify the clade 2 entry phenotype through single point mutations.

We have previously employed network science for comparative genomic approaches to explore complex phenotypes underlying transmission of tick pathogens and bacterial evolution.[15−17] Here we used a similar computational method to visualize sequence similarities between groups of viral sequences.[18] If the clade 2 sarbecoviruses are infecting human cells through a classical RBD-receptor interaction, we hypothesized we may be able to distinguish these viruses from sequences alone, even though the receptor and contact points are completely unknown. Our analysis revealed sequence similarities within a distinct sub-group of clade 2 RBDs that were able to infect human cells in our previous study.[7] We further confirmed the clade 2 subgroups with viral pseudotypes using a series of chimeric and full-length spike proteins. We found that exchanging a short, surface-exposed stretch of amino acids identified within the clade 2 RBD is sufficient to toggle human cell entry. These findings provide additional evidence that animal sarbecoviruses may be able to infect humans through routes that are distinct from known human sarbecoviruses, SARS-CoV and SARS-CoV-2, and underscore the urgency of developing universal sarbecovirus vaccines.

Methods

Network of receptor binding domains

We obtained receptor binding domain (RBD) sequences from the publicly available NCBI database using GenPept. The RBD sequences were retrieved using a Biopython[19] package to parse the GenPept file. After removing duplicate sequences, 165 unique RBD sequences remained, including RBDs found in bat coronaviruses,

SARS (SARS-like) coronaviruses, and Middle East respiratory syndrome-related coronaviruses (MERS).

To create the sequence similarity network (SSN), we first used the Enzyme Similarity Tool (EFI-EST).[20] The EFI-EST performed an all-by-all BLAST search of the 165 RBD sequences uploaded as a single FASTA file to calculate pairwise alignment scores for all sequences. The output of the EFI-EST tool is in XGMML file format. The XGMML, or eXtensible Graph Markup and Modeling Language, is a file format used to describe structured graphs and can be used directly as the input to Cytoscape.[21] Cytoscape is open-source software for integration, visualization, and analysis of biological networks. Cytoscape 3.8.2 was used to convert the information provided in the XGMML file into two tables, one with alignment information for each pair of nodes (sequences), including percent identification, alignment length, and alignment score. The second table provided node descriptions together with their amino acid sequences and was used with the first table to generate the adjacency matrix for the SSN. The adjacency matrix was fed into visone 2.18 software to create the RBD network.[22] We applied a weight threshold to remove extraneous low-weight edges in order to sparsify the network. Sparsification results in a network with fewer edges and nodes for a clearer view while maintaining meaningful structure. We applied 80% sparsification using the backbone layout of visone. This layout greatly reduces the number of weak edges while maintaining the connectedness of the network. We note that the threshold for connecting two nodes by an edge, an alignment score of 5, was deliberately chosen to capture sequence similarity between any two protein sequences. However, all edges with little similarity were removed during sparsification.

Network of receptor binding motifs

We used two different receptor binding motif (RBM) patterns to identify the amino acids in the 165 RBD sequences corresponding to their RBMs. This resulted in 144 unique RBMs. The two patterns used were: TVPHNLTTITKPLKYSYINKCSRLLSDDRTEVPQLV-NANQYSPCVSIVPSTVWE DGDYYRKQLSPLEGGG WLVASGSTVAMTEQLQ and QIAPGQTGVIADY-NYKLPDDFMGCVLAWNTRNIDATSTGNYNYKYR-YLRHGK LRPFERDISNVPFSPDGKPCTPPALNCYWP LNDY GFYTTTGIGYQPYRVVVLS

The first of these is the RBM pattern for MERS-CoV, and the second is the RBM pattern for SARS-CoV. Using the 144 RBMs, we created an SSN for RBMs following the same approach as for the RBD SSN described in the previous section.

Plasmids

Full-length Rs4081 (GenBank KY417143) and Rf1 (GenBank DQ412042) spike sequences were codon optimized, flag tagged, and engineered with silent restriction sites flanking the RBD as previously described.[7] Each spike was synthesized in two fragments of roughly equivalent length with a 15-nucleotide overlap in the middle as well as overlap with the pcDNA3.1+ cloning vector to facilitate one-step cloning with the Infusion system following the manufacturer's instructions (Takara Biosciences). *Artibeus jamaicensis* ACE2 (GenBank XM_037157556.1) was codon optimized, synthesized and cloned into pcDNA3.1+.

Pseudotype cell entry assay

293T cells (RRID CVCL_0045), BHK cells (RRID CVCL_1914), or AJ*i* cells (RRID CVCL_YZ05) were maintained in Dulbeccos Modified Eagles Medium (DMEM), supplemented with fetal bovine serum (FBS), penicillin, streptomycin, and l-glutamine. Cell line were confirmed as mycoplasma-negative and origin species was verified by cytochrome B sequencing.[7] Vesicular stomatitis virus (VSV)-based pseudotyped particles bearing spike glycoproteins were produced as described previously.[7]

Briefly, 293T were seeded in 6-well plates and transfected with plasmids encoding a FLAG-tagged sarbecovirus spike protein. Twenty-four hours post-transfection, cells were infected with VSV-glycoprotein-pseudotyped VSV particles for one hour at 37°C, washed three times with culture media, and incubated for 48 h in DMEM supplemented with 2% FBS. Supernatants were collected, briefly clarified by centrifugation, aliquoted, and stored at −80°C.

Sarbecovirus spike pseudotype entry assays were performed as described previously.[7] Particles were incubated with trypsin diluted in HBSS and phenol red or HBSS with phenol red, only, as a negative control, at 37 °C for 10 min. BHK cells were transfected with ACE2 expression plasmids, 24 h before infection, as previously described.[7] Target cells were infected with trypsin-treated or untreated particles at 1200 RFC for 1 h at 4° and then incubated for approximately 24 h at 37 °C. Luciferase was measured with the Bright-Glo reagent following the manufacturer's instructions (Promega Corporation).

Statistics

The pseudotype entry assay was analyzed in Microsoft Excel, and graphing and statistics were performed in PRISM Graph Pad (version 9). For experiments where protease was or was not added, ordinary two-way ANOVA was performed, comparing all values from both conditions, and Tukey's multiple comparisons test was applied.

Sequence analysis and structural modeling

Web-logo plots were generated from clade 2 sequences using WebLogo (Berkeley).[23,24] The structure for SARS-CoV/Urbani RBD bound to ACE2 (PDB: 2AJF) was

used to predict the structure of Rs4081 using SwissMo-del.[25] Structures were visualized using PyMol.

Ethics statement

No human samples or animals were used in this study.

Role of funding source

Results

Network analysis of betacoronavirus RBD and RBM sequences

To broadly visualize similarities between coronavirus spike sequences, we employed sequence similarity networks (SSN), which display the interrelationships between proteins using pairwise alignments between sequences, similar to phylogenetic trees.[18] Because the entry phenotype we observed in our previous study was with chimeric spike proteins containing the RBD from animal-derived sarbecoviruses, we first generated an SSN using the RBD region corresponding to SARS-CoV amino acids 323-501 (Figure 1a). Each network node corresponds to one viral RBD. In this network, SARS-CoV clustered with related, ACE2-utilizing bat sarbecoviruses, WIV1 and SHC014, while SARS-CoV-2 formed a separate cluster along with pangolin-derived relatives and the recently identified Japanese ACE2-utilizing sarbecovirus Rc0319.[26] Although this RBD-based SSN depicted two clusters for merbecoviruses (blue) and sarbecoviruses (red), several animal-derived viruses from both betacoronavirus lineages formed a separate, bridging cluster (fuchsia), and we did not observe any subclustering by receptor usage (Figure 1a).

Within the MERS-CoV, SARS-CoV and SARS-CoV-2 RBDs, residues that interact with the host receptor are located within a short, continuous, and surface-exposed region known as the receptor binding motif (RBM; Supplementary Figure. 1a). Because the sarbecovirus RBDs are fairly conserved in their N-terminal region but have highly diverse RBMs, we wondered if localized disparity in sequence similarities was confounding our SSN analysis. To assess the relationships using the RBM, we generated a second SSN using just the short stretch of amino acids corresponding to the SARS-CoV RBM (Figure 1b). The resulting network is divided into two separate groups: merbecoviruses (Figure 1b; cluster on the lower left) and sarbecoviruses (Figure 1b; upper right). Within the sarbecoviruses, ACE2-utilizing viruses (clade 1 RBD) formed central clusters (green and fuchsia), with distantly-related bat sarbecoviruses

forming distinct branching clusters to the right and left of clade 1, respectively (Figure 1b). Networks generated using full genomes showed clustering by phylogenetically-accepted genera and sub-genera, but we did not observe similar groups within the sarbecoviruses (Supplementary Figure. 1b,c). Interestingly, in the RBM SSN, the clade 2 bat coronaviruses were further divided into two sub-clusters (Figure 1b; upper right).

Host cell entry phenotypes form clusters in RBM similarity network

For better visualization, we collapsed the repetitive and human patient-derived viral sequences across the nodes within the sarbecovirus SSN and colored the nodes by RBD clades (Figure 1a vs Figure 2a). Within the clade 2 RBDs, the viruses whose RBDs can infect human cells formed a cluster (cluster 2a; in grey) which is distinct from other clade 2 viruses that do not infect human cells (cluster 2b; in purple).

To test the clade 2 subclusters, we generated VSV (Vesicular Stomatitis Virus) reporter pseudotypes bearing chimeric SARS-CoV spike proteins with the RBD of each of the clade 2 viruses shown in the network and subsequently infected 293T cells in the presence or absence of trypsin as previously described[7] (Figure 2b). None of the sarbecovirus pseudotypes exhibited strong entry into 293T cells in the absence of external protease, which was expected as these cells express only very low levels of ACE2, and no cells to date have been permissive for clade 2b RBDs without trypsin (Figure 2b, top panel).[7,27] However, as previously reported, addition of trypsin activated the wildtype SARS-CoV spike as well as chimeric sarbecovirus spikes bearing clade 2 RBDs from the "compatible" 2a cluster assigned in our RBM SSN[7] (Figure 2b). Clade 2b sarbecoviruses formed a cluster distinct from the "compatible" 2a cluster of our network and did not exhibit cell entry following protease treatment. These data suggest that observed ACE2-independent entry phenotypes in our viral entry assay can be mapped to sequence determinants within the putative RBM.

Clade 2b sarbecoviruses infect bat cells independent of bat ACE2

We and others have previously shown that clade 2 viruses are not capable of using ACE2 from human, mouse, civet, pangolin, or from the natural hosts, *Rhinolophus -affinis* or *-sinicus* bats.[7,28,29] To further assess the ACE2-independence of clade 2 viruses, we tested our panel with an *Artibeus jamaicensis* cell line (AJi) that is selectively permissive to clade 2 entry following trypsin treatment, but not to the ACE2-dependent clade 1 viruses.[7] We included the SARS-CoV-2 RBD in these assays, which has been recently suggested to use alternative, ACE2-independent pathways[30] (Figure 2c). As reported before, ACE2-dependent viruses were not able
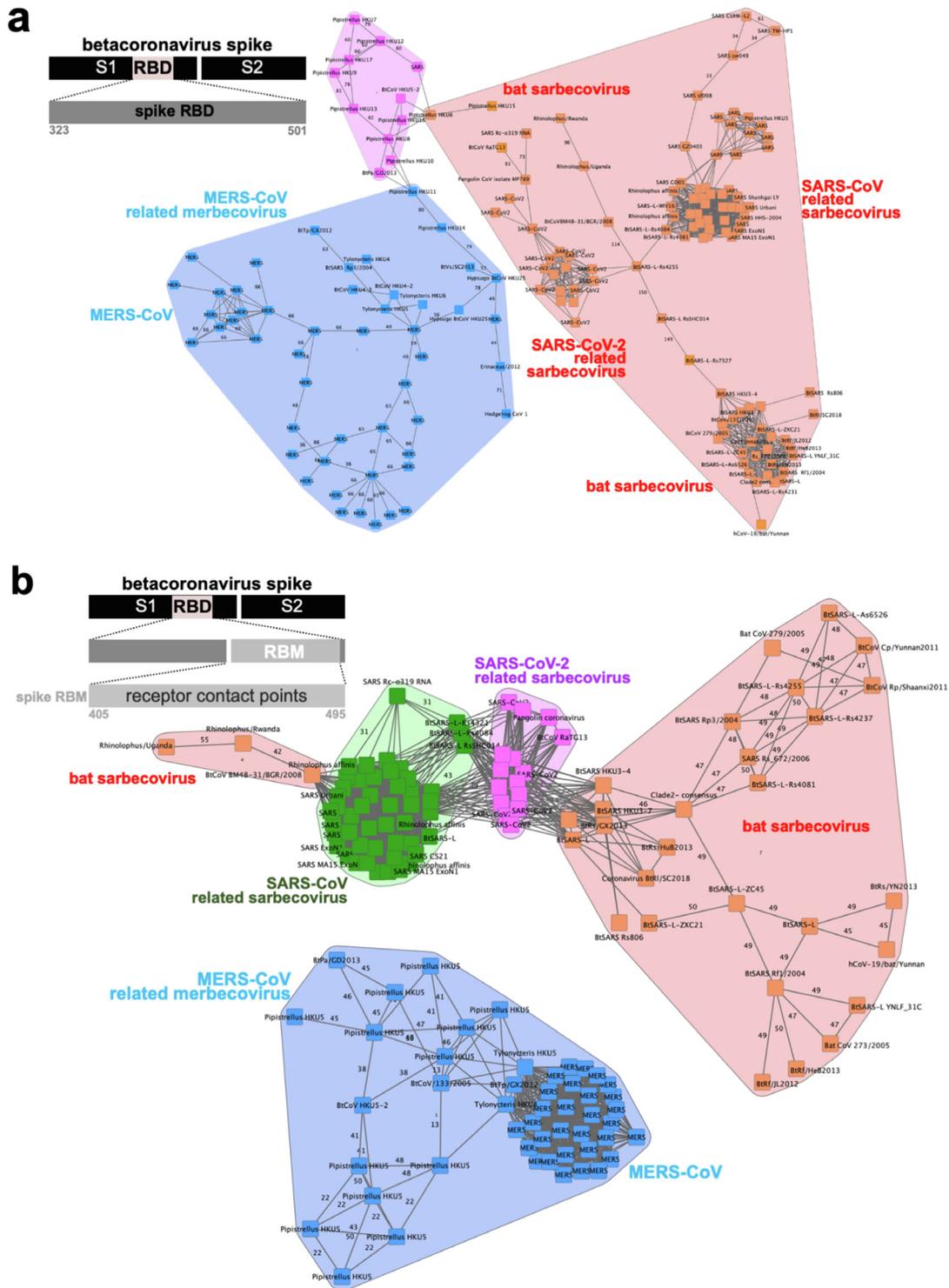
Figure. 1. Sequence similarity network (SSN) for sarbecovirus and merbecovirus spike domains. (**a**) Amino acid SSN of the receptor binding domain (RBD), encompassing spike residues 323–501, from merbeco- (blue) and sarbeco- (red) viruses. Betacoronaviruses that belong to both groups are indicated in fuchsia. (**b**) Amino acid SSN of the receptor binding motif (RBM), encompassing spike residues 405–481, from merbeco- (blue) and sarbeco-viruses. Sarbecoviruses are further divided into SARS-CoV viruses in green, SARS-CoV-2 in fuchsia, and bat sarbecoviruses in red. Numbering in the spike diagrams corresponds to the SARS-CoV/Urbani amino acid sequence.
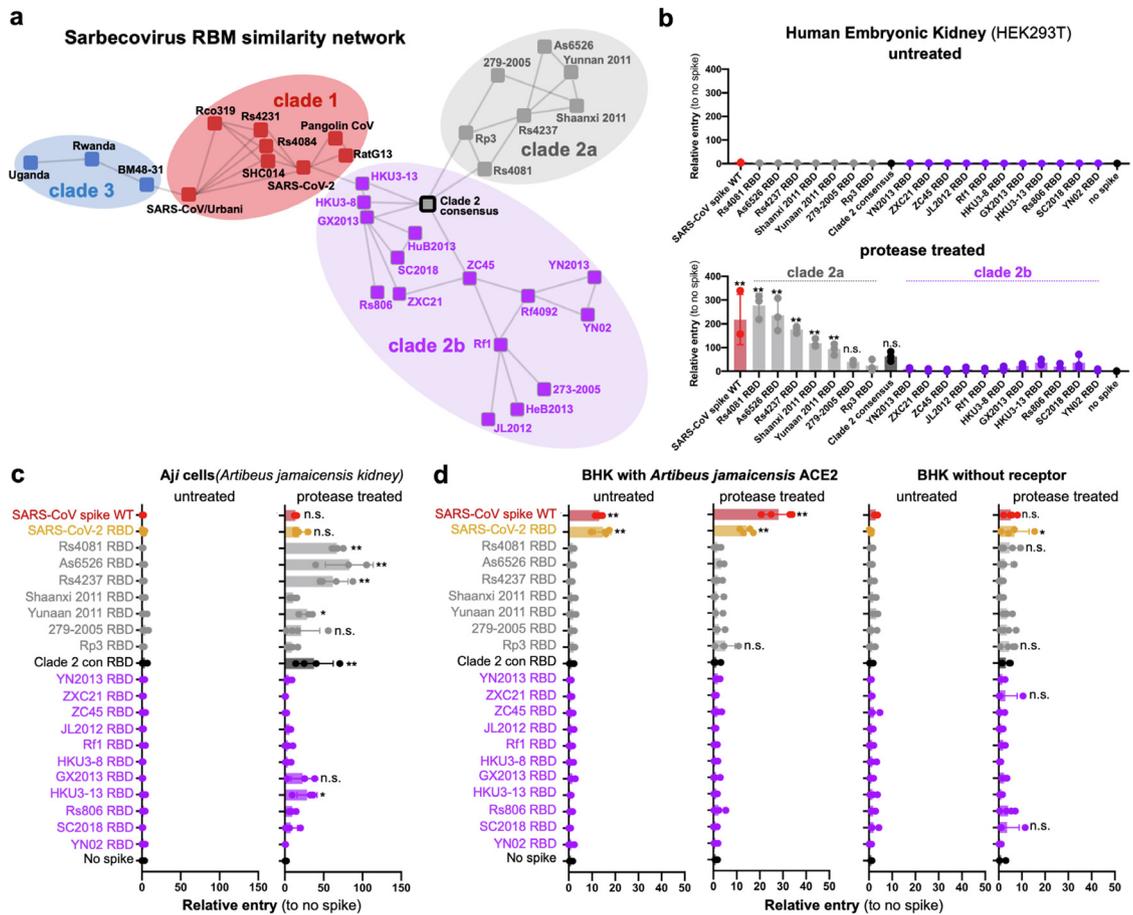
Figure. 2. Validation of phenotypes from the receptor binding motif (RBM) sequence similarity network (SSN). (**a**) Simplified RBM SSN with sarbecovirus receptor binding domain (RBD) clades 1, 2, and 3 indicated in red, gray/purple, and blue, respectively. Sequences from human outbreaks of SARS-CoV and SARS-CoV-2 have each been condensed into single points. Viruses that were functionally tested for cell entry are shown as indicated. (**b**−**d**) Luciferase-reporter particles pseudotyped with SARS-CoV chimeric spike protein containing the RBD from the indicated sarbecoviruses were used to infect 293T cells (**b**), *Artibeus jamaicensis* kidney cells (**c**), or BHK cells transfected with *Artbieus jamaicensis* ACE2 (**d**), in the presence or absence of trypsin. Shown are experimental triplicates (b) or quadruplicates (c, d) from one experiment representative of at least three biological replicates. Asterisks indicate statistical significance from no spike control (**=p-value < 0.001; *=p-value < 0.05; n.s. = not significant; values are not significant unless otherwise indicated).

to infect the AJi cells, while clade 2 viruses could infect following trypsin treatment (Figure 2c). We next tested if these viruses could use *A. jamaicensis* ACE2 in baby hamster kidney (BHK) cells that are non-permissive to sarbecoviruses (Figure 2d). While both SARS-CoV and SARS-CoV-2 could use *A. jamaicensis* ACE2, none of the clade 2 RBDs could infect cells expressing this receptor. Taken together, these data show that some clade 2 viruses are not capable of using *A. jamaicensis* ACE2 but can robustly infect *A. jamaicensis* cells.

Chimeric and full-length sarbecovirus spike proteins exhibit similar entry

While we have previously observed ACE2-independent entry with both chimeric and full-length spike proteins,

most of the sarbecovirus spikes have not been experimentally characterized as full-length proteins. In our earlier study, we only assessed the full-length spike of human-compatible, clade 2 virus As6526. To expand on these previous findings, we synthesized full-length, FLAG-tagged spike genes from the clade 2 viruses, Rs4081 and Rf1, as representative examples of human compatible 2a and incompatible 2b viruses, respectively (Figures. 2 and 3). We generated luciferase-GFP dual reporter pseudotypes bearing either a chimeric SARS-CoV spike with clade 2 RBDs or full-length clade 1, 2, or 3 spikes and infected human embryonic kidney 293T cells (Figure 3). In agreement with our previous study, the full-length and chimeric As6526 spikes exhibited moderate human cell entry.[7] As observed for the chimeric spike proteins, the full-length Rs4081 spike
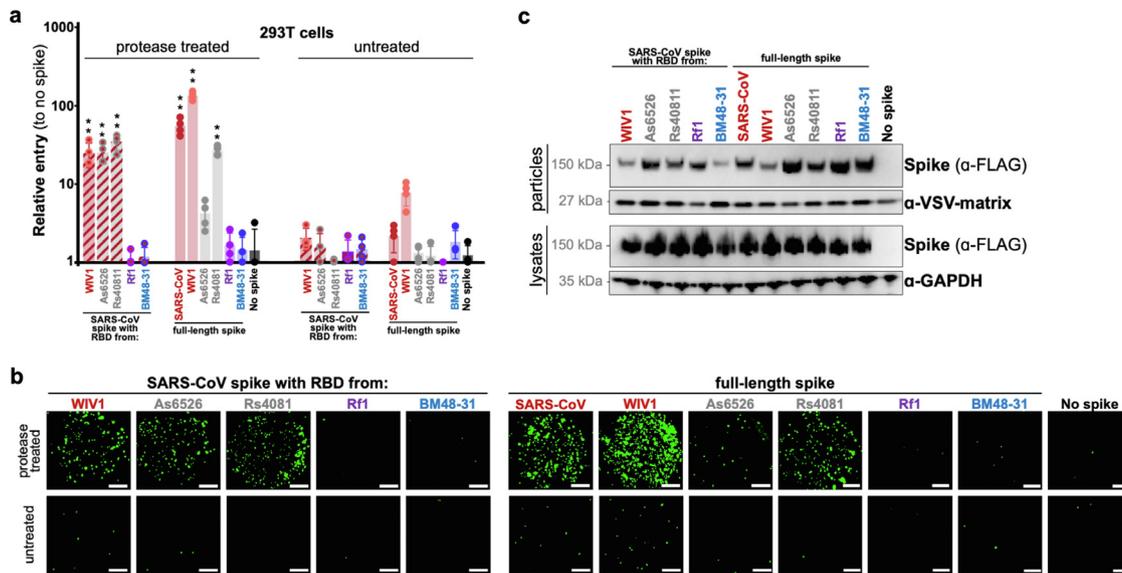
Figure. 3. Full-length spikes are functionally similar to chimeric spikes. (**a**) 293T cells were infected with pseudotypes bearing either the chimeric (hashed bars and semi-filled circles) or full-length spike (solid bars and circles) in the presence or absence of trypsin, and luciferase was measured the next day. Shown are experimental triplicates from one experiment representative of at least three biological replicates. (**=*p*-value < 0.001; *=*p*-value < 0.05; bars without asterisks indicate no statistical significance from no spike control). (**b**) GFP microscopy of cells infected with GFP-luciferase dual-reporter pseudotypes. Scale bar indicated 500um. (**C**) Expression and incorporation of the spike protein.

exhibited notably stronger entry into human cells following protease treatment, while the Rf1 spike was completely incapable of entry (Figure 3a,b). Although pseudotype incorporation of some of the spike proteins seemed to vary slightly, these differences did not track with the observed entry phenotypes (Figure 3c). For example, the WIV1 spike appeared to incorporate into pseudotypes less efficiently than the SARS-CoV spike but infected 293T cells with slightly better efficiency. These findings show that chimeric spikes recapitulate full-length spike phenotypes (Figure 3).

A putative receptor binding motif region in the spike protein mediates ACE2-independent, bat sarbecovirus cell entry

Our network analysis suggested that only the clade 2 RBM (spike residues 405−481) were associated with the human cell entry phenotype (Figures. 1 and 4a). To test whether this region can modulate the entry phenotype, we generated pseudotyped reporter particles with chimeric spike RBDs in which just the RBM was exchanged between Rs4081 (human-compatible) and Rf1 (Figure 4a−c). All chimeric spikes expressed and incorporated into viral particles similarly (Figure 4b). Pseudotypes bearing the wild-type Rs4081 RBD or the Rs4081 RBM were capable of infecting human cells in the presence of trypsin. In contrast, and in support of our previous results, spikes with Rf1 RBDs or RBMs were incapable of infecting human cells under any conditions (Figure 4c).

Web-logo analysis of consensus sequences for human-compatible and incompatible clade 2 RBMs revealed numerous amino acid variations between the groups (Figure 4d), with 12 differences between Rs4081 and Rf1 (Figure 4e). Mapping these variations to the predicted structure of the Rs4081 RBD showed that these residues are surface-exposed and cluster together analogous to the ACE2 contact residues on the SARS-CoV spikes (Figure 4f). These findings demonstrate that a subgroup of sarbecovirus clade 2 RBDs are compatible with human cells and that, while distinct from SARS-1 and -2, this human compatibility likely exhibits some similarities to other coronavirus spike entry mechanisms (Figure 4).

Discussion

The NCBI (National Center for Biotechnology Information) GenBank nucleotide database currently lists over 1.6 million entries for "sarbecovirus." With the recent emergence of SARS-CoV-2, there has been an increase in global viral discovery efforts to identify additional viral threats and new progenitor sources for the sarbecoviruses — expanding the public sequence repositories even further. We have devised a scalable laboratory platform to study the zoonotic potential of the sarbecoviruses, but even with this approach it is not practical to continue testing each newly discovered virus. Thus, there is a need to synthesize trends and learn more broadly from our expanding collection of functional
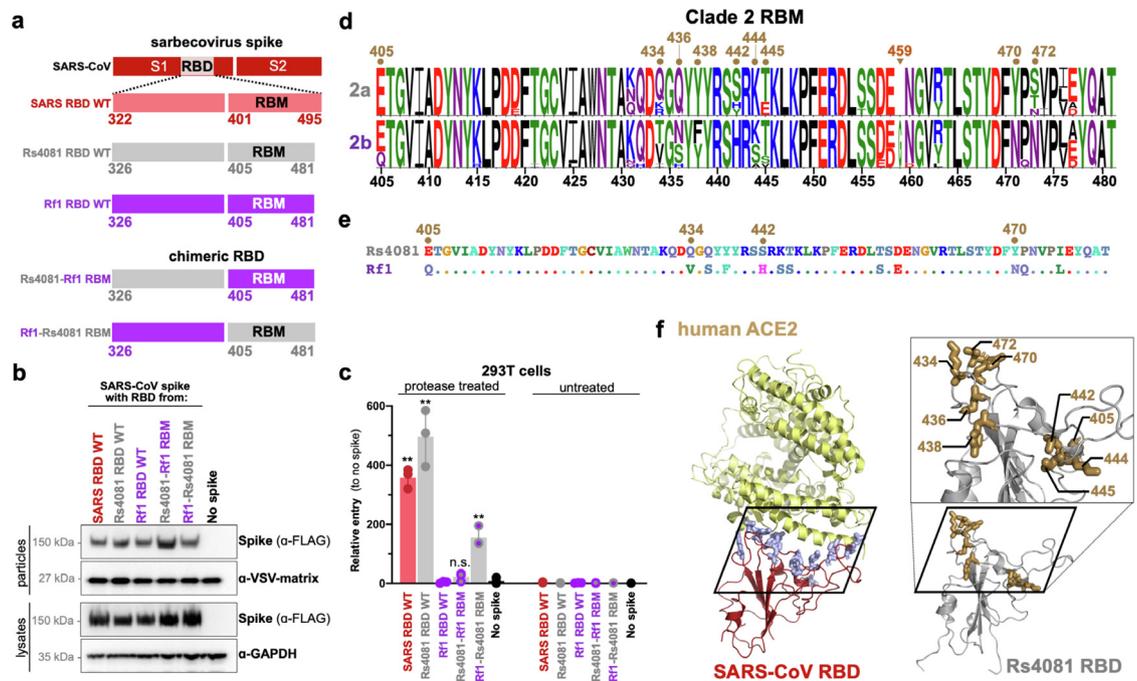
Figure. 4. Identification of a putative ACE2-independent sarbecovirus spike receptor binding motif (RBM). (**a**) RBMs were exchanged between clade 2 RBDs and inserted into the SARS-CoV spike protein backbone. (**b**) Chimeric sarbecovirus spike expression and pseudotype incorporation by western blot. (**c**) VSV-luciferase pseudotype reporter particles were used to infect 293T cells in the presence or absence of trypsin. Solid-colored circles indicate a full-length spike RBD in the SARS-CoV spike background, while mixed-color circles indicate a chimeric spike RBD in the SARS-CoV spike background. Shown are experimental triplicates from one experiment representative of at least three biological replicates. (**=*p*-value < 0.001; *=*p*-value < 0.05; bars without asterisks indicate no statistical significance from no spike control). (**d**) Web-logo plot comparison of sarbecovirus RBD clade 2 amino acid sequences, with varying residues highlighted. (**e**) Amino acid comparison between Rs4081 and Rf1 RBD. (**f**) SARS-CoV spike RBD (red) complexed with human ACE2 (yellow) (PDB: 2AJF) and a representative clade 2 RBD (grey; predicted, SwissModel). SARS-CoV spike residues that contact ACE2 are indicated in blue, and residues that vary between RBD clade 2 are indicated in khaki.

cell-entry data for the sarbecoviruses.[7,31] Here, we used sequence similarity networks that clustered viral sequences by entry phenotype (Figs. 1 and 2). This approach allowed us to distinguish viral spike sequences that exhibit human-cell compatibility from those that cannot infect human cells, as determined by our functional laboratory assays (Figure 2). Our network maps provided a binary, "compatible" or "incompatible" differentiation between clade 2 RBDs, but we did not observe any further grading regarding "strength" of entry. For example, human-compatible Rs4081 and As6526 viruses are on opposite ends of the Clade 2a subcluster despite having RBDs with similar levels of entry (Figure 2). Additional point mutagenesis studies are needed to determine residues within the putative clade 2 RBM that contribute to the entry phenotype. Such information could be taken into consideration to build weighted network maps that allow visualizing the "strength" of entry. As we learn more about sarbecovirus entry mechanisms and the sequence determinants underlying them, functionally-annotated network maps may be used to help predict viral tropism for novel, untested viral sequences.

Sequence similarity networks most accurately clustered viruses by their entry phenotype when only a region corresponding to the viral RBM was used and not more-inclusive viral regions (Figure 1 and Supplementary Figure. 1). The RBM region used for these analyses was determined solely from ACE2-dependent clade 1 viruses and extrapolated for all sarbecoviruses. Because this region also differentiated clade 2 viruses by their compatibility with human cells (Figure 2), our initial network analysis further underscores the importance of this region of amino acids for sarbecovirus entry, in general. We have previously shown that clade 1 sarbecovirus RBMs can be substituted into clade 2 RBDs to impart ACE2-dependent entry.[7] In this current study, we show that ACE2-independent entry can also be toggled using similar chimeric RBD methods (Figure 4). The results from our network analysis and these chimeric RBD experiments suggest that sarbecovirus spikes commonly use a similar region in the spike to interact with their receptors.

While most vaccine design is currently centered around preventing infection by ACE2-dependent sarbecoviruses, a few "universal" sarbecovirus vaccines are

currently in development with the goal of providing broader protection from animal-derived sarbecoviruses. Unfortunately, these vaccines are using sequences from HKU3, SC2018, and ZC45,[32,33] which we have shown both previously and in the present work, do not possess human-cell compatibility, at least not at the same levels as other clade 2 viruses (Figs. 2 and 3).[7] Thus, future universal sarbecovirus vaccines should be designed for sarbecoviruses that pose more immediate risk of zoonotic spillover.

Our previous work with the full-length As6526 spike showed ACE2-independent human-cell compatibility but notably less than the chimeric SARS-CoV spike with the As6526 RBD.[7] We have now shown the Rs4081 spike is capable of entering human kidney (293T; Figs. 2, and 3), human liver (Huh7.5), African green monkey kidney (VeroE6), and the AJ*i* kidney cell line (AJ*i*),[7] while ACE2 from these species clearly does not support clade 2 infection when provided in ACE2-defficient cells like BHKs (Figure 2c,d).[7] Studies by other groups have further demonstrated that clade 2 RBDs are not capable of binding ACE2 from humans, pangolins, civets, and the Rhinolophus bat species that carry these viruses[29,34] [preprint]. Taken together with our previous evidence showing that individual point mutations cannot impart ACE2 binding to clade 2 viruses, it is apparent that at least some clade 2 sarbecoviruses are capable of employing an ACE2-independent cell entry route that is conserved across a range of mammalian species.

Analogous to our findings with ACE2-independent sarbecoviruses, exogenous protease has also been shown to facilitate human cell entry of several bat-derived merbecoviruses, independent of the receptor used by MERS-CoV, dipeptidyl peptidase IV (DPP4).[9] Together, these studies demonstrate additional coronaviruses with zoonotic potential. Because the mechanism of host cell entry dictates viral transmission routes and potential pathogenesis, more research is urgently needed to uncover how these pre-emergent coronaviruses are infecting human cells.

Declaration of interests
The authors have nothing to declare.

Contributors
M.L. and S.B. designed and supervised the study. M.L., S.U., J.E., S.F., and E.K. performed experiments. M.L. verified collected data. E.K., S.B., and M.L. analyzed data. E.K., S.B., and M.L. wrote the paper. All authors contributed to, read, and approved the final manuscript.

Acknowledgements

Data sharing statement
Data from this study are available on request from the corresponding authors.

Supplementary materials
Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ebiom.2022.103990.

References
1 Letko M, Seifert SN, Olival KJ, Plowright RK, Munster VJ. Bat-borne virus diversity, spillover and emergence. *Nat Rev Microbiol.* 2020;18(8):461–471.
2 Becker MM, Graham RL, Donaldson EF, et al. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc Natl Acad Sci U S A.* 2008;105(50):19944–19949.
3 Li F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol.* 2016;3(1):237–261.
4 Ohnuma K, Haagmans BL, Hatano R, et al. Inhibition of middle east respiratory syndrome coronavirus infection by anti-CD26 monoclonal antibody. *J Virol.* 2013;87(24):13892–13899.
5 van Doremalen N, Miazgowicz KL, Milne-Price S, et al. Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. *J Virol.* 2014;88 (16):9220–9232.
6 Yang Y, Du L, Liu C, et al. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. *Proc Natl Acad Sci U S A.* 2014;111(34):12516–12521.
7 Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nat Microbiol.* 2020;5(4):562–569.
8 Hoffmann M, Muller MA, Drexler JF, et al. Differential sensitivity of bat cells to infection by enveloped RNA viruses: coronaviruses, paramyxoviruses, filoviruses, and influenza viruses. *PLoS One.* 2013;8(8):e72942.
9 Menachery VD, Dinnon KH, Yount BL, et al. Trypsin treatment unlocks barrier for zoonotic bat coronaviruses infection. *J Virol.* 2019;94(5):e01774-19.
10 Han DP, Penn-Nicholson A, Cho MW. Identification of critical determinants on ACE2 for SARS-CoV entry and development of a potent entry inhibitor. *Virology.* 2006;350(1):15–25.
11 Lan J, Ge J, Yu J, et al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature.* 2020;581 (7807):215–220.
12 Li F, Li W, Farzan M, Harrison SC. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science.* 2005;309(5742):1864–1868.
13 Li W, Zhang C, Sui J, et al. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 2005;24 (8):1634–1643.
14 Starr TN, Greaney AJ, Hilton SK, et al. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell.* 2020;182(5):1295-310. e20.
15 Khaledian E, Brayton KA, Broschat SL. A systematic approach to bacterial phylogeny using order level sampling and identification of HGT using network. *Sci Microorg.* 2020;8(2).
16 Khaledian E, Gebremedhin AH, Brayton KA, Broschat SL. A network science approach for determining the ancestral phylum of bacteria. In: *Proceedings of the ACM International Conference on Bio-informatics, Computational Biology, and Health Informatics.* 2018398–403.
17 Lockwood S, Brayton KA, Broschat SL. Comparative genomics reveals multiple pathways to mutualism for tick-borne pathogens. *BMC Genom.* 2016;17:481.

18 Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One*. 2009;4(2):e4345.

19 Cock PJ, Antao T, Chang JT, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*. 2009;25(11):1422–1423.

20 Zallot R, Oberg N, Gerlt JA. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. *Biochemistry*. 2019;58(41):4169–4182.

21 Saito R, Smoot ME, Ono K, et al. A travel guide to cytoscape plugins. *Nat Methods*. 2012;9(11):1069–1076.

22 Benkert M, Brandes U, Cornelsen S, et al. Visone software for visual social network analysis. *International Symposium on Graph Drawing*. Springer; 2001:463–464.

23 Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*. 1990;18(20):6097–6100.

24 Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14(6):1188–1190.

25 Waterhouse A, Bertoni M, Bienert S, et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018;46(W1):W296–W303.

26 Murakami S, Kitamura T, Suzuki J, et al. Detection and characterization of bat sarbecovirus phylogenetically related to SARS-CoV-2, Japan. *Emerg Infect Dis*. 2020;26(12):3025–3029.

27 Shulla A, Heald-Sargent T, Subramanya G, Zhao J, Perlman S, Gallagher T. A transmembrane serine protease is linked to the severe acute respiratory syndrome coronavirus receptor and activates virus entry. *J Virol*. 2011;85(2):873–882.

28 Hu B, Zeng LP, Yang XL, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLoS Pathog*. 2017;13:(11) e1006698.

29 Starr TN, Zepeda SK, Walls AC, et al. ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature*. 2022;603(7903):913–918.

30 Jackson CB, Farzan M, Chen B, Choe H. Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol*. 2022;23(1):3–20.

31 Wells HL, Letko M, Ssebide B, et al. The evolutionary history of ACE2 usage within the subgenus Sarbecovirus. *Virus Evol*. 2021;7 (1):veab007.

32 Martinez DR, Schafer A, Leist SR, et al. Chimeric spike mRNA vaccines protect against Sarbecovirus challenge in mice. *Science*. 2021;373(6558):991–998.

33 Tortorici MA, Czudnochowski N, Starr TN, et al. Broad sarbecovirus neutralization by a human monoclonal antibody. *Nature*. 2021;597(7874):103–108.

34 Roelle SM, Shukla N, Pham AT. Expanded ACE2 dependencies of diverse SARS-like coronavirus receptor binding domains. *bioRxiv*. 2021.