ACE2 from *Pipistrellus abramus* bats is a receptor for HKU5 coronaviruses

Nicholas J. Catanzaro[2*], Ziyan Wu[4*], Chengcheng Fan[4], Alexandra Schäfer[2], Boyd L. Yount[2], Pamela J. Bjorkman[4], Ralph Baric[2,3], Michael Letko[1#]

1. Paul G. Allen School for Global Health, Washington State University, Pullman, WA, 99163.
2. Department of Epidemiology, Gillings School of Global Public Heath, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599.
3. Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599
4. Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, 91125

*co-first authors

[#]Corresponding author:

michael.letko@wsu.edu

SUMMARY:

The merbecovirus subgenus of coronaviruses includes Middle East Respiratory Syndrome Coronavirus (MERS-CoV), a zoonotic pathogen transmitted from dromedary camels to humans that causes severe respiratory disease. Viral discovery efforts have uncovered hundreds of merbecoviruses in different species across multiple continents, but few have been studied under laboratory conditions, leaving basic questions regarding their human threat potential unresolved. Viral entry into host cells is a critical step for transmission between hosts. Here, a scalable approach that assesses novel merbecovirus cell entry was developed and used to evaluate receptor use across the entire merbecovirus subgenus. Merbecoviruses are sorted into clades based on the receptor-binding domain of the spike glycoprotein. Receptor tropism is clade-specific, with the clade including MERS-CoV using DPP4 and multiple clades using ACE2, including HKU5 bat coronaviruses. Mutational analysis identified possible structural limitations to HKU5 adaptability and a cryo-EM structure of the HKU5-20s spike trimer revealed only ‘down’ RBDs.

53 **INTRODUCTION:**

54 Middle East Respiratory Syndrome Coronavirus (MERS-CoV) was first discovered in the middle
55 east in 2012 as a cause of severe respiratory illness in humans with high mortality.[1] Field studies
56 have shown that MERS-CoV is endemic in dromedary camels in the region and routinely transmits
57 to humans, sustaining a slow but continual regional outbreak.[2,3] MERS-CoV is the prototypic
58 member of the merbecovirus subgenus of beta-coronaviruses. Virus discovery efforts have
59 uncovered hundreds of merbecoviruses circulating among diverse wildlife across multiple
60 continents.[4-13] As is true with most of the global virome, the majority of all sequenced
61 coronaviruses have never been isolated or molecularly characterized and therefore their threat of
62 spillover into humans remains unknown.

63 Coronaviruses must enter the cells and replicate to enable efficient transmission between
64 species. MERS-CoV cell entry is mediated by the viral spike protein through several viral-host
65 interactions: (1) the N-terminal domain of spike attaches virus to lectins present on the host cell
66 surface, (2) the receptor binding domain (RBD) binds to the host cell receptor molecule, (3) and
67 receptor binding reorganizes structural elements in the S2 subunit of the spike protein that
68 mediate membrane fusion and subsequent cell entry.[14-16] Each function is performed by a
69 separate region of the spike protein, the RBD being responsible for the core interaction with the
70 host receptor.[17-20] While MERS-CoV uses dipeptidyl peptidase IV (DPP4) as its host receptor to
71 infect cells, several studies have shown that some bat merbecoviruses do not use DPP4 and
72 instead use angiotensin-converting enzyme 2 (ACE2) to infect their hosts.[12,21-24]

73 Similar to other beta-coronaviruses, the RBD in merbecovirus spike proteins is a single,
74 contiguous domain located in the N-terminal half of the spike sequence, and it folds as a
75 structurally stable unit that engages with the host receptor (Figure 1A-B).To reduce cost of viral
76 gene synthesis while maintaining the ability to study viral protein function, we previously employed
77 a method to replace the RBD of a sarbecovirus, SARS-CoV, with RBDs from other
78 sarbecoviruses.[25] The resulting chimeric spike gene was used in viral pseudotype assays to

79 measure entry efficiency in cell culture of a panel of sarbecovirus RBDs.[25] Using this approach
80 (called "SarbecoType"), we previously characterized the cell entry for the majority of all published
81 sarbecovirus RBDs, which led to isolation of new sarbecoviruses, identified understudied cell
82 entry pathways, and helped elucidate the molecular determinants of sarbecovirus cell entry.[25-29]
83 We have also produced chimeric sarbecovirus molecular clones that are capable of replicating
84 with a different receptor tropism when given either a different spike gene or RBD from other
85 sarbecoviruses, further demonstrating the feasibility of exchanging the RBD within this
86 subgenus.[30]
87 Here, we took a similar approach, called "MerbecoType," with the merbecovirus subgenus
88 of betacoronaviruses. Similar to sarbecoviruses, RBDs of merbecoviruses can be classified into
89 four different "clades" based on the presence or absence of indels, which exhibit clade-specific
90 trends in receptor use and cell entry. Surprisingly, at least half of the merbecovirus RBD clades
91 can use ACE2 for cell entry, with the entire group of HKU5 viruses exhibiting clear preference for
92 ACE2 from their natural host species: *Pipistrellus abramus* bats and subsequently validated using
93 live virus assays. We also determined the structure for HKU5 spike trimer and discuss similarities
94 with other bat coronavirus spikes. These findings are important for ongoing efforts to identify
95 looming coronavirus threats and in the development of interventions such as broad-spectrum
96 antivirals and universal coronavirus vaccines.
97
98
99
100
101
102
103
104

105 **RESULTS:**

106 ***Mapping the functional receptor binding domain***

107 In order to test a large panel of RBDs from diverse merbecoviruses in viral pseudotype entry
108 assays, MERS-CoV spike expression constructs were engineered to allow for easy replacement
109 of the RBD with an RBD from other merbecoviruses (Figure 1A-C). MERS-CoV was selected as
110 the recipient spike backbone because of its extensive structural characterization and compatibility
111 with cell-culture based models of merbecovirus entry. [1,31-33] Similar to our prior sarbecovirus
112 study,[25] the MERS-CoV spike was codon optimized for human cells, appended with a C-terminal
113 FLAG tag and silent mutations were introduced outside the RBD (Figure 1B). While MERS-CoV
114 uses DPP4 for entry, ACE2 functions as a receptor for some other merbecoviruses including
115 PDF2180 (Figure 1D).[22] Thus the RBD from PDF2180 made an ideal candidate to develop a
116 chimeric MERS-CoV spike and demonstrate successful receptor switching from DPP4 to ACE2
117 usage, generating a platform to explore different merbecovirus RBD receptor preferences in entry.

118 To guide construction and expression of chimeric merbecovirus spikes, we utilized
119 available structures. Unfortunately, amino acid motifs that are conserved across all merbecovirus
120 spike sequences are rare, complicating the task of identifying common RBD exchange junctions
121 across the merbecovirus subgenus. However, closer inspection of published spike sequences
122 revealed several highly conserved glycine residues flanking the RBD (MERS-CoV spike positions
123 372 and 610; Figure 1A, D, E). These two residues flank the RBD region of spike used for
124 structures with host receptor molecules.[22,34] While the structures of RBDs have been determined
125 for some merbecoviruses,[11,22,34,35] regions flanking the RBD are flexible, making them difficult to
126 structurally resolve in RBD:receptor complexes (Figure 1A, C; pink shaded region). To maximize
127 chances for successful spike expression, several chimeras between MERS-CoV and PDF2180
128 were produced: version 1 included the region of PDF2180 spike between the conserved glycines,
129 version 2 was based on structure and version 3 was based on the PDF2180-RBD as defined
130 previously (Figure 1E).[22] DNA fragments encoding the different RBD fragments were cloned into

131 the MERS-CoV spike backbone and used to generate vesicular stomatitis virus (VSV) single-
132 cycle pseudotypes (Figure 1C).
133
134 VSV-based pseudotypes were evaluated using a cell-based entry assay. While Baby Hamster
135 Kidney cells (BHK cells) are not naturally susceptible to coronavirus entry but can support entry
136 after ectopic receptor expression, BHK cells expressing human ACE2 do not always recapitulate
137 natural viral entry for ACE2-dependent viruses with low affinity for the receptor.[29,36] Therefore,
138 293T cells, which exhibit low susceptibility to coronavirus entry, were stably transduced to express
139 human ACE2 or DPP4. Infecting these stable cell lines with pseudotypes bearing chimeric spikes
140 revealed that the PDF2180 RBD was able to use human ACE2 for entry, indicating a successful
141 receptor exchange for MERS-CoV spike (Figure 1D, F; PDF2180 V1). The regions of PDF2180
142 spike between RBD residues 376-381 and 586-609 were essential for receptor-use exchange, as
143 two other spikes with shorter regions of the PDF2180 RBD were unable to use ACE2 (Figure 1A,
144 D-F; PDF2180 V2, V3). Addition of exogenous protease (trypsin) during the assay enhanced viral
145 entry, and afforded more sensitive detection of low affinity interactions. Importantly, trypsin-
146 enhanced entry was still receptor dependent (Figure 1F). Production of a chimeric spike
147 containing an analogous region from the DPP4-dependent HKU4-SM3A virus showed a retention
148 in DPP4 use in MERS-CoV spike,[37] further suggesting this region would allow for assessment of
149 diverse receptor preferences (Figure 1E-F).
150
151 ***Optimizing MERS-CoV spike for pseudotype assays***
152 With an operational RBD-exchange strategy in place, the next goal was to optimize the spike
153 backbone for pseudotype efficiency, which would allow for improved sensitivity to low-level viral
154 entry. After engaging with the host receptor, coronavirus spikes are processed by host-cell
155 proteases to release the spike fusion peptide and mediate membrane fusion.[38,39] While the
156 TMPRSS-family of cell-surface proteases and endosomal cathepsins have been shown to cleave

157 many different coronavirus spikes, including MERS-CoV, SARS-CoV and SARS-CoV-2, furin is
158 another protease that plays an important role in spike biogenesis and cell entry, depending on the
159 availability of an appropriate cleavage site.[40-43] MERS-CoV spike contains furin sites at the S1/S2
160 and S2' junctions, although the role of furin in entry is less clear as removal of these sites has no
161 effect on viral pseudotype entry *in vitro.*[38,44] Because previous studies have shown that removal
162 of the furin site from SARS-CoV-2 spike results in increased pseudotype efficiency,[45-47] several
163 versions of chimeric merbecovirus spikes were produced with either one or both furin sites
164 mutated. Additionally, C-terminal truncations in the cytoplasmic tails of MERS-CoV, SARS-CoV
165 and SARS-CoV-2 spikes increase surface expression of spike in transfected cells, resulting in
166 improved pseudotype efficiency.[48-50] Therefore, the protease site mutants were also tested with
167 or without a 16-amino acid truncation in the C-terminal end of MERS-CoV spike (Figure S2A-C).
168 Surprisingly, these modifications to spike had little effects on entry in 293T cells for MERS-CoV
169 WT, PDF2180 RBD or HKU4/SM3A RBD (Figure S2B-C). While not statistically significant, spikes
170 containing a C-terminal truncation and removal of the S2' furin site performed slightly better than
171 the other spikes; therefore, this set of modifications was included for further experiments.
172
173 ***A pseudotype panel encompassing merbecovirus RBD diversity***
174 The 2012 MERS-CoV outbreak renewed efforts in coronavirus sequencing from human patients
175 and the dromedary camel intermediate host species, as well as virus discovery efforts. [51-54]
176 Genbank has over 2000 entries for "merbecovirus" and "MERS-like" viruses; of these 2000
177 entries, but only about 842 entries contain full-length spike, and of those entries, approximately
178 half are from species other than humans and camels (Figure 2A). Many coronavirus Genbank
179 listings do not always contain species identifiers in the metadata, which left several MERS-CoV
180 entries remaining even after excluding human and camel viruses. Further eliminating viruses titled
181 as "MERS-CoV" from the list resulted in 65 spike sequences, which encompassed 34 unique RBD
182 sequences (Figure 2A; Table 1, Figure S1). Some merbecoviruses are likely missing from this list

183 as coronavirus nomenclature has evolved over the years. However, even with these
184 considerations, this curated panel of merbecoviruses spans 18 years of virus discovery (some
185 sequences in this study pre-date the MERS-CoV outbreak), 6 countries and 3 continents (Table
186 1).
187
188 Phylogenetic analysis of the RBD amino acid sequences (as defined between the two glycine
189 residues highlighted in Figure 1) has revealed several genetically-clustered groups based on
190 characteristic indels and the sequences surrounding them (Figure 2B, C, Figure S1). The groups
191 were labeled as clades 1 through 4: clade 1 includes MERS-CoV and HKU4 discovered between
192 2006-2012,[1,55,56] clade 2 includes HKU5 also discovered in 2006,[55] clade 3 incudes the African
193 merbecoviruses, NeoCoV and PDF2180, discovered in 2017,[13] and clade 4 includes the
194 contemporary hedgehog merbecoviruses, "ErinCoVs," discovered in Germany in 2014 and most
195 recently identified in Italy in 2020 (Figure 2B-C, Table 1).[9,57-59] RBDs for these merbecoviruses
196 were synthesized and cloned into the optimized MERS-CoV spike backbone plasmid. All chimeric
197 spike plasmids were stably expressed in 293T cells and used to produce pseudotyped viral-like
198 particles for downstream experiments (Figure 2D). Concentrated supernatants were positive for
199 the presence of spike by western blot analysis; however, variations observed in spike levels did
200 not correlate with viral entry phenotypes observed in downstream assays (Figure 2D).
201
202 ***Merbecovirus entry in common laboratory cell lines***
203 Vero cells (from African Green Monkey kidneys) are common laboratory cell lines in coronavirus
204 research and are used in attempts to isolate and propagate merbecoviruses and other viruses.
205 [1,60] VeroE6 cells were poorly permissive for entry of the pseudotyped merbecovirus RBD panel,
206 with addition of trypsin only mildly improving entry for some of the viruses in a calde-independent
207 manner (Figure 2E). In contrast, Huh-7.5 cells, a human hepatocellular carcinoma cell line, were
208 more susceptible to clade 1 merbecovirus RBDs (Figure 2E). Addition of trypsin improved entry

209 for some of the clade 4 viruses for both cell lines (e.g., ErinCoV-1-18 and ErinCov-12-19) (Figure
210 2E).
211
212 ***Merbecovirus compatibility with known human coronavirus receptors in human cells***
213 The diverse panel of merbecovirus pseudotypes was tested on 293T cell lines stably transduced
214 to express known human coronavirus receptors: amino peptidase N (APN), DPP4, or Angiotensin-
215 ACE2. In the absence of trypsin, untransduced 293T cells exhibited poor susceptibility to the
216 entire panel, while APN cells were susceptible to entry with a positive control: human coronavirus
217 229E, which is known to use APN as an entry receptor (Figure 2F).[61] Cells transduced with human
218 DPP4 exhibited clear susceptibility to clade 1 merbecovirus RBDs but not the other clades (Figure
219 2F), while cells expressing human ACE2 exhibited low level susceptibility to SARS-CoV-2 spike
220 and clade 3 merbecovirus RBDs (Figure 2F). Exogenous trypsin moderately increased entry in
221 untransduced 293T cells for human coronaviruses MERS-CoV and SARS-CoV-2, which uses
222 human DPP4 and human ACE2, respectively, suggesting low levels of the human receptors were
223 present on 293T cells (Figure 2F). In the presence of trypsin, clade 1 RBDs entered cells through
224 DPP4, while clades 2, and 3 RBDs entered cells expressing ACE2 (Figure 2F). APN cells
225 remained susceptible to HCoV-229E entry in the presence of trypsin. While entry signal under
226 10x is considered borderline permissive in this assay, and difficult to interpret, there were many
227 viruses that presented low, but consistently detectable levels of ACE2 compatibility (Figure 2F).
228
229 ***Species-specific receptor use across the merbecovirus RBD clades***
230 Human cells expressing human ACE2 exhibited low level susceptibility to a wide range of
231 merbeocviruses (Figure 2). To investigate if the merbecovirus RBD panel can use receptor
232 orthologues from their cognate host species, ACE2 and DPP4 were synthesized from European
233 hedgehogs (*Erinaceus europaeus*), the natural host species for the ErinCoVs (Table 1;
234 merbecovirus RBD clade 4), and ACE2 was synthesized from Japanese house bats (*Pipistrellus*

235 *abramus*): the natural host species for the HKU5 viruses (Table 1; merbecovirus RBD clade 2).
236 DPP4 and ACE2 from European hedgehogs were inefficient as entry receptors and did not
237 mediate entry for any of the merbecovirus clades in the absence of trypsin (Figure 2F). However,
238 293T cells transfected with ACE2 from Japanese house bats were highly permissive for entry by
239 the entire HKU5 cluster of clade 2 merbecovirus RBDs – in the presence or absence of exogenous
240 trypsin (Figure 2F).

241 To further reduce the background signal in the entry assays, 293T cells were exchanged
242 with BHK cells, which are not susceptible to any sarbecovirus or merbecvirus due to significant
243 hamster receptor incompatibilities and poor cell surface expression[62,63]. Following transfection
244 with *Pipstrellus abramus* ACE2, BHK cells became highly permissive for HKU5 RBD-mediated
245 entry (Figure 2G). Notably, all HKU5 sequences derive exclusively from *Pipistrellus abramus*
246 samples (Table 1).

247

248 ***Confirmation of HKU5 receptor use with stable cell lines and full-length spikes***

249 The previous experiments were performed with human receptor stable cell lines and cells
250 transiently transfected with *Pipistrellus abramus* ACE2. To achieve more consistent entry results
251 with *Pipistrellus abramus* ACE2, a lentiviral vector was produced to generate stable cell lines,
252 following the successful methodologies applied to generate the human receptor cell lines[29,64]
253 (Figure 3A). Unlike the parental cells, VeroE6 transduced with *Pipistrellus abramus* ACE2 were
254 highly susceptible to the entire HKU5 RBD panel as well as the clade 3 viruses, NeoCoV and
255 PDF2180, but remained poorly susceptible to clade 1, clade 4 and the clade 2 RBDs that were
256 not identified in *Pipistrellus abramus* bats (compare Figure 2E compared with 3B). To further
257 confirm receptor use for the HKU5 viruses, full-length spike sequences from HKU5-20s and
258 HKU5-21s were synthesized and used to produce viral pseudotypes. For comparison,
259 pseudotypes were also produced with the chimeric MERS-CoV-based spikes used in the previous
260 experiments (Figure 3C). Full-length HKU5-20s and HKU5-21s spikes were expressed and

261 incorporated into pseudotypes (Figure 3D). Again, while detection of spike varied across the
262 concentrated pseudotype panel, it did not correspond with the observed entry results (Figure 3D-
263 E). BHK cells transduced with *Pipistrellus abramus* ACE2 were susceptible to both full-length and
264 chimeric HKU5-20s and -21s spikes but not MERS-CoV, a DPP4-dependent spike (Figure 3E, F).
265 Altogether, the clear and species-specific entry observed for all HKU5 RBDs and full-length HKU5
266 spikes exclusively in cells expressing *Pipstrellus abramus* ACE2 strongly support a hypothesis
267 that ACE2 is the natural receptor for these viruses in natural hosts.
268
269 ***Confirmation of HKU5 receptor use with a replication-competent HKU5 molecular clone***
270 To confirm if *Pipistrellus abramus* ACE2 could function as a receptor for true virus, we generated
271 a replication competent molecular clone of HKU5 with green fluorescent protein (GFP) cloned in
272 place of Orf5 (Figure 3F).[24] Vero cells, 293T cells or BHK cells stably transduced with *Pipistrellus*
273 *abramus* all strongly supported effecient viral replication in the absence of trypsin, as evidenced
274 by GFP expression during replication, accumulation of infectious virus as measured by plaque
275 assays, and production of viral spike and nucleocapsid proteins in infected cells (Figure 3G-I).
276 Detection of distinct spike glycoprotein products at ~180 and 75 kDa strongly support spike
277 proteolytic processing and cleavage due to the presence of a putative furin cleavage site.
278
279 ***Consensus sequences mimic clade-specific entry phenotypes***
280 To further assess clade-specific receptor use, consensus sequences were generated for the
281 RBDs from clades 1, 2 and 4 and used to produce chimeric spikes. PDF2180 was selected as a
282 representative of clade 3 because clade 3 only contains two representative RBD sequences
283 (Figure 4A). Pseudotypes bearing chimeric consensus RBD spikes generally reflected clade-
284 specific entry phenotypes observed for individual RBDs from previous experiments (Figure 4B-
285 C), suggesting that features common amino acids, structural features and motifs within each RBD
286 clade dictate entry.

287 ***Mutations in the HKU5 RBD modulate spike activity and species compatibility***

288 A previous study reported structures of clade 3 merbecovirus RBDs from PDF2180 and NeoCoV
289 bound to ACE2 from *Pipstrellus pipstrellus* bats.[22] Although this study demonstrated HKU5 spike
290 could not use *Pipistrellus pipistrellus* or any of 46 other bat ACE2s for entry, these RBD-ACE2
291 structures were used as a template to model predicted structures for HKU5 RBDs and *Pipistrellus*
292 *abramus* ACE2 (Figure 5A-C). In the experimental structure, the interface between the clade 3
293 RBD, PDF2180, with *Pipistrellus pipistrellus* ACE2 involves multiple interactions with two surface-
294 exposed loops on the RBD (Figure 5A), which we refer to here as Loop 1 and Loop 2. Curiously,
295 the Loop 2 is highly polymorphic in merbecoviruses, with several HKU5 viruses, but not all,
296 exhibiting a 4 amino acid deletion in this region (Figure 5B-C; Figure S1B). To assess if an
297 analogous interface is compatible with interactions between HKU5 RBDs and bat ACE2, several
298 sterically bulky or charged amino acid mutations were introduced at two points in the middle of
299 Loops 1 and 2 (Figure 5). For these mutation experiments, HKU5-20s RBD was selected because
300 it includes both loops and shows strong entry with *Pipistrellus abramus* ACE2, and HKU5-21s
301 RBD was selected because it contains a 4 amino acid deletion in loop 2, while still strongly
302 exhibiting entry using *Pipistrellus abramus* ACE2 (Figure 2, 3A, Figure S1C). For experimental
303 consistency, BHK cells and 293T cells were transduced with a lentivector to constitutively express
304 *Pipistrellus abramus* ACE2 and were infected in parallel assays to reduce effects from non-
305 specific variation in the cell line background. Overall, mutations in Loop 2, but not Loop 1, of the
306 HKU5 RBD disrupted entry with *Pipistrellus abramus* ACE2 (Figure 5E-G). While most of the
307 mutations only reduced spike entry with bat ACE2, mutations in Loop 2 of the HKU5-20s RBD,
308 but not the HKU5-21s RBD, increased entry with cells expressing human ACE2 (Figure 5G). While
309 this difference was relatively small, the statistically-significant increase in entry was observed
310 across different cell line backgrounds. Importantly, each mutation's effect on viral entry was
311 observed across both the BHK (hamster) and 293T-based (human) cell lines, suggesting the

312 observed effects derive from the transduced ACE2 rather than other non-specific cell line variation
313 (Figure 5).
314
315 ***HKU5-20s spike structure reveals all 'down' RBDs***
316 To further assess the interaction between HKU5-20s spike and *Pipistrellus abramus* ACE2
317 (paACE2), we attempted to solve a single-particle cryo-EM structure of a complex between a 2P
318 stabilized[65] KU5-20s trimeric ectodomain and soluble paACE-2 (both monomeric and dimeric
319 paACE2-Fc), preparing grids at room temperature or at 37 ˚C. All grids yielded structures of the
320 HKU5-20s spike trimer only, with no identifiable density for paACE2, suggesting further
321 optimization of the binding conditions is need for determining a receptor-spike complex structure
322 by cryo-EM. Indeed, prior work assessing merbecovirus spike interactions with *Pipistrellus* ACE2
323 were only in the context of purified RBD with receptor, not with a spike trimer ectodomain.[22]
324 We processed data from a grid incubated with paACE2-Fc plus HKU5-20s spike trimer
325 ectodomain to derive a 2.8 Å structure of the unbound HKU5 trimer (Figure S3, Table S1). The
326 structure revealed the expected S1 and S2 subunit organization as seen in other CoV spike
327 trimers (Figure 6A-D). Although a distribution of 'up' and 'down' RBDs is usually found for
328 sarbecovirus spike trimers,[66,67] we found only 'down' RBDs in the HKU5-20s trimer structure
329 (Figure 6A,B). To examine the degree of sequence conservation across 10 representative
330 merbecovirus spikes (see Methods), we plotted amino acid sequence conservation on surface
331 representations of structures of the HKU5-20s spike protomer and the HKU5-20s RBD (Figure
332 6D). We found more conserved regions in S2 and at the base of the RBD (Figure 6D). The well-
333 resolved densities for the three 'down' RBDs allowed building of an RBD model including
334 coordinates for Loops 1 and 2 (Figure 6E). These results provide additional structural insights for
335 HKU5-20s spike trimer over an existing HKU4 RBD structure,[11] and will be instrumental in future
336 antigen design.
337

338 **DISCUSSION:**

339 Compared to sarbecoviruses, which share ≥90% nucleotide sequence identity in their
340 RBDs,[25,68] the diversity in merbecovirus RBD sequences is greater, initially complicating the RBD
341 exchange strategy utilized in this study (Figure 1, 6E; Figure S1). Wide variability between
342 merbecovirus host species (hedgehogs, bats, camels, humans and pangolins) also poses
343 unforeseen challenges when studying cross-species interactions. The chimeric spike approach
344 presented herein aims to eliminates some of the interspecies variability by using the human-
345 compatible MERS-CoV spike as a backbone in experiments performed in human cells and other
346 cells that support MERS-CoV infection. Our studies support the hypothesis that this approach
347 reduces off-target phenotypes relating to cell attachment and protease processing. Addition of
348 exogenous protease bypasses host-cell protease-processing, allowing for assessment of just the
349 RBD interaction in the presence or absence of receptor. As now seen in this study and many
350 others, chimeric coronavirus spikes faithfully re-create entry of both full spikes and wild-type
351 viruses (Figure 1-3).[25-30,36,64,69]

352 Merbecovirus RBDs can be classified into four clades based on sequence, characteristic
353 indels, and entry phenotype, but not geographic origin (Figure 2, 4; Figure S1, Table 1). Results
354 from this study show that clade 1 viruses can use DPP4 from multiple species, clade 2 viruses
355 have species-specific ACE2 preference, and clade 3 can use ACE2 from a wider range of host
356 species. Recently, a merbecovirus found in Russian bats was shown to use ACE2 with a restricted
357 host range, further demonstrating species-specific ACE2 interactions among merbecoviruses.[70]
358 No clear entry was observed with clade 4 viruses and receptors tested here, but nevertheless,
359 human-cell entry was observed with ErinCoVs in the presence of trypsin (Figure 2E-G, 3B).
360 Although the data presented here suggest ACE2 and DPP4 are not the receptor for clade 4
361 merbecoviruses, it is possible these viruses use these receptors in other host species not tested
362 herein, and/or that the approach taken in this study is not suitable for detecting their entry. For
363 example, if a region of ErinCoV spike outside of the RBD boundaries tested here interacts with

364 the host receptor or regulates RBD binding, then the chimeric spike entry assays would miss this
365 interaction. Additionally, while the chimeric spike strategy helps to reduce unforeseen entry
366 barriers, dependence on extra-cellular factors remains possible, but is difficult to anticipate or
367 control.[71] Alternatively, clade 4 merbecovirus RBDs may use a receptor that is neither DPP4 nor
368 ACE2, analogous to sarbecovirus RBD clade 2 viruses, which also can infect human cells through
369 an unidentified route (Figure 2, 4).[25-28]

370 MERS-CoV spike can rapidly adapt to sub-optimal interactions with DPP4 from non-
371 cognate species,[64] and the ongoing SARS-CoV-2 pandemic underscores the ability of spike to
372 adapt and correct sub-optimal ACE2 interactions while simultaneously escaping the immune
373 response [72-75]. Other research has shown that clade 3 merbecoviruses gain even broader host
374 ACE2 tropism with a single point mutation in the spike RBD, T510F, significantly increasing
375 tropism for human ACE2.[21,22] Mutagenesis experiments presented here identified a region of
376 HKU5 RBDs that also modifies ACE2 tropism (Figure 5). Structurally, T510F is on loop 1 of the
377 RBD, while the effective mutations identified in this study are located on the adjacent loop 2
378 (Figure 5A-C). Only HKU5-20s, which possesses a larger second loop in its RBD, was able to
379 gain consistent and statistically significant entry using human ACE2, with mutations. By contrast,
380 HKU5-21s, with a shorter surface loop, did not measurably shift in ACE2 tropism (Figure 5E-G;
381 Figure S1B). These results suggest that deletions within loop 2 of some HKU5 viruses may limit
382 their evolutionary adaptability and that have a larger second loop, found in spikes such from
383 viruses such as HKU5-20s, may allow for a wider range of ACE2 interactions. These data also
384 suggest HKU5 spike may interact with a different region of ACE2 than has been reported for other
385 merbecoviruses. Deep mutagenesis scanning and structural studies with both types of HKU5
386 RBDs are needed to further understand the HKU5-RBD interface, although even the low -level
387 receptor compatibility reported here for these viruses is concerning and underscores the need for
388 broadly protective coronavirus vaccines[69,76].

389 Structural studies of other coronavirus spike trimers have revealed a distribution of
390 different conformations with respect to the position of the RBDs.[11,67,77] Coronavirus RBDs on a
391 spike trimer can be in an 'up' state, with receptor contacts being surface-exposed and capable of
392 interacting with host cell receptors, or in the 'down' state, with the receptor contacts buried and
393 inaccessible.[11,67,77] MERS-CoV spike has been observed with one or two RBDs down,[78] SARS-
394 CoV spike has been observed with two or three 'down' RBDs,[78] and SARS-CoV-2 spike has been
395 observed with two or three 'down' RBDs.[66,67,79,80] Analogously, the spike trimer from HCoV-229E,
396 an alphacoronavirus with a tri-partite RBD, has been observed with all 'down' RBDs or some 'up'
397 RBDs.[77] Because the RBD is the primary target of neutralizing antibodies, transitions between
398 'up' and 'down' RBD states is believed to play a role in immune evasion.[67,78] We only observed
399 HKU5-20s spike ectodomains with all 'down' RBDs (Figure 6). Structures of spike trimers from
400 bat merbecoviruses PDF2180 and NeoCoV, as well as well as the bat sarbecovirus RsSHC014,
401 have also been observed exclusively with all 'down' RBDs.[22,81] Unlike the human respiratory
402 coronaviruses, all bat coronaviruses have been identified in fecal or other gut-derived samples
403 and are presumed to be gastrointestinal in their hosts.[6,12,13] The low pH in the stomach and
404 protease-rich gastrointestinal tract presents a unique and harsh environment for viruses, which
405 may select for more closed confirmations in coronavirus spikes that have this tissue tropism. How
406 these closed spikes transition to a state that can interact with the receptor is yet to be determined
407 but may involve extracellular factors and/or pH changes.[66]

408 HKU5 viruses were first sequenced in 2006, shown to be DPP4-independent in 2014, and
409 demonstrated to replicate in Vero cells in the presence of exogenous trypsin in 2020.[12,24,55] To
410 date, more than 40 ACE2 orthologues from bats and other species have been tested with HKU5
411 viruses;[6,22,24,82] however, only ACE2 from *Pipistrellus abramus* has proven to mediate entry for
412 these viruses (Figure 2-5). The RBD from HKU5 viruses shares ~50% amino acid identity with
413 MERS-CoV and clade 1 RBDs, while the RBD from NeoCoV is only ~30-40% identical to MERS-

414 CoV and HKU5 coronaviruses (Figure S1). This low sequence identity between NeoCoV and
415 HKU5 RBDs has confounded efforts to predict the receptor for HKU5 from sequences alone.

416 A previous study identified ACE2 as the receptor for the clade 3 merbecoviruses such as
417 PDF2180 and closely-related NeoCoV but was unable to identify the receptor for the clade 2
418 HKU5 viruses.[22] This earlier study assessed the ability of HKU5 to interact with ACE2 from
419 *Pipistrellus pipistrellus* bats and several other Yangochiropteran bats, but not the true host species
420 for HKU5. The clear preference of HKU5 for *Pipistrellus abramus* ACE2, but not other ACE2
421 species orthologues tested clearly demonstrates a tighter species specificity found within clade 2
422 RBDs (Figure 2-3). *Pipistrellus abramus* ACE2 varies from previously tested *P. pipistrellus* and *P.*
423 *khuli* ACE2s at residues 328 and 329, in close proximity to the interaction site previously reported
424 with NeoCoV/PDF2180 RBDs, potentially explaining why HKU5 RBDs do not bind this orthologue
425 (Figure 5A-C).[22] Residue N329 in *Pipistrellus pipistrellus* ACE2 is also N-glycosylated,[21] while the
426 corresponding residue in *Pipistrellus abramus* ACE2, R328, is not, which may also explain the
427 striking restriction in receptor use. Similar to the merbecovirus RBD clades, clade 1
428 sarbecoviruses can use ACE2 from diverse species, while clade 3 sarbecovirus RBDs only
429 interact with ACE2 from their cognate species (Figure 4C).[36,69]

430 A clade 1 RBD pangolin merbecovirus has been shown to use human DPP4 and was
431 isolated in Vero81 cells.[83] While Vero cells are frequently used in virus isolation attempts,[1,60,84]
432 data from our study suggests that Huh-7.5 cells may be a more suitable cell line for isolating
433 DPP4-dependent merbecoviruses and perhaps even some clade 4 merbecoviruses. The
434 mechanisms underlying why these cells seem to allow more of these viruses to enter is yet to be
435 determined. Furthermore, MERS-CoV grown in Vero cells has been shown to quickly accumulate
436 tissue-culture adaptations in the spike protein driving the need for alternative cell culture models.[33]

437 Merbecoviruses represent yet another type of beta-coronavirus found in wildlife with
438 potential human cell compatibility and a possibility to cause severe disease in humans. Japanese
439 house bats (*Pipistrellus abramus*) are the natural host for HKU5 coronaviruses and are

440 synanthropic, roosting in human dwellings and structures, and therefore pose repeat opportunities
441 for human exposure.[55,85] Luckily, clade 2 merbecovirus RBDs exhibit low human ACE2
442 compatibility (Figure 2-5), although efforts should still be made to reduce contact with these
443 animals. The development of replication competent HKU5 and PDF2180 recombinant viruses will
444 support the development of pan-coronavirus and pan-merbecovirus vaccines and small molecule
445 countermeasures to protect the health of populations globally. The work presented here will serve
446 as a guide to understanding merbecovirus cell entry, subsequent spillover risk and support One
447 Health Preparedness.
448
449
450 **METHODS:**
451 Biosafety
452 Prior approval for experiments using full-length, recombinantly derived HKU5 was obtained from
453 the Institutional Biosafety Committee of the University of North Carolina at Chapel Hill. All
454 experiments using live HKU5 virus were performed under Biosafety Level 3 conditions with
455 personnel wearing full-body personal protective equipment and HEPA-filtered respiratory
456 protection.
457
458 Cells and viruses
459 293T, VeroE6, Huh-7.5 and BHK-21 cells were maintained in DMEM supplemented with 10%
460 FBS, penicillin-streptomycin, and L-glutamine and maintained at 37°C, 5% CO_2. Cells stably
461 expressing coronavirus receptors were generated as previously described and maintained under
462 selection with either 1µg/mL puromycin (BHK and 293T) or 5µg/mL puromycin (VeroE6).[29] All cell
463 lines were species-confirmed with cytochrome sequencing and verified mycoplasma negative
464 with MycoSniff PCR Kit (MP Bio). The molecular clone derived HKU5-GFP was generated as

465 previously described.[24] Briefly, genomic cDNA sequences were ligated, *in vitro* transcribed, and
466 electroporated into BHK cells. Stocks of recombinant HKU5-GFP were then propagated in Vero
467 cells stably expressing *Pipistrellus abramus* ACE2 at 37°C and 5.0% CO_2. For multistep growth
468 curve analysis, cells were infected at an MOI of 0.01. Samples of infected cell culture supernatant
469 were collected at 0, 4, 24, 48, and 72 h.p.i. and stored at -80 for titration via plaque assay. For
470 fluorescent microscopy and western blot experiments, cells were infected with HKU5-GFP at an
471 MOI of 1.0. At 24 h.p.i., GFP-fluorescence was monitored using an inverted optical microscope
472 (Olympus, IX73), and lysates were harvested for western blot analysis as described below.
473
474 Phylogenetic analysis
475 As described previously, the amino acid sequences for spike receptor binding domains aligned
476 using Clustal multiple sequence alignment software and default parameters. A maximum
477 likelihood phylogenetic tree was produced with PhyML v. 3.0 [86]. The 'WAG' matrix +G model of
478 amino acid substitution was selected by the Smart Model Selection method with 1000 bootstrap
479 replicates[87]. A tree was then visualized as a cladogram with FigTree v1.4.4
480 (https://github.com/rambaut/figtree). Clade consensus sequences were generated with the aid of
481 Geneious (Dotmatics, Inc.).
482
483 Plasmids
484 MERS-CoV/EMC12 spike (accession number JX869059) was codon optimized, appended with
485 C-terminal FLAG tag, and silent mutations were introduced near spike amino acids 341 and 617
486 to generate AflII and NotI restriction digest sites. RBDs were codon optimized and synthesized
487 (IDT DNA) as previously described for sarbecoviruses[25]. Lyophilized RBD fragments were
488 resuspended in water and used in downstream Infusion based cloning (Takara Bio) reactions to
489 assemble full-length chimeric spike proteins. *Erinaceus europaeus* ACE2, transcript variant 1
490 (genbank XM_060183012.1), *Pipistrellus abramus* ACE2 (genbank GQ262782.1) and *Erinaceus*

491 *europaeus* DPP4 (genbank XM_060177812.1) were codon optimized, split in half and
492 synthesized for Infusion based DNA assembly. Splitting genes into multiple parts reduces
493 synthesis time, costs and overall synthesized product complexity. A complete list of accession
494 numbers for the spike panel can be found in Table 1.
495
496 *Pipistrellus abramus* ACE2 cell lines
497 *Pipistrellus abramus* ACE2 was sub-cloned from expression plasmid into a lentiviral expression
498 vector with PCR and used to produce lentiviral particles in 293T cells [64,88]. Supernatants were
499 filtered, aliquoted and frozen at -80°C. BHK cells, 293 T cells and VeroE6 cells were transduced
500 with a lentiviral expression vector as previously described [29,64]. Cells were maintained for 3
501 passages under selective pressure before use in experiments.
502
503 Pseudotype entry assay
504 Pseudotype assays were performed as previously described[26,27]. For human ACE2, APN, and
505 DPP4 experiments, 293T cells stably transduced with potential coronavirus receptors were
506 seeded in black 96-well plates and subsequently infected 24 hours later with equal volumes of
507 viral pseudotypes. Infections were performed on ice to prevent unintended, premature trypsin
508 activity. Pseudotypes were combined with either HBSS or HBSS containing trypsin (not TPCK-
509 treated) to achieve a final concentration of 100µg/mL trypsin. Cells were washed once with cold
510 PBS, centrifuged with pseudotypes at 4°C, 1200 x g for 1 hour and incubated at 37°C overnight.
511 Luciferase was measured using Bright-Glo reagent (Promega) approximately 18 hours post-
512 infection. Relative entry was calculated by dividing raw luciferase values for each spike by the
513 signal for no-spike pseudotypes. To compensate for plate reader background error, plates were
514 measured 4 times, analyzed individually and then averaged across all four measurements.
515 Relative entry values for all four replicates were averaged and plotted as heatmaps using
516 Microsoft Excel.

517 Statistics and reproducibility

518 Each figure shows data that is representative of four technical replicates. During this study,
519 experiments were performed with different batches of pseudotypes and cells and at different
520 times. Therefore, the results shown are representative of the observed biological replicates.
521 Statistical significance for mutation comparisons in Figure 5 were determined by 2-way ANOVA
522 with Tukey correction for multiple comparisons (GraphPad).

523

524 Structure prediction and docking

525 MERS-CoV spike RBD bound to DPP4 (PDB ID: 4L72) [35], the MERS-CoV spike trimer (PDB ID:
526 5X59)[78], and the co-structure of NeoCoV-RBD with P. pipistrellus ACE2 (PDB ID: 7WPZ)[22], were
527 visualized in PyMol (version 2.4.0). Structures for HKU5-20s and HKU5-21s RBDs were predicted
528 with SwissModel using PDB IDs 8SAK and 5XGR, respectively[34,89]. *Pipistrellus abramus* ACE2
529 was modelled with similar methods using PDB ID 7C8J as a guide[90].

530

531 Protein Expression for cryo-EM

532 A trimeric HKU5-20s trimeric ectodomain (residues 24-1301 of the GenBank sequence
533 AGP04933) with 2P stabilizing mutations,[65] a mutated furin cleavage site between S1 and S2, a
534 C-terminal TEV site, foldon trimerization motif, 8xHis tag, and AviTag) was expressed as
535 described.[91] A gene encoding a 8xHis-tagged soluble paACE2 (*Pipistrellus abramus* ACE2)
536 construct (residues 20-613 of the Genbank sequence ACT66266) or a soluble paACE2-Fc
537 (residues 20-613 of Genbank sequence ACT66266 fused to human IgG1 Fc) were expressed and
538 purified as described for human ACE2 and human ACE2-Fc[92] by nickel-NTA (ACE2) or protein A
539 (paACE2-Fc) chromatography followed by size-exclusion chromatography using a Superdex 200
540 column (Cytiva) as described.[93] Peak fractions were identified by SDS-PAGE, and those
541 containing spike trimer, paACE2, or paACE2-Fc were pooled, concentrated, and stored at 4 ˚C
542 until use.

543 Cryo-EM sample preparation

544 paACE2-spike trimer and paACE2-Fc–spike trimer complexes were prepared by incubating
545 purified paACE2 or paACE2-Fc with HKU5-20s trimer at a molar ratio of 3:1 (paACE2 or pACE2-
546 Fc:spike) at 37°C for 30 minutes to a final concentration of ~2 mg/mL. Cryo-EM grids were
547 prepared using a Mark IV Vitrobot (ThermoFisher) at 37 °C and 100% humidity. Sample (3 µL)
548 was applied to 300 mesh Quantifoil R1.2/1.3 grids (Electron Microscopy Sciences) that had been
549 freshly glow discharged with PELCO easiGLOW (Ted Pella) for 1 min at 20 mA, blotted for 3
550 seconds with Whatman No.1 filter paper, and grids were vitrified in liquid ethane. Tweezers were
551 preheated to 37˚C before picking up grids.

552

553 Cryo-EM data collection and processing

554 A single-particle cryo-EM dataset for HKU5-20s spike 2P with paACE2-Fc was collected using
555 SerialEM automated data collection software[94] on a 200 keV Talos Arctica cryo-electron
556 microscope (Thermo Fisher Scientific) equipped with a K3 camera (Gatan) (Table S1). Movies
557 were recorded with 40 frames, a defocus range of -1 to -3 µm, and a total dosage of 60 e-/Å^2
558 using a 3x3 beam image shift pattern with 3 exposures per hole in the super-resolution mode with
559 a pixel size of 0.4345 Å. The dataset was motion corrected with patch motion correction using a
560 binning factor of 2, and CTF parameters were estimated using Patch CTF in cryoSPARC v4.5.3.[95]
561 Particle picking was done with blob picker in cryoSPARC using a particle diameter of 100 to 200
562 Å, and movies and picked particles were inspected before extraction. Particles were extracted
563 and classified using 2D classification in cryoSPARC. Particles from selected classes were used
564 for template picker with a particle diameter of 150 Å. Picked particles were inspected before
565 extraction and 2D classification. After exposure curation, the remaining particles from selected
566 classes were used for ab initio modeling with 4 volumes, which were further refined with
567 heterogeneous refinement in cryoSPARC. Subsequent homogeneous and non-uniform

568 refinements were carried out, and the final reconstruction was obtained from a non-uniform
569 refinement conducted with C3 symmetry in cryoSPARC.

570 Structure figures were made using ChimeraX.[96] For plotting sequence conservation onto
571 protein structure surfaces, sequence alignments of 10 representative merbecovirus spikes were
572 performed using Geneious Prime. Conservation scores for analogous residues (HKU5-20s
573 residues 24-1231 for spike protomer and residues 381-614 for RBD) were calculated and mapped
574 onto the HKU5-20s protomer or RBD by using ConSurf.[97] The following 10 merbecovirus spike
575 sequences were used in the sequence alignment:

576 - GenBank AFS88936 (MERS-CoV-EMC/2012)
577 - GenBank AGP04933 (HKU5-20s)
578 - GenBank AGP04934 (HKU5-21s)
579 - GenBank AGY29650 (NeoCoV)
580 - GenBank AHY61337 (BtVs-BetaCoV/SC2013)
581 - GenBank ARJ34226 (PDF2180)
582 - GenBank AUM60014 (Bat-CoV/H.savii)
583 - GenBank AUM60024 (Bat-CoV/P.khulii)
584 - GenBank QRN68031 (ErinCoV-19/2018)
585 - GenBank USL83011 (MOW15-22/2015)

586 MERS-CoV/EMC12 was selected because of high sequence identity with HKU5 spikes;[12]
587 NeoCoV, PDF2180 and MoW-15-22 spikes were selected because they have been shown to use
588 ACE2;[21,22,70] BatCoV/PKhuli, H.Savii, and SC2013 were selected as other clade 2 RBDs to
589 compare with HKU5 RBD (Figure S1), and the clade 4 RBD from ErinCoV-19-18 was selected
590 because of low sequence identity with other merbecovirus RBD clades (Figure S1).

591

592

593 Western blot

594 As described previously,[25] 293T cells transfected with spike plasmids were lysed in 1% SDS lysis
595 buffer and stored at -80°C until use. Lysates were clarified, boiled, reduced and separated on
596 10% Bis-Tris gels in MOPS buffer (Invitrogen NuPage).[25] Pseudotypes were concentrated over
597 iodixanol (Opti-Prep; Sigma) gradients as previously described [25,28,29]. Membranes were probed
598 with FLAG:HRP conjugated monoclonal antibody (ThermoFisher), GAPDH monoclonal
599 (ThermoFisher), MERS-CoV N (Sino Biological; catalog number: 40068-RP02; 1:2000), MERS-
600 CoV S2 (Invitrogen; catalog number: 66008-1-Ig; 1:5000) and reprobed with a secondary goat-
601 anti-mouse:HRP poloyclonal mix (ThermoFisher), goat-anti-rabbit-HRP (Cell Signaling
602 Technology; catalog number: 7074S; 1:5000), or goat anti-mouse-HRP (Cell Signaling
603 Technology; catalog number: 7076S; 1:5000).

604

605

606 Data availability

607 Data that support the findings of this study are available from the corresponding author upon
608 request. Accession numbers for all spike sequences used in this study can be found in Table 1.
609 Atomic models and cryo-EM maps generated from cryo-EM studies of the HKU5-20s spike have
610 been deposited at the Protein Data Bank (PDB) and Electron Microscopy Data Bank (EMDB)
611 under accession codes PDB 9D4T and EMBD-46569, respectively

612

613 Reagent availability

614 All biological reagents are available upon request.

Acknowledgements

We would like to thank Dr. Stephanie Seifert, Dr. Alex Cohen, Dr. Tyler Starr and Dr. Thomas Gallagher for providing valuable insights and comments on the work and manuscript. Research reported in this publication was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (NIAID/NIH) under Award Numbers R01AI179720 to M.L., R01 AI167966 and CEIRS Contract 75N93021C00014 to R.S.B., and P01AI100148 to P.J.B. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health

Conflict of Interest Statement

RSB is a member of advisory boards for VaxArt, Takeda and Invivyd, and has collaborative projects with Gilead, J&J, and Hillevax, focused on unrelated projects.

628 **REFERENCES:**

629 1 Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A.
630 Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J*
631 *Med* **367**, 1814-1820, doi:10.1056/NEJMoa1211721 (2012).
632 2 Dudas, G., Carvalho, L. M., Rambaut, A. & Bedford, T. MERS-CoV spillover at the
633 camel-human interface. *Elife* **7**, doi:10.7554/eLife.31257 (2018).
634 3 Dudas, G. & Rambaut, A. MERS-CoV recombination: implications about the reservoir
635 and potential for adaptation. *Virus Evol* **2**, vev023, doi:10.1093/ve/vev023 (2016).
636 4 Moreno, A. *et al.* Detection and full genome characterization of two beta CoV viruses
637 related to Middle East respiratory syndrome from bats in Italy. *Virol J* **14**, 239,
638 doi:10.1186/s12985-017-0907-1 (2017).
639 5 Luo, C. M. *et al.* Discovery of Novel Bat Coronaviruses in South China That Use the
640 Same Receptor as Middle East Respiratory Syndrome Coronavirus. *J Virol* **92**,
641 doi:10.1128/JVI.00116-18 (2018).
642 6 Lau, S. K. *et al.* Genetic characterization of Betacoronavirus lineage C viruses in bats
643 reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus
644 HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East
645 respiratory syndrome coronavirus. *J Virol* **87**, 8638-8650, doi:10.1128/JVI.01055-13
646 (2013).
647 7 Ithete, N. L. *et al.* Close relative of human Middle East respiratory syndrome coronavirus
648 in bat, South Africa. *Emerg Infect Dis* **19**, 1697-1699, doi:10.3201/eid1910.130946
649 (2013).
650 8 He, W. T. *et al.* Virome characterization of game animals in China reveals a spectrum of
651 emerging pathogens. *Cell* **185**, 1117-1129 e1118, doi:10.1016/j.cell.2022.02.014 (2022).
652 9 De Sabato, L. *et al.* Can Coronaviruses Steal Genes from the Host as Evidenced in
653 Western European Hedgehogs by EriCoV Genetic Characterization? *Viruses* **12**,
654 doi:10.3390/v12121471 (2020).
655 10 Chen, J. *et al.* A bat MERS-like coronavirus circulates in pangolins and utilizes human
656 DPP4 and host proteases for cell entry. *Cell* **186**, 850-863 e816,
657 doi:10.1016/j.cell.2023.01.019 (2023).
658 11 Wang, Q. *et al.* Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of
659 human receptor CD26. *Cell Host Microbe* **16**, 328-337, doi:10.1016/j.chom.2014.08.009
660 (2014).
661 12 Yang, Y. *et al.* Receptor usage and cell entry of bat coronavirus HKU4 provide insight
662 into bat-to-human transmission of MERS coronavirus. *Proc Natl Acad Sci U S A* **111**,
663 12516-12521, doi:10.1073/pnas.1405889111 (2014).
664 13 Anthony, S. J. *et al.* Further Evidence for Bats as the Evolutionary Source of Middle East
665 Respiratory Syndrome Coronavirus. *MBio* **8**, doi:10.1128/mBio.00373-17 (2017).
666 14 Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human
667 coronavirus-EMC. *Nature* **495**, 251-254, doi:10.1038/nature12005 (2013).
668 15 Park, J. E. *et al.* Proteolytic processing of Middle East respiratory syndrome coronavirus
669 spikes expands virus tropism. *Proc Natl Acad Sci U S A* **113**, 12262-12267,
670 doi:10.1073/pnas.1608147113 (2016).
671 16 Li, W. *et al.* Identification of sialic acid-binding function for the Middle East respiratory
672 syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci U S A* **114**, E8508-E8517,
673 doi:10.1073/pnas.1712592114 (2017).
674 17 Li, F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral*
675 *Res* **100**, 246-254, doi:10.1016/j.antiviral.2013.08.014 (2013).
676 18 Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol*
677 **3**, 237-261, doi:10.1146/annurev-virology-110615-042301 (2016).

19 Li, F., Li, W., Farzan, M. & Harrison, S. C. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* **309**, 1864-1868, doi:10.1126/science.1116480 (2005).

20 Li, W. *et al.* Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J* **24**, 1634-1643, doi:10.1038/sj.emboj.7600640 (2005).

21 Ma, C. *et al.* Broad host tropism of ACE2-using MERS-related coronaviruses and determinants restricting viral recognition. *Cell Discov* **9**, 57, doi:10.1038/s41421-023-00566-8 (2023).

22 Xiong, Q. *et al.* Close relatives of MERS-CoV in bats use ACE2 as their functional receptors. *Nature* **612**, 748-757, doi:10.1038/s41586-022-05513-3 (2022).

23 Ma, C. *et al.* Identification of ACE2 as the Entry Receptor for Two Novel European Bat Merbecoviruses. *biorrXiv*.

24 Menachery, V. D. *et al.* Trypsin Treatment Unlocks Barrier for Zoonotic Bat Coronavirus Infection. *J Virol* **94**, doi:10.1128/JVI.01774-19 (2020).

25 Letko, M., Marzi, A. & Munster, V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology*, doi:10.1038/s41564-020-0688-y (2020).

26 Guo, H. *et al.* Isolation of ACE2-dependent and -independent sarbecoviruses from Chinese horseshoe bats. *J Virol* **97**, e0039523, doi:10.1128/jvi.00395-23 (2023).

27 Guo, H. *et al.* ACE2-Independent Bat Sarbecovirus Entry and Replication in Human and Bat Cells. *mBio* **13**, e0256622, doi:10.1128/mbio.02566-22 (2022).

28 Khaledian, E. *et al.* Sequence determinants of human-cell entry identified in ACE2-independent bat sarbecoviruses: A combined laboratory and computational network science approach. *EBioMedicine* **79**, 103990, doi:10.1016/j.ebiom.2022.103990 (2022).

29 Seifert, S. N. *et al.* An ACE2-dependent Sarbecovirus in Russian bats is resistant to SARS-CoV-2 vaccines. *PLoS Pathog* **18**, e1010828, doi:10.1371/journal.ppat.1010828 (2022).

30 Becker, M. M. *et al.* Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. *Proc Natl Acad Sci U S A* **105**, 19944-19949, doi:10.1073/pnas.0808116105 (2008).

31 Almazan, F. *et al.* Engineering a replication-competent, propagation-defective Middle East respiratory syndrome coronavirus as a vaccine candidate. *mBio* **4**, e00650-00613, doi:10.1128/mBio.00650-13 (2013).

32 Mou, H. *et al.* The receptor binding domain of the new Middle East respiratory syndrome coronavirus maps to a 231-residue region in the spike protein that efficiently elicits neutralizing antibodies. *J Virol* **87**, 9379-9383, doi:10.1128/JVI.01277-13 (2013).

33 Scobey, T. *et al.* Reverse genetics with a full-length infectious cDNA of the Middle East respiratory syndrome coronavirus. *Proc Natl Acad Sci U S A* **110**, 16157-16162, doi:10.1073/pnas.1311542110 (2013).

34 Han, X. *et al.* Structure of the S1 subunit C-terminal domain from bat-derived coronavirus HKU5 spike protein. *Virology* **507**, 101-109, doi:10.1016/j.virol.2017.04.016 (2017).

35 Wang, N. *et al.* Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4. *Cell Res* **23**, 986-993, doi:10.1038/cr.2013.92 (2013).

36 Roelle, S. M., Shukla, N., Pham, A. T., Bruchez, A. M. & Matreyek, K. A. Expanded ACE2 dependencies of diverse SARS-like coronavirus receptor binding domains. *PLoS Biol* **20**, e3001738, doi:10.1371/journal.pbio.3001738 (2022).

37 Lau, S. K. P. *et al.* Isolation of MERS-related coronavirus from lesser bamboo bats that uses DPP4 and infects human-DPP4-transgenic mice. *Nat Commun* **12**, 216, doi:10.1038/s41467-020-20458-9 (2021).

38 Millet, J. K. & Whittaker, G. R. Host cell entry of Middle East respiratory syndrome coronavirus after two-step, furin-mediated activation of the spike protein. *Proc Natl Acad Sci U S A* **111**, 15214-15219, doi:10.1073/pnas.1407087111 (2014).
39 Millet, J. K. & Whittaker, G. R. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res* **202**, 120-134, doi:10.1016/j.virusres.2014.11.021 (2015).
40 Glowacka, I. *et al.* Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J Virol* **85**, 4122-4134, doi:10.1128/JVI.02232-10 (2011).
41 Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* **181**, 271-280 e278, doi:10.1016/j.cell.2020.02.052 (2020).
42 Shirato, K., Kawase, M. & Matsuyama, S. Middle East respiratory syndrome coronavirus infection mediated by the transmembrane serine protease TMPRSS2. *J Virol* **87**, 12552-12561, doi:10.1128/JVI.01890-13 (2013).
43 Shirato, K., Kawase, M. & Matsuyama, S. Wild-type human coronaviruses prefer cell-surface TMPRSS2 to endosomal cathepsins for cell entry. *Virology* **517**, 9-15, doi:10.1016/j.virol.2017.11.012 (2018).
44 Matsuyama, S. *et al.* Middle East Respiratory Syndrome Coronavirus Spike Protein Is Not Activated Directly by Cellular Furin during Viral Entry into Target Cells. *J Virol* **92**, doi:10.1128/JVI.00683-18 (2018).
45 Wang, Z. *et al.* Loss of furin site enhances SARS-CoV-2 spike protein pseudovirus infection. *Gene* **856**, 147144, doi:10.1016/j.gene.2022.147144 (2023).
46 Hoffmann, M., Kleine-Weber, H. & Pohlmann, S. A Multibasic Cleavage Site in the Spike Protein of SARS-CoV-2 Is Essential for Infection of Human Lung Cells. *Mol Cell* **78**, 779-784 e775, doi:10.1016/j.molcel.2020.04.022 (2020).
47 Bestle, D. *et al.* TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Sci Alliance* **3**, doi:10.26508/lsa.202000786 (2020).
48 Fukuma, A. *et al.* Inability of rat DPP4 to allow MERS-CoV infection revealed by using a VSV pseudotype bearing truncated MERS-CoV spike protein. *Arch Virol* **160**, 2293-2300, doi:10.1007/s00705-015-2506-z (2015).
49 Petit, C. M. *et al.* Genetic analysis of the SARS-coronavirus spike glycoprotein functional domains involved in cell-surface expression and cell-to-cell fusion. *Virology* **341**, 215-230, doi:10.1016/j.virol.2005.06.046 (2005).
50 Chen, H. Y. *et al.* Cytoplasmic Tail Truncation of SARS-CoV-2 Spike Protein Enhances Titer of Pseudotyped Vectors but Masks the Effect of the D614G Mutation. *J Virol* **95**, e0096621, doi:10.1128/JVI.00966-21 (2021).
51 Adney, D. R. *et al.* Replication and shedding of MERS-CoV in upper respiratory tract of inoculated dromedary camels. *Emerg Infect Dis* **20**, 1999-2005, doi:10.3201/eid2012.141280 (2014).
52 Alagaili, A. N. *et al.* Middle East respiratory syndrome coronavirus infection in dromedary camels in Saudi Arabia. *MBio* **5**, e00884-00814, doi:10.1128/mBio.00884-14 (2014).
53 Chu, D. K. *et al.* MERS coronaviruses in dromedary camels, Egypt. *Emerg Infect Dis* **20**, 1049-1053, doi:10.3201/eid2006.140299 (2014).
54 Chu, D. K. *et al.* Middle East respiratory syndrome coronavirus (MERS-CoV) in dromedary camels in Nigeria, 2015. *Euro Surveill* **20**, doi:10.2807/1560-7917.ES.2015.20.49.30086 (2015).
55 Woo, P. C. *et al.* Molecular diversity of coronaviruses in bats. *Virology* **351**, 180-187, doi:10.1016/j.virol.2006.02.041 (2006).

56 van Boheemen, S. *et al.* Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* **3**, doi:10.1128/mBio.00473-12 (2012).
57 Lau, S. K. P. *et al.* Identification of a Novel Betacoronavirus (Merbecovirus) in Amur Hedgehogs from China. *Viruses* **11**, doi:10.3390/v11110980 (2019).
58 Monchatre-Leroy, E. *et al.* Identification of Alpha and Beta Coronavirus in Wildlife Species in France: Bats, Rodents, Rabbits, and Hedgehogs. *Viruses* **9**, doi:10.3390/v9120364 (2017).
59 Corman, V. M. *et al.* Characterization of a novel betacoronavirus related to middle East respiratory syndrome coronavirus in European hedgehogs. *J Virol* **88**, 717-724, doi:10.1128/JVI.01600-13 (2014).
60 Kiesslich, S. & Kamen, A. A. Vero cell upstream bioprocess development for the production of viral vectors and vaccines. *Biotechnol Adv* **44**, 107608, doi:10.1016/j.biotechadv.2020.107608 (2020).
61 Yeager, C. L. *et al.* Human aminopeptidase N is a receptor for human coronavirus 229E. *Nature* **357**, 420-422, doi:10.1038/357420a0 (1992).
62 Peck, K. M. *et al.* Glycosylation of mouse DPP4 plays a role in inhibiting Middle East respiratory syndrome coronavirus infection. *J Virol* **89**, 4696-4699, doi:10.1128/JVI.03445-14 (2015).
63 van Doremalen, N. *et al.* Host species restriction of Middle East respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. *J Virol* **88**, 9220-9232, doi:10.1128/JVI.00676-14 (2014).
64 Letko, M. *et al.* Adaptive Evolution of MERS-CoV to Species Variation in DPP4. *Cell Rep* **24**, 1730-1737, doi:10.1016/j.celrep.2018.07.045 (2018).
65 Pallesen, J. *et al.* Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc Natl Acad Sci U S A* **114**, E7348-E7357, doi:10.1073/pnas.1707304114 (2017).
66 Zhou, T. *et al.* Cryo-EM Structures of SARS-CoV-2 Spike without and with ACE2 Reveal a pH-Dependent Switch to Mediate Endosomal Positioning of Receptor-Binding Domains. *Cell Host Microbe* **28**, 867-879 e865, doi:10.1016/j.chom.2020.11.004 (2020).
67 Costello, S. M. *et al.* The SARS-CoV-2 spike reversibly samples an open-trimer conformation exposing novel epitopes. *Nat Struct Mol Biol* **29**, 229-238, doi:10.1038/s41594-022-00735-5 (2022).
68 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, doi:10.1038/s41564-020-0771-4 (2020).
69 Starr, T. N. *et al.* ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature*, doi:10.1038/s41586-022-04464-z (2022).
70 Ma, C.-B. *et al.* Multiple independent acquisitions of ACE2 usage in MERS-related coronaviruses. *biorxiv* (2024).
71 Earnest, J. T. *et al.* The tetraspanin CD9 facilitates MERS-coronavirus entry by scaffolding host cell receptors and proteases. *PLoS Pathog* **13**, e1006546, doi:10.1371/journal.ppat.1006546 (2017).
72 Zhang, S. *et al.* Bat and pangolin coronavirus spike glycoprotein structures provide insights into SARS-CoV-2 evolution. *Nat Commun* **12**, 1607, doi:10.1038/s41467-021-21767-3 (2021).
73 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, doi:10.1038/s41594-020-0468-7 (2020).
74 Wrobel, A. G. *et al.* Evolution of the SARS-CoV-2 spike protein in the human host. *Nat Commun* **13**, 1178, doi:10.1038/s41467-022-28768-w (2022).

75 Calvaresi, V. *et al.* Structural dynamics in the evolution of SARS-CoV-2 spike glycoprotein. *Nat Commun* **14**, 1421, doi:10.1038/s41467-023-36745-0 (2023).
76 Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310 e1220, doi:10.1016/j.cell.2020.08.012 (2020).
77 Song, X. *et al.* Cryo-EM analysis of the HCoV-229E spike glycoprotein reveals dynamic prefusion conformational changes. *Nat Commun* **12**, 141, doi:10.1038/s41467-020-20401-y (2021).
78 Yuan, Y. *et al.* Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains. *Nat Commun* **8**, 15092, doi:10.1038/ncomms15092 (2017).
79 Toelzer, C. *et al.* Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein. *Science* **370**, 725-730, doi:10.1126/science.abd3255 (2020).
80 Yan, R. *et al.* Structural basis for the different states of the spike protein of SARS-CoV-2 in complex with ACE2. *Cell Res* **31**, 717-719, doi:10.1038/s41422-021-00490-0 (2021).
81 Tse, A. L. *et al.* Distinct pathway for evolution of enhanced receptor binding and cell entry in SARS-like bat coronaviruses. *bioRxiv*, doi:10.1101/2024.06.24.600393 (2024).
82 Lau, S. K. P. *et al.* Replication of MERS and SARS coronaviruses in bat cells offers insights to their ancestral origins. *Emerg Microbes Infect* **7**, 209, doi:10.1038/s41426-018-0208-9 (2018).
83 Xia, L. Y. *et al.* Isolation and characterization of a pangolin-borne HKU4-related coronavirus that potentially infects human-DPP4-transgenic mice. *Nat Commun* **15**, 1048, doi:10.1038/s41467-024-45453-2 (2024).
84 Ogando, N. S. *et al.* SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J Gen Virol* **101**, 925-940, doi:10.1099/jgv.0.001453 (2020).
85 Wang, W. *et al.* Airport noise disturbs foraging behavior of Japanese pipistrelle bats. *Ecol Evol* **12**, e8976, doi:10.1002/ece3.8976 (2022).
86 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, doi:10.1093/sysbio/syq010 (2010).
87 Lefort, V., Longueville, J. E. & Gascuel, O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol* **34**, 2422-2424, doi:10.1093/molbev/msx149 (2017).
88 Letko, M., Booiman, T., Kootstra, N., Simon, V. & Ooms, M. Identification of the HIV-1 Vif and Human APOBEC3G Protein Interface. *Cell Rep* **13**, 1789-1799, doi:10.1016/j.celrep.2015.10.068 (2015).
89 Tse, L. V. *et al.* A MERS-CoV antibody neutralizes a pre-emerging group 2c bat coronavirus. *Sci Transl Med* **15**, eadg5567, doi:10.1126/scitranslmed.adg5567 (2023).
90 Liu, K. *et al.* Cross-species recognition of SARS-CoV-2 to bat ACE2. *Proc Natl Acad Sci U S A* **118**, doi:10.1073/pnas.2020216118 (2021).
91 Fan, C. *et al.* Neutralizing monoclonal antibodies elicited by mosaic RBD nanoparticles bind conserved sarbecovirus epitopes. *Immunity* **55**, 2419-2435 e2410, doi:10.1016/j.immuni.2022.10.019 (2022).
92 Chan, K. K. *et al.* Engineering human ACE2 to optimize binding to the spike protein of SARS coronavirus 2. *Science* **369**, 1261-1265, doi:10.1126/science.abc0870 (2020).
93 Barnes, C. O. *et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682-687, doi:10.1038/s41586-020-2852-1 (2020).
94 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol* **152**, 36-51, doi:10.1016/j.jsb.2005.07.007 (2005).

95 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* **14**, 290-296, doi:10.1038/nmeth.4169 (2017).
96 Pettersen, E. F. *et al.* UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci* **30**, 70-82, doi:10.1002/pro.3943 (2021).
97 Ashkenazy, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res* **44**, W344-350, doi:10.1093/nar/gkw408 (2016).
98 Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res* **33**, W299-302, doi:10.1093/nar/gki370 (2005).

FIGURES and FIGURE LEGENDS

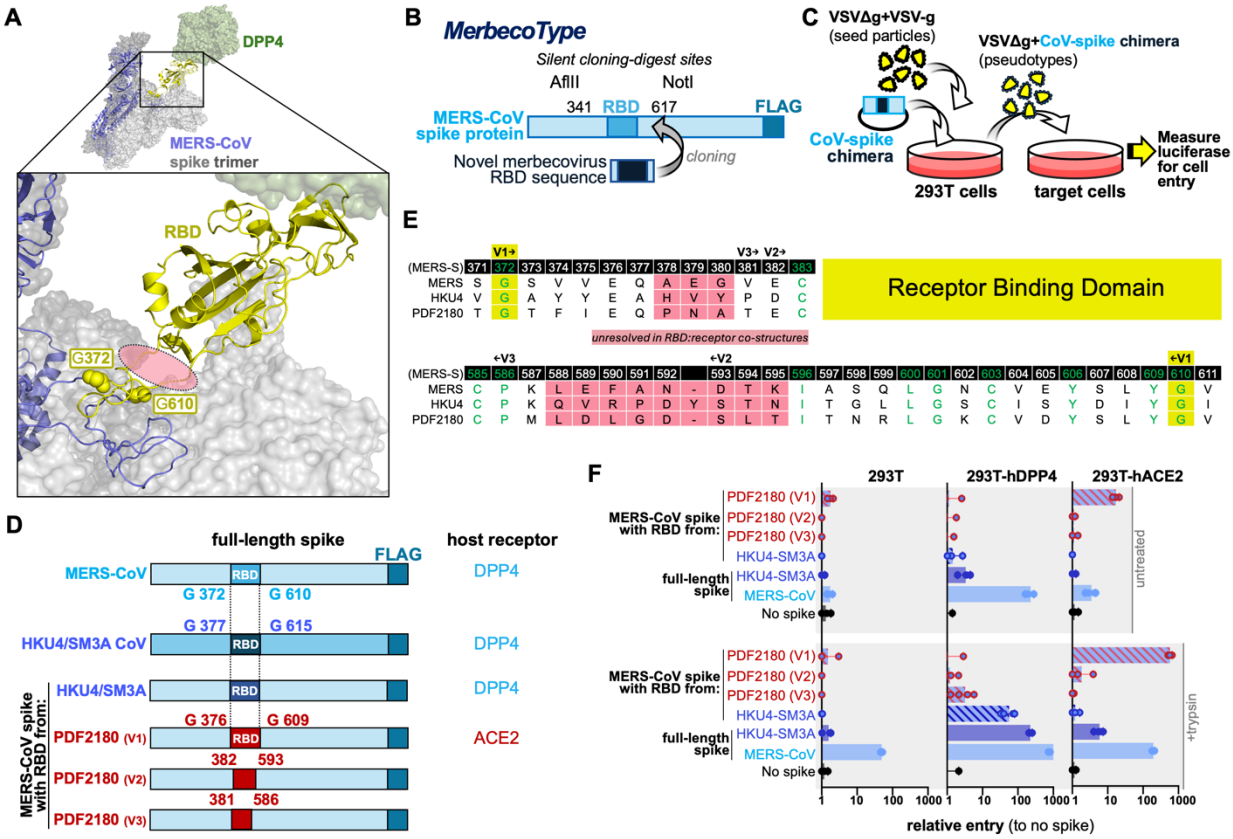


Figure 1. Identification of the functional RBD in merbecoviruses. (A) The receptor binding domain in merbecoviruses engages the host receptor, DPP4 (PDB ID: 4L72, 5X59). Conserved glycines marking the functional RBD boundaries are indicated in yellow spheres. **(B)** Overview of MERS-CoV spike design for RBD replacement. **(C)** Chimeric spikes are used to generate single-cycle pseudotyped VSV reporter particles. **(D)** Overview of chimeric spike designs. **(E)** Amino acid sequence alignment of region flanking the RBD in merbecoviruses. **(F)** 293T cell lines expressing indicated receptors were infected with pseudotypes and luciferase was measured for entry. Cells were infected in quadruplicate. For each graph, individual replicates are plotted as points, the mean is shown as a bar and lines indicate standard deviation.

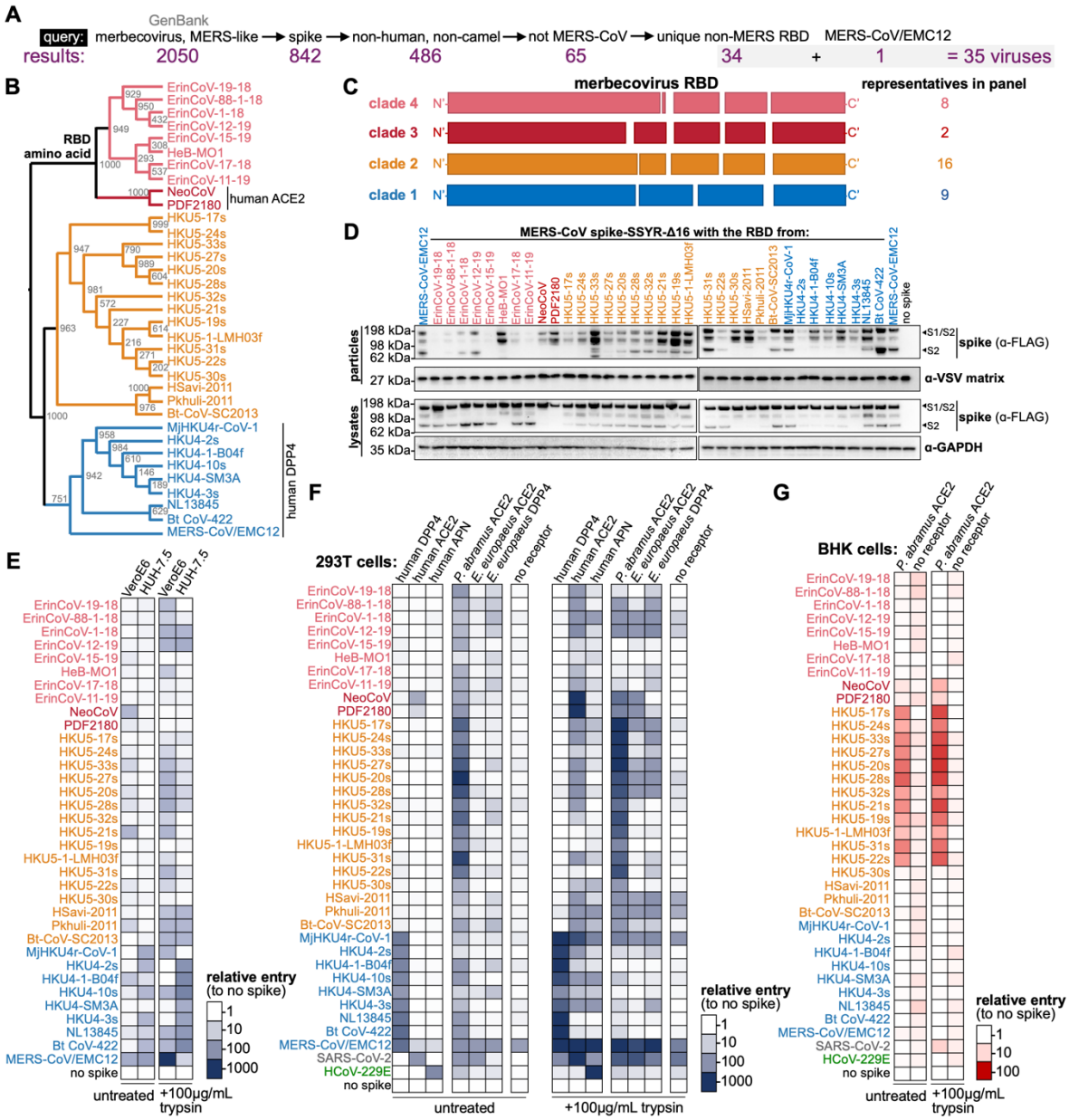


Figure 2. Merbecovirus RBDs sort into clades based on sequence and entry phenotype. **(A)** Total sequence diversity in merbecoviruses encompasses approximately 34 RBD sequences. **(B)** Cladogram showing the relationship of RBD amino acid sequences. Known receptors indicated. **(C)** Schematic overview of RBD clades, indicating the approximate size and location of indels characteristic to each merbecovirus RBD clade. **(D)** Western blot for FLAG (spike) expression of chimeric spikes. **(E)** VeroE6 cells or Huh-7.5 cells were infected with VSV pseudotype panel bearing indicated chimeric spikes, in quadruplicate. Luciferase was measured for cell entry. **(F)** 293T cells transduced to express human APN, human DPP4 or human ACE2 or transfected with *Erinaceus europaeus* ACE2, *Erinaceus europaeus* DPP4 or *Pipistrellus abramus* ACE2 were infected in the absence or presence of 100μg/mL trypsin. **(G)** BHK-21 cells were transiently transfected with empty vector or *Pipistrellus abramus* ACE2, and infected with or without exogenous trypsin. All cells were infected in quadruplicate. Heatmaps represent mean relative entry as compared to pseudotypes without spike.

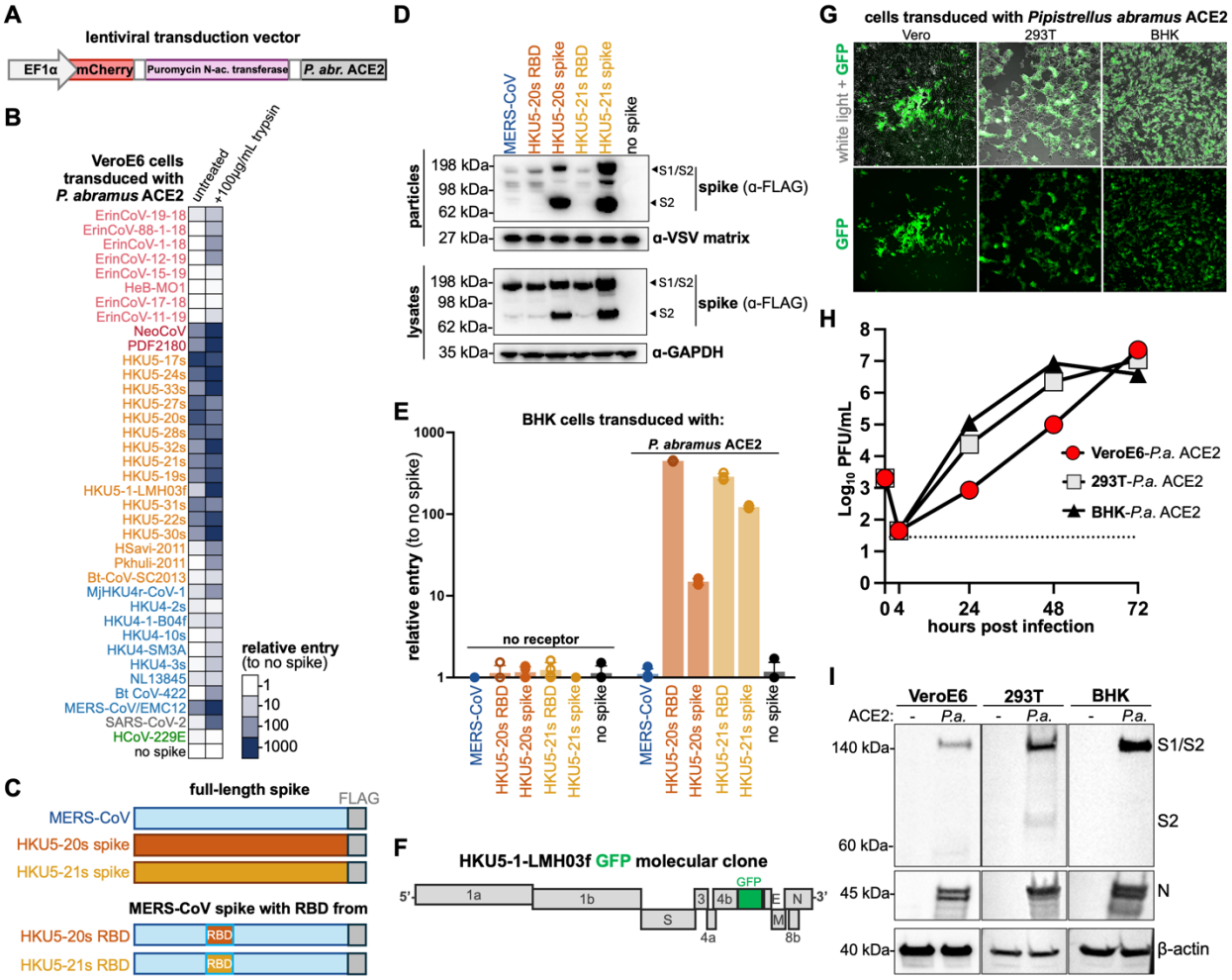


Figure 3. *Pipistrellus abramus* ACE2 rescues entry of full-length HKU5 spike in non-susceptible cells. (A) *Pipistrellus abramus* ACE2 expression cassette used in lentiviral transduction. **(B)** VeroE6 cells were transduced with *Pipistrellus abramus* ACE2 and infected with chimeric spike pseudotypes. **(C)** Schematic overview of full length and chimeric spikes. **(D)** Spike expression in cell lysates and incorporation in pseudotyped particles. **(E)** BHK-21 cells were transduced to express *Pipistrellus abramus* ACE2 and infected with pseudotypes bearing the indicated spikes. Cells were infected in quadruplicate. For each graph, individual replicates are plotted as points, the mean is shown as a bar and lines indicate standard deviation. **(F)** Overview of HKU5 molecular clone with GFP reporter in place of Orf5. **(G)** Light and fluorescent microscopy of indicated cell lines at 24-hours post-infection. **(H)** Growth curve of viral replication as measured by PFU of supernatants at indicated time points. **(I)** Western blot of infected cells for viral and host proteins at 24-hours post-infection.

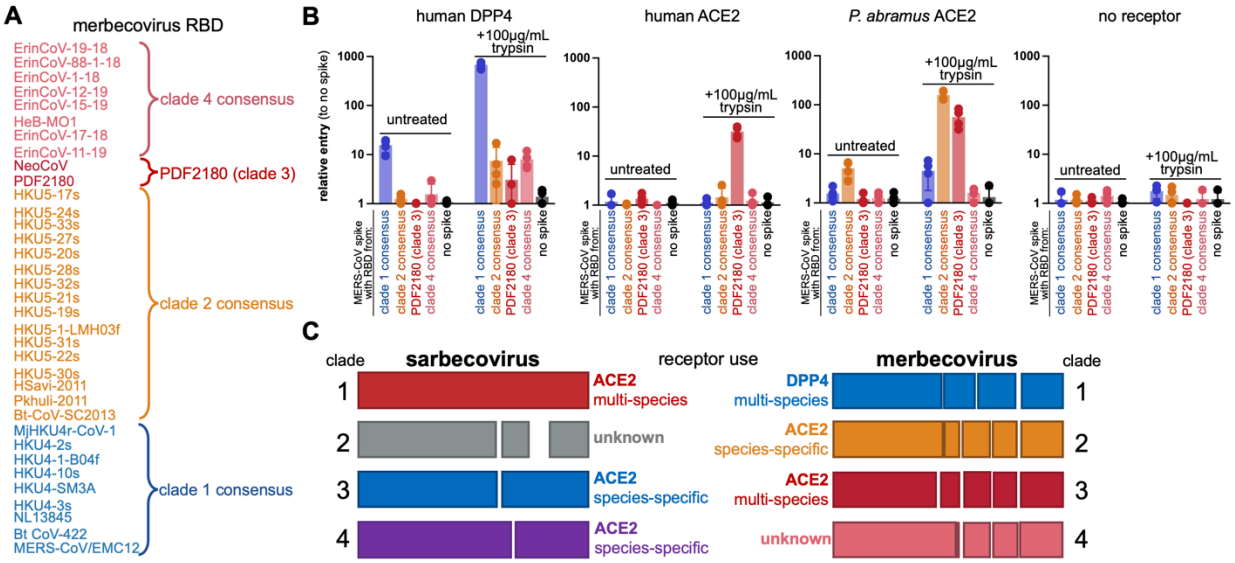


Figure 4. Clade consensus RBDs mimic clade-specific entry (A) Consensus sequences were generated for clades 1, 2, and 4. **(B)** 293T cells transfected with indicated receptors were infected with chimeric spikes bearing clade consensus RBDs. **(C)** Schematic overview of betacoronavirus RBD clades, indel patterns and receptor use.

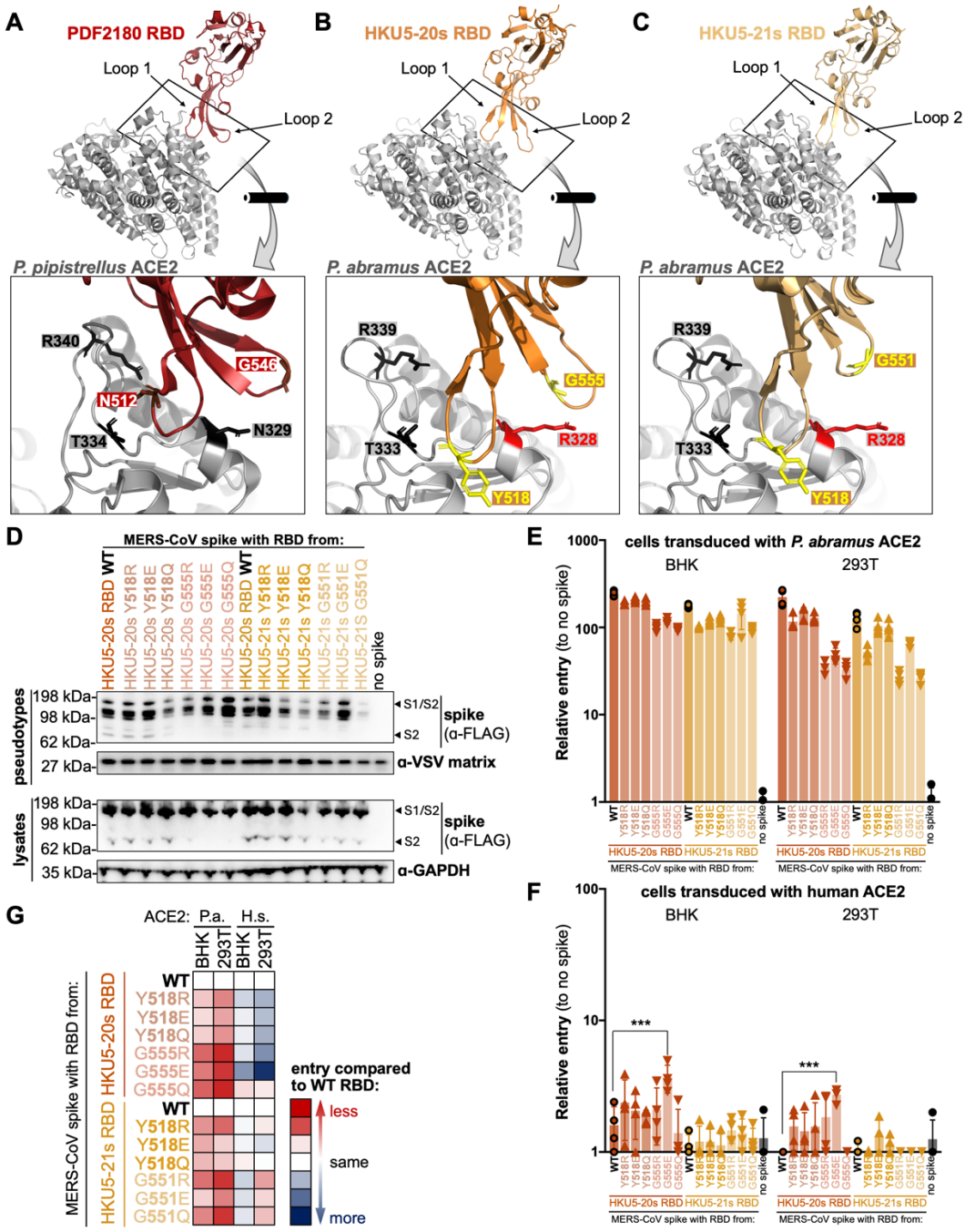


Figure 5. Structure-guided mutagenesis at the HKU5:ACE2 interface. (A) Co-structure for PDF2180 RBD and *Pipistrellus pipistrellus* ACE2 (PDB ID: 7wpz). **(B)** Predicted structures for HKU5-20s RBD and *Pipistrellus abramus* ACE2 (based on PDB IDs: 5xgr and 7c8j, respectively). **(C)** Predicted structure for HKU5-21s RBD and *Pipistrellus abramus* ACE2 (based on PDB ID: 8sak and 7c8j, respectively). **(D)** Mutant spike expression in cell lysates and incorporation in pseudotyped particles. **(E)** BHK or 293T cells transduced with either *Pipistrellus abramus* (P.a.) or **(F)** human (H.s.) ACE2 were infected with indicated pseudotypes. For each graph, four individual replicates are plotted as points, the mean is shown as a bar and lines indicate standard deviation. (***) Indicates P-value <0.001. **(G)** Entry values for mutant RBDs as compared to parental wild-type RBD.

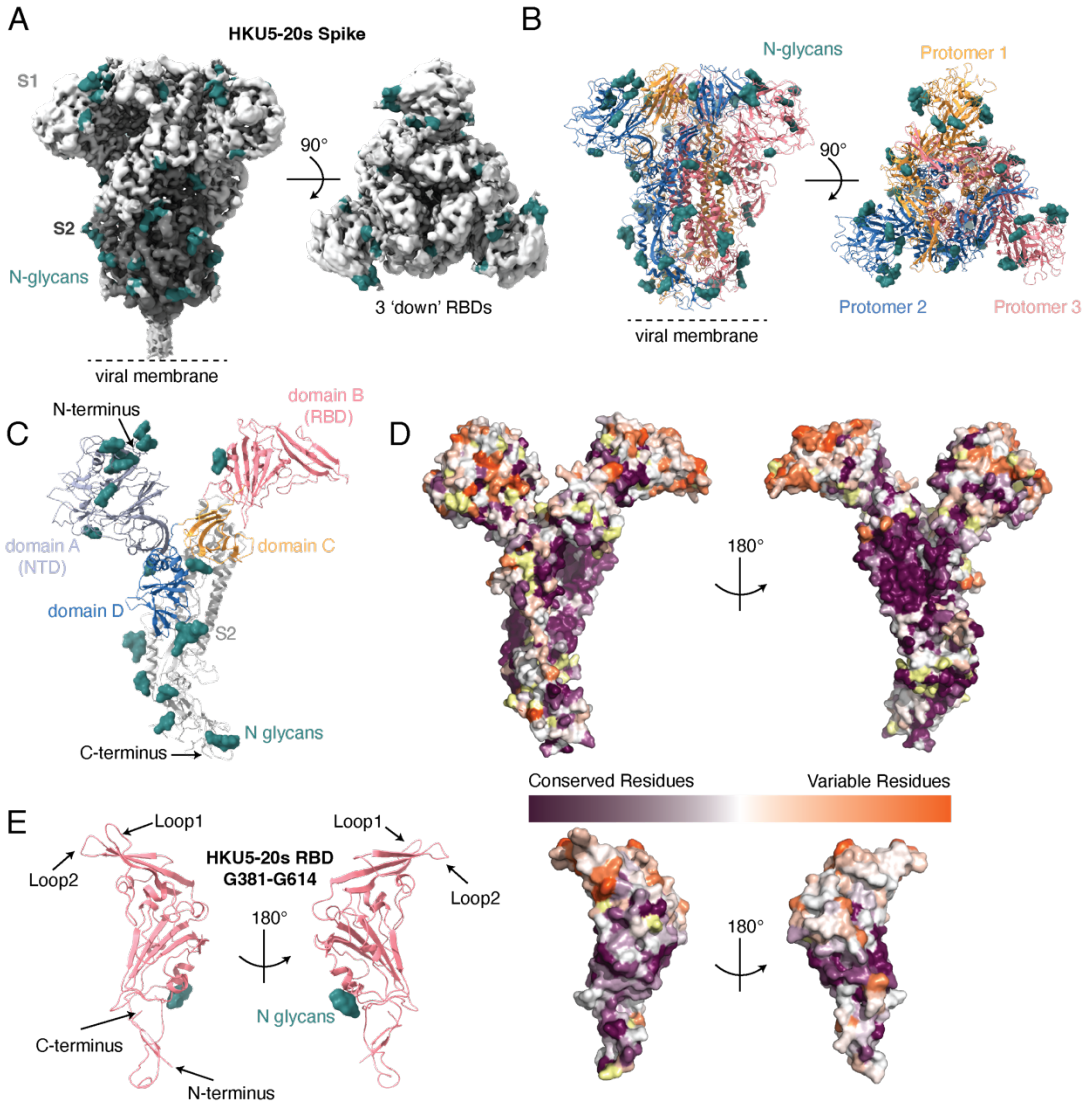


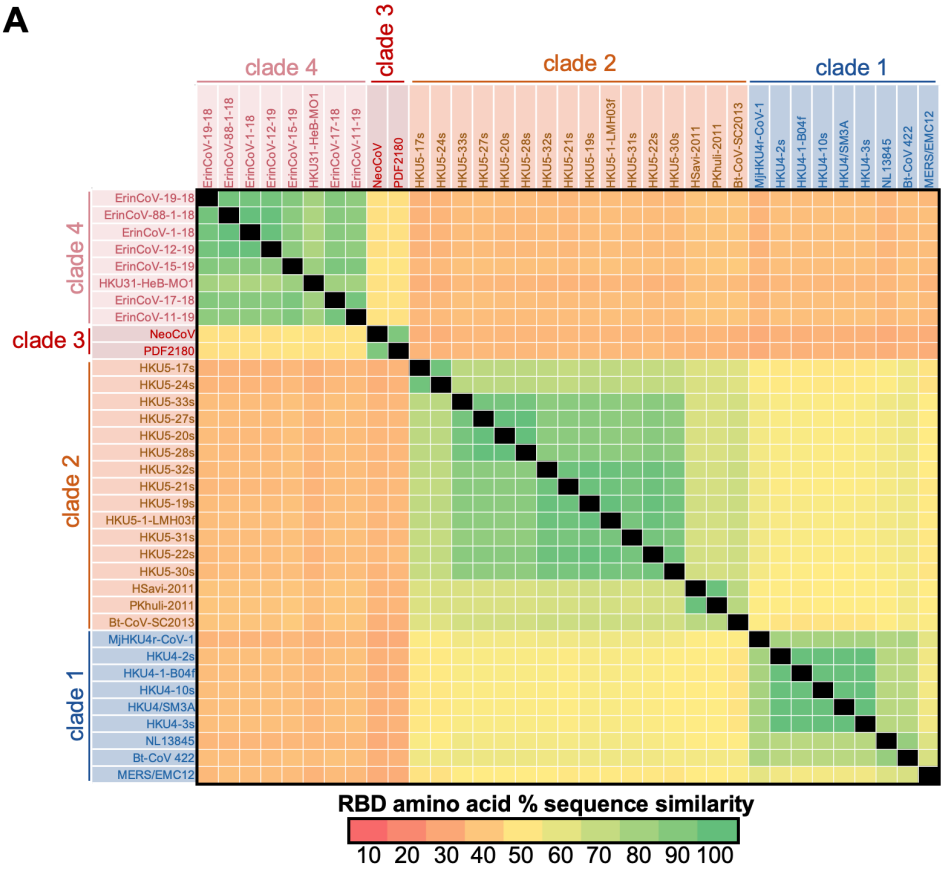
Figure 6. Structure of HKU5-20s spike trimer. (A) 2.8 Å EM density of structure of HKU5-20s spike with three 'down' RBDs is shown from the side (left) and top (right). **(B)** Cartoon diagram of HKU5-20s spike trimer with different colors for protomers shown from the side (left) and top (right). **(C)** Cartoon diagram of spike protomer showing domains A-D. **(D)** Amino acid sequence conservation of 10 merbecovirus spike sequences calculated by ConSurf Database[98] plotted on a surface representation of an HKU5-20s protomer (top) and HKU5-20s RBD (bottom). Yellow patches in the purple to orange sequence conservation gradient indicate than <10% of the sequences were available for the calculation at that position. **(E)** Cartoon diagram of the HKU5-20s RBD from two orientations.

Table 1. Diverse merbecovirus panel encompassing global diversity

	Virus	Accession number	Receptor Clade	Host species	Location
1	ErinCoV-19-18	MW246796	4	*Erinaceus europaeus*	Italy
2	ErinCoV-88-1-18	MW246795	4	*Erinaceus europaeus*	Italy
3	ErinCoV-1-18	MW246797	4	*Erinaceus europaeus*	Italy
4	ErinCoV-12-19	MW246800	4	*Erinaceus europaeus*	Italy
5	ErinCoV-15-19	MW246802	4	*Erinaceus europaeus*	Italy
6	HKU31-HeB-MO1	OM451213	4	*Erinaceus amurensis*	China
7	ErinCoV-17-18	MW246798	4	*Erinaceus europaeus*	Italy
8	ErinCoV-11-19	MW246799	4	*Erinaceus europaeus*	Italy
9	NeoCoV	KC869678	3	*Neoromicia capensis*	South Africa
10	PDF2180	KX574227	3	*Pipistrellus hesperidus*	Uganda
11	HKU5-17s	AGP04930	2	*Pipistrellus abramus*	Hong Kong
12	HKU5-24s	AGP04936	2	*Pipistrellus abramus*	Hong Kong
13	HKU5-33s	KC522104	2	*Pipistrellus abramus*	Hong Kong
14	HKU5-27s	AGP04938	2	*Pipistrellus abramus*	Hong Kong
15	HKU5-20s	AGP04933	2	*Pipistrellus abramus*	Hong Kong
16	HKU5-28s	AGP04939	2	*Pipistrellus abramus*	Hong Kong
17	HKU5-32s	AGP04942	2	*Pipistrellus abramus*	Hong Kong
18	HKU5-21s	AGP04934	2	*Pipistrellus abramus*	Hong Kong
19	HKU5-19s	AGP04932	2	*Pipistrellus abramus*	Hong Kong
20	HKU5-1-LMH03f	NC009020	2	*Pipistrellus abramus*	China
21	HKU5-31s	AGP04941	2	*Pipistrellus abramus*	Hong Kong
22	HKU5-22s	AGP04935	2	*Pipistrellus abramus*	Hong Kong
23	HKU5-30s	AGP04940	2	*Pipistrellus abramus*	Hong Kong
24	HSavi-2011	MG596802	2	*Hypsugo savii*	Italy
25	PKhuli-2011	MG596803	2	*Pipistrellus kuhlii*	Italy
26	Bt-CoV-SC2013	KJ473821	2	*Vespertilio superans*	China
27	MjHKU4r-CoV-1	OQ786862	1	*Manis javanica*	China (Smuggled Malayan pangolins)
28	HKU4-2s	AGP04916	1	*Tylonycteris pachypus*	Hong Kong
29	HKU4-1-B04f	NC009019	1	*Tylonycteris pachypus*	Hong Kong
30	HKU4-10s	AGP04924	1	*Tylonycteris pachypus*	Hong Kong
31	HKU4/SM3A	MW218395	1	*Tylonycteris pachypus*	Hong Kong
32	HKU4-3s	AGP04917	1	*Tylonycteris pachypus*	Hong Kong
33	NL13845	MG021451	1	*Tylonycteris pachypus*	China
34	Bt-CoV 422	MG021452	1	*Tylonycteris pachypus*	China
35	MERS/EMC12	JX869059	1	*Homo sapiens*	Saudi Arabia

Supplemental Information

A



B

NeoCoV RBD residues that interact with *P. pipistrellus* ACE2

		LOOP 1									LOOP 2						
	NeoCoV aa:	**504**	**505**	**506**	**508**	**509**	**510**	**511**	**512**	**513**	**532**	**539**	**540**	**546**	**547**	**548**	**550**
Clade 3	NeoCoV	N	C	N	G	A	T	N	K	D	N	L	W	G	V	P	R
	PDF2180	N	C	N	G	G	F	N	K	D	N	V	W	G	V	P	R
Clade 2	HKU5-17s	E	C	Y	L	G	Y	S	A	K	Q	Q	S	P	S	G	F
	HKU5-24s	E	C	Y	Q	G	H	S	S	K	Q	Q	S	P	S	R	F
	HKU5-33s	E	C	Y	F	T	A	Y	G	K	S	Q	S	P	D	G	F
	HKU5-27s	E	C	Y	G	T	A	Y	G	K	S	Q	S	P	D	G	F
	HKU5-20s	E	C	Y	S	T	A	Y	G	K	S	Q	S	P	D	G	F
	HKU5-28s	E	C	Y	S	T	A	Y	G	K	S	Q	S	P	D	G	F
	HKU5-32s	E	C	Y	T	S	A	Y	G	K	R	Q	S	-	-	-	-
	HKU5-21s	E	C	Y	T	S	A	Y	G	K	R	Q	S	-	-	-	-
	HKU5-19s	E	C	Y	T	S	A	Y	G	K	R	Q	S	-	-	-	-
	HKU5-1-LMH03f	E	C	Y	T	S	A	Y	G	K	R	Q	S	-	-	-	-
	HKU5-31s	E	C	Y	T	T	E	Y	G	K	R	Q	S	-	-	-	-
	HKU5-22s	E	C	Y	T	S	A	Y	G	K	R	Q	S	-	-	-	-
	HKU5-30s	E	C	Y	I	S	A	Y	G	K	R	Q	S	-	-	-	-
	Hsavi-2011	Q	C	Y	F	S	A	Y	S	D	S	Q	T	N	-	-	K
	Pkhuli-2011	Q	C	Y	F	S	A	Y	S	D	S	Q	T	N	-	-	K

Figure S1. Sequence Identity matrix of merbecovirus RBDs. (A) Amino acid sequences for each spike RBD were aligned and compared. Black squares indicate identical residue. **(B)** Comparison of clade 3 RBD residues that interact with *P. pipistrellus* ACE2 with clade 2 sequences.

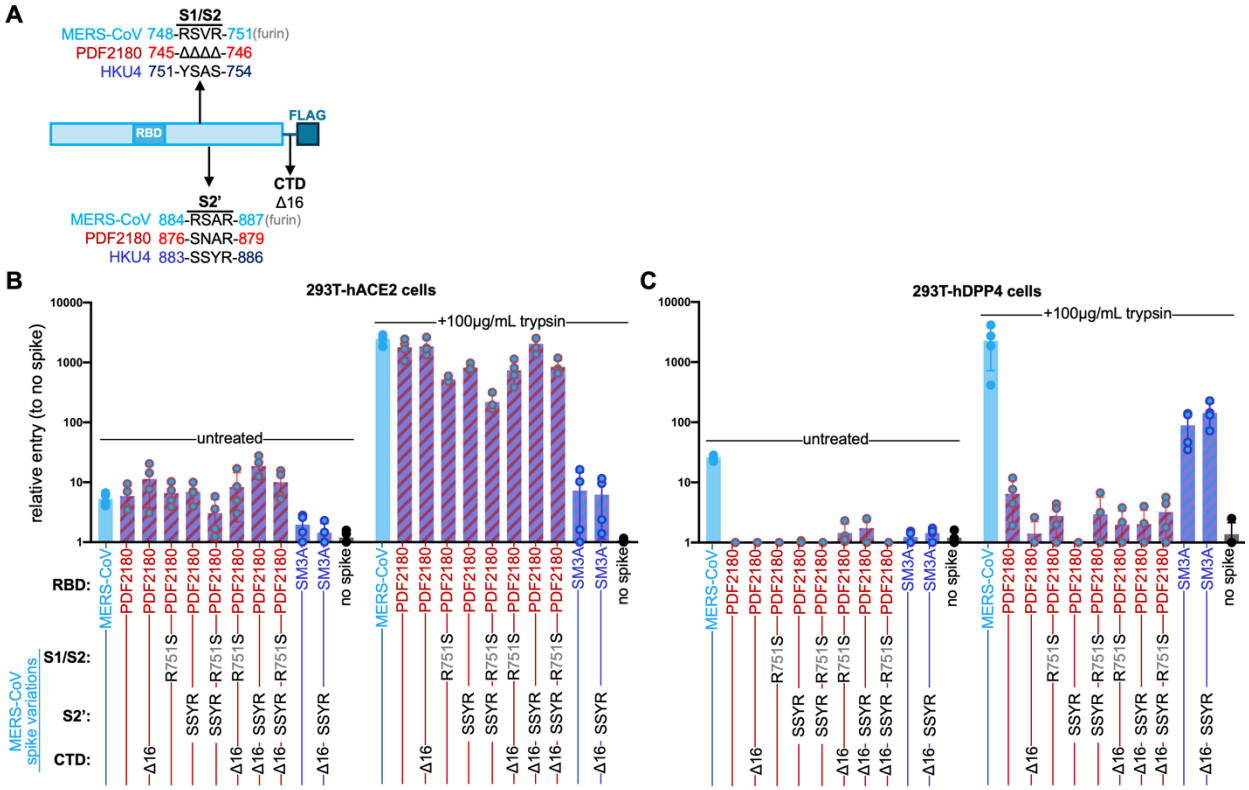


Figure S2. Optimizing MERS-CoV spike for MerbecoType. (A) Overview of S1/S2 and S2' furin sites in MERS-CoV spike. **(B)** 293T cells transduced with human ACE2 or **(C)** human DPP4 were infected with indicated spikes and luciferase was measured for entry. Cells were infected in quadruplicate. For each graph, individual replicates are plotted as points, the mean is shown as a bar and lines indicate standard deviation.

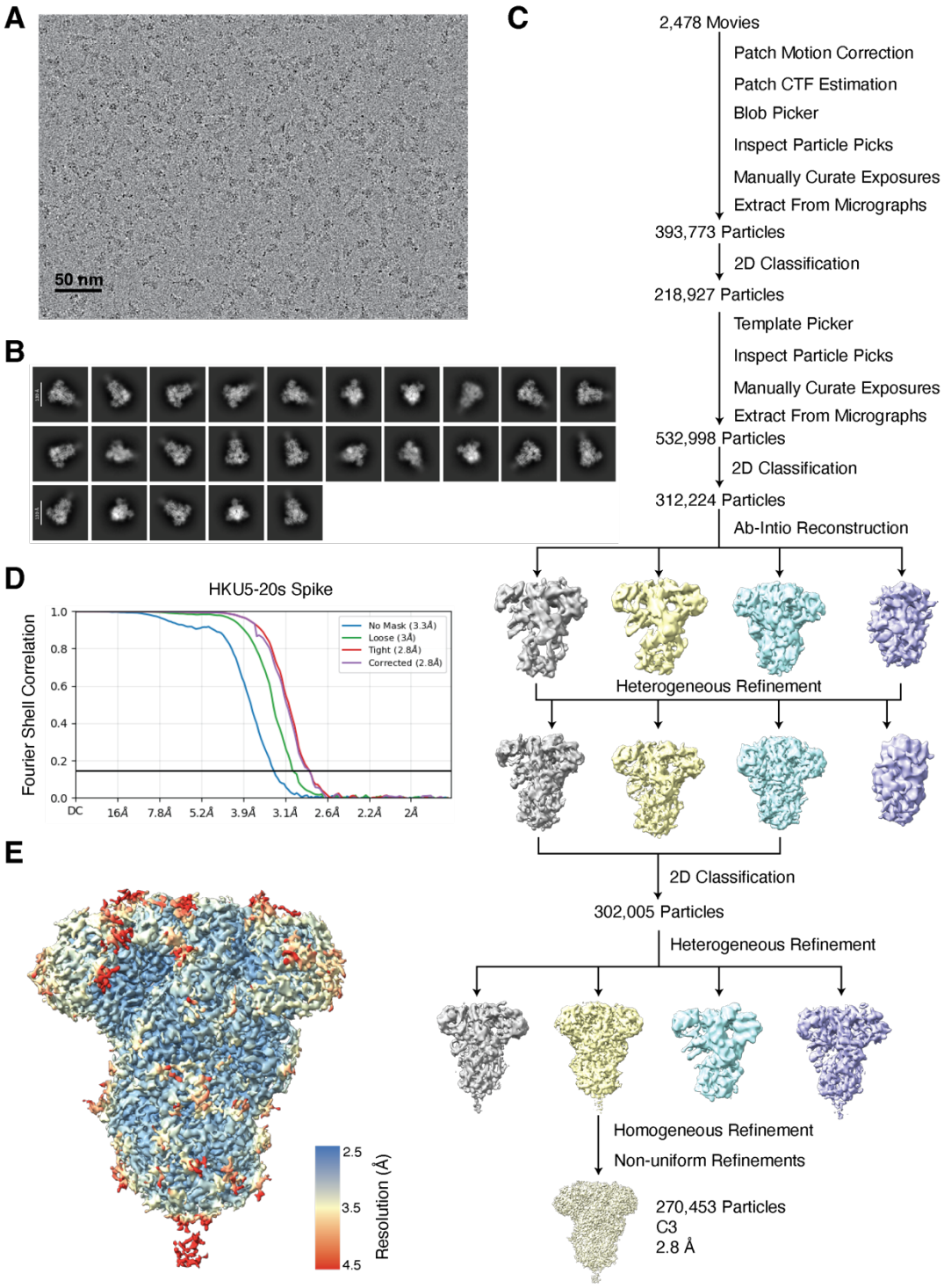


Figure S3. Cryo-EM workflow. (A) Representative micrograph. **(B)** Selected 2D classes. **(C)** Workflow of data processing. **(D)** Fourier shell correlation (FSC) plot of the final reconstruction. (E) Final reconstruction of HKU5-20s spike, colored from blue to red for local resolutions from high to low, respectively.

Data Collection and processing	
Microscope	Talos Arctica
Camera	Gatan K3
Magnification	45,000x
Voltage (keV)	200
Exposure (e/Å^2)	60
Pixel size (Å)	0.869
Defocus Range (μm)	-1 to -3
Initial Particle Images (no.)	532,998
Final Particle Images (no.)	270,453
Symmetry Imposed	C3
Map Resolution (Å)	2.8
FSC Threshold	0.143
Map Resolution Range (Å)	2.8-3.3
Refinement	
Initial Model Used	PDB 6NB3
Model Resolution (Å)	3.5
FSC Threshold	0.143
Model composition	
non-hydrogen atoms	28,440
protein residues	3,548
ligands	63
Average B-factors (Å^2)	
protein	60.51
ligands	96.61
R.M.S. deviations	
Bond length (Å)	0.002
Bond angles (°)	0.526
Validation	
MolProbity score	1.6
Clashscore	6.7
Rotamer outliers (%)	1.3
Ramachandran plot	
Ramachandran favored (%)	97.4
Ramachandran allowed (%)	2.6
Ramachandran outliers (%)	0
PDB ID	9D4T
EMDB ID	EMD-46569

TABLE S1. Cryo-EM statistics for HKU5-20s Spike structure