Functional assessment of cell entry and receptor use for merbecoviruses

Michael Letko[1#]

1. Paul G. Allen School for Global Health, Washington State University, Pullman, WA, 99163.

[#]Corresponding author:

michael.letko@wsu.edu

ABSTRACT:

The merbecovirus subgenus of coronaviruses includes Middle East Respiratory Syndrome Coronavirus (MERS-CoV), which is a zoonotic respiratory pathogen that transmits from dromedary camels to humans and causes severe respiratory disease. Viral discovery efforts have uncovered hundreds of merbecoviruses in different species across multiple continents, but few of these viruses have been isolated or studied under laboratory conditions, leaving basic questions regarding their threat to humans unresolved. Viral entry into host cells is considered an early and critical step for transmission between hosts. In this study, a scalable approach to assessing novel merbecovirus cell entry was developed and used to measure receptor use across the entire merbecovirus subgenus. Merbecoviruses are sorted into four clades based on the receptor binding domain of the spike glycoprotein. Receptor tropism is clade-specific, with only one clade using DPP4 and multiple clades using ACE2, including the entire HKU5 cluster of bat coronaviruses.

53 **INTRODUCTION:**

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

63 In order to transmit to between species, coronaviruses must be able to enter the cells of
64 their hosts. MERS-CoV cell entry is mediated by the viral spike protein and occurs in three steps:
65 (1) spike attaches virus to the cell surface, (2) binds to the host cell receptor molecule, (3) and
66 exposes structural elements in the base of the spike protein that mediate membrane fusion and
67 subsequent cell entry[14-16]. Each function is performed by a separate region of the spike protein,
68 with the Receptor Binding Domain (RBD) forming the core interaction with the host receptor[17-20].
69 While MERS-CoV uses the receptor dipeptidyl peptidase IV (DPP4) to infect cells, several studies
70 have shown that some bat merbecoviruses do not use DPP4 and instead use angiotensin
71 converting enzyme 2 (ACE2) to infect their hosts[21-23].

72 Similar to other beta-coronaviruses, the receptor binding domain (RBD) in merbecovirus
73 spike proteins is a single, contiguous domain located in the N-terminal half of the spike sequence,
74 and folds as a structurally stable unit that engages with the host receptor (Fig. 1a-b). Gene
75 synthesis is required to produce fragments of viral genomes that have not been isolated, with
76 sequence length dictating labor and cost. To reduce cost of viral gene synthesis while maintaining
77 the ability to study viral protein function, we previously employed a method to replace the RBD of
78 the sarbecovirus, SARS-CoV, with RBDs from any other sarbecovirus[24]. The resulting chimeric

79 spike gene was used in viral pseudotype assays to measure the entry efficiency in cell culture of
80 a panel of sarbecovirus RBDs[24]. Using this approach (called "SarbecoType"), we previously
81 characterized the cell entry for the majority of all published sarbecovirus RBDs, which lead to
82 isolation of new sarbecoviruses, identification of understudied cell entry pathways, and improved
83 understanding for molecular determinants of sarbecovirus cell entry[24-28].

84 Here, a similar approach called "MerbecoType" was taken with the merbecovirus
85 subgenus of betacoronaviruses. Similar to sarbecoviruses, RBDs of merbecoviruses can be
86 classified into four different "clades" based on the presence or absence of indels, and exhibit
87 clade-specific trends in receptor use and cell entry. Surprisingly, at least half of the merbecovirus
88 RBD clades can use ACE2, with the entire group of HKU5 viruses exhibiting clear preference for
89 ACE2 from their natural host species: *Pipistrellus ambramus* bats. These findings will be important
90 for future efforts to identify looming coronavirus threats and in the development of interventions
91 such as broad-spectrum antivirals and universal coronavirus vaccines.
92
93
94
95
96
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 receptor binding domains from diverse merbecoviruses in viral
108 pseudotype entry assays, the first objective was to produce a MERS-CoV spike expression
109 construct that would allow for easy replacement of the receptor binding domain (RBD) with the
110 RBD from other merbecoviruses (Fig. 1a-c). MERS-CoV was selected as the spike backbone
111 because of its extensive structural characterization and compatibility with cell-culture based
112 models of merbecovirus entry. Similar to our prior sarbecovirus study, the MERS-CoV spike was
113 codon optimized for human cells, appended with a C-terminal FLAG tag and silent mutations were
114 introduced outside of the RBD (Fig. 1b)[24]. While MERS-CoV uses DPP4 for entry, ACE2 has been
115 shown to be the receptor for some other merbecoviruses including PDF2180 (Fig. 1d)[22]. Thus the
116 RBD from PDF2180 made an ideal candidate to develop a chimeric MERS-CoV spike: a
117 successful exchange of DPP4 use in MERS-CoV spike to ACE2 was considered indicative of a
118 functional spike that would allow for further exploration of merbecovirus RBD receptor
119 preferences. Unfortunately, there are very few amino acid motifs that are conserved across all
120 merbecovirus spike sequences, complicating the task of identifying a functional region common
121 to all merbecoviruses. However, closer inspection of published spike sequences revealed several
122 highly conserved glycine residues flanking the RBD (MERS-CoV spike positions 372 and 610;
123 Fig. 1a, d, e). These two residues are just beyond the RBD region of spike used for co-crystal
124 structures with host receptor molecules. While the RBD has been determined structurally for some
125 merbecoviruses, regions flanking the RBD are flexible, difficult to resolve with x-ray
126 crystallography and not included in RBD:receptor based co-structures (Fig. 1a, c; pink shaded
127 region). Several chimeras between MERS-CoV and PDF2180 were produced: version 1 included
128 the region of PDF2180 spike between the conserved glycines, version 2 was based on structure
129 and version 3 was based on the PDF2180-RBD as defined previously[22] (Fig. 1e). DNA fragments
130 encoding the different RBD fragments were cloned into the MERS-CoV spike backbone and used

131 to generate vesicular stomatitis virus (VSV) single-cycle pseudotypes (Fig. 1c). While Baby
132 Hamster Kidney cells (BHKs) are not naturally susceptible to coronavirus entry but can support
133 entry if provided the receptor, BHKs expressing human ACE2 do not always recapitulate natural
134 viral entry for ACE2-dependent viruses with low affinity for the receptor[28,29]. Therefore, 293T,
135 which exhibit low susceptibility to coronavirus entry, were stably transduced to express human
136 ACE2 or DPP4. Infecting these stable cell lines with pseudotypes bearing chimeric spikes,
137 revealed that the chimeric spike containing the entire region of PDF2180 spike between the
138 conserved glycines was able to use human ACE2, indicating a successful receptor exchange for
139 MERS-CoV spike (Fig. 1d, f; PDF2180 V1). The regions of PDF2180 spike between residues 376-
140 381 and 586-609 were essential for receptor-use exchange, as two other spikes with shorter
141 regions of the PDF2180 RBD were unable to use ACE2 (Fig. 1a, d-f; PDF2180 V2, V3). Addition
142 of exogenous protease (trypsin) enhanced viral entry, allowing more sensitive detection of low
143 affinity interactions. Importantly, trypsin-enhanced entry was still receptor dependent (Fig. 1f).
144 Production of a chimeric spike containing an analogous region from the DPP4-dependent
145 HKU4/SM3A virus showed a retention in DPP4 use in MERS-CoV spike further suggesting this
146 region would allow for assessment of diverse receptor preferences (Fig. 1e-f).
147
148 ***Optimizing MERS-CoV spike for pseudotyping***
149 With a working receptor-exchange in the spike construct, the next goal was to optimize the spike
150 backbone for pseudotype efficiency, which would allow for improved sensitivity to low-level viral
151 entry. After engaging with the host receptor, coronavirus spikes are processed by host-cell
152 proteases to release the spike fusion peptide and mediate membrane fusion. While the TMPRSS-
153 family of cell-surface proteases and endosomal cathepsins have been shown to cleave many
154 different coronavirus spikes, including MERS-CoV, SARS-CoV and SARS-CoV-2, furin is another
155 protease that plays a role both spike-biogenesis and cell entry, depending on the availability of an
156 appropriate cleavage site[30-33]. MERS-CoV spike contains two furin sites, at the S1/S2 and S2'

157 junctions, although the role of furin in entry is less clear as removal of these sites has no effect
158 on viral replication *in vitro*[34,35]. Because previous studies have shown that removal of the furin site
159 from SARS-CoV-2 spike results in increased pseudotype efficiency, several versions of spike were
160 produced with either one or both sites mutated[36-38]. Additionally, C-terminal truncations in MERS-
161 CoV, SARS-CoV and SARS-CoV-2 spikes have shown to increase cell-surface expression of
162 spike in transfected cells, resulting in improved pseudotype efficiency[39-41]. Therefore, the protease
163 site mutants were also tested with or without a 16-amino acid truncation in the C-terminal end of
164 MERS-CoV spike (Extended Data Fig. 1a-c). Surprisingly, these modifications to spike had very
165 little effect on entry in 293T cells for MERS-CoV WT, PDF2180 RBD or HKU4/SM3A RBD
166 (Extended Data Fig. 1b-c). While not statistically significant, spikes containing a C-terminal
167 truncation and removal of the S2' furin site performed slightly better than the other spikes,
168 therefore, this set of modifications was included for further experiments.
169
170 ***A panel of merbecoviruses encompassing total RBD diversity***
171 The 2012 MERS-CoV outbreak lead to an influx of coronavirus sequencing and discovery
172 research, with sequences from hundreds of MERS-CoV patients as well as dromedary camels
173 after they were discovered as the intermediate species[42-45]. Genbank has over 2000 entries for
174 "merbecovirus" and "MERS-like" viruses; of these 2000 entries, only approximately 842 entries
175 contain full-length spike and of those entries, only approximately half are from species other than
176 humans and camels (Figure 2A). Unfortunately, coronavirus genbank listings do not always
177 contain universal identifiers, which left many entries for "MERS-CoV" even after excluding human
178 and camel viruses. Further eliminating viruses titled as "MERS-CoV" from the list resulted 65
179 spike sequences which encompassed 36 unique RBD sequences (Fig. 2a; table 1). Some
180 merbecoviruses are likely missing from this list as coronavirus nomenclature from 10-15 years
181 ago when the majority of these merbecovirus genomes were published was not well defined.
182 Phylogenetic analysis of the RBD amino acid sequences (as defined between the two glycine

183 residues established in figure 1) revealed several genetically clustered groups based on
184 characteristic indels and the sequences surrounding them (Fig. 2b, c). The groups were labeled
185 as clades 1 through 4, based on the order of their discovery: clade 1 includes MERS and HKU4
186 discovered between 2006-2012[1,46,47], clade 2 includes HKU5 also discovered around 2006[46],
187 clade 3 incudes the African merbecoviruses discovered in 2017[13] and clade 4 includes the
188 contemporary hedgehog merbecoviruses discovered in europe[13] (Fig. 2b, c, table 1). The RBDs
189 for these merbecoviruses was synthesized and cloned into the MERS-CoV spike backbone
190 plasmid. All chimeric spike plamsmids were stably expressed in 293T cells used to produce
191 pseudotyped viral-like particles for downstream experiments (Fig. 2d).
192
193 ***Merbecovirus entry in common laboratory cell lines***
194 Vero cells are common laboratory cell lines in coronavirus research and are used in attempts to
195 isolate and propagate viruses, including merbecoviruses. VeroE6 cells were poorly susceptible to
196 the merbecovirus RBD panel, with addition of trypsin only mildly improving entry for some of the
197 clade 2 and 4 viruses (Fig. 3a). In contrast, Huh-7.5 cells were susceptible to clade 1
198 merbecovirus RBDs (Fig. 3b). Notably, addition of trypsin improved entry for some of the clade 4
199 viruses for both cell lines (eg, ErinCov-12-19) (Fig. 3).
200
201 ***Merbecovirus compatibility with known human coronavirus receptors in human cells***
202 The diverse panel of merbecovirus pseudotypes was tested on 293T cell lines stably transduced
203 to express the known human coronavirus receptors: amino peptidase N (APN), dipeptidyl
204 peptidase IV (DPP4), or Angiotensin converting enzyme-2 (ACE2). In the absence of trypsin,
205 untransduced 293T cells exhibited poor susceptibility to the entire panel and APN cells were only
206 susceptible to entry with the positive control: human coronavirus 229E, which is known to use
207 APN (Fig. 4a, Extended Data Fig. 2a-b). Cells transduced with human DPP4 exhibited clear
208 susceptibility to clade 1 merbecovirus RBDs but not the other clades (Figure 4A, Extended Data

209 Fig. 2c), while cells expressing human ACE2 exhibited low level susceptibility to SARS-CoV-2
210 spike and clade 3 merbecovirus RBDs (Fig. 4a, Extended Data Fig. 2d). Exogenous trypsin
211 moderately increased entry in untransduced 293Ts for human coronaviruses MERS-CoV, which
212 uses human DPP4 and SARS-CoV-2, which uses human ACE2, suggesting low levels of the
213 human receptors were present on 293Ts (Extended Data Fig. s2a). To reduce the background
214 signal and asses only the increase in entry resulting from the transduced receptor, the relative
215 entry signal was normalized against the no-receptor cell line (Fig. 4). In the presence of trypsin,
216 DPP4 only mediated clade 1 entry, while ACE2 cells exhibited low susceptibility to clade 1, 2, and
217 4 and high susceptibility to clade 3 merbecovirus RBDs (Fig. 4b). APN cells remained susceptible
218 to HCoV-229E entry in the presence of trypsin. While entry signal under 10x is not considered as
219 robust in this assay, there were a large number of viruses that seemed to have low level ACE2
220 compatibility (Fig. 4b; Extended Data Fig. 2d).
221
222 ***Species-specific receptor use across the merbecovirus RBD clades***
223 Human cells expressing human ACE2 exhibited low level susceptibility to a wide range of
224 merbeocviruses. To investigate if these viruses are compatible with receptors from their cognate
225 host species, ACE2 and DPP4 was synthesized from European hedgehogs (*Erinaceus*
226 *europaeus*), the natural host species for the ErinCoVs (table 1; merbecovirus RBD clade 4), and
227 ACE2 was synthesized from Japanese house bats (*Pipistrellus abramus*): the natural host species
228 for the HKU5 viruses (table 1; merbecovirus RBD clade 2). 293T cells transfected with hedgehog
229 DPP4 or ACE2 were not particularly susceptible to any of the clades, with or without trypsin
230 (Figure 5). However, 293T transfected with ACE2 from Japanese house bats exhibited clear
231 susceptibility to the entire HKU5 cluster of clade 2 merbecovirus RBDs – with or without trypsin
232 (Fig. 5; Extended data Fig. 2). In a confirmatory assay, *Pipstrellus abramus* ACE2 specifically
233 mediated only HKU5 RBD entry in BHK cells, which are otherwise not susceptible to entry with
234 any ACE2 or DPP4-compatible betacoronavirus (Fig. 6). Taken together, this clear and HKU5-

235 specificific entry strongly suggest ACE2 is the natural receptor for the HKU5 viruses (Fig. 5 and
236 6).
237
238
239 **DISCUSSION:**
240 Compared to sarbecoviruses, which share 90% nucleotide sequence similarity and above,
241 the diversity in merbecoviruses is higher, initially complicating the chimeric spike approach utilized
242 in this study (Fig. 1). Wide variability between merbecovirus host species (hedgehogs, bats,
243 camels, humans and pangolins) also poses unforeseen challenges when studying cross-species
244 interactions. The chimeric spike approach presented here aims to eliminate some of the
245 interspecies challenges because using the human-compatible MERS-CoV spike in experiments
246 performed in human cells should reduce off-target phenotypes relating to cell attachment and
247 protease processing. Addition of exogenous protease further reduces entry phenotypes not
248 related to the RBD. Importantly, multiple studies have now shown that chimeric coronavirus spikes
249 faithfully re-create entry of both full spikes and wild-type virus[24-29,48].
250 Merbecovirus RBDs can be classified into four clades based on sequence, characteristic
251 indels, and entry phenotype (Fig. 2; Extended Data Fig. 4, table 1): clade 1 viruses can use DPP4
252 from multiple species, clade 2 viruses have species-specific ACE2 preference, and clade 3 can
253 use ACE2 from a wider range of host species. No clear entry was observed with clade 4 viruses
254 and known receptors, but nevertheless, human-cell entry was observed with ErinCoVs in the
255 presence of trypsin (Fig. 3-5, Extended Data Fig. 2-3). Because clade 4 entry seemed most
256 apparent when ACE2 was provided, it is possible these viruses use ACE2 from their host species
257 and that the approach taken in this study is not suitable for detecting their entry. For example,
258 hedgehogs have at least two reported transcript variants for ACE2 and three transcript variants
259 for DPP4 – only one of each was tested here (see methods). Additionally, while the chimeric spike
260 helps to reduce unintended entry barriers, barriers likely remain, including receptor compatibility

261 with other critical host cell-membrane factors, which is difficult to anticipate or control for[49].
262 Alternatively, clade 4 merbecovirus RBDs may use a receptor that is neither DPP4 nor ACE2,
263 analogous to sarbecovirus RBD clade 2 viruses, which have also can infect human cells through
264 an unidentified route (Extended Data Fig. 4)[24-27].

265 MERS-CoV spike can rapidly adapt to sub-optimal interactions with DPP4[50] and the
266 ongoing SARS-CoV-2 pandemic underscores the ability of spike to adapt and correct sub-optimal
267 ACE2 interactions while simultaneously escaping the immune response[51-54]. Other research has
268 shown that clade 3 merbecoviruses gain even broader host ACE2 tropism with single point
269 mutations at 305 and 510 – with the mutation T510F significantly increasing tropism for human
270 ACE2[21,22]. Thus, even low-level receptor compatibility observed for the merbecoviruses is
271 concerning and underscores the need for broadly protective coronavirus vaccines.

272 HKU5 viruses were first sequenced in 2006, shown to be DPP4-independent in 2014, and
273 demonstrated to replicate in vero cells in the presence of exogenous trypsin in 2019[12,46,55]. A
274 previous study identified ACE2 as the receptor for the clade 3 merbecoviruses such as PDF2180
275 and closely-related NeoCoV, but was unable to identify the receptor for the clade 2 HKU5
276 viruses[22]. However, this earlier study assessed the ability of HKU5 to interact with ACE2 from
277 *Pipistrellus pipistrellus* bats and several other Yangochiropteran bats, which are not the true host-
278 species for HKU5. The clear preference of HKU5 for *Pipistrellus abramus* ACE2, and the
279 simultaneous poor use of *Pipistrrellus abramus* ACE2 use from other clade 2 RBDs further
280 demonstrates the tight species specificity found within clade 2 RBDs (Fig. 5 and 6). *Pipistrellus*
281 *abramus* ACE2 varies from previously tested *P. pipistrellus* and *P. khuli* ACE2s at residues 328
282 and 329, in close proximity to the interaction previously reported with NeoCoV/PDF2180 RBDs[22].
283 Similarly, some sarbecoviruses are capable of using ACE2 from diverse species, while other
284 sarbecoviruses can only interact with ACE2 from their own cognate species (Extended Data Fig.
285 4)[29,48].

286 A clade 1 RBD pangolin merbecovirus has been shown to use human DPP4 and was
287 isolated in Vero81 cells[56]. While Vero cells are frequently used in virus isolation attempts, data
288 from this study suggests that Huh-7.5 cells may be a more suitable cell line for isolating DPP4-
289 dependent merbecoviruses and perhaps even some clade 4 merbecoviruses. Further, MERS-
290 CoV grown in Vero cells has been shown to quickly accumulate tissue-culture adaptations in the
291 spike protein driving the need for alternative cell culture models[57].
292 Merbecoviruses are yet another type of beta-coronavirus found in wildlife with potential
293 human cell compatibility and a possibility to cause severe disease in humans. Japanese house
294 bats (*Pipistrellus abramus*) are the natural host for HKU5 coronaviruses and are synanthropic,
295 roosting in human dwellings and structures, and therefore pose repeat opportunities for human
296 exposure[46,58]. Luckily, clade 2 merbecovirus RBDs exhibit low human ACE2 compatibility,
297 although efforts should still be made to reduce contact with these animals (Fig. 4, Extended Data
298 Fig. 2). Developing universal vaccines capable of thwarting multiple viruses is a daunting but
299 essential challenge. The work presented here will serve as a guide to understanding
300 merbecovirus cell entry and subsequent spillover risk.
301
302
303
304
305
306
307
308
309
310
311

312 **METHODS:**

313 Cells

314 293T, VeroE6, Huh-7.5 and BHK-21 cells were maintained in DMEM supplemented with 10%
315 FBS, penicillin-streptomycin, and L-glutamine and maintained at 37°C, 5% CO_2. Cells stably
316 expressing coronavirus receptors were previously described[28].

317

318 Plasmids

319 MERS-CoV/EMC12 spike (accession number JX869059) was codon optimized, appended with
320 C-terminal FLAG tag, and silent mutations were introduced near spike amino acids 341 and 617
321 to generate AflII and NotI restriction digest sites. RBDs were codon optimized and synthesized
322 (IDT DNA) as previously described for sarbecoviruses[24]. Lyophilized RBD fragments were
323 resuspended in water and used in downstream Infusion based cloning (Takara Bio) reactions to
324 assemble full-length chimeric spike proteins. *Erinaceus europaeus* ACE2, transcript variant 1
325 (genbank XM_060183012.1), *Pipistrellus abramus* ACE2 (genbank GQ262782.1) and *Erinaceus*
326 *europaeus* DPP4 (genbank XM_060177812.1) were codon optimized, split in half and
327 synthesized for Infusion based DNA assembly. Splitting genes into multiple parts reduces
328 synthesis time, costs and overall synthesized product complexity. A complete list of accession
329 numbers for the spike panel can be found in table 1.

330

331 Westernblot

332 As described previously, 293T cells transfected with spike plasmids were lysed in 1% SDS lysis
333 buffer and stored at -80°C until use. Lysates were clarified, boiled, reduced and separated on
334 10% Bis-Tris gels in MOPS buffer (Invitrogen NuPage).[24] Membranes were probed with
335 FLAG:HRP conjugated monoclonal antibody (ThermoFisher) or GAPDH monoclonal
336 (ThermoFisher) and reporbed with a secondary goat-anti-mouse:HRP poloyclonal mix
337 (ThermoFisher).

338 Pseudotype entry assay

339 Pseudotype assays were performed as previously described[25,26]. For human ACE2, APN, and
340 DPP4 experiments, 293T cells stably transduced with coronavirus receptors were seeded in black
341 96-well plates and subsequently infected 24 hours with equal volumes of viral pseudotypes.
342 Pseudotypes were combined with either HBSS or HBSS containing trypsin (not TPCK-treated) to
343 achieve a final concentration of 100µg/mL trypsin. Cells were washed once with cold PBS,
344 centrifuged with pseudotypes at 4°C, 1200xG for 1 hour and incubated at 37°C overnight.
345 Luciferase was measured using Brioght-Glo reagent (Promega) approximately 18 hours post-
346 infection. Relative entry was calculated by dividing raw lucifefrase values for each spike by the
347 signal for no-spike pseudotypes. Normalized entry was calculated for the receptors by dividing
348 the relative entry values for each spike and receptor by the signal for cells without receptor.

349

350 Phylogenetic analysis

351 As described previously, the amino acid sequences for spike receptor binding domains aligned
352 using Clustal multiple sequence alignment software and default parameters. A maximum
353 likelihood phylogenetic tree was produced with PhyML v. 3.0[59]. The 'WAG' matrix +G model of
354 amino acid substitution was selected by the Smart Model Selection method with 1000 bootstrap
355 replicates[60]. A tree was then visualized as a cladogram with FigTree v1.4.4
356 (https://github.com/rambaut/figtree).

357

358 Structural depictions

359 MERS-CoV spike RBD bound to DPP4 (PDB ID: 4L72)[61] and the MERS-CoV spike trimer (PDB
360 ID: 5X59)[62] were visualized in PyMol (version 2.4.0).

361

362

363

Acknowledgements

I would like to thank Dr. Stephanie Seifert, Dr. Pamela Bjorkman, Dr. Alex Cohen, Ziyan Wu and Dr. Tyler Starr 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 Number R01AI179720. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health.

390 **REFERENCES:**

391 1 Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A.
392 Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J*
393 *Med* **367**, 1814-1820, doi:10.1056/NEJMoa1211721 (2012).
394 2 Dudas, G., Carvalho, L. M., Rambaut, A. & Bedford, T. MERS-CoV spillover at the
395 camel-human interface. *Elife* **7**, doi:10.7554/eLife.31257 (2018).
396 3 Dudas, G. & Rambaut, A. MERS-CoV recombination: implications about the reservoir
397 and potential for adaptation. *Virus Evol* **2**, vev023, doi:10.1093/ve/vev023 (2016).
398 4 Moreno, A. *et al.* Detection and full genome characterization of two beta CoV viruses
399 related to Middle East respiratory syndrome from bats in Italy. *Virol J* **14**, 239,
400 doi:10.1186/s12985-017-0907-1 (2017).
401 5 Luo, C. M. *et al.* Discovery of Novel Bat Coronaviruses in South China That Use the
402 Same Receptor as Middle East Respiratory Syndrome Coronavirus. *J Virol* **92**,
403 doi:10.1128/JVI.00116-18 (2018).
404 6 Lau, S. K. *et al.* Genetic characterization of Betacoronavirus lineage C viruses in bats
405 reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus
406 HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East
407 respiratory syndrome coronavirus. *J Virol* **87**, 8638-8650, doi:10.1128/JVI.01055-13
408 (2013).
409 7 Ithete, N. L. *et al.* Close relative of human Middle East respiratory syndrome coronavirus
410 in bat, South Africa. *Emerg Infect Dis* **19**, 1697-1699, doi:10.3201/eid1910.130946
411 (2013).
412 8 He, W. T. *et al.* Virome characterization of game animals in China reveals a spectrum of
413 emerging pathogens. *Cell* **185**, 1117-1129 e1118, doi:10.1016/j.cell.2022.02.014 (2022).
414 9 De Sabato, L. *et al.* Can Coronaviruses Steal Genes from the Host as Evidenced in
415 Western European Hedgehogs by EriCoV Genetic Characterization? *Viruses* **12**,
416 doi:10.3390/v12121471 (2020).
417 10 Chen, J. *et al.* A bat MERS-like coronavirus circulates in pangolins and utilizes human
418 DPP4 and host proteases for cell entry. *Cell* **186**, 850-863 e816,
419 doi:10.1016/j.cell.2023.01.019 (2023).
420 11 Wang, Q. *et al.* Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of
421 human receptor CD26. *Cell Host Microbe* **16**, 328-337, doi:10.1016/j.chom.2014.08.009
422 (2014).
423 12 Yang, Y. *et al.* Receptor usage and cell entry of bat coronavirus HKU4 provide insight
424 into bat-to-human transmission of MERS coronavirus. *Proc Natl Acad Sci U S A* **111**,
425 12516-12521, doi:10.1073/pnas.1405889111 (2014).
426 13 Anthony, S. J. *et al.* Further Evidence for Bats as the Evolutionary Source of Middle East
427 Respiratory Syndrome Coronavirus. *MBio* **8**, doi:10.1128/mBio.00373-17 (2017).
428 14 Raj, V. S. *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human
429 coronavirus-EMC. *Nature* **495**, 251-254, doi:10.1038/nature12005 (2013).
430 15 Park, J. E. *et al.* Proteolytic processing of Middle East respiratory syndrome coronavirus
431 spikes expands virus tropism. *Proc Natl Acad Sci U S A* **113**, 12262-12267,
432 doi:10.1073/pnas.1608147113 (2016).
433 16 Li, W. *et al.* Identification of sialic acid-binding function for the Middle East respiratory
434 syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci U S A* **114**, E8508-E8517,
435 doi:10.1073/pnas.1712592114 (2017).
436 17 Li, F. Receptor recognition and cross-species infections of SARS coronavirus. *Antiviral*
437 *Res* **100**, 246-254, doi:10.1016/j.antiviral.2013.08.014 (2013).
438 18 Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol*
439 **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 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).

25 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).

26 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).

27 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).

28 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).

29 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).

30 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).

31 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).

32 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).

33 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).

34 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).

35 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).

36 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).

37 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).

38 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).
39 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).
40 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).
41 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).
42 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).
43 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).
44 Chu, D. K. *et al.* MERS coronaviruses in dromedary camels, Egypt. *Emerg Infect Dis* **20**, 1049-1053, doi:10.3201/eid2006.140299 (2014).
45 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).
46 Woo, P. C. *et al.* Molecular diversity of coronaviruses in bats. *Virology* **351**, 180-187, doi:10.1016/j.virol.2006.02.041 (2006).
47 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).
48 Starr, T. N. *et al.* ACE2 binding is an ancestral and evolvable trait of sarbecoviruses. *Nature*, doi:10.1038/s41586-022-04464-z (2022).
49 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).
50 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).
51 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).
52 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).
53 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).
54 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).
55 Menachery, V. D. *et al.* Trypsin Treatment Unlocks Barrier for Zoonotic Bat Coronavirus Infection. *J Virol* **94**, doi:10.1128/JVI.01774-19 (2020).
56 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).

57 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).
58 Wang, W. *et al.* Airport noise disturbs foraging behavior of Japanese pipistrelle bats. *Ecol Evol* **12**, e8976, doi:10.1002/ece3.8976 (2022).
59 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).
60 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).
61 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).
62 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).

FIGURE LEGENDS

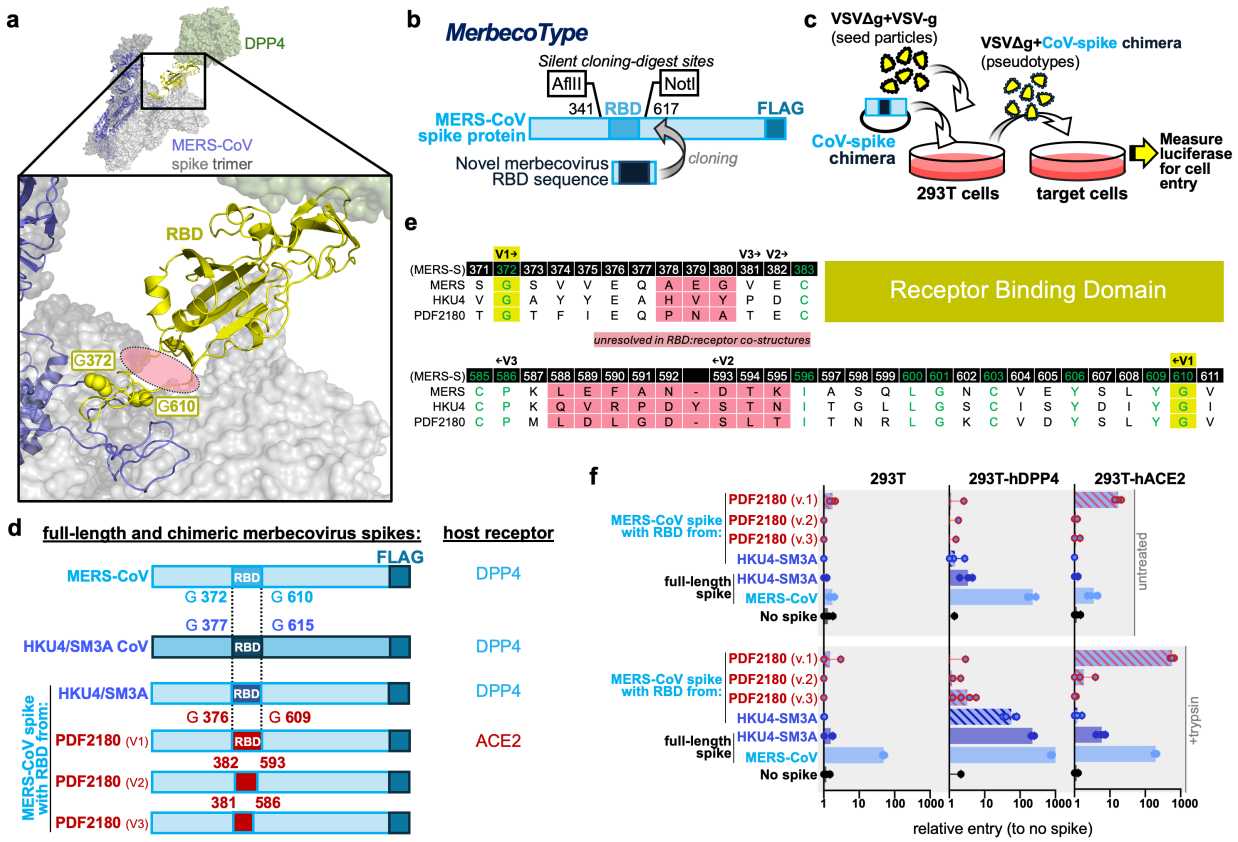
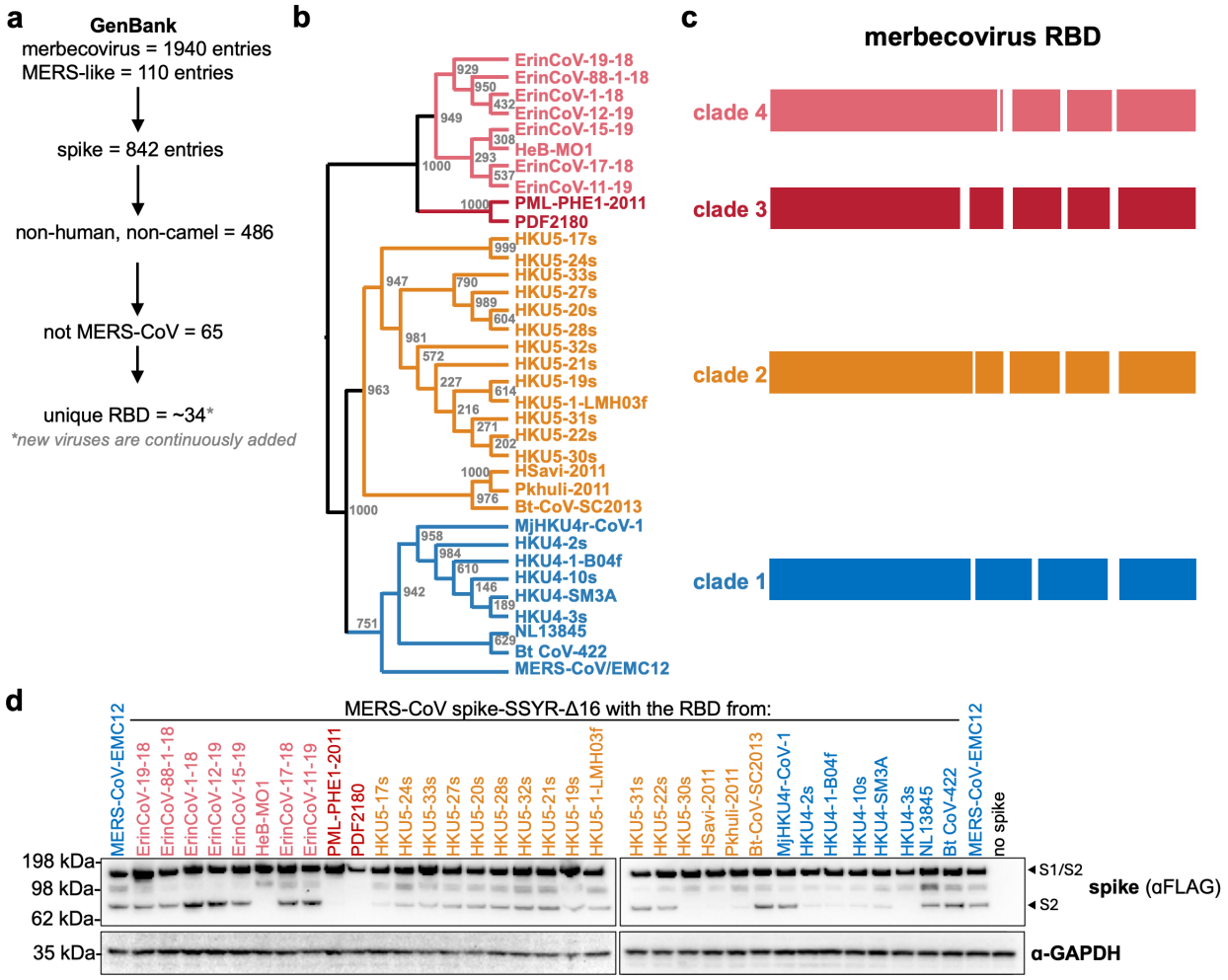


Fig. 1. Identification of the functional RBD in merbecoviruses. (a) The receptor binding domain in merbecoviruses engages the host receptor, DPP4 (PDB ID: 4L72, 5X59). Conserved glycines marking the functional RBD boundaries are indicated in yellow spheres. **(b)** Overview of MERS-CoV spike design for 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.



583

584 **Fig. 2. Diversity in merbecovirus RBDs. (a)** Total sequence diversity in merbecoviruses

585 encompasses approximately 34 RBD sequences **(b)** Cladogram showing the relationship of

586 RBD amino acid sequences **(c)** Schematic overview of RBD clades, indicating the approximate

587 size and location of indels characteristic to each merbecovirus RBD clade. **(d)** Westernblot for

588 FLAG (spike) expression of chimeric spikes.

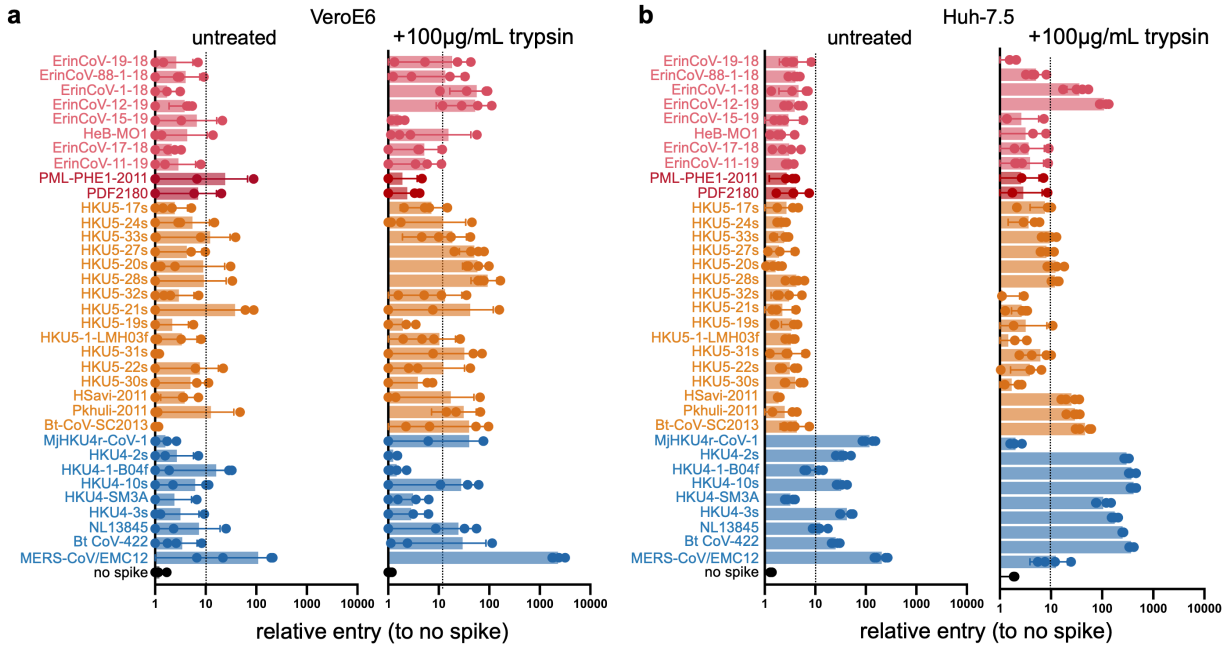


Fig. 3. Susceptibility of common cell lines to merbecovirus entry. (a) VeroE6 cells or **(b)** Huh-7.5 cells were infected with VSV pseudotype panel bearing indicated chimeric spikes, in quadruplicate. Luciferase was measured for cell entry. Dashed line indicates arbitrary reliability cut-off at 10-fold signal over background.

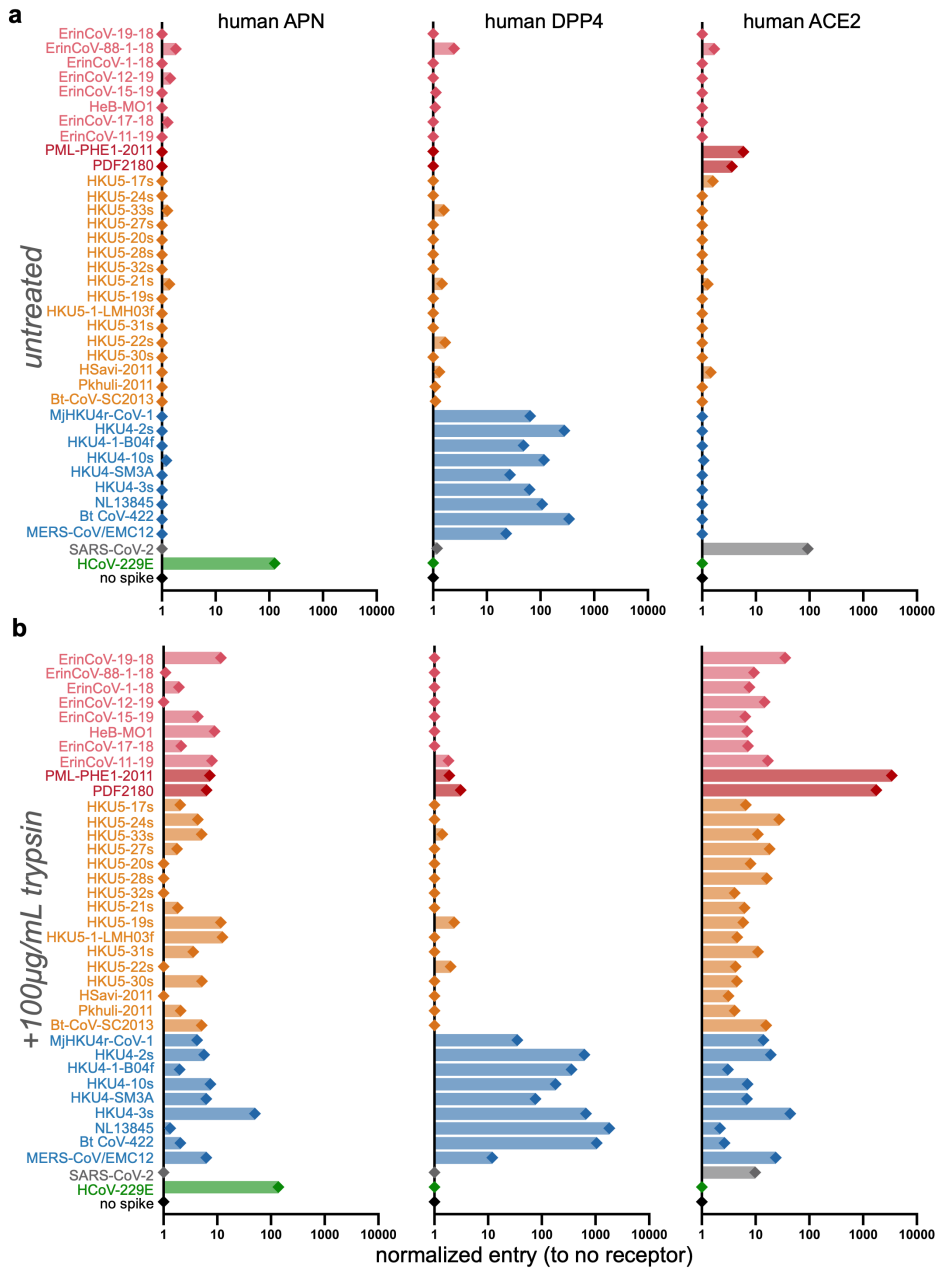


Fig. 4. Clade-specific use of human coronavirus receptors 293T cells transduced with known human coronavirus receptors were infected in **(a)** the absence of trypsin or **(b)** in the presence of 100µg/mL trypsin. Entry relative to no spike was compared to entry in cells without receptor to produce normalized entry values. Spike sequences for MERS-CoV/EMC12, SARS-CoV-2 and HCoV-229E are full-length, S1/S2 domains from these viruses, all other spikes are MERS-CoV spike chimeras with the indicated RBD.

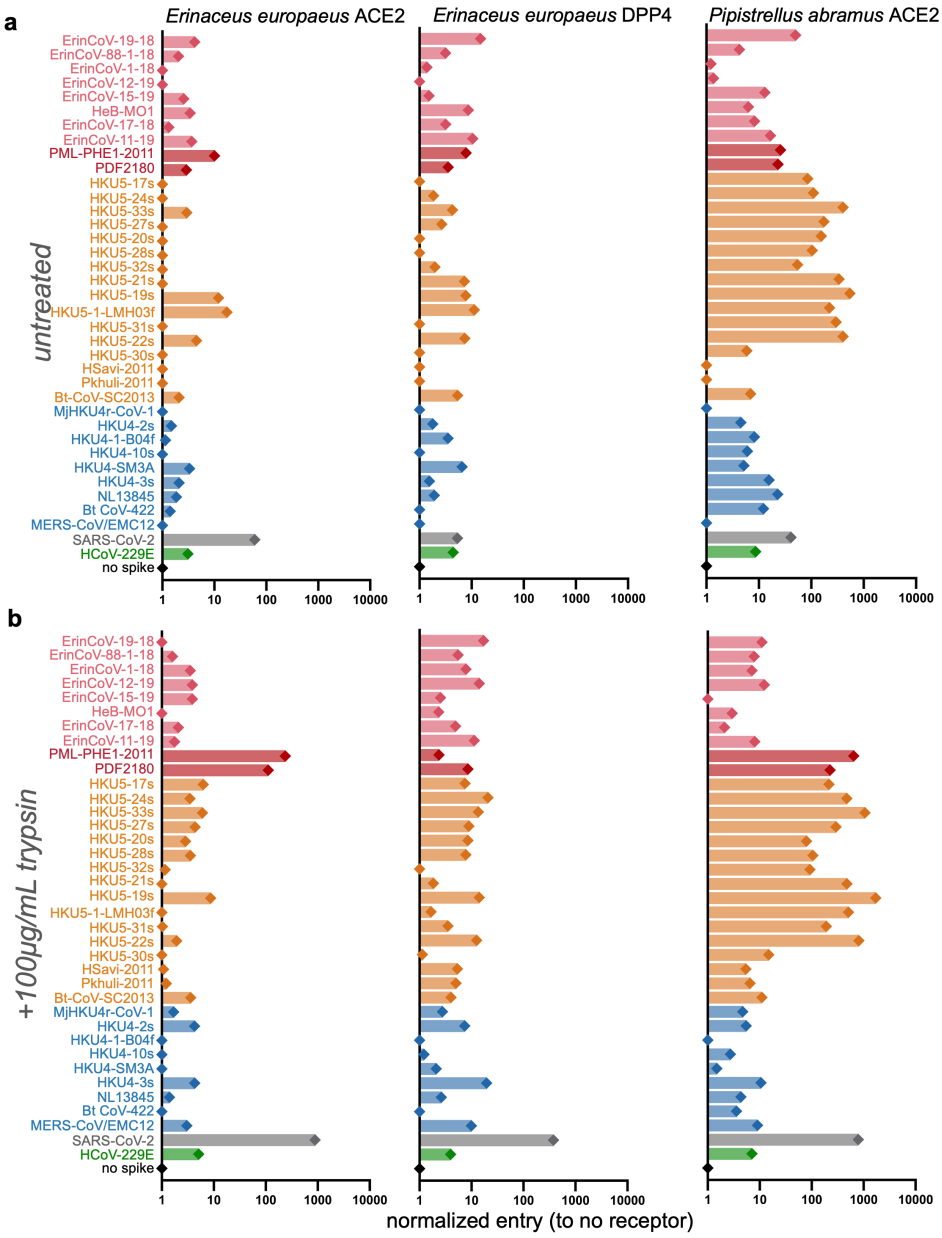
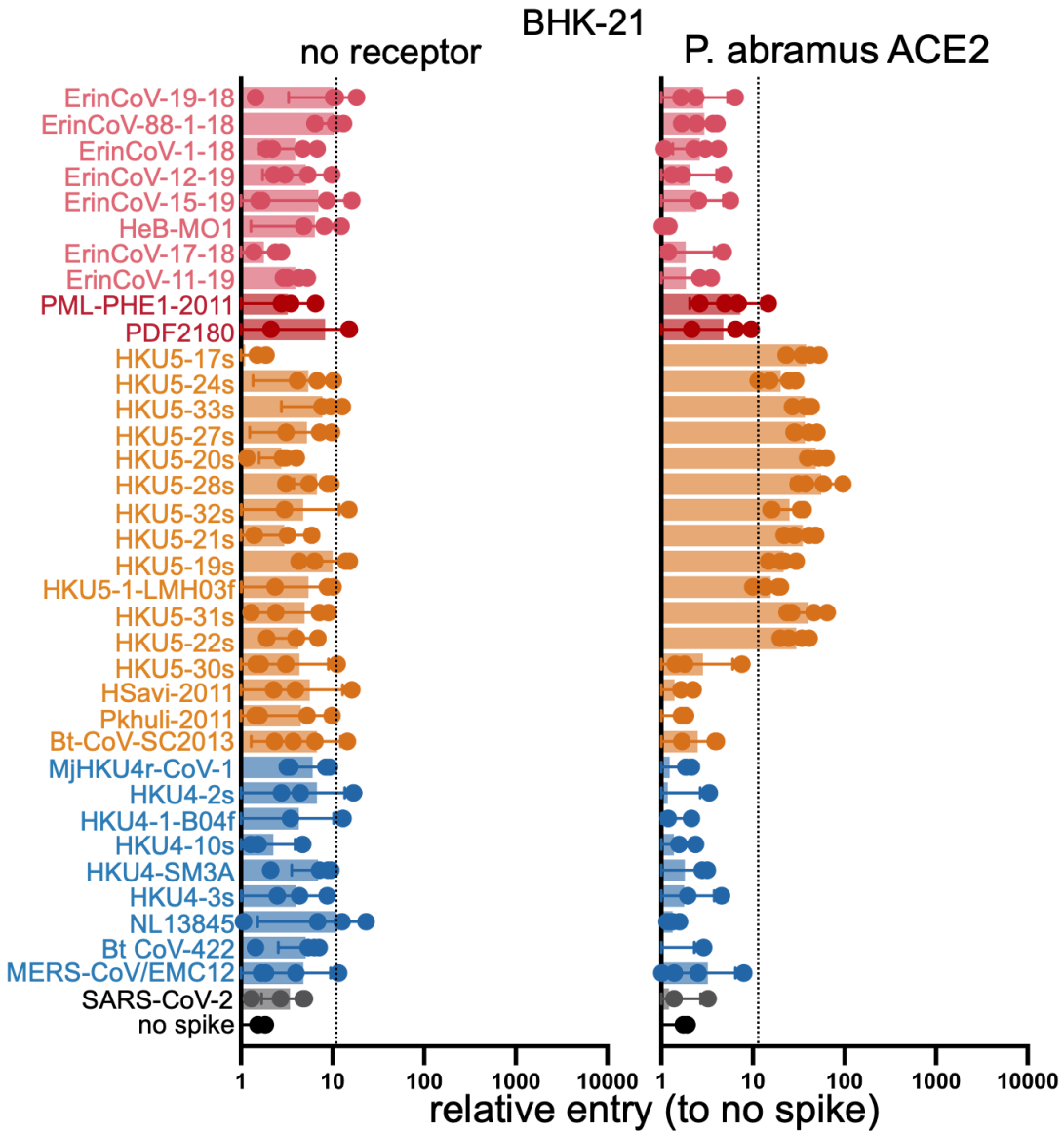


Fig. 5. ACE2 from Japanese house bats mediates HKU5 entry in human cells. 293T cells transiently transfected with species' orthologues of coronavirus receptors were infected in **(a)** the absence of trypsin or **(b)** in the presence of 100µg/mL trypsin. Entry relative to no spike was compared to entry in cells without receptor to produce normalized entry values. Spike sequences for MERS-CoV/EMC12, SARS-CoV-2 and HCoV-229E are full-length, S1/S2 domains from these viruses, all other spikes are MERS-CoV spike chimeras with the indicated RBD.



610

611 **Fig. 6. ACE2 from Japanese house bats specifically mediates HKU5 entry.** BHK-21 cells

612 were transiently transfected with empty vector or *Pipistrellus abramus* ACE2 and infected in the

613 absence of exogenous trypsin. Spike sequences for MERS-CoV/EMC12 and SARS-CoV-2 are

614 full-length, S1/S2 domains from these viruses, all other spikes are MERS-CoV spike chimeras

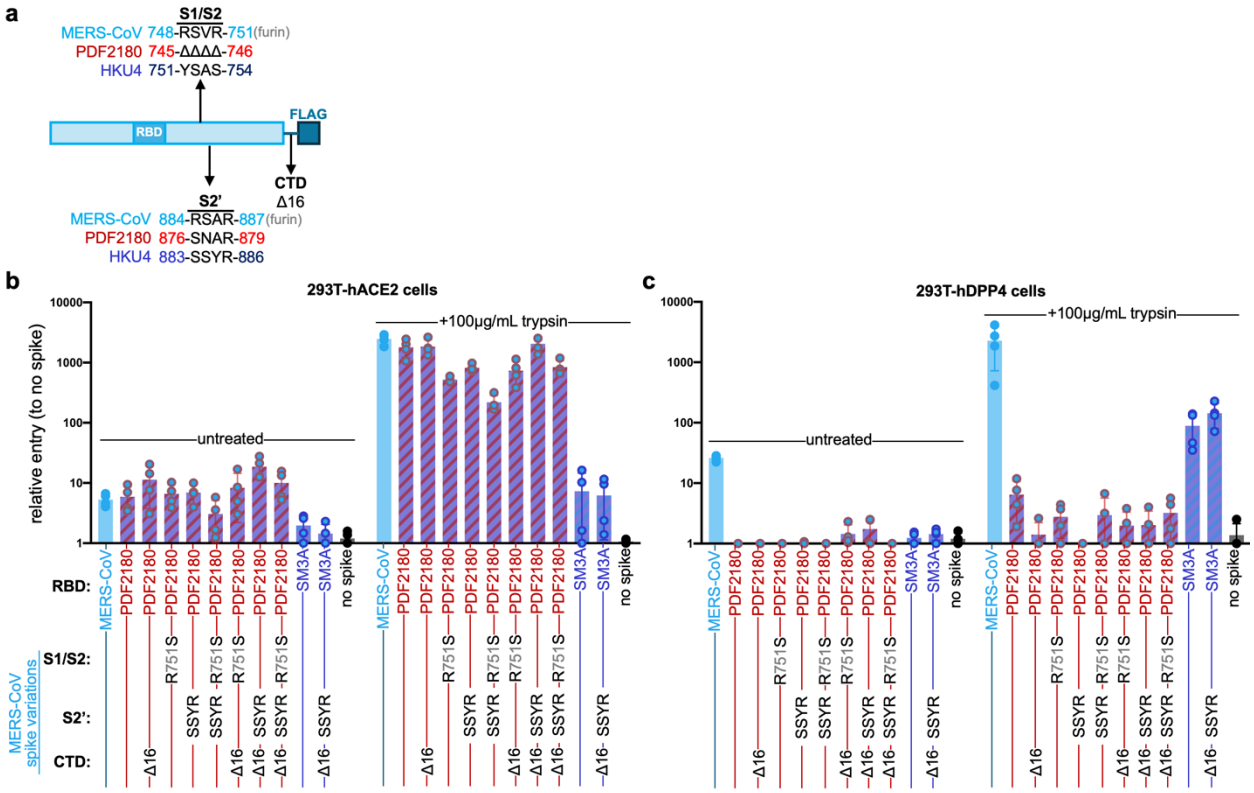
615 with the indicated RBD. Dashed line indicates arbitrary reliability cut-off at 10-fold signal over

616 background.

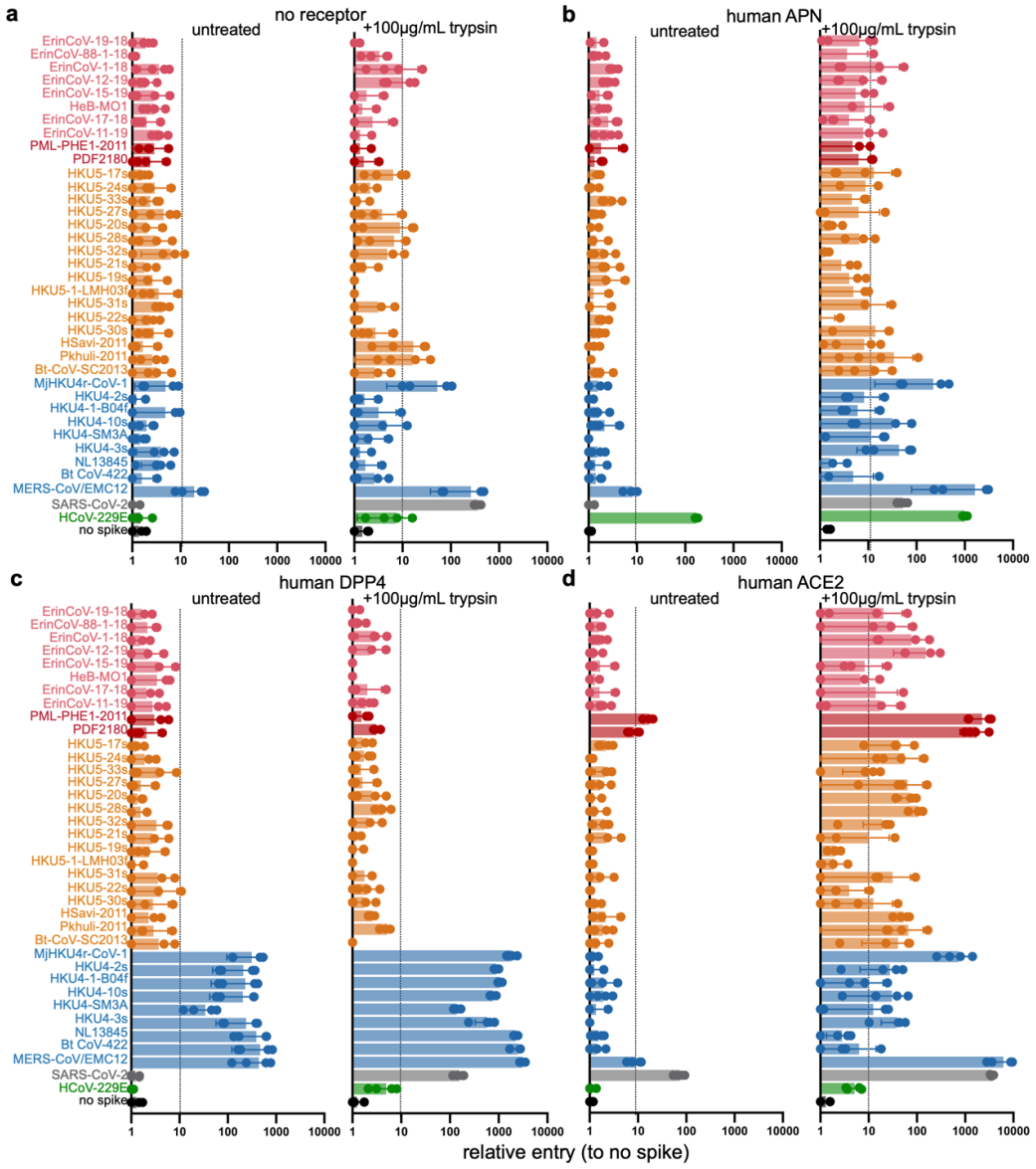
Table 1. Diverse merbecovirus panel encompassing global diversity

	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	PML-PHE1-2011	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

Extended Data Figures

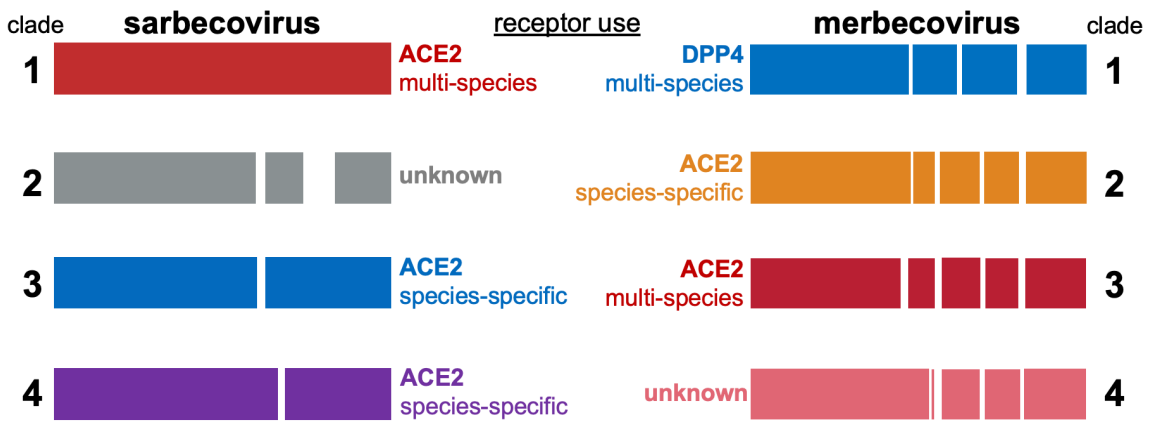


Extended Data Fig. 1. Optimizing MERS-CoV spike for MerbecoType. (a) Overview of S1/S2 and S2' furin sites in MERS-CoV spike. **(b)** 293T cells transduced with human ACE2 or (**c**) human DPP4 were infected with indicated spikes and luciferase was measured for entryCells were infected in quadruplicate.



Extended Extended Data Fig. 2. Relative data used in graphs for figure 4 – human receptors. 293Tcells transduced to express **(a)** no receptor **(b)** human APN **(c)** human DPP4 or **(d)** human ACE2 were infected in the absence or presence of 100µg/mL trypsin, in quadruplicate. Luciferase was measured for entry and relative entry was calculated by comparing to entry signal from bald pseudotypes. Spike sequences for MERS-CoV/EMC12, SARS-CoV-2 and HCoV-229E are full-length, S1/S2 domains from these viruses, all other spikes are MERS-CoV spike chimeras with the indicated RBD. Dashed line indicates arbitrary reliability cut-off at 10-fold signal over background.

Extended Data Fig. 3. Relative data used in graphs for figure 5 – animal receptors. 293T cells transfected to express **(a)** *Erinaceus europaeus* ACE2 **(b)** *Erinaceus europaeus* DPP4 or **(c)** *Pipistrellus abramus* ACE2 were infected in the absence or presence of 100µg/mL trypsin, in quadruplicate. Luciferase was measured for entry and relative entry was calculated by comparing to entry signal from bald pseudotypes. Spike sequences for MERS-CoV/EMC12, SARS-CoV-2 and HCoV-229E are full-length, S1/S2 domains from these viruses, all other spikes are MERS-CoV spike chimeras with the indicated RBD. Dashed line indicates arbitrary reliability cut-off at 10-fold signal over background.



Extended Data Fig. 4. Overview of beta-coronavirus subgenus RBD clades. RBDs from different beta-coronavirus subgenera share similar trends of clade assortment based on indels and clade-specific entry phenotypes.