1   **Permissivity of DPP4 Orthologs to MERS-Coronavirus is Governed by Glycosylation and**

2   **Other Complex Determinants**

3

4   Kayla M. Peck[a], Trevor Scobey[b], Jesica Swanstrom[b], Kara L. Jensen[b], Christina L. Burch[a], Ralph

5   S. Baric[b,d,*], and Mark T. Heise[c,d,*]

6

7   Departments of Biology[a], Epidemiology[b], Genetics[c], and Microbiology and Immunology[d],

8   University of North Carolina-Chapel Hill, Chapel Hill, North Carolina.

9

10   Running title: DPP4 orthologs have complex interactions with MERS-CoV

11

12   Keywords: MERS-coronavirus, DPP4, orthologs, host range expansion, animal models

13
14
15
16   *Corresponding Authors: Ralph S. Baric (rbaric@email.unc.edu) and Mark T. Heise
17   (mark_heisem@med.unc.edu) contributed equally.
18
19   University of North Carolina
20   Department of Genetics
21   Burnett-Womack Bldg., CB#7292
22   Chapel Hill, NC, 27599
23   919-966-4026
24

25

26

27   Abstract word count: 272
28   Text word count: 5865
29

30

31   **ABSTRACT**

32   Middle East respiratory syndrome coronavirus (MERS-CoV) utilizes dipeptidyl peptidase 4

33   (DPP4) as an entry receptor. While bat, camel, and human DPP4 support MERS-CoV infection,

34   several DPP4 orthologs, including mouse, ferret, hamster, and guinea pig, do not. Previous work

35   revealed that glycosylation of mouse DPP4 plays a role in blocking MERS-CoV infection. Here,

36   we test whether glycosylation also acts as a determinant of permissivity for ferret, hamster, and

37   guinea pig DPP4. We find that while glycosylation plays an important role in these orthologs,

38   additional sequence and structural determinants impact their ability to act as functional receptors

39   for MERS-CoV. These results provide insight into DPP4 species-specific differences impacting

40   MERS-CoV host range and better inform our understanding of virus-receptor interactions

41   associated with disease emergence and host susceptibility.

42

43   **IMPORTANCE**

44   MERS-CoV is a recently emerged zoonotic virus that is still circulating in the human population

45   with a ~35% mortality rate. With no available vaccines or therapeutics, the study of MERS-CoV

46   pathogenesis is crucial its control and prevention. However, *in vivo* studies are limited because

47   MERS-CoV cannot infect wildtype mice due to incompatibilities between the virus spike and the

48   mouse host cell receptor, mouse DPP4 (mDPP4). Specifically, mDPP4 has a nonconserved

49   glycosylation site that acts as a barrier to MERS-CoV infection. Thus, one mouse model strategy

50   has been to modify the mouse genome to remove this glycosylation site. Here, we investigate

51   whether glycosylation acts as a barrier to infection for other nonpermissive small animal species,

52   namely ferret, guinea pig, and hamster. Understanding the virus-receptor interactions for these

53 DPP4 orthologs will help in the development of additional animal models while also revealing

54 species-specific differences impacting MERS-CoV host range.

55

56 **INTRODUCTION**

57   Coronaviruses are a diverse family of viruses that infect a wide range of hosts, including

58 both mammalian and avian species. Phylogenetic studies suggest that over the last 800 years,

59 several zoonotic coronaviruses have expanded their host range into humans, resulting in four

60 antigenically distinct strains that are still circulating in the human population (1-2). Whereas

61 most human coronaviruses cause only mild symptoms in healthy adults, two recent emergence

62 events have resulted in severe disease in humans. Severe acute respiratory syndrome coronavirus

63 (SARS-CoV) emerged from its zoonotic reservoir in 2003 and infected over 8,000 people with a

64 9% mortality rate before being controlled by public health measures (3). Although SARS-CoV

65 is no longer circulating in the human population, several SARS-like coronaviruses persist in bats,

66 can use the human receptor for entry, and are poised for emergence (4-5). In 2012, Middle East

67 respiratory syndrome coronavirus (MERS-CoV) emerged from its zoonotic reservoir and

68 continues to cause human infection. As of July 2017, MERS-CoV has infected 2,037 people with

69 a ~35% mortality rate (6). The exact origins of MERS-CoV remain unclear. However, most data

70 suggest that the virus originated in bats and spread into human populations using camels as an

71 intermediate host species (7-8). Understanding how MERS-CoV and other coronaviruses evolve

72 and spread will allow us to combat MERS-CoV infection while also developing approaches for

73 dealing with future pandemic coronavirus strains.

74   One of the key determinants of viral host range is the interaction between the virus spike

75 protein and the host cell receptor. For MERS-CoV, the host cell receptor is dipeptidyl peptidase

76   IV (DPP4) (9), a ubiquitously expressed cell surface protein that functions in immune

77   homeostasis. Interestingly, whereas MERS-CoV can utilize the bat, camel, and human DPP4

78   molecules for entry (10-12), it is unable to infect cells using the DPP4 molecules from traditional

79   small animal models, including mice, ferrets, guinea pigs, and hamsters (13-17). The inability of

80   MERS-CoV to infect these species *in vivo* and *in vitro* is primarily due to spike-receptor

81   incompatibilities and not due to other species-specific host cell factors (13, 18-19). Because

82   MERS-CoV is unable to infect traditional small animal models, our ability to study MERS-CoV

83   pathogenesis is restricted, as is the ability to evaluate drugs and vaccine efficacy. To overcome

84   this obstacle, several mouse models have recently been developed in order to study MERS-CoV

85   pathogenesis *in vivo*. These include an adeno-associated virus delivery of human DPP4 (hDPP4)

86   (18), overexpression or knock-in of hDPP4 (20-23), and the use of gene editing to generate mice

87   carrying a chimeric mouse DPP4 (mDPP4) gene (24). These mouse models greatly improve our

88   ability to study MERS-CoV infection, motivating us to investigate the potential for alternate

89   species to act as models for MERS-CoV infection. Importantly, the development of vaccines and

90   therapeutics relies on positive results from more than one animal species. For MERS-CoV, the

91   disparate findings reported for nonhuman primate systems, including the rhesus macaque and

92   common marmoset (25-28), emphasize the need for additional animal models.

93   In our previous studies, we found that receptor incompatibilities between MERS-CoV

94   and mDPP4 can be alleviated with just two amino acid substitutions (13). Introducing the

95   mutations A288L and T330R (residues numbered relative to mDPP4) allows mDPP4 to act as a

96   functional receptor for MERS-CoV infection. The A288L mutation helps stabilize a hydrophobic

97   core in the MERS-CoV receptor binding domain (RBD) shown to be important for proper

98   binding to DPP4 (29). Conversely, the T330R mutation knocks out a nonconserved glycosylation

99   site present in mDPP4. This glycosylation site is a substantial barrier to MERS-CoV infection,

100   with mDPP4 only acting as a functional receptor when the glycan is absent (19). Taken together,

101   these two mutations informed the development of the aforementioned chimeric mDPP4 mouse

102   model (24) and improve our understanding of the biochemical mechanisms that govern the

103   interactions between DPP4 and the MERS-CoV RBD.

104   Here we investigate whether glycosylation acts as a broader determinant of permissivity

105   using alternate small animal models including the ferret, hamster, and guinea pig. Overall, we

106   find that while glycosylation of these DPP4 orthologs plays an important role in blocking

107   MERS-CoV infection, additional species-specific determinants contribute to the inability of each

108   of these receptors to support MERS-CoV infection. Further investigation is required to identify

109   these determinants and establish whether they lie at the level of the sequence or structure.

110   Improving our understanding of the interactions between DPP4 and the MERS-CoV RBD can

111   help inform the development of alternate small animal models as well as help reveal how

112   coronaviruses are able to emerge into novel species.

113

114   **RESULTS**

115   Susceptibility to MERS-CoV varies among a number of host species, making detailed

116   analyses of orthologous DPP4 receptors essential for elucidating fundamental mechanisms that

117   govern coronavirus species specificity. DPP4 orthologs are classified as permissive (human, bat,

118   and camel) or non-permissive (mouse, ferret, hamster, and guinea pig) (Figure 1A). Based on

119   mean fluorescent cell count, hDPP4, bDPP4, and cDPP4 all support higher levels of infection

120   compared to nonpermissive orthologs ($p < 0.05$, Student's t-test, Figure 1B). Orthologs

121   categorized as nonpermissive have levels of infection that are not significantly different from

122   when no DPP4 is present (Figure 1B). Our previous work found that mDPP4 could support

123   MERS-CoV infection with just two amino acid substitutions (13) and that knocking out a

124   glycosylation site present on blade IV of mDPP4 was crucial for this result (19). Here, we

125   investigate alternate nonpermissive DPP4 orthologs to determine whether they can act as

126   backbones to support MERS-CoV infection and if so, what determinants are responsible for

127   blocking infection in the wildtype molecules.

128   To assess the potential for fDPP4, haDPP4, and gpDPP4 to act as functional receptors for

129   MERS-CoV, we compared the overall structures of each molecule. As the crystal structures for

130   these orthologs have not yet been solved, we generated homology models using I-TASSER (30).

131   Each structure is predicted to have a highly similar backbone topology to hDPP4 (Figure 2A).

132   The root mean square deviation (RMSD) scores obtained for mDPP4, fDPP4, haDPP4 and

133   gpDPP4 aligned to hDPP4 (PDB code 4L72) are 0.644, 0.616, 0.378, and 0.604, respectively,

134   with a smaller number indicating greater structural homology. This can be compared to amino

135   acid sequence identity values of 85%, 88%, 85%, and 87%, respectively. The high sequence and

136   predicted structural similarities between hDPP4 and these orthologs suggest that they can likely

137   act as backbones to support MERS-CoV infection, consistent with previous DPP4 ortholog work

138   (15, 17).

139   To gain an intuition on specific differences present between permissive and

140   nonpermissive DPP4 orthologs, we can assess their relatedness through phylogenetics.

141   Unfortunately, analysis of the DPP4 gene tree does not reveal a pattern of permissivity on a

142   global sequence scale (i.e., permissive and nonpermissive species do not form distinct

143   monophyletic groups) (31-32). However, it is possible that a signature may exist at a local scale.

144   Our previous work in mDPP4 revealed that glycosylation on blade IV of the DPP4 molecule can

145   act as a barrier to MERS-CoV infection (19). Thus, we investigated whether glycosylation

146   (designated by the motifs NXS or NXT) acts as a broader determinant of DPP4 ortholog

147   permissivity. A total of eight N-linked glycans are known to be present in the extracellular

148   domain of hDPP4 (29). Without solved crystal structures for DPP4 orthologs, it is unknown how

149   many glycans are truly present on the surface of these molecules. However, predictive software

150   (33) estimates 7–11 putative glycosylation sites for each of the DPP4 orthologs analyzed in this

151   study (Table 1). Aligning the sequences of these orthologs species revealed the presence of

152   putative glycosylation sites on blade IV of fDPP4, haDPP4, and gpDPP4 in a region that is

153   crucial for interactions with the MERS-CoV RBD (Figure 2B, Table 1). Interestingly, haDPP4

154   encodes a glycosylation site at the residues aligning to the site identified in our previous mDPP4

155   studies. In contrast, fDPP4 encodes a glycosylation site slightly upstream whereas gpDPP4

156   encodes a glycosylation site slightly downstream from the mDPP4 site (Figure 2B). All three

157   glycosylation sites lie on blade IV, within a region of DPP4 that interacts with the MERS-CoV

158   RBD (Figure 2C). Of note, the downstream gpDPP4 glycosylation site is also present within the

159   permissive bDPP4 receptor used in this study (from *Pipistrellus pipistrellus*), causing us to

160   hypothesize that host restriction might be mediated by differences in the sequence or structure of

161   gpDPP4 that are independent of this glycosylation site.

162   To investigate whether the putative glycosylation sites identified in fDPP4, haDPP4, and

163   gpDPP4 play a role in hindering MERS-CoV infection, we assessed receptor permissivity for

164   glycosylation knockout mutants. Each DPP4 ortholog knockout includes a mutation that changes

165   the N of the glycosylation NXT (or NXS) motif to an alanine, designated by "-gly" (specific

166   mutations are fDPP4 N331A, haDPP4 N332A, and gpDPP4 N548A with residue numbering

167   relative to its own sequence). Results show that removing glycosylation from these three DPP4

168   orthologs, confirmed by a ~2.5 kDa downward shift in protein mobility via Western blot analysis

169   (Figure 3B), did not result in an increase in infection (Figure 3A). In fact, levels of infection

170   supported by all three glycosylation knockout molecules were not significantly different from

171   their respective wildtype DPP4 molecules ($p < 0.05$, Student's t-test, Figure 3C).

172   Immunofluorescence assay (IFA) and flow cytometry confirmed that the DPP4 variants were

173   expressed at the surface of the cell (Figure 4), ruling out the possibility of nonpermissivity

174   resulting from a lack of proper expression of the DPP4 ortholog glycosylation knockouts.

175   Although removing the blade IV glycosylation site on its own is not enough to confer

176   permissivity to fDPP4, haDPP4, and gpDPP4 (Figure 3A), we have previously shown in mDPP4

177   that a second determinant acts in conjunction with the blade IV glycosylation site to impact the

178   interactions between DPP4 and the MERS-CoV RBD (13, 19). Specifically, introducing select

179   human amino acids into the mDPP4 sequence on blade V (A288L) acts to support MERS-CoV

180   infection when combined with the blade IV glycosylation knockout (N328A). In the case of

181   haDPP4, the L288 amino acid identity is conserved between the hamster and human sequences,

182   suggesting that different determinants are responsible for blocking MERS-CoV infection

183   between haDPP4 and mDPP4. This is consistent with prior studies which identified several

184   amino acid differences between hDPP4 and haDPP4 in blade V that were required to support

185   MERS-CoV infection (17). Therefore, we asked whether specific changes in blade V of haDPP4

186   could allow it to act as a functional MERS-CoV receptor when combined with the blade IV

187   glycosylation knockout mutation. We found that three mutations on blade V (E289A, V291M,

188   and T293I, residue numbering relative to haDPP4) conferred MERS-CoV receptor function to

189   haDPP4 when introduced in conjunction with the blade IV glycosylation knockout mutation

190   (Figure 5A). Two of the identified mutations overlap with previous work while a third site

191 knocks out the same blade IV glycosylation site but by mutating a different residue of the

192 glycosylation site motif (Table 2). Testing combinations of these mutations in sets that included

193 less than the four identified here, including testing haDPP4 E289A, V291M, and T293I (denoted

194 haDPP4 (289)) in a glycosylation positive background, did not result in an increase in infection

195 (data not shown). These results demonstrate that the presence of the blade IV glycosylation site

196 plays an important role in regulating the MERS-CoV receptor function of haDPP4. However,

197 species-specific differences in blade V of haDPP4, which differ from those found in mDPP4, are

198 also responsible for haDPP4's inability to act as a functional receptor for MERS-CoV infection.

199 Given our findings with haDPP4, we then went on to assess whether the same principles

200 applied to fDPP4. Specifically, we tested whether additional changes in blade V would act in

201 concert with the blade IV glycosylation knockout mutation to enhance the MERS-CoV receptor

202 function. Previous studies found that swapping fDPP4 amino acids 246 to 503 with hDPP4

203 amino acids 247 to 504 allowed fDPP4 to support MERS-CoV infection, but no single mutations

204 were identified that could recapitulate the susceptible phenotype (15). Because the mDPP4 and

205 haDPP4 data suggest that changes on *both* blades IV and V of DPP4 are likely required to confer

206 infection, we mutated residues on blades V and tried them in combination with the blade IV

207 glycosylation knockout mutation (N331A, residue numbering relative to fDPP4). We found that

208 no set of mutations on blade V conferred permissiveness to MERS-CoV when made only with

209 the blade IV glycosylation knockout mutant (data not shown). Given this results, we expanded

210 our search and made additional mutations on blade IV of fDPP4. By identifying residues not

211 present in permissive DPP4 orthologs, we generated a chimeric fDPP4 that changed 10 residues

212 on blade IV, including the glycosylation knockout mutation (N331A), and 9 residues on blade V

213 to the equivalent human amino acid identities, indicated by their starting residues of 330 and

214 278, respectively, numbering relative to the fDPP4 sequence (Figure 5B). Results show that the

215 combination of these 19 amino acid changes on blades IV and V showed a significant increase in

216 infection ($p < 0.05$, Student's t-test, Figures 5B and 5D). Western blot analysis confirmed high

217 expression of each DPP4 variant. The larger downward shift seen for the fDPP4 (278) receptor is

218 likely due to the large number of amino acid changes made on blade V (Figure 5C).

219 Combinations of individual mutations from these blocks did not result in an increase in infection

220 (data not shown). These data reveal that unlike mDPP4 and haDPP4, additional determinants in

221 blade IV contribute to the ability of fDPP4 to act as a functional MERS-CoV receptor.

222   In addition to haDPP4 and fDPP4, we also investigated the determinants of gpDPP4.

223 Whereas haDPP4 and fDPP4 do not share their glycosylation sites with permissive molecules,

224 gpDPP4 shares its downstream glycosylation site with bDPP4 (Figure 2B). Because of this, we

225 knocked out the blade IV glycosylation site in bDPP4 to determine 1) whether it is truly

226 glycosylated and 2) its impact on the permissivity of bDPP4. Results show that removing

227 glycosylation from bDPP4 caused no statistically significant change in its ability to support

228 MERS-CoV infection (Figures 6A and 6D). Western blot analysis confirmed that the bDPP4

229 glycosylation motif is a true glycosylation site when expressed in human cells, as evidenced by a

230 downward shift in the glycosylation knockout protein band (Figure 6B). Additionally, flow

231 cytometry and IFA confirmed surface expression of each variant tested in this assay (Figures 4B-

232 C, 6C). Low apparent expression of bDPP4 is likely due to inefficient binding of the α-hDPP4

233 antibody and bDPP4, an observation that is supported by previous work (35). Unfortunately, the

234 gpDPP4 plasmid was not stable, and the gene was repeatedly lost during site-directed

235 mutagenesis. This instability prevented further investigation of the determinants responsible for

236 gpDPP4's inability to support MERS-CoV infection. Therefore, future studies are needed to

237 determine whether changes on both blades IV and V or just blade IV of gpDPP4 can confer

238 permissivity when combined with the blade IV glycosylation knockout mutation, as was shown

239 with fDPP4 and haDPP4, respectively.

240

241 **DISCUSSION**

242   The interactions between the virus spike and host cell receptor are crucial for mediating

243 infection and for acting as a barrier to nonpermissive species. However, viruses frequently

244 overcome species barriers to expand their host range. Glycosylation plays an important role in

245 mediating permissivity for many viruses, both in the context of the virus spike protein and of the

246 host cell receptor. For example, glycosylation of the H1N1 influenza virus hemagglutinin protein

247 is highly conserved and has been shown to be crucial for host cell receptor binding specificity

248 (36-37). Conversely, glycosylation of the parvovirus host cell receptor (transferrin receptor type-

249 1) in canines was found to confer resistance to infection by previously circulating feline

250 parvovirus strains (38). In fact, dogs did not become susceptible to parvovirus until a lineage

251 (now canine parvovirus) was able to overcome this glycosylation through specific mutations in

252 the capsid protein (38-40).

253   Glycosylation has previously been identified as a key determinant of coronavirus host

254 range. For example, some lineages of group 1 coronaviruses use aminopeptidase N (APN) as

255 their host cell receptor and APN orthologs have different glycosylation profiles that influence the

256 species-specificity of coronavirus infection. Porcine and feline APN orthologs are glycosylated

257 and can support infection by porcine and feline coronaviruses, respectively (41). However,

258 adding a glycosylation site into human APN near residue 290 can abrogate its ability to support

259 human coronavirus 229E (HCoV-229E) infection (41). Similarly, removing the aligning

260   glycosylation site in mouse APN can allow it to act as a functional receptor for HCoV-229E

261   (42). As another example, SARS-CoV utilizes angiotensin-converting enzyme 2 (ACE2) as its

262   functional receptor. Whereas rat ACE2 is not permissive to SARS-CoV, introduction of a

263   glycosylation site combined with a point mutation allows rat ACE2 to support SARS-CoV

264   infection (43). Taken together, these results highlight glycosylation as a key determinant of

265   receptor species-specificity for numerous coronaviruses.

266   Here we investigate the extent to which glycosylation of DPP4 can act to mediate the

267   host range of MERS-CoV. Our previous work revealed the importance of glycosylation in

268   blocking MERS-CoV infection in the context of mDPP4 (19). Here, we find that glycosylation is

269   not the only determinant that mediates MERS-CoV infection in the nonpermissive orthologs

270   fDPP4, haDPP4, and gpDPP4. This is particularly surprising in the case of haDPP4; this protein

271   not only has a glycosylation site in the same location as mDPP4 (Figure 2B), but has the same

272   amino acid identity as hDPP4 at the secondary blade V residue (288) that was identified to be

273   important for mDPP4. Instead, we found that three amino acid substitutions in blade V of

274   haDPP4 affect the molecule's ability to mediate MERS-CoV infection when coupled with the

275   blade IV glycosylation knockout mutation (Figure 5A). For fDPP4, we found that extensive

276   changes on both blades IV and V were required to support MERS-CoV infection (Figure 5B),

277   indicating that there are likely species-specific determinants that are not shared with mDPP4 and

278   haDPP4. Finally, the blade IV glycosylation site in gpDPP4 is conserved with the permissive

279   bDPP4 ortholog (Figure 2B). We found that removing this glycosylation site in bDPP4 did not

280   affect its ability to support MERS-CoV infection (Figure 6A). This suggests that the impact of

281   the bDPP4 blade IV glycosylation site on receptor permissivity may be species-specific and

282   further research should work to elucidate additional changes that allow gpDPP4 to act as a

283   functional receptor for MERS-CoV. These changes can then be made independently and in

284   combination with the blade IV glycosylation site to determine its impact within the gpDPP4

285   backbone. Overall, we find that glycosylation in DPP4 orthologs is a substantial barrier to

286   MERS-CoV infection, particularly when combined with species-specific changes in blades V

287   and/or IV. The biochemical mechanism of these additional determinants can be investigated in

288   future studies.

289   Comparing the data we have so far on the residues that are important for mediating

290   permissivity in hDPP4, mDPP4, haDPP4, and fDPP4, it is difficult to discern an obvious pattern.

291   The most evident trend is that at least one change is required on *both* blades IV and V of DPP4

292   (Table 2). This indicates that there are two key points of interaction between DPP4 and the

293   MERS-CoV RBD which is consistent with previous structural work examining the binding

294   interface of MERS-CoV RBD and hDPP4 (29, 34, 44). We show that these two points of

295   interaction are also important for allowing the virus to utilize a new species receptor. Further

296   understanding of the host range expansion of MERS-CoV will come with identifying which

297   mutations can confer permissivity to DPP4 receptors of currently nonpermissive species.

298   Additionally, solving the crystal structures of nonpermissive DPP4 orthologs will reveal the

299   specific interactions between these molecules and MERS-CoV and help to elucidate what

300   prevents MERS-CoV from successfully utilizing these molecules as functional receptors. The

301   inability to identify a small number of changes that confer permissivity to DPP4 orthologs

302   suggests that extensive remodeling of the MERS-CoV RBD might be required for the virus to

303   infect these nonpermissive species. Access to the crystal structures of these DPP4 orthologs

304   would enable us to model changes in the MERS-CoV RBD that would allow the virus to

305   successfully infect nonpermissive species, thereby enhancing our ability to study MERS-CoV

306   pathogenesis and provide robust small animal models for evaluating vaccines and therapeutics.

307   As of now, our results indicate that the generation of a transgenic small animal model in these

308   alternate species would require extensive genomic editing due to the number of mutations

309   needed for each DPP4 to support MERS-CoV infection.

310   The importance of glycosylation in blocking MERS-CoV infection may vary between

311   species. To gain a better intuition about the extent of glycosylation among DPP4 orthologs, we

312   constructed a phylogenetic tree of a subset of full-length DPP4 protein sequences (Figure 7). In

313   the tree, shaded colors indicate the general organism group that each species belongs to - blue:

314   reptiles and amphibians; green: avian species; orange: other mammals; red: Chiroptera (bats);

315   purple: ungulates; gray: rodents; pink: primates (Figure 7). The DPP4 gene tree is slightly

316   discordant with the species tree, notably with the horse and African savanna elephant DPP4

317   sequences not clustering with other ungulate (purple) orthologs. Plotted adjacent to the

318   phylogenetic tree are glycosylation sites that are either upstream (column 1), at the same site

319   (column 2), or downstream (column 3) of the glycosylation site that is present in mDPP4

320   (residues 328-330). Permissivity data is indicated in the far right column, with green squares

321   indicating permissive species and red squares indicating nonpermissive species, based on either

322   *in vitro* or *in vivo* data (10-17).

323   The small number of known permissive and nonpermissive species makes it difficult to

324   map potential shifts in permissivity on the phylogenetic tree. Because permissive and

325   nonpermissive species do not cluster together on the DPP4 tree (Figure 7), it is more likely that

326   local differences between each protein is more important than global similarity. Despite our

327   inability to identify a distinct pattern, a few interesting observations do emerge. First, all non-

328   human primates lack glycosylation sites near residue 330 (equivalent to residue 336 in humans)

329   (Figure 7). Current data would postulate that all non-human primates are susceptible to MERS-

330   CoV infection, but testing this hypothesis *in vitro* would help reveal whether any of these

331   orthologs are nonpermissive. If so, a mechanism other than glycosylation at our location of

332   interest would be responsible for blocking MERS-CoV infection. Second, whereas other

333   permissive DPP4 orthologs have glycosylation sites in this region (horse and common pipistrelle

334   DPP4), of note is the lack of a glycosylation site in the nonpermissive domestic pig, sheep, and

335   cattle DPP4 molecules (Figure 7). This suggests that glycosylation in this region is not the

336   primary explanation for why these species do not support MERS-CoV infection. One example

337   could be the incompatibility of host cell proteases which are required for proper cleavage of the

338   MERS-CoV spike prior to cell entry. Previous works has shown that spike cleavage status

339   determines cell tropism (45) and that adaptation to cellular proteases may have played an

340   important role in the cross-species transmission of MERS-CoV from bats to humans (35).

341   Follow-up studies could investigate the mechanism of nonpermissivity in these species

342   specifically and whether or not it lies at the level of the receptor. Third, glycosylation sites in the

343   designated region are prevalent in rodents, other mammals, and avian species (Figure 7). For

344   example, eight other species have putative glycosylation sites aligning to the glycosylation site

345   identified in mDPP4. The majority of these are present in the rodent (gray) group, however two

346   are present within the Chiroptera group (black flying fox and large flying fox) (Figure 7).

347   However, the diversity of glycosylation profiles suggests a lack of strong conservation of the site

348   seen in mDPP4. Further research to determine whether these other orthologs are permissive can

349   help elucidate whether this region plays a broader role in the MERS-CoV infection phenotype. In

350   general, while some trends are seen when looking at the glycosylation profile across many

351   species (e.g. upstream glycosylation sites in the other mammal (orange) group), more data on

352 species permissivity will help determine whether DPP4 phylogenetic relationships can help

353 inform receptor binding dynamics.

354   Here, we demonstrate that glycosylation is an important barrier to MERS-CoV infection,

355 yet other species-specific determinants are also responsible. Removing the glycosylation sites of

356 fDPP4, haDPP4, and gpDPP4 orthologs did not result in an increase in infection. Rather, a set of

357 19 and 4 amino acid substitutions in fDPP4 and haDPP4, respectively, were required before a

358 significant increase in infection could be detected. Future work can focus on revealing the

359 importance of these additional substitutions in mediating MERS-CoV infection. By further

360 elucidating the mechanisms by which DPP4 orthologs block MERS-CoV infection, we can better

361 understand the constraints on coronavirus host range expansion and predict coronavirus

362 emergence into new species in the future. Furthermore, this work can inform the development of

363 better animal models for studying MERS-CoV pathogenesis. Our previous work on mDPP4

364 helped generate a successful transgenic mouse using the CRISPR-Cas9 gene editing technique

365 (24). The success of this model emphasizes the importance of understanding receptor-virus

366 interactions and is a potent example of the direct application of these studies.

367

368 **MATERIALS AND METHODS**

369 **Viruses and cells**

370   MERS-CoV was isolated from a molecular clone as described in Scobey et al. (46).

371 Recombinant MERS-CoV tagged with tomato red fluorescent protein (rMERS-CoV-RFP) was

372 engineered based on the EMC2012 substrain and was shown to infect and replicate similarly

373 compared to wildtype MERS-CoV (46). Recombinant viruses were passaged once (P1) on Vero

374 cells to generate a working stock of rMERS-CoV-RFP. All recombinant viruses were produced

375   and studied under BSL3 conditions, wearing Tyvek suits, apron and booties, PAPR, and double

376   gloved as described in the Laboratory Safety Plan.

377   Measures of rMERS-CoV-RFP infectivity, receptor protein expression levels, and

378   immunofluorescence assays (IFA) were performed in human embryonic kidney (HEK) 293T

379   cells transfected with cDNAs encoding wildtype or mutant DPP4 orthologs. Cells were seeded at

380   1×10^6 cells per well in 6-well plates pretreated with 1 mL of 10 µg/mL poly-L-lysine (Sigma)

381   in 1X DPBS (Gibco). All cells were grown at 37 ºC with 5% CO_2.

382

383   **DPP4 ortholog constructs**

384   DPP4 orthologs were ordered on plasmids (cDPP4, hDPP4, mDPP4) or constructed using

385   Gibson assembly (bDPP4, fDPP4, haDPP4, gpDPP4) (47). DPP4 orthologs were transferred into

386   the 945ΔRRE expression vector, a lentiviral vector derived from pTK945, using the NotI and

387   SpeI restriction sites. DPP4 mutants were generated by overlap PCR mutagenesis and verified by

388   Sanger sequencing. All plasmids were stored at -20 ºC. Accession numbers for DPP4 sequences

389   used in this study are as follows: bDPP4, KC249974.1; cDPP4, XM_006176808.2; fDPP4,

390   DQ266376.1; gpDPP4, XM_013142395.1; haDPP4, NM_001310571.1; hDPP4, NM_001935.3;

391   mDPP4, NM_010074.3.

392

393   **DPP4 ortholog transfections**

394   DPP4 orthologs were transfected into HEK 293T cells, as these cells express little if any

395   endogenous hDPP4 (13). Transfections followed a 2X BES protocol. For each DPP4 ortholog,

396   15 µg of the construct was mixed with 15 µg of pcDNA3.1, 125 µL of 1 M $CaCl_2$, and dH_2O up

397   to final volume of 500 µL. Then, 500 µL of 2X BES was added to each sample in a dropwise

398   manner, vortexing between drops. Samples were incubated at room temperature for 45 min.

399   During this incubation, media was removed from 6-well plates seeded with 1×10^6 HEK 293T

400   cells/well and 1 mL of fetal bovine serum (FBS)-free media was added to each well. After the 45

401   min incubation, 200 μL of each transfection reaction was added to individual wells, dropwise

402   and in concentric circles. Plates were incubated at 37 ℃ for 2 hours before 150 μL of FBS was

403   added to each well and the cultures incubated overnight at 37 ℃. To measure transfection

404   efficiencies, venus protein-tagged 945 plasmid was transfected in parallel with each experiment

405   (with efficiencies typically between ~80-100%).

406

407   **Infection of DPP4-transfected cells by MERS-CoV**

408   At ~20 hours post-transfection, media was removed and each well infected with 200 uL

409   of rMERS-CoV-RFP at a multiplicity of infection (MOI) of 5 for 45 min at 37 ℃. Next, 2 mL of

410   3% FBS media was added to each well and the plates incubated at 37 ℃ for 24 hours. Cells were

411   visualized 24 hours post-infection in the BSL3 using an Olympus IX51 microscope at a

412   magnification of 10X and wavelength of 541 nm. Each well was imaged using a Hamamatsu

413   ORCA-Flash4.0 LT camera. Successful virus infection is indicated by the presence of red

414   fluorescence.

415

416   **Fluorescent cell count assay**

417   For quantification of levels of MERS-CoV infection, transfected cells were infected as

418   above, except at an MOI of 0.1. After the 45 min incubation period, virus samples were removed

419   and the cells washed twice with 1 mL 1X DPBS. Next, 2 mL of 3% FBS media was added to

420   each well and infections incubated at 37 ℃ for 72 hours. At 72 hours post-infection, the number

421   of red fluorescent cells was determined. Briefly, three randomly selected microscope fields were

422   counted for each well. The average number of red cells was taken from these three fields and

423   multiplied by a factor of 537 to approximate the number of red cells in the entire well. Infection

424   in the presence of each DPP4 ortholog was measured as described in triplicate, thus 9 fields were

425   measured to determine significance between two DPP4 treatments using a Student's t-test.

426

427   **Protein analysis and Western blotting**

428   Cells were transfected with DPP4 constructs as described above. At 48 hours post-

429   transfection, media was removed from each well and cells were washed once with 2 mL 1X

430   DPBS. Next, 200 µL of ice cold AV lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 0.5%

431   deoxycholine, 1% Igepal, 0.1% SDS) was added to each well. Cells were incubated at room

432   temperature for 10 min with the lysis buffer. Lysates were then transferred to 1.5 mL

433   microcentrifuge tubes and centrifuged at maximum speed for 2 min. Supernatants were

434   transferred to new tubes and 200 µL of a 10mM EDTA/0.9% SDS stock added to each tube.

435   Lysates were stored at -80 ℃.

436   Protein concentration for each sample was measured using a Bradford assay. Briefly, an

437   aliquot of each sample was diluted 1:5 and 10 µL added to a 96-well plate in triplicate. Bovine

438   serum albumin (BSA) was used as a standard with 10 µL of a dilution series added to the 96-well

439   plate in triplicate. To each well, 200 µL of Coomassie Plus protein assay reagent

440   (ThermoScientific) was added and the plate incubated at room temperature in the dark for 10

441   min. Absorbance readings were measured at a wavelength of 595 nm and the protein

442   concentration of each sample determined by comparison to the BSA dilution standard curve.

443   Proteins were separated using either a 6% (for DPP4) or an 8% (for actin) polyacrylamide

444   gel. For each sample, 15 µg of the protein lysate was mixed with 6X loading dye and heated to

445   95 ºC for 10 min. Samples were loaded into the gel and separated by electrophoresis in 1X

446   running buffer (3.03g Tris HCl, 14.4g glycine, 1g SDS in 1L dH$_2$O) at 55 mA for ~2 hours.

447   Proteins were transferred onto an Immun-Blot PVDF membrane (BioRad) at 15V for 45 min

448   using 1X dry transfer buffer (5.82g Tris base, 2.92g glycine, 100 mL methanol in 1L dH$_2$O). The

449   membranes were blocked in 1X phosphate-buffered saline with Tween 20 (PBST) with 5% milk

450   at room temperature for 1 hour with shaking. Membranes were then treated with 1º antibody:

451   goat α-DPP4 (R&D systems) at 1:1000 or goat α-actin (Santa Cruz Biotechnology) at 1:1000 in

452   1X PBST and incubated at 4 ºC overnight with shaking. Membranes were washed three times

453   with 1X PBST (10 min per wash) before treatment with 2º antibody. Both DPP4 and actin blots

454   used rabbit α-goat HRP at 1:10,000 in 1X PBST with 5% milk and were incubated at room

455   temperature for 1 hour with shaking. Membranes were washed three times with 1X PBST (10

456   min per wash) and developed using Amersham ECL Western Blotting detection reagents (GE

457   Healthcare). Membranes were incubated at room temperature for 10 min before imaging using

458   Amersham Hyperfilm ECL (GE Healthcare).

459

460   **Immunofluorescence Assay (IFA)**

461   For IFA analyses, transfections were performed in 50 mm glass-bottom dishes seeded

462   with 2×10^6 HEK 293T cells. At 24 hours post-transfection, cells were washed once with 1X

463   DPBS and then fixed with 2 mL 2% paraformaldehyde (PFA) in 1X DPBS for 30 min at room

464   temperature. Cells were washed once with 2 mL of 100 mM glycine in 1X DPBS and incubated

465   in 2 mL of 100 mM glycine in 1X DPBS for 15 min at room temperature. Cells were washed

466   twice with 1X DPBS before blocking. To each plate, 200 μL of 3% BSA and 10% normal

467   donkey sera in IFA wash (0.05% Tween-20 in 1X DPBS) was added and incubated at room

468   temperature for 1 hour with shaking. Next, cells were incubated with goat α-DPP4 polyclonal

469   antibody (R&D systems) at 1:50 in 3% BSA in IFA wash at room temperature for 1 hour with

470   shaking. Cells were washed three times with IFA wash and then incubated with 2º donkey-anti-

471   goat Alexa Fluor 488 (Life Technologies) at 1:500 for 1 hour at room temperature in the dark

472   with shaking. Cells were washed again three times with IFA wash. To each plate, three drops of

473   ProLong Gold antifade reagent (Invitrogen) were added to the cells before imaging at 20X

474   magnification. Green fluorescence indicates DPP4 staining while blue is DAPI staining for

475   nuclei.

476

477   **Flow cytometry analysis**

478   Transfections were performed as detailed above. At ~20 hours post-transfection, media

479   was removed and cells were washed with 1X DPBS. Cells were removed from plates using non-

480   enzymatic cell dissociation buffer (Gibco). Cells were then resuspended in complete media and

481   centrifuged (all spins at $500 \times g$ for 5 minutes). Supernatants were removed and cells were

482   washed three times with 1X DPBS. Cells were counted and resuspended at a concentration of 5

483   x 10^6 cells/mL. To each well of a 96-well plate (Corning), 200 μL of cell suspension was added

484   in duplicate, plates were spun to pellet cells, and the supernatant was discarded. Cells were

485   suspended in 100 μL Flow Cytometry Staining Buffer (eBiosciences) with 8 μg/mL goat α-DPP4

486   polyclonal antibody (R&D Systems) and incubated at 4 ºC for 1 hour. Cells were washed twice

487   with Staining Buffer and then incubated with 1 μL/well of FITC-conjugated donkey α-goat

488   secondary antibody (Life Technologies) in 100 μl Staining Buffer for 1 hour at 4 ºC. Cells were

489   washed twice with Staining Buffer, and resuspended in 150 μL Staining Buffer. At least 10,000

490   events per well were acquired on the Guava easyCyte HT flow cytometer (Millipore) for each

491   read. The background signal of secondary antibody only (no primary antibody) was used to

492   subtract out the background. Data were analyzed using FlowJo software version 10.3 (Tree Star).

493

494   **Phylogenetic analysis**

495   Amino acid sequences were retrieved from GenBank and aligned using MAFFT (48).

496   The phylogenetic tree was generated using maximum likelihood and 100 bootstrap replicates

497   with the PhyML package. Smart model selection (SMS) was implemented to determine the best

498   model parameters (49). The tree was visualized using EvolView (50). Only bootstrap values ≥50

499   are displayed.

500

501   **Statistics**

502   All quantitative data are presented as means ± one standard deviation. All statistics were

503   performed using the R programming language version 3.1.2.

504

505   **ACKNOWLEDGMENTS**

511

512   **REFERENCES**

513   1.   Pfefferle S, Oppong S, Drexler JF, Gloza-Rausch F, Ipsen A, Seebens A, Muller MA,

514   Annan A, Vallo P, Adu-Sarkodie Y, Kruppa TF, Drosten C. 2009. Distant relatives of

515   severe acute respiratory syndrome coronavirus and close relatives of human coronavirus

516   229E in bats, Ghana. *Emerg. Infect. Dis.* 15:1377–84

517   2.   Huynh J, Li S, Yount B, Smith A, Sturges L, Olsen JC, Nagel J, Johnson JB,

518   Agnihothram SS, Gates JE, Frieman MB, Baric RS, Donaldson EF. 2012. Evidence

519   supporting a zoonotic origin of human coronavirus strain NL63. *J. Virol.* 86:12816–25

520   3.   Cherry JD. 2004. The chronology of the 2002–2003 SARS mini pandemic. *Paediatr.*

521   *Respir. Rev.* 5:262–69

522   4.   Menachery VD, Yount BL, Debbink K, Agnihothram SS, Gralinski LE, Plante JA,

523   Graham RL, Scobey T, Ge XY, Donaldson EF, Randell SH, Lanzavecchia A, Marasco

524   WA, Shi ZL, Baric RS. 2015. A SARS-like cluster of circulating bat coronaviruses shows

525   potential for human emergence. *Nat Med.* 21(12):1508-13

526   5.   Menachery VD, Yount BL, Sims AC, Debbink K, Agnihothram SS, Gralinksi LE,

527   Graham RL, Scobey T, Plante JA, Royal SR, Swanstrom J, Sheahan TP, Pickles RJ, Corti

528   D, Randell SH, Lanzavecchia A, Marasco WA, Baric RS. 2016. SARS-like WIV1-CoV

529   poised for human emergence. *PNAS.* 113(11):3048-53

530   6.   WHO (World Health Organization) 2017. Middle East respiratory syndrome coronavirus

531   (MERS-CoV). Mar. 27, WHO, Geneva. http://www.who.int/emergencies/mers-cov/en/

532   7.   Reusken CBEM, Haagmans BL, Muller MA, Gutierrez C, Godeke GJ, Meyer B, Muth D,

533   Raj VS, Smits-De Vries L, Corman VM, Drexler JF, Smits SL, El Tahir YE, De Sousa R,

534   van Beek J, Nowotny N, van Maanen K, Hidalgo-Hermoso E, Bosch BJ, Rottier P,

535   Osterhaus A, Gortazar-Schmidt C, Drosten C, Koopmans MP. 2013. Middle East

536   respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a

537   comparative serological study. *Lancet Infect. Dis*. 13:859–66

538 8. Gossner C, Danielson N, Gervelmeyer A, Berthe F, Faye B, Kaasik Aaslav K, Adlhoch

539   C, Zeller H, Penttinen P, Coulombier D. 2014. Human-dromedary camel interactions and

540   the risk of acquiring zoonotic Middle East respiratory syndrome coronavirus infection.

541   *Zoonoses Public Health.* 63(1):1-9

542 9. Raj VS, Mou H, Smits SL, Dekkers DH, Muller MA, Annan A, Vallo P, Adu-Sarkodie

543   Y, Kruppa TF, Drosten C. 2013. Dipeptidyl peptidase 4 is a functional receptor for the

544   emerging human coronavirus-EMC. *Nature.* 495:251–54

545 10. de Wit E, Prescott J, Baseler L, Bushmaker T, Thomas T, Lackemeyer MG, Martellaro C,

546   Milne-Price S, Haddock E, Haagmans BL, Feldmann H, Munster VJ. 2013. The Middle

547   East respiratory syndrome coronavirus (MERS-CoV) does not replicate in Syrian

548   hamsters. *PLoS ONE.* 8:e69127

549 11. Barlan A, Zhao J, Sarkar MK, Li K, McCray PB Jr, Perlman S, Gallagher T. 2014.

550   Receptor variation and susceptibility to Middle East respiratory syndrome coronavirus

551   infection. *J. Virol.* 88:4953–61

552 12. Eckerle I, Corman VM, Muller MA, Lenk M, Ulrich RG, Drosten C. 2014. Replicative

553   capacity of MERS coronavirus in livestock cell lines. *Emerg. Infect. Dis.* 20:276–79

554 13. Cockrell AS, Peck KM, Yount BL, Agnihothram SS, Scobey T, Curnes NR, Baric RS,

555   Heise MT. 2014. Mouse dipeptidyl peptidase 4 is not a functional receptor for Middle

556   East respiratory syndrome coronavirus infection. *J. Virol.* 88:5195–99

557   14. Coleman CM, Matthews KL, Goicochea L, Frieman MB. 2014. Wild-type and innate

558   immune-deficient mice are not susceptible to the Middle East respiratory syndrome

559   coronavirus. *J. Gen. Virol.* 95:408–12

560   15. Raj VS, Smits SL, Provacia LB, van den Brand JM, Wiersma L, Ouwendijk WJ,

561   Bestebroer TM, Spronken MI, van Amerongen G, Rottier PJ, Fouchier RA, Bosch BJ,

562   Osterhaus AD, Haagmans BL. 2014. Adenosine deaminase acts as a natural antagonist

563   for dipeptidyl peptidase 4-mediated entry of the Middle East respiratory syndrome

564   coronavirus. *J. Virol.* 88:1834–38

565   16. van Doremalen N, Miazgowicz KL, Milne-Price S, Bushmaker T, Robertson S, Scott D,

566   Kinne J, McLellan JS, Zhu J, Munster VJ. 2014. Host species restriction of Middle East

567   respiratory syndrome coronavirus through its receptor, dipeptidyl peptidase 4. *J. Virol*.

568   88:9220–32

569   17. van Doremalen N, Miazgowicz KL, Munster VJ. 2016. Mapping the specific amino acid

570   residues that make hamster DPP4 functional as a receptor for Middle East respiratory

571   syndrome coronavirus. *J. Virol.* 90: 5499-5502

572   18. Zhao JC, Li K, Wohlford-Lenane C, Agnihothram SS, Fett C, Zhao J, Gale MJ Jr, Baric

573   RS, Enjuanes L, Gallagher T, McCray PB Jr, Perlman S. 2014. Rapid generation of a

574   mouse model for Middle East respiratory syndrome. *PNAS.* 111:4970–75

575   19. Peck KM, Cockrell AS, Yount BL, Scobey T, Baric RS, Heise MT. 2015. Glycosylation

576   of mouse DPP4 plays a role in inhibiting Middle East respiratory syndrome coronavirus

577   infection. *J. Virol.* 89:4696–99

578   20. Agrawal AS, Garron T, Tao X, Peng BH, Wakamiya M, Chan TS, Couch RB, Tseng CT.

579   2015. Generation of transgenic mouse model of Middle East respiratory syndrome-

580   coronavirus infection and disease. *J. Virol.* 89:3659–70

581   21. Pascal KE, Coleman CM, Mujica AO, Kamat V, Badithe A, Fairhurst J, Hunt C, Strein J,

582   Berrebi A, Sisk JM, Matthews KL, Babb R, Chen G, Lai KM, Huang TT, Olson W,

583   Yancopoulos GD, Stahl N, Frieman MB, Kyratsous CA. 2015. Pre- and postexposure

584   efficacy of fully human antibodies against Spike protein in a novel humanized mouse

585   model of MERS-CoV infection. *PNAS.* 112(28):8738-8743

586   22. Li K, Wohlford-Lenane C, Perlman S, Zhao J, Jewell AK, Reznikov LR, Gibson-Corley

587   KN, Meyerholz DK, McCray PB Jr. 2016. Middle East respiratory syndrome coronavirus

588   causes multiple organ damage and lethal disease in mice transgenic for human dipeptidyl

589   peptidase 4. *J. Infect. Dis.* 213(5):712-722

590   23. Coleman CM, Sisk JM, Halasz G, Zhong J, Beck SE, Matthews KL, Venkataraman T,

591   Rajagopalan S, Kyratsous CA, Frieman MB. 2017. CD8+ T cells and macrophages

592   regulate pathogenesis in a mouse model of Middle East respiratory syndrome. *J. Virol.*

593   91:e01825-16

594   24. Cockrell AS, Yount BL, Scobey T, Jensen K, Douglas M, Beall A, Tang XC, Marasco

595   WA, Heise MT, Baric RS. 2016. A mouse model for MERS coronavirus-induced acute

596   respiratory distress syndrome. *Nature Microbiol.* 2:16226

597   25. de Wit E, Rasmussen AL, Falzarano D, Bushmaker T, Feldmann F, Brining DL, Fischer

598   ER, Martellaro C, Okumura A, Chang J, Scott D, Benecke AG, Katze MG, Feldmann H,

599   Munster VJ. 2013. Middle East respiratory syndrome coronavirus (MERS-CoV) causes

600   transient lower respiratory tract infection in rhesus macaques. *PNAS.* 110:16590-603

601 26. Falzarano D, de Wit E, Rasmussen AL, Feldmann F, Okumara A. Scott DP, Brining D,

602   Bushmaker T, Martellaro C, Baseler L, Benecke AG, Katze MG, Munster VJ, Feldmann

603   H. 2013. Treatment with interferon-a2b and ribavirin improves outcome in MERS-CoV-

604   infected rhesus macaques. *Nat. Med.* 19:1313-17

605 27. Falzarano D, de Wit E, Feldmann F, Rasmussen AL, Okumara A, Peng X, Thomas MJ,

606   van Doremalen N, Haddock E, Nagy L, LaCasse R, Liu T, Zhu J, McLellan JS, Scott DP,

607   Katze MG, Feldmann H, Munster VJ. 2014. Infection with MERS-CoV causes lethal

608   pneumonia in the common marmoset. *PLoS Pathog.* 10:e1004250

609 28. Chan JFW, Yao Y, Yeung ML, DDeng W, Bao L, Jia L, Li F, Xiao C, Gao H, Yu P, Cai

610   JP, Chu H, Zhou J, Chen H, Qin C, Yuen KY. 2015. Treatment with lopinavir/ritonavir or

611   interferon-B1b improves outcome of MERS-CoV infection in a nonhuman primate model

612   of common marmoset. *J. Infect. Dis*. 212(12):1904-13

613 29. Wang N, Shi X, Jiang L, Zhang S, Wang D, Tong P, Guo D, Fu L, Cui Y, Liu X, Arledge

614   KC, Chen YH, Zhang L, Wang X. 2013. Structure of MERS-CoV spike receptor-binding

615   domain complexed with human receptor DPP4. *Cell Res.* 23:986–93

616 30. Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC*

617   *Bioinformatics* 9:40

618 31. Cui J, Eden JS, Holmes EC, Wang LF. 2013. Adaptive evolution of bat dipeptidyl

619   peptidase 4 (dpp4): implications for the origin and emergence of Middle East respiratory

620   syndrome coronavirus. *Virol. J.* 10:304

621 32. Peck KM, Burch CL, Heise MT, Baric RS. 2015b. Coronavirus host range expansion and

622   Middle East respiratory syndrome coronavirus emergence: biochemical mechanisms and

623   evolutionary perspectives. *Annu. Rev. Virol* 2:95-117

624   33. Gupta R, Jung E, Brunak S. 2004. NetNGlyc 1.0 Server. Center for biological sequence

625   analysis, technical university of Denmark. http://www.cbs.dtudk/services/NetNGlyc.

626   34. Song W, Wang Y, Wang N, Wang D, Guo J, Fu L, Shi X. 2014. Identification of residues

627   on human receptor DPP4 critical for MERS-CoV binding and entry. *Virology* 471-

628   473:49-53

629   35. Wang CC, Chen JR, Tseng YC, Hsu CH, Hung YF, Chen SW, Chen CM, Khoo KH,

630   Cheng TJ, Cheng YS, Jan JT, Wu CY, Ma C, Wong CH. 2009. Glycans on influenza

631   hemagglutinin affect receptor binding and immune response. *PNAS.* 106(43):18137-42

632   36. Yang Y, Du L, Liu C, Wang L, Ma C, Tang J, Baric RS, Jiang S, Li F. 2014. Receptor

633   usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human

634   transmission of MERS-Coronavirus. *PNAS.* 111(34):12516-21

635   37. Jayaraman A, Koh X, Li J, Raman R, Viswanathan K, Shriver Z, Sasisekharan R. 2012.

636   Glycosylation at Asn91 of H1N1 haemagglutinin affects binding to glycan receptors.

637   *Biochem J.* 444:429-435

638   38. Kaelber JT, Demogines A, Harbison CE, Allison AB, Goodman LB, Ortega AN, Sawyer

639   SL, Parrish CR. 2012. Evolutionary reconstructions of the trasnferrin receptor of

640   Caniforms supports canine parvovirus being a re-emerged and not a novel pathogen in

641   dogs. *PLoS Pathog.* 8(5):e1002666

642   39. Goodman LB, Lyi SM, Johnson NC, Cifuente JO, Hafenstein SL, Parrish CR. 2010.

643   Binding site on the transferrin receptor for the parvovirus capsid and effects of altered

644   affinity on cell uptake and infection. *J. Virol.* 84(10):4969-78

645 40. Allison AB, Organtini LJ, Zhang S, Hafenstein SL, Holmes EC, Parrish CR. 2015. Single

646   mutations in the VP2 300 loop region of the three-fold spike of the carnivore parvovirus

647   capsid can determine host range. *J. Virol.* 90(2):753-67

648 41. Wentworth DE and Holmes KV. 2001. Molecular determinants of species specificity in

649   the coronavirus receptor aminopeptidase N (CD13): influence of N-linked glycosylation.

650   *J. Virol.* 75(20):9741-52

651 42. Tussell SM, Schittone SA, Holmes KV. 2007. Mutational analysis of aminopeptidase N,

652   a receptor for several group 1 coronaviruses, identifies key determinants of viral host

653   range. *J. Virol.* 81(3):1261-73

654 43. Li W, Zhang C, Sui J, Kuhn JH, Moore MJ, Luo S, Wong S, Huang I, Xu K, Vasilieva N,

655   Murakami A, He Y, Marasco WA, Guan Y, Choe H, Farzan M. 2005. Receptor and viral

656   determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J.* 24(8):1634-43

657 44. Lu G, Hu Y, Want Q, Qi Jianxun, Gao F, Li Y, Zhang Y, Zhang W, Yuan Y, Bao J,

658   Zhang B, Shi Y, Yan J, Gao GF. 2013. Molecular basis of binding between novel human

659   coronavirus MERS-CoV and its receptor CD26. *Nature*. 500:227-231

660 45. Park JE, Li K, Barlan A, Fehr AR, Perlman S, McCray PB Jr, Gallagher T. 2016.

661   Proteolytic processing of Middle East respiratory syndrome coronavirus spikes expands

662   virus tropism. *PNAS.* 113(43):12262-7

663 46. Scobey T, Yount BL, Sims AC, Donaldson EF, Agnihothram SS, Menachery VD,

664   Graham RL, Swanstrom J, Bove PF, Kim JD, Grego S, Randell SH, Baric RS. 2013.

665   Reverse genetics with a full-length infectious cDNA of the Middle East respiratory

666   syndrome coronavirus. *PNAS.* 110:16157–62

667   47. Gibson DB, Young L, Chuang RY, Venter JC, Hutcison CA 3rd, Smith HO. Enzymatic

668   assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* 5:343-5

669   48. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid

670   multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.*

671   (30):3059-3066

672   49. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New

673   algorithms and methods to estimate maximum-likelihood phylogenies: assessing the

674   performance of PhyML 3.0. *Syst Biol.* 59:307-21

675   50. Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. 2012. EvolView, an online tool for

676   visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* 40:W569-

677   72.

678

679   **FIGURE LEGENDS**

680

681   **FIG 1**. Permissivity of DPP4 orthologs to MERS-CoV. (A) Seven DPP4 orthologs were tested

682   for their ability to support infection by rMERS-CoV-RFP. DPP4 constructs were transfected into

683   HEK 293T cells and infected at an MOI of 5 ~20 hours post-transfection. Cells were imaged for

684   fluorescence ~24 hpi. hDPP4, human DPP4; cDPP4, camel DPP4; bDPP4, bat DPP4; mDPP4,

685   mouse DPP4; fDPP4, ferret DPP4; haDPP4, hamster DPP4; gpDPP4, guinea pig DPP4. (B)

686   Mean fluorescent cell counts of MERS-CoV infection utilizing various DPP4 orthologs. Cells

687   were infected at an MOI of 0.1 and the number of infected cells counted at ~72 hpi. Each DPP4

688   ortholog was measured in triplicate. Only hDPP4, bDPP4, and cDPP4 have significantly higher

689   levels of infection compared to no DPP4 ($p < 0.05$, Student's t-test, indicated by *). All DPP4

690   orthologs have significantly lower levels of infection compared to hDPP4 ($p < 0.05$, Student's t-

691   test). Levels of infection are not significantly different between bDPP4 and cDPP4. Error bars

692   indicate mean values ± one standard deviation.

693

694   **FIG 2**. Sequence and structural comparison of nonpermissive DPP4 orthologs. (A) Structural

695   comparison of threaded molecules (30) mDPP4 (orange), fDPP4 (green), haDPP4 (blue), and

696   gpDPP4 (purple) overlaid on hDPP4 (yellow) complexed with the MERS-CoV RBD (red) (PDB

697   code 4L72). (B) Sequence alignment of permissive (human, camel, bat; blue) and nonpermissive

698   (mouse, ferret, hamster, guinea pig; red) DPP4 amino acid sequences. Residue 330 is numbered

699   relative to mDPP4. Boxes represent glycosylation sites that are either unique to nonpermissive

700   species (black) or shared with a permissive species (gray). (C) hDPP4 (yellow) complexed with

701   the MERS-CoV RBD (red) (PDB code 4L72). Residues aligning to the ferret (green), hamster

702   and mouse (blue), and guinea pig (purple) glycosylation sites are highlighted. Dashed line circles

703   indicate the regions of the DPP4 molecule that corresponds to blades IV and V.

704

705   **FIG 3**. DPP4 ortholog glycosylation knockout mutants. (A) Neither wildtype nor glycosylation

706   knockout DPP4 molecules for ferret (fDPP4), hamster (haDPP4), or guinea pig (gpDPP4)

707   support infection by MERS-CoV. (B) Successful removal of glycosylation is supported by a ~2.5

708   kDa downward shift seen via Western blot. Top blot represents DPP4 and the bottom blot

709   represents β-actin as a control. (C) Fluorescent cell counts of MERS-CoV infection utilizing

710   various DPP4 orthologs and their respective glycosylation knockout mutants. Cells were infected

711   at an MOI of 0.1 and number of infected cells counted at 72 hpi. Each DPP4 ortholog was

712   measured in triplicate. Only hDPP4 has significantly higher levels of infection compared to no

713   DPP4 ($p < 0.05$, Student's t-test, indicated by *). The remaining DPP4 orthologs have infection

714   levels that are not significantly different from no DPP4 or each other. Error bars indicate mean

715   values ± one standard deviation.

716

717   **FIG 4.** DPP4 and mutant variants are expressed on the surface of cells as evidence by

718   immunofluorescence assay (A) and flow cytometry (B, C). (A) Cells were transfected with each

719   DPP4 ortholog, fixed, and probed with primary goat-anti-DPP4 polyclonal antibody (R&D

720   Systems) at 1:50 and secondary donkey-α-goat Alexa Fluor 488 (Life Technologies) at 1:500.

721   Cells were imaged at 20X for DAPI (300 ms exposure) and DPP4 (1.5 s exposure). (B) DPP4

722   expression frequencies (blue outlined histogram) by DPP4 construct after subtraction of

723   background from replicate wells stained with secondary donkey-α-goat IgG (H+L) Alexa Fluor

724   488 antibody only (grey shaded histogram). Percentage values represent the average across two

725   duplicate wells. (C) Geometric mean fluorescence intensity of the DPP4 positive populations for

726   each DPP4 construct.

727

728   **FIG 5.** Many amino acid changes are required to make fDPP4 and haDPP4 permissive to

729   MERS-CoV infection. (A) Removing glycosylation on its own does not confer permissivity to

730   haDPP4. However, combining three amino acid changes on blade V (starting at residue 289)

731   with the glycosylation knockout mutant on blade IV (N332A) results in high levels of MERS-

732   CoV infection. Sequences show the alignment between hDPP4 and haDPP4 with the blue boxes

733   indicating the amino acids that were swapped from hDPP4 into haDPP4. (B) Removing

734   glycosylation on its own does not confer permissivity to fDPP4. However, introducing a set of

735   amino acid changes on blade V (starting at residue 278) and blade IV (starting at residue 330)

736   allows fDPP4 to support MERS-CoV infection (fDPP4 (278)(330)). Sequences show the

737   alignment between hDPP4 and fDPP4 with the blue boxes indicating the amino acids that were

738   swapped from hDPP4 into fDPP4. Note that fDPP4 –gly is a negative control and only includes

739   the single point mutation N331A. (C) Western blot analysis of fDPP4 and haDPP4 and

740   designated variants for DPP4 and β-actin expression. Successful glycosylation knockout is

741   indicated by a downward shift of ~2.5 kDa. (D) Fluorescent cell counts of MERS-CoV infection

742   utilizing DPP4 orthologs. Cells were infected at an MOI of 0.1 and number of red cells counted

743   at 72 hpi. Each DPP4 ortholog was measured in triplicate. hDPP4, fDPP4 (278)(330), and

744   haDPP4 (289),-gly have levels of infection that are significantly greater than no DPP4 ($p < 0.05$,

745   Student's t-test, indicated by *). fDPP4 –gly and haDPP4 –gly infection levels are not

746   significantly different from no DPP4. Error bars indicate mean values ± one standard deviation.

747

748   **FIG 6.** Bat and guinea pig DPP4 share the same glycosylation site downstream of the site

749   identified to be important in mDPP4 (Figure 2B). (A) Removing the glycosylation site from

750   bDPP4 shows no decrease in infection while removing glycosylation from gpDPP4 resulted in

751   no increase in infection. (B) Western blot analysis of bDPP4 and gpDPP4 and their respective

752   glycosylation knockout mutants for DPP4 and β-actin expression. Successful glycosylation

753   knockout is indicated by a downward shift of ~2.5 kDa. (C) Fluorescent cell counts of MERS-

754   CoV infection utilizing various DPP4 orthologs. Cells were infected at an MOI of 0.1 and

755   number of infected cells counted at 72 hpi. Each DPP4 ortholog was measured in triplicate.

756   hDPP4, bDPP4, and bDPP4 – gly have levels of infection that are significantly higher than no

757   DPP4 ($p < 0.05$, Student's t-test, indicated by *). gpDPP4 and gpDPP4 –gly infection levels are

758   not significantly different from no DPP4. (D) DPP4 and mutant variants are expressed on the

759   surface of cells, visible by immunofluorescence. Cells were transfected with each DPP4

760   ortholog, fixed, and probed with primary goat-anti-DPP4 polyclonal antibody (R&D Systems) at

761   1:50 and secondary donkey-anti-goat Alexa Fluor 488 (Life Technologies) at 1:500. Cells were

762   imaged at 20X for DAPI (300 ms exposure) and DPP4 (1.5 s exposure). Error bars indicate mean

763   values ± one standard deviation.

764

765   **FIG 7.** DPP4 protein phylogenetic tree based on amino acid sequences. Shaded colors indicate

766   the group each species falls in. Blue: reptiles and amphibians; Green: avian species; Orange:

767   other mammals; Red: Chiroptera (bats); Purple: ungulates; Gray: rodents; Pink: primates.

768   Colored circles to the right of the species names indicate whether the sequence has a

769   glycosylation site upstream (first column), at the same site (second column), or downstream

770   (third column) of the NXT glycosylation site in mDPP4 (residues 332-334). Numbers inside the

771   circle designate how many amino acids upstream (or downstream) the N of the NXT or NXS

772   glycosylation site is. For the second column, a circle indicates that there is a glycosylation site

773   aligning to the site present in mDPP4. Squares in the rightmost column indicate permissive

774   (green) or non-permissive (red) species, as determined from either *in vivo* or *in vitro* studies.

775

776   **TABLES**

hDPP4		85	92	150		219	229			281	**321**						520
bDPP4		82	89	147		216	226	272			**315**		332	393	432	490	
mDPP4		83	90	144			223				**315**	328					514
fDPP4		84	91		178		228		280		**320**				437		519
haDPP4		83	90	148			227	275			**319**	332					518
gpDPP4	32	295	302	360			439			491	**531**		548				730

777

778   **Table 1:** Known and putative N-linked glycans in DPP4 orthologs

779 Numbers indicate the residues known (hDPP4) or predicted (all other DPP4 orthologs) (33) to be

780 N-linked glycosylated for each DPP4 molecule. All residues are numbered relative to the given

781 DPP4 molecule, with columns representing residues that align. Out of the eight N-linked glycans

782 in hDPP4, only four are conserved across all orthologs. Bold font indicates sites that align to

783 within two residues of a contact residue in hDPP4 (see 29).

784

Ortholog	Important residues									Source
Human DPP4	267			294	295		336			(34)
Mouse DPP4				294			336*			(13, 24)
Hamster DPP4		291			295		336*	341	346	(16, 17)
			293		295	297	334*			This study
Ferret DPP4	279 - 295						331-341			This study
	246 - 505									(15)

785

786 **Table 2:** Residues identified to be important for MERS-CoV permissivity

787 Experimental studies and the residues that have been identified as important for mediating

788 permissivity to MERS-CoV among various DPP4 orthologs *in vitro*. All residues are relative to

789 the aligning residue in hDPP4. Residues are either on blade IV (blue shading) or on blade V

790 (gray shading). * indicates residues that knock out a glycosylation site in the mouse or hamster

791 DPP4 molecules.

792

A

B

A

mouse   ferret   hamster   guinea pig

Blade V

Blade IV
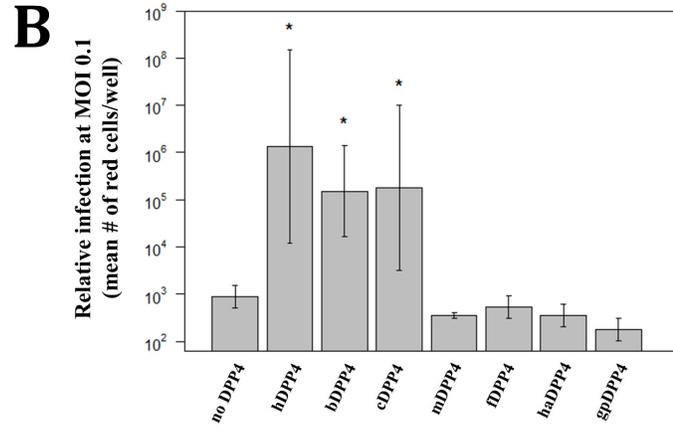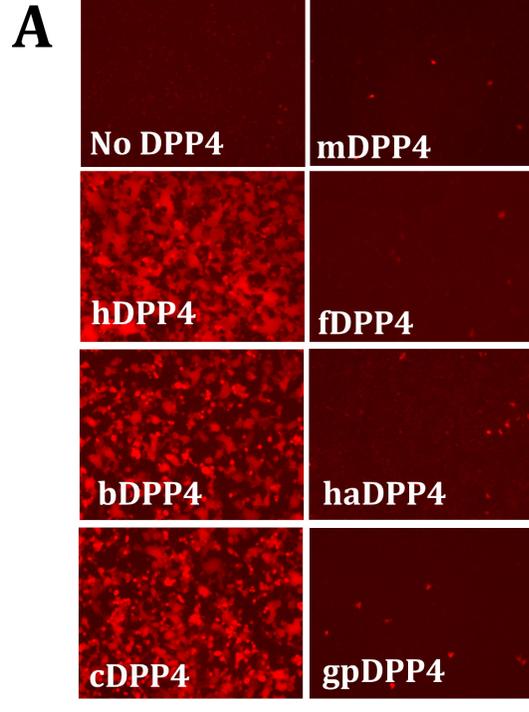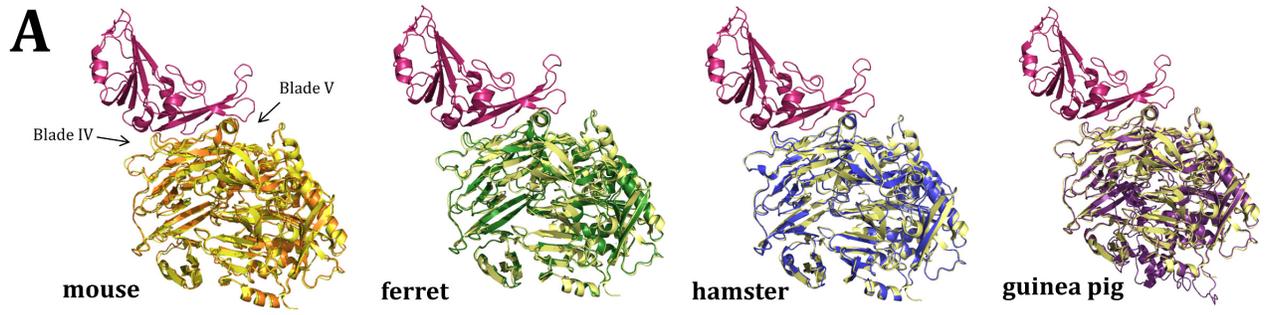
B

330

human	C	D	Y	D	E	S	S	G	R	W	N	C	L	V	A	R	Q	H
camel	C	D	Y	D	E	S	T	G	R	W	A	S	S	V	G	R	Q	H
bat	C	D	Y	D	A	P	N	S	K	W	N	C	S	V	P	R	Q	H
mouse	C	D	Y	D	K	I	N	L	T	W	N	C	P	S	E	Q	Q	H
ferret	C	D	Y	N	N	D	T	N	S	W	R	K	P	E	A	Q	E	H
hamster	C	D	Y	N	K	T	N	L	T	W	N	C	P	L	A	Q	Q	H
guinea pig	C	D	Y	D	K	S	R	L	G	W	N	C	S	L	K	N	E	H

▢ Glycosylation site

C

Blade V

Blade IV

A

No DPP4

hDPP4

fDPP4

haDPP4

gpDPP4

fDPP4 -gly

haDPP4 -gly

gpDPP4 -gly

B

No DPP4

hDPP4

fDPP4

fDPP4 -gly

haDPP4

haDPP4 -gly

gpDPP4

gpDPP4 - gly

C

Relative infection at MOI 0.1
(mean # of red cells/well)

10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2

*

no DPP4

hDPP4

fDPP4

fDPP4 -gly

haDPP4

haDPP4 -gly

gpDPP4

gpDPP4 -gly

A

hDPP4   fDPP4   haDPP4   gpDPP4

no DPP4   fDPP4 -gly   haDPP4 -gly   gpDPP4 -gly

C

Geometric mean fluorescent intensity

no DPP4, hDPP4, bDPP4, fDPP4, fDPP4 -gly, haDPP4, haDPP4 -gly, gpDPP4, gpDPP4 -gly

B

hDPP4 66%   bDPP4 49%   fDPP4 78%   fDPP4 -gly 82%   haDPP4 59%   haDPP4 -gly 68%   gpDPP4 86%   gpDPP4 -gly 84%

% of Max

DPP4 - Alexa Fluor 488

A

haDPP4 -gly

haDPP4 (289), -gly

289
hDPP4 QITAP ASMLI GDHY
haDPP4 QVTAP ESVLT GDHY

332
hDPP4 CDYDES S GRWNCLV
haDPP4 CDYNK T N LTWNCPL

B

fDPP4 -gly

fDPP4 (278)(330)

278
hDPP4 SS VTNATSIQITAPASMLI GD
fDPP4 DP NTNATSVEITPPDAIST GD

330
hDPP4 CDYDE SSGRWNCLV ARQHIEMS
fDPP4 CDYNN NDTNSWRKPE AQEHTEMS

C

fDPP4
fDPP4 (278)
fDPP4 -gly
fDPP4 (278)(330)
haDPP4
haDPP4 (289)
haDPP4 -gly
haDPP4 (289), -gly

D

Relative infection at MOI 0.1
(mean # of red cells/well)

10^9
10^8
10^7
10^6
10^5
10^4
10^3
10^2

*

*

*

no DPP4
hDPP4
fDPP4 -gly
fDPP4 (278)(330)
haDPP4 -gly
haDPP4 (289)-gly

A

No DPP4   hDPP4

bDPP4   bDPP4 -gly

gpDPP4   gpDPP4 -gly

B

No DPP4   hDPP4   bDPP4   bDPP4 -gly   gpDPP4   gpDPP4 -gly

C

bDPP4   bDPP4 -gly

gpDPP4   gpDPP4 -gly

D

Relative infection at MOI 0.1
(mean # of red cells/well)

no DPP4   hDPP4   bDPP4   bDPP4-gly   gpDPP4   gpDPP4 -gly

upstream · same · downstream

Taxon		
Zebrafish		2
Northern pike		
Blackstripe livebearer	3	
Common garter snake		
Chinese soft-shelled turtle		
Chinese alligator		
American alligator		
Japanese quail		2
Chicken		2
Mallard	2	2
Swan goose	2	2
Peregrine falcon		
Saker falcon		
Common starling		2
Medium ground-finch	2	
Zebra finch	2	
Barn owl		
Rock pigeon	2	2
Chimney swift		
Little egret		
Rhinoceros hornbill	2	
Golden eagle		2
White-tailed eagle		2
Bald eagle		2
Platypus		
Gray short-tailed opossum		
Tasmanian devil		4
African savanna elephant		
Star-nosed mole		
European shrew		
Horse	2	
Egyptian rousette		
Large flying fox		
Black flying fox		
Jamaican fruit-eating bat		
Natal long-fingered bat		
Common pipistrelle (bat)	4	
Davids Myotis (vesper bat)	3	
Little brown bat	3	
Brandts bat	3	
Cattle		
Bison		
Wild yak		
Goat		
Mouflon sheep		
Sheep		
Minke whale		
Sperm whale		
Killer whale	3	
Domestic pig		
Alpaca		
Arabian camel		
Bactrian camel		
Wild Bactrian camel		
Domestic Cat	2	
Cheetah	2	
Dog	3	
Pacific walrus	2	
Domestic ferret	2	
Giant panda	2	
Chinese tree shrew		4
American pika		
Ords kangaroo rat		
Egyptian jerboa	3	
Mouse		
Prairie deer mouse	3	
Prairie vole		
Chinese hamster	3	
Golden hamster	3	
Damara mole rat		4
Naked mole rat		4
Long-tailed chinchilla	3	
Degu		4
Domestic guinea pig		4
Alpine marmot		
Small-eared galago		
Coquerels sifaka		
Gray mouse lemur		
Philippine tarsier		
White-tufted-ear marmoset		
Bolivian squirrel monkey		
Mas night monkey		
Northern white-cheeked gibbon		
Sumatran orangutan		
Western lowland gorilla		
Humans		
Chimpanzee		
Pygmy chimpanzee		
Golden snub-nosed monkey		
Angola colobus		
Green monkey		
Olive baboon		
Mandrill		
Sooty mangabey		
Pig-tailed macaque		
Rhesus macaque		
Crab-eating macaque		