1   Recurrent Loss of APOBEC3H Activity during Primate Evolution

2

3   Erin I. Garcia[a,b] and Michael Emerman[a,b,#]

4   [a] Department of Microbiology, University of Washington, Seattle, USA

5   [b] Divisions of Human Biology and Basic Sciences, Fred Hutchinson Cancer Research

6   Center, Seattle, USA

7

8   **Running Head:** Loss of APOBEC3H Activity during Evolution

9

10   [#]To whom correspondence should be addressed: Michael Emerman

11   Division of Human Biology, Fred Hutchinson Cancer Research Center, 1100 Fairview

12   Ave. N, P.O. Box 10924, Seattle, WA 98109-1024.

13   Phone: (206) 667-5058. Fax: (206) 667-6523.

14   E-mail: memerman@fredhutch.org

15   **Word count**

16   Abstract: 177

17   Importance: 104

18   Text: 5,544

19   Figures: 6, Tables: 2

20

21   **Abstract**

22   Genes in the *APOBEC3* family encode cytidine deaminases that provide a barrier

23   against viral infection and retrotransposition. Of all *APOBEC3* genes in humans,

24   *APOBEC3H* (*A3H*) is the most polymorphic: some haplotypes encode stable and active

25   A3H proteins, while others are unstable and inactive. Such variation in human A3H

26   affects interactions with the lentiviral antagonist Vif, which counteracts A3H via

27   proteasomal degradation. In order to broaden our understanding of A3H-Vif interactions

28   as well as its evolution in Old World monkeys, we characterized A3H variation within

29   four African green monkey (AGM) subspecies. We found that A3H is highly polymorphic

30   in AGMs and has lost antiviral activity in multiple Old World monkeys. This loss of

31   function was partially related to protein expression levels but was also influenced by

32   amino acid mutations in the N-terminus. Moreover, we demonstrate that the evolution of

33   A3H in the primate lineages leading to AGMs was not driven by Vif. Our work suggests

34   that activity of A3H is evolutionarily dynamic and may have a negative effect on host

35   fitness, resulting in its recurrent loss in primates.

36

37   **Importance**

38   Adaptation of viruses to their hosts is critical for transmission of viruses between

39   different species. Previous studies had identified changes in a protein from the

40   APOBEC3 family that influenced species-specificity of simian immunodeficiency viruses

41   (SIVs) in African green monkeys. We studied the evolution of a related protein in the

42   same system, APOBEC3H, which has experienced a loss of function in humans. This

43   evolutionary approach revealed that recurrent loss of APOBEC3H activity has taken

44   place during primate evolution suggesting that APOBEC3H places a fitness cost on

45   hosts. The variability of APOBEC3H activity between different primates highlights the

46   differential selective pressures on the *APOBEC3* gene family.

47

Introduction

The seven members of the *APOBEC3 (A3)* gene family in primates encode cytidine deaminases involved in innate immune defense against retroviruses and retroelements (1-3). Four A3 enzymes are known to potently restrict the replication of lentiviruses like simian immunodeficiency virus (SIV) or human immunodeficiency virus (HIV): A3D, A3F, A3G, and A3H (4). A3 proteins are packaged into budding virions and cause G-to-A hypermutation of viral DNA, although deamination-independent modes of restriction have also been characterized (5). Hypermutation of viral DNA results in detrimental mutations that render the virus inactive and thus protects new cells from infection. However, lentiviruses have evolved a mechanism to evade A3 restriction by encoding a viral antagonist, Vif, which binds and targets A3 proteins for proteasomal degradation via a cellular E3 ubiquitin ligase complex (6). Given that A3-Vif interactions often act in a species-specific manner, adaptation of Vif to host A3 proteins is important for successful adaptation of lentiviruses to their hosts (7-11).

Various A3 proteins differ in their ability to restrict viral infection. For instance, A3G is the most potent A3-mediated inhibitor of HIV-1, while A3A and A3B have limited antiviral potential (4, 12). In humans, the A3H protein is especially remarkable because multiple polymorphisms drastically impact its antiviral activity (13-16). Two independent mutations have occurred in human evolution that destabilized A3H (13); haplotypes that encode an R105G mutation or a deletion of amino acid 15 make unstable proteins (haplotypes I, III, IV, VI) that have lost antiviral activity, while those without these changes make stable proteins (haplotypes II, V, VII) that potently restrict HIV-1 (13-15, 17). Stability and antiviral activity of human A3H has been further linked to subcellular localization: unstable/inactive proteins are more nuclear, while stable/active proteins remain cytoplasmic (18). A3H haplotype I has been associated with breast and lung cancer (19), as has another nuclear A3, A3B (20)**,** suggesting that in some cases A3 activity may be detrimental for the host. While such events have occurred in humans, examples of gains or losses of A3 activity over evolutionary time in other primates have been less explored.

The *A3H* genetic polymorphisms present in human populations also impact the interactions between human A3H and HIV-1 Vif (17, 21-23). Stable A3H proteins are

79   only partially susceptible to degradation by Vif from the LAI isolate of HIV-1 and not at

80   all by HIV-1 NL4-3 Vif (21, 23). However, studies using human cohorts encoding

81   different haplotypes of *A3H* have shown that Vif proteins from primary virus strains

82   isolated from patients with *A3H* haplotype II are able to antagonize stable proteins,

83   while those from patients with unstable haplotypes cannot. This suggests that there is

84   selection *in vivo* for Vif strains that counteract the stably expressed forms of A3H in

85   infected people (22, 24, 25). Furthermore, the cross-species transmissions that led to

86   adaptation of SIV from monkeys to chimpanzees to humans, giving rise to HIV-1,

87   involved adaptation of Vif to antagonize the A3 proteins found in each host (7, 8, 10, 11)

88   including adaptation of Vif from SIVcpz to antagonize human A3H (11).

89   The study of the evolution of lentiviral-host interactions within Old World monkeys

90   has provided insights into the longer-term dynamics of the evolutionary arms race

91   between host antiviral proteins and their lentiviral targets (7, 9, 26). African green

92   monkeys (AGMs), in particular, provide a unique opportunity to assess the evolutionary

93   forces governing interactions between lentiviruses and their hosts since the genus

94   *Chlorocebus* encompasses four geographically distinct subspecies (vervets, sabaeus,

95   tantalus, and grivets) each of which are infected with species-specific subtypes of

96   SIVagm (27-29). These SIV subtypes are named by the subspecies they infect:

97   SIVagm.ver, SIVagm.sab, SIVagm.tan, and SIVagm.gri (29). Furthermore, this system

98   is particularly powerful because although these subspecies are closely genetically

99   related, enough divergence exists that has allowed species-specific lentiviral infections

100   to occur.

101   Previous studies have demonstrated that many genes involved in antiviral

102   immunity are polymorphic in AGMs and some changes at the protein level are critical for

103   interactions with viral antagonists (8, 30, 31). A3G in particular was discovered to

104   encode species-specific polymorphisms (8). Amino acid changes in the Vif binding

105   domain of AGM A3G (sites 128 and 130) confer protection against SIVagm strains from

106   other subspecies. For example, K128E, found in grivet monkeys, is resistant to all Vifs

107   except SIVagm.gri, while D130H, found in sabaeus monkeys, is resistant to Vifs from

108   SIVagm.ver and SIVagm.tan (8). This demonstrates that Vif continues to drive the

4

109   evolution of A3G and contributes to the species-specificity of lentiviruses in AGM

110   populations. It is unknown whether A3H has a similar role in these primates.

111   In this study, we asked whether we could identify a host-virus "arms-race"

112   between A3H in AGMs and Vif proteins encoded by the SIVs that infect these species.

113   Surprisingly, we found that although A3H is highly polymorphic in AGMs, the antiviral

114   activity of A3H has been largely lost in AGMs. The reduced antiviral activity of AGM

115   A3H is in part caused by lower protein expression levels, although amino acid changes

116   also lower antiviral activity independent of protein levels. By reconstructing ancestral

117   A3H proteins spanning evolution in AGMs and in other Old World monkeys, we find that

118   there has been recurrent loss of A3H activity in some, but not all, Old World monkeys.

119   While higher expression levels generally increase viral inhibition, we also identified

120   amino acids that affect A3H restriction without increasing protein expression, which map

121   to regions implicated in RNA binding (32-34). Thus, our data support a model where

122   A3H antiviral activity has been repeatedly lost throughout evolution. This argues that

123   there is a longer scale dynamic between the cost and benefit for A3H function in

124   primates that is not necessarily driven by interactions with its antagonist.

125

126 **Results**

127 **A3H is highly polymorphic in AGM subspecies, but all tested alleles have low**

128 **antiviral activity.**

129   Polymorphisms in human *A3H* are known to affect its antiviral activity as well as

130 its interactions with Vif (13, 15, 17, 22, 23). Similarly, polymorphisms in *A3G* in AGMs

131 impact interactions with Vif, suggesting an ongoing and ancient genetic conflict between

132 A3G and SIV Vif in AGMs (8). These observations prompted us to explore the genetic

133 landscape of A3H in African green monkeys and other Old World monkeys to determine

134 if A3H evolution has been driven by Vif over a broad evolutionary scale. First, we

135 sequenced *A3H* in 50 AGM samples collected from all four subspecies infected with a

136 species-specific SIV: vervet, sabaeus, tantalus, and grivet monkeys. The mitochondrial

137 DNA of these animals had also been previously sequenced and was confirmed to

138 cluster by AGM subspecies (8). Sequence analysis of 80 independent *A3H* genes

139 revealed 34 single-nucleotide polymorphisms (SNPs), 6 of which are synonymous

140 changes at a frequency higher than 5%. 28 SNPs are nonsynonymous; 23 are found in

141 more than one individual (Table 1). Most polymorphisms are represented in all four

142 subspecies (Figure 1A and B), except for low frequency variants only found in one AGM

143 and two nonsynonymous mutations found in two vervet sequences at amino acid

144 residues 113 and 116. Phylogenetic analysis showed that *A3H* sequences from all AGM

145 subspecies are paraphyletic (Figure 1A), similar to a previous study examining *A3G* in

146 these species (8).

147   Nonsynonymous SNPs are spread throughout the protein (Table 1), although

148 one group is clustered between amino acids 127 – 139. The changes are tightly linked

149 and result in divergence of the phylogenetic tree between two clades (Figure 1A, red

150 arrow). Haplotypes with a CGRELP motif compose one clade on the tree, while the

151 other has an SRQKRQ motif. Residues 127 and 128 are located on the $\alpha4$ helix, which

152 has been implicated in A3H-Vif binding (35, 36). Furthermore, the regions of *A3H* with

153 the greatest number of nonsynonymous mutations are in the predicted loops 1 and 7,

154 and the $\alpha6$ helix (Table 1), which were recently shown by structural studies to be

155 involved in an interaction between A3H and a co-crystalized RNA (32-34). These results

6

156   indicate that, similar to *A3G*, *SAMHD1*, and other antiviral genes in AGMs (7, 30, 31),

157   there is extensive polymorphism in AGM *A3H*.

158   The presence of numerous polymorphisms suggests there may be functional

159   consequences for either antiviral activity or Vif antagonism. To determine whether or not

160   nonsynonymous polymorphisms in *A3H* impact antiviral activity, we tested the ability of

161   A3H protein variants to restrict lentiviruses. We cloned haplotypes, numbered 1 – 11,

162   which are representative of each clade in the phylogenetic tree into a mammalian

163   expression vector for functional analysis (Figure 1B). These haplotypes also represent

164   all four subspecies of AGMs. Out of the 28 total nonsynonymous SNPs found in AGMs

165   (Table 1), 11 unique protein sequences were tested and 20 polymorphic sites were

166   characterized (Table 2).

167   A3H expressing plasmids were co-transfected into HEK293T cells for single-

168   round infectivity experiments with HIVΔ*env*Δ*vif* or SIVagm.TANΔ*env*Δ*vif* proviruses and

169   a VSV-G expression plasmid for pseudotyping to measure antiviral activity. Viral

170   supernatants were collected, normalized for RT activity, and used to infect SupT1 cells.

171   We found that none of the AGM A3H variants restricted lentiviruses as potently as the

172   A3H from rhesus macaque (Figure 2A), an Old World monkey that has been previously

173   characterized for its A3H activity (4, 13, 14). Rhesus A3H restricted viral infection of

174   HIVΔ*vif* by approximately 21-fold, while AGM A3H variants restrict viral infection by no

175   more than 3-fold, and some not at all (Figure 2A). Although activity against HIV-1Δ*vif*

176   correlates with activity against other lentiviruses, we also validated this result using

177   SIVagm.TANΔ*vif*, a strain originally isolated from tantalus monkeys. We found that the

178   AGM A3H haplotypes were also poorly restrictive against SIVagm.TANΔ*vif* compared to

179   the activity of rhesus macaque A3H (Figure 2A). This indicates A3H variants encoded

180   by AGMs have poor antiviral activity against at least two separate lentiviruses and is not

181   due to species-specificity (Figure 2B).

182   The differences in restriction could be explained by changes in expression since

183   we found that no variant of AGM A3H is as strongly expressed as rhesus macaque A3H

184   in HEK293T cells (Figure 2B, top). This lower expression of AGM A3H proteins relative

185   to rhesus macaque A3H was not due to the species of origin of the cells used for

186   transfection since when we transfected AGM-derived Cos7 cells, the protein expression

187 levels of AGM A3H were similarly poor relative to the expression of the rhesus A3H

188 (Figure 2B, bottom). These data indicate that lower protein expression levels are

189 correlated with less potent antiviral activity, similar to unstable haplotypes of human

190 A3H.

191   To ask whether lower antiviral activity of AGM A3H proteins was due to low

192 expression levels alone, we selected one of the most potent AGM A3H proteins,

193 haplotype 1 (Figure 2A), and increased the amount of transfected plasmid from 200ng

194 to 1000ng in parallel to rhesus macaque A3H. At the highest plasmid concentration,

195 haplotype 1 was able to restrict HIVΔ*vif* 16-fold, while rhesus A3H restricted HIVΔ*vif* 19-

196 fold at the lowest concentration used (Figure 2C, top). This corresponds to

197 approximately equal levels of protein expression at these amounts of plasmid

198 transfected (Figure 2C, bottom; compare band for AGM A3H at the highest level of

199 plasmid transfected with the level of Rhesus A3H at the lowest level of plasmid

200 transfected). This demonstrates that AGM A3H has not inherently lost its function, since

201 as the protein concentration increases, the antiviral activity correspondingly becomes

202 more potent.

203   In order to more fully explore the relationship between expression levels and

204 antiviral activity of AGM A3H, we codon-optimized both AGM *A3H* haplotype 1 and

205 rhesus macaque *A3H* sequences to remove rare codons that might negatively affect

206 protein translation efficiency. Based on codon-usage statistics in primates (see

207 Methods) 100 out of 211 codons were replaced to more frequent codons in the codon-

208 optimized AGM haplotype 1 *A3H* and 106 out of 211 were replaced in the codon-

209 optimized rhesus macaque *A3H*. We found that codon-optimization of both AGM A3H

210 and of rhesus macaque A3H increased their expression levels relative to the native

211 codons in each gene (Figure 3A—compare WT to CO (codon-optimized)). Moreover,

212 after codon-optimization, AGM A3H haplotype 1 and rhesus macaque A3H are

213 expressed to similar levels (Figure 3A and Figure 3B, bottom). However, while codon

214 optimization increased the antiviral activity of both AGM and rhesus macaque A3H over

215 that of wild-type (compare Figure 3B to 2D), codon-optimized rhesus macaque A3H still

216 restricts viral infection 10-fold better than codon-optimized AGM A3H (Figure 3B, top).

8

217   This shows that while increasing the expression level of the protein is sufficient to

218   improve its antiviral activity, other factors also influence protein function.

219   Studies in humans have shown that inactive A3H proteins also localize to the

220   nucleus (13, 18). We therefore asked whether AGM A3H proteins were expressed to

221   lower levels with low antiviral activity due to a change in localization. WT AGM A3H

222   haplotype 1 and WT rhesus macaque A3H were transfected into HeLa cells and

223   visualized using immunofluorescent microscopy. However, both AGM and rhesus

224   macaque A3H were mainly present in the cytoplasm (Figure 3C), demonstrating that a

225   drastic change in localization was not responsible for the lack of potent antiviral activity

226   of AGM A3H.

227   Our results suggest that the extensive diversity observed in AGM A3H results in

228   lower antiviral activity linked to both protein levels and to other functional differences

229   due to amino acid divergence between species.

230

231   **Reconstruction of ancestral A3H proteins demonstrates the loss of activity in**

232   **more recent evolution of AGMs and other primates**

233   We previously explored the evolutionary dynamic of hominoid A3H by

234   reconstructing the ancestor of human/chimpanzee A3H and found that the predicted

235   A3H protein at the human/chimpanzee ancestor had higher antiviral activity than either

236   the extant chimpanzee or human proteins (13). This suggests that there has been a loss

237   of some activity in both lineages over their evolution. Due to the finding that all tested

238   AGM A3H proteins are poorly active relative to the rhesus macaque A3H, we wanted to

239   reconstruct the ancestral history of A3H leading to the modern AGM lineage. Thus, in

240   order to gain statistical power in the ancestral sequence predictions at each node, we

241   determined the *A3H* sequence from a broader panel of Old World monkeys in sister

242   clades including De Brazza's monkey, Allen's Swamp monkey, Wolf's guenon,

243   mustached guenon, talapoin, and patas monkey (Figure 4A).

244   We tested the A3H activity of the closest sister species to the AGMs, patas

245   monkeys, and the A3H activity of a sister species to the rhesus, the sooty mangabey

246   (Figure 4A). Upon transfection into HEK293T cells, the protein expression level of patas

247   monkey A3H was lower in comparison to rhesus macaque A3H, similar to AGM A3H

9

248   (Figure 4B, bottom). Patas monkey A3H correspondingly had low antiviral activity when

249   tested against HIVΔ*vif*; that is, while AGM and rhesus macaque A3H restricted viral

250   infection 6-fold and 63-fold, respectively, patas monkey A3H restricted HIVΔ*vif* infection

251   only 3-fold (Figure 4B, top). However, active A3H proteins from sooty mangabey and

252   the human A3H haplotype II restrict viral infection 20-fold and 40-fold, respectively.

253   These data show that the antiviral activity is low in a species closely related to AGMs,

254   but a relative of the rhesus macaque and humans encode more active A3H proteins.

255   This finding further suggests that changes in A3H activity may have deeper evolutionary

256   origins in primate evolution.

257   In order to determine whether A3H antiviral activity was gained in the rhesus

258   macaque/sooty mangabey lineage, or was lost in the AGM/patas monkey lineage, we

259   reconstructed the *A3H* ancestors at various nodes in the Old World monkey phylogeny

260   (Figure 4A). These included the common ancestor of AGMs and patas monkeys (node

261   1), AGMs, patas monkeys, and its sister clade (node 2), as well as the common

262   ancestor of rhesus macaque, pig-tailed macaque, and sooty mangabey (node 3) and

263   common ancestor of both groups (node 4). Each ancestor was constructed using

264   maximum likelihood with FastML (37) based on the primate species phylogeny.

265   Although the majority of codons were reconstructed with a probability >99%, site 207

266   was ambiguous in the common ancestor of all tested Old World monkeys and two

267   codons were possible – encoding either an isoleucine (node 4a) or a threonine (node

268   4b) at position 207. In this case, both possible ancestors were generated using point-

269   mutagenesis.

270   We then tested the predicted A3H protein at the reconstructed ancestral nodes of

271   the Old World monkey phylogenetic tree for antiviral activity and protein expression

272   level. All ancestors inhibited viral infection between 17- and 33-fold (Figure 4C). Thus,

273   this result suggests that activity was lost within the AGM/patas monkey clade rather

274   than specifically gained in the rhesus macaque/sooty mangabey clade. Moreover, the

275   node 1 ancestor representing the common ancestor of AGM and patas monkeys

276   restricted viral infection 33-fold, which was more potent than its descendants, AGMs (6-

277   fold) and patas monkeys (3-fold), despite having similar protein expression levels

278   (Figure 4C). Therefore, loss of activity in A3H in AGM and patas monkeys that occurred

279   after the common ancestor at node 1, which diverged at least 4 million years ago (27),

280   included mutations that both decreased protein expression levels and led to the losses

281   of antiviral activity that are independent of protein expression.

282

283   **Multiple amino acid mutations are responsible for loss of A3H antiviral activity.**

284   The ancestral A3H protein at node 1 representing the common ancestor of AGMs

285   and patas monkeys (Figure 4A) has stronger antiviral activity than its extant

286   descendants (Figure 4C). Therefore, we wanted to trace the amino acid mutations that

287   resulted in the subsequent loss along the branches leading from node 1 to AGMs and

288   patas monkeys. The sequence alignment of the node 1 ancestor with AGM haplotype 1

289   and patas monkey A3H reveals 4 (sites 18, 20, 48, and 171) and 3 (sites 17, 25, and

290   51) amino acid differences respectively (Figure 5A). One site in AGMs, S20N, created a

291   protein sequence identical to AGM A3H haplotype 8. We introduced the other residues

292   found in the node 1 ancestor into AGM and patas monkey A3H backgrounds to test the

293   expression level and antiviral activity of each mutant.

294   Notably, no single amino acid mutation in AGM A3H increased the antiviral

295   activity compared to wild-type AGM A3H and all single mutants inhibited viral infection

296   around 5-fold relative to the no A3 control (Figure 5B). In contrast, the double mutants

297   H18R/S20N, H18R/Q48K, and S20N/Q48K, antiviral activity increased to around 13-fold

298   which is still less than the node 1 ancestor (17-fold; Figure 5B). Similarly, patas monkey

299   A3H with single mutations at sites 25 (R25P) and 51 (E51K) did not improve restriction,

300   while a H17R mutation slightly increased the antiviral activity compared to wild-type (3-

301   fold to 5-fold; Figure 5B). However, this single mutation at position 17 did not make

302   patas monkey A3H comparable to the node 1 ancestor. Inserting both H17R and R25P

303   mutations in patas monkey A3H similarly did not further improve restriction (Figure 5B).

304   Taken together, the inability of single point mutations to restore antiviral activity to its

305   ancestral state emphasizes that changes at multiple sites have functional

306   consequences in AGM and patas monkey A3H.

307   Of interest, site 48 is the only residue that is fixed in all AGMs sequenced for this

308   study. This demonstrates that it occurred first, whereas the other sites are polymorphic

309   and have not yet become fixed in the species (Table 1). To determine whether site 48

11

310   alone is sufficient to decrease antiviral activity we added a K48Q mutation in the node 1

311   background and tested its antiviral activity. However, node 1 K48Q has similar antiviral

312   activity and expression (Figure 5C, bottom), showing that epistasis between multiple

313   amino acids may play an essential role for viral restriction by A3H.

314   Overall, these data establish that amino acids in the N-terminal portion of A3H

315   are important for antiviral activity. While no tested combinations increase the antiviral

316   activity of AGM or patas monkey A3H equivalent to the node 1 ancestor, all changes

317   (minus one) are before the catalytic domain (Figure 5A). Two out of 3 residues

318   responsible for the loss of activity in patas monkey A3H, sites 17 and 25, are also

319   polymorphic in AGMs (Table 1), suggesting that the loss began at a shared common

320   ancestor not sampled by our analysis. Significantly, many Old World monkeys encode

321   additional mutations near amino acid 15, whose loss in human A3H is known to affect

322   protein stability (13). For example, residues 18 – 23 have been deleted in De Brazza's

323   monkey (Figure 5A). Such a large deletion in this region would likely impact the ability of

324   A3H to inhibit viral infection in this species. Similarly, talapoin A3H encodes a histidine

325   at position 18 instead of an arginine, similar to AGM A3H, suggesting that its activity

326   may also be lost (Figure 5A). In contrast, rhesus macaque and sooty mangabey A3H

327   encode active ancestral amino acids at such residues, such as two arginines at

328   positions 17 and 18 (Figure 5A), which may result in the potent antiviral activity of these

329   modern proteins. This implies that loss of A3H function has been lost independently at

330   various points of evolutionary history in both hominoids and Old World monkeys.

331

332   **Lack of evidence that evolution in A3H leading to AGM has been driven by Vif.**

333   The variability in A3H function throughout Old World monkey evolution suggests

334   an outside selective pressure is driving its loss. Antiviral restriction factors are often

335   rapidly evolving and undergo positive selection (38), which is defined by an excess of

336   nonsynonymous mutations compared to synonymous mutations. Evolutionary conflicts

337   between host restriction factors and viral proteins to either maintain or escape

338   interactions result in an accumulation of nonsynonymous mutations at binding

339   interfaces. A previous study found that A3H is under positive selection (14). However,

340   the study used fewer primate sequences which can bias the analysis. Therefore, we

12

341   wanted to re-test positive selection in primate A3H using additional sequences we

342   obtained from Old World monkeys. Using the PAML (phylogenetic analysis by maximum

343   likelihood) program (39), we calculated the number of nonsynonymous mutations (dN)

344   over the number of synonymous mutations (dS) for the entire *A3H* gene as well as

345   individual codons. In agreement with previous data, we found that the *A3H* gene is

346   under positive selection and with a dN/dS ratio of 1.3 in all primates. Additionally,

347   models that allow codons to evolve under positive selection fit the data significantly

348   better than models that do not for all primate clades (Figure 6A). We evaluated

349   individual sites within the gene and found that in all primates, a total of 19 residues are

350   under positive selection with a posterior probability > 0.98. One site, position 90, has

351   been implicated in Vif-binding interactions, but only had a posterior probability > 0.98 in

352   one codon model (F3x4; Figure 6A). Since selection can sometimes be driven by the

353   inclusion of specific clades in the analysis, we also condensed our analysis to Old World

354   monkeys alone and found that 5 sites remain under positive selection (blue arrows,

355   Figure 6B), all of which are polymorphic in AGMs (Table 1). No positively selected

356   residues in Old World monkey A3H are in the putative Vif-binding region (35, 36). Thus,

357   we conclude that multiple sites are under diversifying selection in primates but are likely

358   not driven by lentiviral Vif.

359   In order to test the hypothesis that the changes in A3H in the lineages leading to

360   the AGMs were not driven by the Vif protein of the lentiviruses that infect AGMs, we

361   tested the restriction capabilities of the node 1 and node 2 ancestors against HIV-1

362   proviruses expressing SIVagm Vif. We also included the codon-optimized AGM A3H

363   since, as it is isolated from an AGM, should be susceptible to degradation by all Vifs

364   encoded by SIVagm. Of interest, the node 2 ancestor encodes an aspartic acid (D) at

365   amino acid 100, which has previously been important for differences in Vif-biding for

366   SIVcpz and HIV-1, while the node 1 ancestor and all AGM A3Hs encode an asparagine

367   (N) (35, 36). The antiviral activity of all ancestors, as well as codon-optimized AGM

368   A3H, was counteracted by each Vif, resulting in a rescue of viral infection (Figure 6C).

369   Because there are no differences in the ability of Vif to rescue viral infection, these data

370   suggest that Vif is not the primary force on A3H evolution in AGMs, which is instead

371   driven by different selective pressures.

13

372

373   **Discussion**

374   We show evidence for the recurrent functional loss of APOBEC3H in primates.

375   We found that a decrease in protein expression levels, as well as amino acid mutation

376   in the N-terminal region, results in lower antiviral activity. Using molecular reconstruction

377   of ancestral *A3H* sequences, we found that the most recent common ancestor of AGMs

378   and patas monkeys likely encoded an active A3H, similar to other common ancestors

379   throughout evolutionary history. This suggests that the recurrent loss is a more recent

380   event in primate evolution. Selective pressure by Vif is does not appear to be a primary

381   force behind the evolution of A3H in the AGM clade, but, as loss has occurred both in

382   humans and in other Old world monkeys, there may be a fitness cost to encoding this

383   mutator protein over long evolutionary time periods.

384

385   **Molecular evolution of A3H protein impacts expression levels and antiviral**

386   **activity.**

387   While increasing the amount of A3H present in cells did increase its capability to

388   inhibit viral infection, greater antiviral activity does not perfectly correlate with higher

389   protein expression. Surprisingly, codon-optimization increased both the expression level

390   and antiviral activity of A3H. However, codon-optimized AGM A3H haplotype 1 was not

391   as potent as codon-optimized rhesus macaque A3H, demonstrating that both protein

392   expression level and amino acid differences have functional consequences (Figure 3A

393   and B). In support of this idea, while AGM A3H haplotypes 1 and 8 have the fewest

394   amino acid changes from the most recent common ancestor with another species

395   (Table 2, Figure 5A). Multiple mutations are required to allow the inactive extant protein

396   to act like its active ancestor. Position 48 is fixed in AGMs, indicating that this change

397   likely occurred first, while subsequent amino acid changes concurrently altered the

398   antiviral activity of the protein. However, a K48Q mutation in the node 1 ancestor does

399   not decrease its function (Figure 5C), demonstrating that epistasis between observed

400   amino acid changes more likely lead to its loss. For instance, the AGM A3H double

401   mutant H18R/Q48K inhibits viral infection more comparable to the node 1 ancestor

402   while the individual mutants do not, indicating that both residues are required to gain

403   antiviral activity (Figure 5B). Other AGM A3H haplotypes have also accumulated

15

404   additional nonsynonymous mutations (Figure 2A), indicating that additional genetic drift

405   may be actively driving A3H to become less antiviral in AGMs, as many animals encode

406   proteins with different expression levels and antiviral activity (Figure 2B and C).

407

408   **Why has primate A3H maintained partial activity rather than a complete loss?**

409   Loss of protein activity can be indicative of a fitness cost. In the case of human

410   A3H, haplotypes encoding unstable proteins have been linked to greater cancer risk,

411   likely due to its nuclear localization and proximity to host DNA (19). Moreover, human

412   A3H haplotypes III and IV, which have a deletion at amino acid 15, have little to no

413   antiviral activity. We observe that AGM A3H haplotypes vary greatly in their antiviral

414   restriction and can restrict viral infection anywhere from 17 – 70% (Figure 2A).

415   However, despite this immense variation in AGM A3H antiviral activity, we observed no

416   large deletions or premature stop codons, which would indicate that the gene itself is

417   being lost. Why has the antiviral activity of A3H not been lost completely in AGMs? It is

418   possible that A3H has been co-opted for a non-antiviral function in Old World monkeys,

419   thus it is preserved. Alternatively, A3H could be retained through linkage to protective

420   A3G haplotypes. A3G and A3H are located close together on chromosome 19 in AGMs,

421   meaning A3H haplotypes that have ultimately lost most of their function may still be

422   passed onto offspring, particularly if the A3G allele encodes a protein providing a

423   selective advantage. In support of this idea, we noticed poorly active A3H proteins in a

424   monkey encoding protective A3G, as characterized in a separate study (8). This

425   individual, V005, encodes A3H haplotype 11 as well as an A3G haplotype that cannot

426   be antagonized by Vif proteins from SIVagm.ver or SIVagm.tan. Since A3G is a more

427   potent antiviral and this protein protects individuals from two strains of SIVagm, the

428   protection it supplies may supersede any detrimental effects incurred by the presence of

429   an inactive A3H haplotype.

430

431   **Loss of activity of A3H due to recurrent mutations in a putative RNA binding**

432   **domain**

433   Two mutations in human A3H gave rise to inactive protein variants: R105G and a

434   deletion of amino acid 15. Of interest, amino acid 15 is positioned within loop 1 of the

16

435   A3H protein. Recent work has revealed that loop 1, 7, and α6 are important for binding

436   to an RNA duplex (32-34). We similarly identified that amino acids 18 and 20, found

437   within loop 1, are important for increasing the antiviral activity of AGM A3H (Figure 5B).

438   The analogous locations of such residues may suggest that these changes impact the

439   antiviral activity of A3H by affecting its capability to bind to viral RNA. Dual amino acid

440   changes at residues 18 and 20 in AGMs increase the antiviral activity of A3H close to

441   the levels of a recent common ancestor. An additional mutation close to the catalytic

442   site, Q48K, together with either H18R or S20N in AGMs, further increased antiviral

443   activity. Furthermore, we found that amino acid 17 similarly increased the antiviral

444   activity of patas monkey A3H. This residue is polymorphic in AGMs, suggesting that this

445   change may have occurred in a common ancestor not tested in this study. We also find

446   that other Old World monkeys have changes within loop 1, such as a six amino acid

447   deletion of residues 18 - 23 in De Brazza's monkey and an R18H mutation in talapoin

448   (Figure 5A). The diversity within loop 1 of Old World monkey A3H is indicative that A3H

449   activity was possibly lost multiple times independently. RNA-binding has been shown to

450   play an important role in the antiviral activity of A3H (32, 33) and thus loss of RNA-

451   binding may result in functional loss.

452

453   **Vif does not appear to play a role in evolution of A3H along lineage leading to**

454   **AGMs**

455   AGMs are highly polymorphic in a number of antiviral genes that are specific to

456   infection with lentiviruses in both AGMs explicitly and other primates (31), further

457   highlighting the adaptation to SIV within this particular species. Indeed, we found that

458   multiple sites in A3H are under positive selection in primates, which is suggestive of an

459   evolutionary arms race between A3H and another protein. Evidence of A3H-Vif

460   interactions are evident in hominoids, as SIVcpz Vif adaptation to stable human A3H

461   was crucial for transmission of SIVcpz (11) to humans and HIV-1 Vif is highly variable in

462   its ability to target stable human A3H proteins (17, 21-23). Interactions between Vif and

463   other A3 proteins have also been well-characterized in chimpanzees and humans (7,

464   35, 36). Multiple A3s in chimpanzees protect them from infection by SIVs in Old World

465   monkeys. Therefore, Vif must adapt to antagonize multiple A3 proteins such as A3D

17

466 and A3G (10). However, we find that Vif did not play a role in A3H's evolution in AGMs

467 since ancestral proteins were comparably susceptible to antagonism by Vif. Loss of

468 A3H function may facilitate cross-species transmission of SIVagm strains between AGM

469 subspecies (40) due a diminished A3 repertoire. Although many A3 proteins in the

470 family may be redundant, encoding a diverse range of A3s is likely important to achieve

471 maximum protection against lentiviruses.

472 Since lentiviruses have infected simian primates for millions of years (7), it is

473 unlikely that the lack of residues altering A3H-Vif interactions in AGMs stems from

474 recent infection of Old World monkeys. It is possible that the changes are the result of

475 genetic drift or driven by a different viral pathogen in these primates. Conversely,

476 positively selected residues may indicate evolutionary toggling to preserve or eliminate

477 protein function. This is supported by the finding that many amino acid residues found to

478 increase the antiviral activity of A3H in AGMs are under positive selection (Figure 6A,

479 6B). It is also possible that the relative importance of different A3 proteins may change

480 dependent on the evolutionary history of a species, driven by the redundancy of the

481 protein family. In the Old World monkeys studied here, inactive A3H proteins may

482 impart an increased risk for host genome mutations, thus its function was lost due to the

483 balance between viral protection and host fitness.

484

485 Our data implies that A3H function was lost prior to the divergence of different

486 SIVagm strains due to its inactivity in all AGMs, not just specific subspecies. If this is the

487 case, Vif proteins from the ancestral virus may not have required adaptation to escape

488 the antiviral activity of A3H and instead evolved in response to pressure from the more

489 potent A3G. Expansion of the primate *A3* locus provides flexibility of this antiviral protein

490 family to take different trajectories throughout evolution. We have demonstrated that

491 A3H activity is fluid throughout the evolutionary history of primates. In addition to

492 previous work in humans, the A3H homolog in felines, APOBEC3Z3, was recently

493 shown have a similar functional loss (41), demonstrating that loss of A3H and its

494 homologs are frequent throughout a variety of animal species. Such widespread loss of

495 function is suggestive of a potential fitness cost to hosts, although the presence of

496 modern and active A3H proteins exemplifies the importance of encoding a diverse *A3*

497   locus in primates.

498   **Materials and methods**

499   **APOBEC3H cDNA Amplification and Sequencing.**

500   APOBEC3H cDNAs were cloned by nested RT-PCR (QIAGEN One-step RT PCR Kit)

501   or PCR (Accuprime Pfx) from RNA or gDNA isolated from AGM peripheral blood

502   mononuclear cells or cell lines. Sample origins and extractions have been previously

503   described (8, 30). For each sample, PCR products were amplified and sequenced using

504   primers designed to amplify African green monkey A3H (Forward:

505   CACGAATTCGCCACCATGTATCCATACGATGTTCCAGATTACGC

506   TGCTCTGCTAACAGCCAAA Reverse: CACGAGCTCATCTTGAGTTGAGTGT).

507   Primers for gDNA amplification were designed to target intronic regions. Heterozygous

508   sequences were cloned using pGEM T-Easy vector system (Promega) and TOPO TA

509   cloning kit (Invitrogen) to analyze individual clones. Additional primate sequences were

510   obtained from gDNA isolated from immortalized cell lines using QIAGEN DNeasy Blood

511   & Tissue Kit. The following cell lines from Coriell Cell Repositories (Camden, NJ) were

512   used: patas monkey (*Erythrocebus patas;* ID no. 6254), De Brazza's monkey

513   (*Cercopithecus neglectus*; PR01144), Wolf's guenon (*Cercopithecus wolfi*; PR01241),

514   mustached guenon (*Cercopithecus cephus*, PR00527), Allen's swamp monkey

515   (*Allenopithecus nigroviridis;* PR00198), and Francois' Leaf monkey (*Trachypithecus*

516   *francoisi*; PR01099). Talapoin (*Miopithecus talapoin;* OR755) cells were obtained from

517   Frozen Zoo (San Diego, CA).

518

519   **Expression Constructs and Plasmids.**

520   Primate *A3H* genes were cloned from cDNA. A 5' hemagluttinin (HA) tag was added via

521   PCR (Forward:

522   GTGGTGGAATTCATGTATCCATACGATGTTCCAGATTACGCTGCTCTGCT Reverse:

523   CTAGACTCGAGTCATCTTGAGTT). The products were digested using EcoRI/XhoI

524   restriction enzymes and ligated into a mammalian pcDNA 3.1 vector (Invitrogen,

525   #V79020). Site-directed mutagenesis was completed with the QuikChange II Site-

526   Directed Mutagenesis Kit (Agilent, #200524) to construct all ancestral genes and

527   mutants. The *A3H* gene from patas monkeys was generated by gene synthesis (IDT)

528   and cloned into the pcDNA 3.1 backbone. *A3H* genes from AGM and rhesus macaque

20

529   were codon-optimized based on usage frequencies in primates (human and rhesus

530   macaque) to remove rare codons within the gene in Geneious (Biomatters Ltd.). Codon-

531   optimized sequences were generated by gene synthesis and cloned into pcDNA 3.1.

532

533   **Cell Lines, Transfections, and Western Blot Analysis.**

534   HEK293T, HeLa, and Cos7 cell lines (ATCC) were maintained in Dulbecco's modified

535   Eagle's medium (DMEM) with 10% fetal bovine serum (Corning, #35-015-CV) and 100

536   μg/mL penicillin/streptomycin (Gibco, #15140-122) at 37°C. SupT1 cells (ATCC) were

537   maintained in RPMI 1640 with 10% fetal bovine serum and 100 μg/mL

538   penicillin/streptomycin in the same conditions. Transfections were done in serum-free

539   DMEM with TransIT-LT1 transfection reagent (Mirus Bio, #MIR 2305) at a

540   reagent:plasmid DNA ratio of 3:1. For western blot analysis, cells were lysed in ice-cold

541   NP40 buffer (0.5% NP40, 20mM NaCl, 50mM Tris pH 7.5) with protease inhibitors

542   (Roche Complete Mini, EDTA-free tablets, #11836170001). Lysates were quantified

543   using a Pierce BCA Protein Assay Kit (Thermo Scientific, #23225) and 10 μg of protein

544   was resolved by SDS-PAGE, transferred to a PVDF membrane, and probed with anti-

545   HA (BioLegend, #901503) and anti-actin (Sigma, #A2066) or anti-vinculin (Proteintech,

546   #66305-1) antibodies at a 1:2000 dilution. Anti-mouse or anti-rabbit secondary

547   antibodies were used at a 1:5000 dilution (Santa Cruz Biotechnology, sc-2005, sc-

548   2004).

549

550   **Immunofluorescence.**

551   HeLa cells were seeded onto 18mm (VWR, #48380 046) coverslips seeded with 4 x 10^4

552   cells/mL and transfected with 500 ng of A3H-expressing plasmids the next day. 48

553   hours after transfection, the coverslips were fixed in 2% paraformaldehyde,

554   permeabilized in 0.5% PBS/Triton-X, and blocked in PBS/BGS. HA-tagged proteins

555   were detected using the same HA antibody used for western blots at a 1:1000 dilution

556   followed by an anti-mouse AF488 antibody at 1:400 (Invitrogen, #A11001). Nuclei were

557   stained in SlowFade Gold antifade reagent with DAPI mounting media (Life

558   Technologies, #S36939). Images were taken on a Nikon E800 microscope.

559

21

Phylogenetic Analysis.

560
561 *A3H* AGM sequences were analyzed phylogenetically using a Bayesian Monte Carlo

562 Markov chain (MCMC) approach implemented in BEAST v1.7.1. Sequence alignments

563 were constructed using MAFFT align function in Geneious (Biomatters Ltd.) and

564 underwent 10,000,000 MCMC generations using HKY85 substitution model, gamma

565 site heterogeneity model, estimated base frequencies, and constant population size

566 coalescent as the tree prior.

567

Positive Selection Analysis.

568
569 The 27 primate *A3H* sequence alignment was analyzed using HyPhy GARD analysis to

570 ensure there was no recombination in the gene (42). The species phylogeny (27) was

571 input into the CODEML sites model of PAML (39) along with the nucleotide alignment to

572 detect positive selection at individual sites. The p-value was calculated by twice the

573 difference in log-likelihood between models M7 and M7 as well as M8 and M8a with two

574 degrees of freedom. Analysis was conducted with both the F3x4 and F61 codon

575 frequency models with omega values of 0.4 and 1.5. Data from F3x4 and F61 models

576 are shown in Table 1. Positively selected sites were categorized as those with an M8

577 Bayes Empirical Bayes posterior probability greater than 98%.

578

Ancestral Reconstruction.

579
580 The ancestors at specific nodes within the Old World monkey clade were reconstructed

581 using the FASTML webserver (fastml.tau.ac.il; last accessed March 2018; (37)). The 27

582 primate *A3H* sequence alignment was used in conjunction with the species tree to

583 generate marginal reconstructions of codon sequences.

584

Single-Cycle Infectivity Assays.

585
586 HEK293T cells were plated in 1mL in 12-well plates at 1.25×10^5 cells/mL. After cells

587 reached between 50 – 70% confluency, they were co-transfected with 250 ng of A3H or

588 empty expression plasmid, 600 ng of proviral plasmid, and 100 ng of L-VSV-G

589 (vesicular stomatitis virus glycoprotein) for pseudotyping in 100 μL serum-free medium

590 with TransIT-LT1 transfection reagent (Mirus Bio). Supernatants containing virus were

22

591   harvested after 48 hours and clarified through 0.2-micron filters. Viral titers were

592   determined by measuring reverse-transcriptase (RT) activity by qPCR as described

593   previously (43). In short, viral supernatants were lysed in 2X lysis buffer (0.25% Triton

594   X-100, 50 mM KCl, 100 mM Tris-HCl, 40% glycerol) in the presence of 4U RNase

595   inhibitor (Fermentas, #EO0382). qRT-PCR reactions were set up with an MS2 RNA

596   template using the Takyon Rox SYBR MasterMix dTTP Blue kit (Eurogentec, #UF-

597   RSMT-B0101) alongside a standard curve made with a stock virus of previously

598   determined titers. The primers used to amplify duplicate reactions were:

599   TCCTGCTCAACTTCCTGTCGAG (forward) and CACAGGTCAAACCTCCTAGGAATG

600   (reverse). qRT-PCR was performed on an ABI QuantStuido5 Real Time PCR machine.

601   2000 mU/mL was used to infect SupT1 cells plated at 2×10^4 cells/well in a 96-well

602   plate in media supplemented with 20 μg/mL DEAE-Dextran. Infections were done in

603   triplicate for 48 – 64 hours. Luciferase activity was measured with Bright-Glo Luciferase

604   Assay Reagent (Promega) on a LUMIstar Omega luminometer.

605

606   **Accession Numbers.**

607   The GenBank accession numbers for new Old World monkey *A3H* sequences, including

608   De Brazza's, Wolf's Guenon, mustached guenon, patas monkey, talapoin, Francois'

609   Leaf monkey, Allen's swamp monkey, and AGM haplotype, reported here are

610   MH231602 – MH231609.

Acknowledgements

References

1. **Harris RS, Dudley JP.** 2015. APOBECs and virus restriction. Virology **479-480:**131-145.

2. **Harris RS, Liddament MT.** 2004. Retroviral restriction by APOBEC proteins. Nat Rev Immunol **4:**868-877.

3. **Stavrou S, Ross SR.** 2015. APOBEC3 Proteins in Viral Immunity. J Immunol **195:**4565-4570.

4. **Hultquist JF, Lengyel JA, Refsland EW, LaRue RS, Lackey L, Brown WL, Harris RS.** 2011. Human and rhesus APOBEC3D, APOBEC3F, APOBEC3G, and APOBEC3H demonstrate a conserved capacity to restrict Vif-deficient HIV-1. J Virol **85:**11220-11234.

5. **Belanger K, Savoie M, Rosales Gerpe MC, Couture JF, Langlois MA.** 2013. Binding of RNA by APOBEC3G controls deamination-independent restriction of retroviruses. Nucleic Acids Res **41:**7438-7452.

6. **Simon V, Bloch N, Landau NR.** 2015. Intrinsic host restrictions to HIV-1 and mechanisms of viral escape. Nat Immunol **16:**546-553.

7. **Compton AA, Emerman M.** 2013. Convergence and divergence in the evolution of the APOBEC3G-Vif interaction reveal ancient origins of simian immunodeficiency viruses. PLoS Pathog **9:**e1003135.

8. **Compton AA, Hirsch VM, Emerman M.** 2012. The host restriction factor APOBEC3G and retroviral Vif protein coevolve due to ongoing genetic conflict. Cell Host Microbe **11:**91-98.

9. **McCarthy KR, Kirmaier A, Autissier P, Johnson WE.** 2015. Evolutionary and Functional Analysis of Old World Primate TRIM5 Reveals the Ancient Emergence of Primate Lentiviruses and Convergent Evolution Targeting a Conserved Capsid Interface. PLoS Pathog **11:**e1005085.

10. **Etienne L, Bibollet-Ruche F, Sudmant PH, Wu LI, Hahn BH, Emerman M.** 2015. The Role of the Antiviral APOBEC3 Gene Family in Protecting Chimpanzees against Lentiviruses from Monkeys. PLoS Pathog **11:**e1005149.

11. **Zhang Z, Gu Q, de Manuel Montero M, Bravo IG, Marques-Bonet T, Haussinger D, Munk C.** 2017. Stably expressed APOBEC3H forms a barrier for cross-species transmission of simian immunodeficiency virus of chimpanzee to humans. PLoS Pathog **13:**e1006746.

12. **Bishop KN, Holmes RK, Sheehy AM, Davidson NO, Cho SJ, Malim MH.** 2004. Cytidine deamination of retroviral DNA by diverse APOBEC proteins. Curr Biol **14:**1392-1396.

13. **OhAinle M, Kerns JA, Li MM, Malik HS, Emerman M.** 2008. Antiretroelement activity of APOBEC3H was lost twice in recent human evolution. Cell Host Microbe **4:**249-259.

14. **OhAinle M, Kerns JA, Malik HS, Emerman M.** 2006. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. J Virol **80:**3853-3862.

15. **Wang X, Abudu A, Son S, Dang Y, Venta PJ, Zheng YH.** 2011. Analysis of human APOBEC3H haplotypes and anti-human immunodeficiency virus type 1 activity. J Virol **85:**3142-3152.

16. **Harari A, Ooms M, Mulder LC, Simon V.** 2009. Polymorphisms and splice variants influence the antiretroviral activity of human APOBEC3H. J Virol **83:**295-303.

17. **Refsland EW, Hultquist JF, Luengas EM, Ikeda T, Shaban NM, Law EK, Brown WL, Reilly C, Emerman M, Harris RS.** 2014. Natural polymorphisms in human APOBEC3H and HIV-1 Vif combine in primary T lymphocytes to affect viral G-to-A mutation levels and infectivity. PLoS Genet **10:**e1004761.

677   18.   **Li MM, Emerman M.** 2011. Polymorphism in human APOBEC3H affects a phenotype
678   dominant for subcellular localization and antiviral activity. J Virol **85:**8197-8207.
679   19.   **Starrett GJ, Luengas EM, McCann JL, Ebrahimi D, Temiz NA, Love RP, Feng Y,**
680   **Adolph MB, Chelico L, Law EK, Carpenter MA, Harris RS.** 2016. The DNA cytosine
681   deaminase APOBEC3H haplotype I likely contributes to breast and lung cancer
682   mutagenesis. Nat Commun **7:**12918.
683   20.   **Burns MB, Temiz NA, Harris RS.** 2013. Evidence for APOBEC3B mutagenesis in
684   multiple human cancers. Nat Genet **45:**977-983.
685   21.   **Li MM, Wu LI, Emerman M.** 2010. The range of human APOBEC3H sensitivity to
686   lentiviral Vif proteins. J Virol **84:**88-95.
687   22.   **Ooms M, Brayton B, Letko M, Maio SM, Pilcher CD, Hecht FM, Barbour JD, Simon**
688   **V.** 2013. HIV-1 Vif adaptation to human APOBEC3H haplotypes. Cell Host Microbe
689   **14:**411-421.
690   23.   **Ooms M, Letko M, Binka M, Simon V.** 2013. The resistance of human APOBEC3H to
691   HIV-1 NL4-3 molecular clone is determined by a single amino acid in Vif. PLoS One
692   **8:**e57744.
693   24.   **Naruse TK, Sakurai D, Ohtani H, Sharma G, Sharma SK, Vajpayee M, Mehra NK,**
694   **Kaur G, Kimura A.** 2016. APOBEC3H polymorphisms and susceptibility to HIV-1
695   infection in an Indian population. J Hum Genet **61:**263-265.
696   25.   **Sakurai D, Iwatani Y, Ohtani H, Naruse TK, Terunuma H, Sugiura W, Kimura A.**
697   2015. APOBEC3H polymorphisms associated with the susceptibility to HIV-1 infection
698   and AIDS progression in Japanese. Immunogenetics **67:**253-257.
699   26.   **Fregoso OI, Ahn J, Wang C, Mehrens J, Skowronski J, Emerman M.** 2013.
700   Evolutionary toggling of Vpx/Vpr specificity results in divergent recognition of the
701   restriction factor SAMHD1. PLoS Pathog **9:**e1003496.
702   27.   **Perelman P, Johnson WE, Roos C, Seuanez HN, Horvath JE, Moreira MA, Kessing**
703   **B, Pontius J, Roelke M, Rumpler Y, Schneider MP, Silva A, O'Brien SJ, Pecon-**
704   **Slattery J.** 2011. A molecular phylogeny of living primates. PLoS Genet **7:**e1001342.
705   28.   **Wertheim JO, Worobey M.** 2007. A challenge to the ancient origin of SIVagm based on
706   African green monkey mitochondrial genomes. PLoS Pathog **3:**e95.
707   29.   **Peeters M, Courgnaud V.** 2002. Overview of Primate Lentiviruses and Their Evolution
708   in Non-human Primates in Africa. HIV Sequence Compendium 2002**:**2 - 23.
709   30.   **Spragg CJ, Emerman M.** 2013. Antagonism of SAMHD1 is actively maintained in
710   natural infections of simian immunodeficiency virus. Proc Natl Acad Sci U S A
711   **110:**21136-21141.
712   31.   **Svardal H, Jasinska AJ, Apetrei C, Coppola G, Huang Y, Schmitt CA, Jacquelin B,**
713   **Ramensky V, Muller-Trutwin M, Antonio M, Weinstock G, Grobler JP, Dewar K,**
714   **Wilson RK, Turner TR, Warren WC, Freimer NB, Nordborg M.** 2017. Ancient
715   hybridization and strong adaptation to viruses across African vervet monkey populations.
716   Nat Genet **49:**1705-1713.
717   32.   **Ito F, Yang H, Xiao X, Li SX, Wolfe A, Zirkle B, Arutiunian V, Chen XS.** 2018.
718   Understanding the Structure, Multimerization, Subcellular Localization and mC
719   Selectivity of a Genomic Mutator and Anti-HIV Factor APOBEC3H. Sci Rep **8:**3763.
720   33.   **Shaban NM, Shi K, Lauer KV, Carpenter MA, Richards CM, Salamango D, Wang J,**
721   **Lopresti MW, Banerjee S, Levin-Klein R, Brown WL, Aihara H, Harris RS.** 2018. The
722   Antiviral and Cancer Genomic DNA Deaminase APOBEC3H Is Regulated by an RNA-
723   Mediated Dimerization Mechanism. Mol Cell **69:**75-86 e79.
724   34.   **Bohn JA, Thummar K, York A, Raymond A, Brown WC, Bieniasz PD, Hatziioannou**
725   **T, Smith JL.** 2017. APOBEC3H structure reveals an unusual mechanism of interaction
726   with duplex RNA. Nat Commun **8:**1021.

727 35. **Nakashima M, Tsuzuki S, Awazu H, Hamano A, Okada A, Ode H, Maejima M,**
728   **Hachiya A, Yokomaku Y, Watanabe N, Akari H, Iwatani Y.** 2017. Mapping Region of
729   Human Restriction Factor APOBEC3H Critical for Interaction with HIV-1 Vif. J Mol Biol
730   **429:**1262-1276.
731 36. **Ooms M, Letko M, Simon V.** 2017. The Structural Interface between HIV-1 Vif and
732   Human APOBEC3H. J Virol **91**.
733 37. **Ashkenazy H, Penn O, Doron-Faigenboim A, Cohen O, Cannarozzi G, Zomer O,**
734   **Pupko T.** 2012. FastML: a web server for probabilistic reconstruction of ancestral
735   sequences. Nucleic Acids Res **40:**W580-584.
736 38. **Daugherty MD, Malik HS.** 2012. Rules of engagement: molecular insights from host-
737   virus arms races. Annu Rev Genet **46:**677-700.
738 39. **Yang Z.** 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol
739   **24:**1586-1591.
740 40. **Bell SM, Bedford T.** 2017. Modern-day SIV viral diversity generated by extensive
741   recombination and cross-species transmission. PLoS Pathog **13:**e1006466.
742 41. **Konno Y, Nagaoka S, Kimura I, Takahashi Ueda M, Kumata R, Ito J, Nakagawa S,**
743   **Kobayashi T, Koyanagi Y, Sato K.** 2018. A naturally occurring feline APOBEC3 variant
744   that loses anti-lentiviral activity by lacking two amino acid residues. J Gen Virol
745   doi:10.1099/jgv.0.001046.
746 42. **Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD.** 2006. GARD:
747   a genetic algorithm for recombination detection. Bioinformatics **22:**3096-3098.
748 43. **Vermeire J, Naessens E, Vanderstraeten H, Landi A, Iannucci V, Van Nuffel A,**
749   **Taghon T, Pizzato M, Verhasselt B.** 2012. Quantification of reverse transcriptase
750   activity by real-time PCR as a fast and accurate method for titration of HIV, lenti- and
751   retroviral vectors. PLoS One **7:**e50859.
752
753

754　**Figure Legends**

755　**Figure 1. Sequence and Phylogenetic Analysis of *A3H* in African green monkeys.**

756　**A.** The evolutionary relationship between 80 full-length AGM *A3H* genes was inferred by

757　Bayesian MCMC phylogenetic reconstruction. Red asterisks (*) show nodes that have a

758　posterior probability > 0.5. Colored boxes demonstrate the subspecies of origin (red =

759　Vervet, yellow = Sabaeus, blue = Tantalus, and green = Grivet). Blue asterisks (*)

760　denote cloned haplotypes, numbered 1 – 11. **B.** Number of AGM individuals encoding

761　haplotypes 1 – 11, color coded by subspecies similar to *A3H* phylogeny.

762　**Figure 2. Antiviral activity of A3H is lower in AGMs than in another Old World**

763　**monkey. A.** Single-cycle infectivity assays were performed in the presence or absence

764　of A3 proteins against HIVΔ*vif* (black) and SIVagmΔ*vif* (white). Rhesus macaque was

765　included as a positive control. Relative infection was normalized to viral infectivity in the

766　absence of A3 proteins. Averages of three replicates, each with triplicate infections ($\pm$

767　SEM) are shown. **B.** Western blot analysis of HA-tagged AGM A3H protein expression

768　in human (HEK293T) and AGM (Cos7) cell lines. The different size bands for different

769　AGM A3H haplotypes is reproducible. β-Actin is shown as a loading control. **C.** Top:

770　Single-cycle infectivity assay of HIVΔ*vif* in the presence of increasing amounts of A3H-

771　expressing plasmids. AGM A3H haplotype 1 (black circles) and rhesus macaque A3H

772　(open circles) are compared. Relative infection was normalized to viral infectivity in the

773　absence of A3 proteins. Averages of three replicates, each with triplicate infections ($\pm$

774　SEM) are shown. Bottom: Western blot analysis of protein expression levels with the

775　same amounts of plasmid in the top panel. β-Actin is shown as a loading control.

776　**Figure 3. Codon-optimization increases protein expression and antiviral activity.**

777　**A.** Western blot analysis for the expression of AGM A3H haplotype 1, codon-optimized

778　haplotype 1 A3H, rhesus macaque A3H, and codon-optimized rhesus macaque A3H.

779　Vinculin was used as a protein loading control. Quantification was done relative to

780　rhesus macaque A3H (normalized to 1). **B.** Top: Single-cycle infectivity assay of HIVΔ*vif*

781　in the presence of increasing amounts of A3H plasmid comparing codon-optimized

782　AGM haplotype 1 A3H (black squares) and codon-optimized rhesus macaque A3H

783　(open squares). Relative infection was normalized to viral infectivity in the absence of

784　A3 proteins. Averages of three replicates, each with triplicate infections ($\pm$ SEM) are

28

785   shown. Bottom: Western blot analysis of protein expression level with amounts of

786   plasmid added in panel B. β-Actin is shown as a loading control. **C.** Subcellular

787   localization of rhesus macaque and AGM A3H haplotype 1 in HeLa cells. A3H proteins

788   were detected with an anti-HA antibody (green) and DAPI staining was used to detect

789   the nucleus (blue). Image is representative of n = 90 total images over 3 replicates.
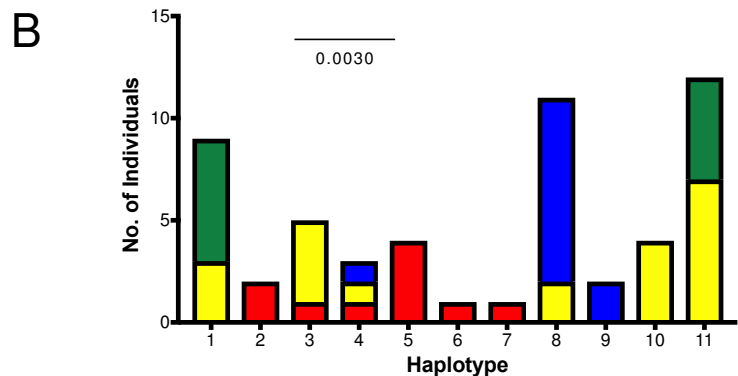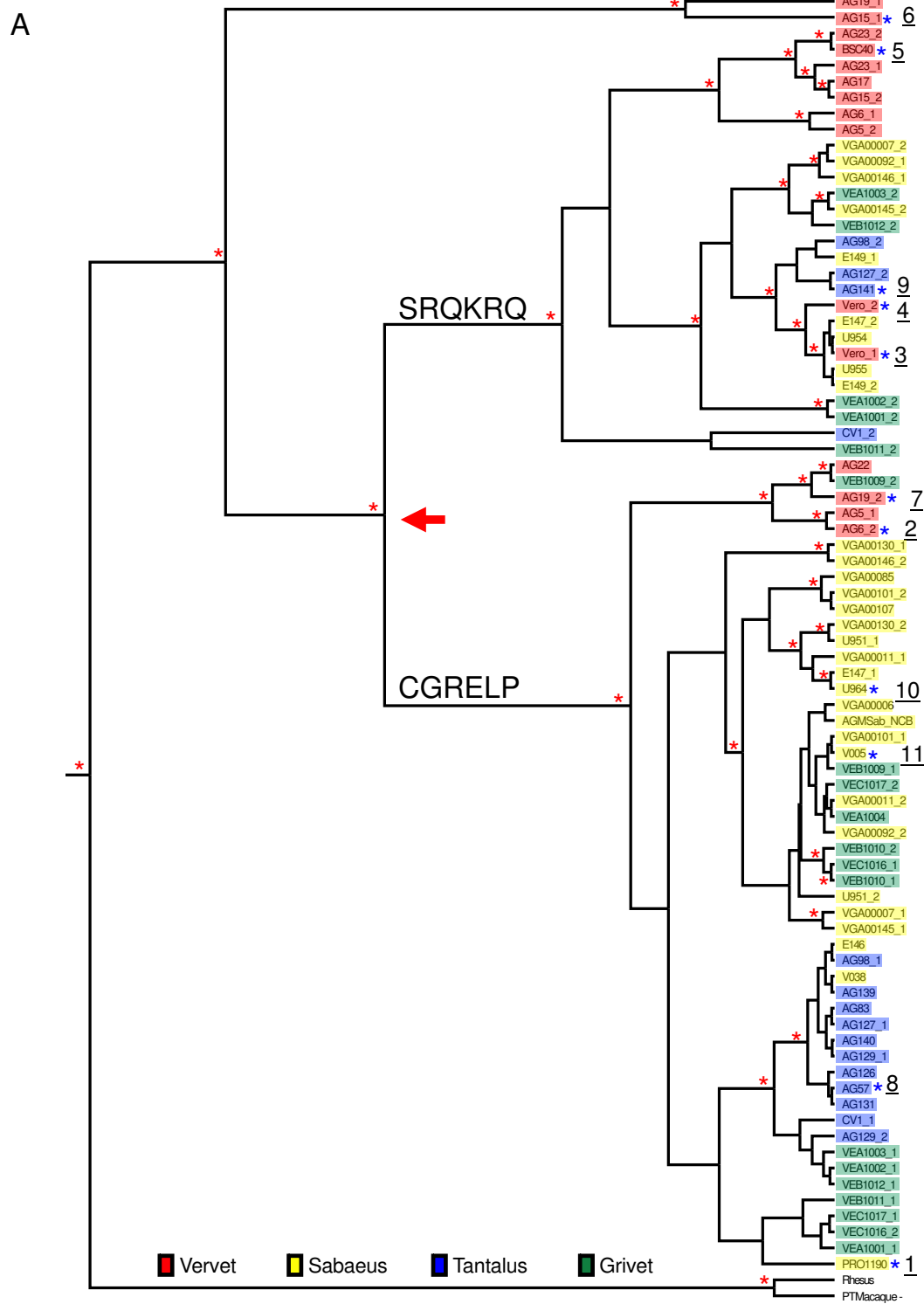
790   **Figure 4**. **AGM ancestors encode potent antiviral proteins. A.** A phylogeny, depicted

791   as a cladogram, based on the accepted species tree of all sequenced Old World

792   primates included in the study (27). Blue circles denote active antiviral proteins; red

793   circles denote inactive antiviral proteins. Ancestral nodes are labeled with numbers (1 –

794   4). **B.** Top: Single-cycle infectivity assay for HIVΔvif against extant primate A3H

795   proteins. Relative infection was normalized to viral infectivity in the absence of A3

796   proteins. Averages of three replicates, each with triplicate infections (± SEM) are

797   shown. Bottom: Western blot analysis of the protein expression level. Vinculin is used

798   as a loading control. **C.** Top: Single-cycle infectivity assay for HIVΔ*vif* against ancestral

799   A3H proteins and their extant descendants. Relative infection was normalized to viral

800   infectivity in the absence of A3 proteins. Averages of three replicates, each with

801   triplicate infections (± SEM) are shown. Dotted line at 10% is an arbitrary reference

802   point. Bottom: Western blot analysis of the protein expression level. Vinculin is used as

803   a loading control.

804   **Figure 5. Multiple amino acid mutations required for an increase in antiviral**

805   **activity. A.** Schematic of the A3H protein. Black bars outline the A3H catalytic site.

806   Numbered amino acid residues that are different between AGM A3H, patas A3H, and

807   the node 1 ancestor are outlined in red on the protein sequence alignment, ancestral

808   residues are colored blue. **B.** Top: Single-cycle infectivity assay for HIVΔ*vif* against

809   extant mutants. Relative infection was normalized to viral infectivity in the absence of A3

810   proteins. Averages of two replicates, each with triplicate infections (± SEM) are shown.

811   Bottom: Western blot analysis of protein expression levels of HA-tagged extant mutants

812   made in the AGM and patas A3H backgrounds. Vinculin is used as a loading control. **C.**

813   Top: Single-cycle infectivity assay for HIVΔ*vif* against AGM haplotype 1, node 1

814   ancestor, and node 1 mutant A3H. Relative infection was normalized to viral infectivity

815   in the absence of A3 proteins. Averages of two replicates, each with triplicate infections

816 (± SEM) are shown. Dotted line at 10% is an arbitrary reference point. Bottom: Western

817 blot analysis of protein expression levels of HA-tagged proteins made in the AGM and

818 patas A3H backgrounds. Vinculin is used as a loading control. **D.** A partial species

819 phylogeny, shown as a cladogram, accompanied by amino acid residues across Old

820 World monkeys at sites found to be important for antiviral activity.

821 **Figure 6. Evolution of A3H is not driven by Vif. A.** Results of positive selection

822 analyses of primate A3H. The last column lists sites under positive selection with dN/dS

823 >1 with a posterior probability > 0.98 under M8 Bayes Empirical Bayes (BEB)

824 implemented in PAML model 8. Sites are relative to African green monkey A3H. Sites

825 that had a posterior probability > 0.98 in both codon models (F3x4 and F61) are bolded.

826 **B.** A3H from individual PR01190 was modeled onto the pig-tailed macaque A3H

827 structure previously described by Bohn et al, 2017 (PDB 5W3V). Locations of positively

828 selected sites are denoted by red triangles. Blue triangles denote sites also under

829 positive selection in Old World primates. Site 208 is part of a region deleted in the

830 crystallized protein and is therefore not resolved in the model. **C.** Single-cycle infectivity

831 assay done with HIVΔ*vif* and HIV-1 expressing SIVagm Vif in the presence of codon-

832 optimized AGM A3H haplotype 1 (CO AGM), node 1 ancestor, and node 2 ancestor.

833 Relative infection was normalized to viral infectivity in the absence of A3 proteins.

834 Averages of three replicates, each with triplicate infections (± SEM) are shown.
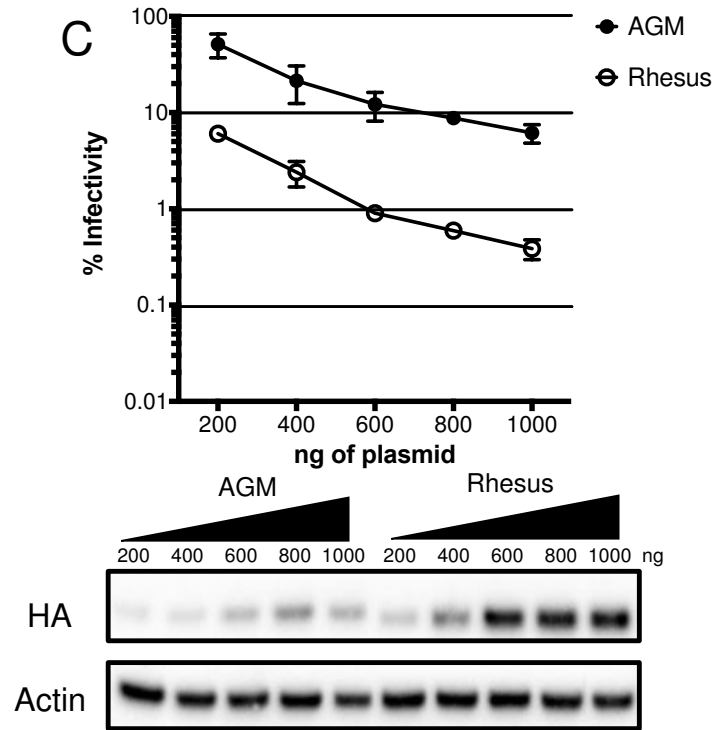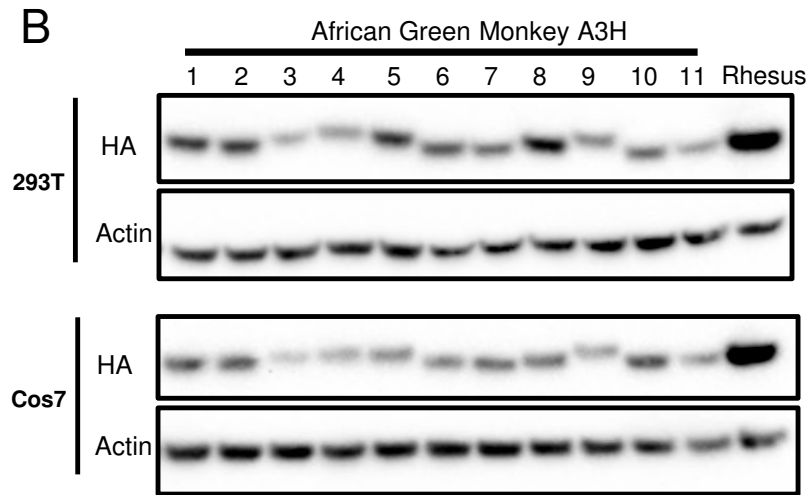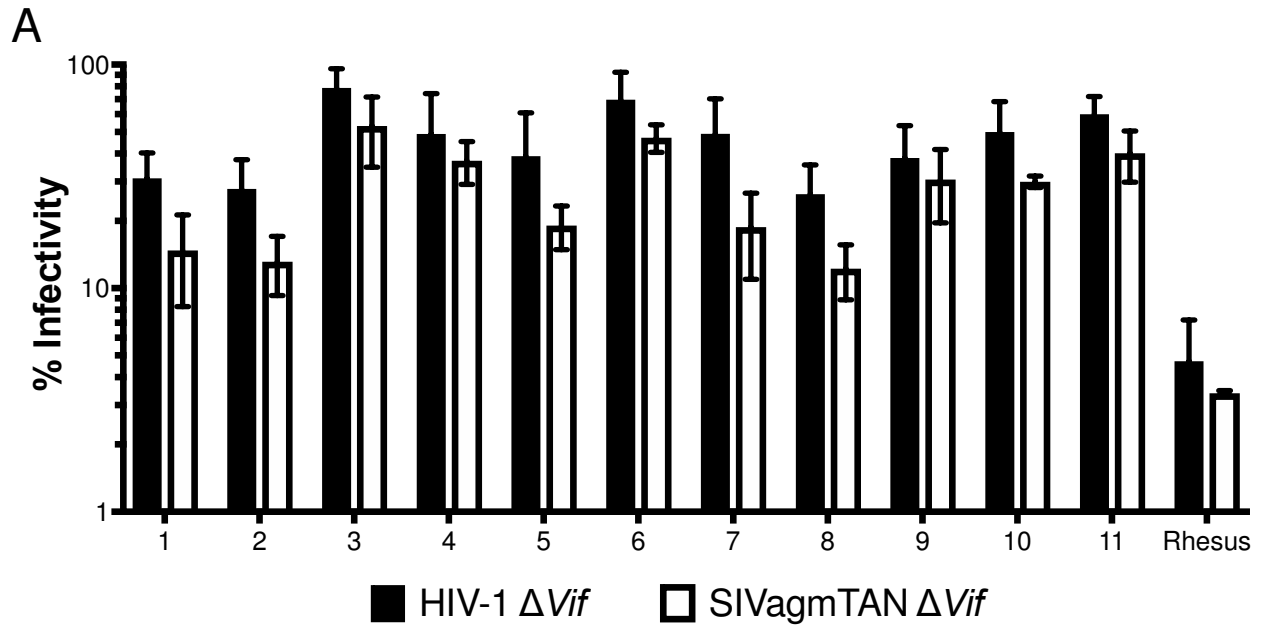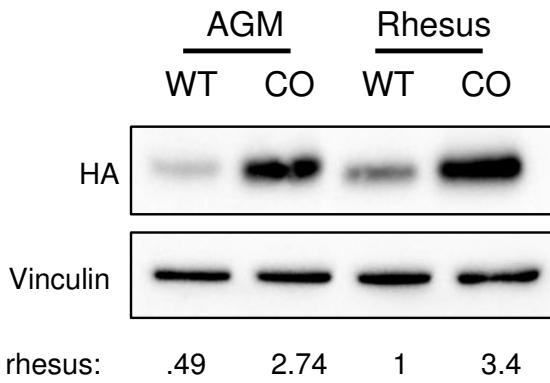
Figure 1

Figure 2

Figure 3

A

AGM | **Rhesus**

WT   CO   WT   CO

HA

Vinculin

Relative to rhesus: .49   2.74   1   3.4

B

C

Figure 4

Figure 5

A

HAE   PC--C

```
Node 1              MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAE…  NSRAIKRRLERI…
Patas               MALLTAKTFSLQFNNKHRVNKPYYRRKALLCYQLTPQNGSTPTRGHLKNKEKDHAE…  NSRAIKRRLERI…
AGM                 MALLTAKTFSLQFNNKRHVSKPYYPRKALLCYQLTPQNGSTPTRGHLQNKKKDHAE…  NSQAIKRRLERI…
                                    17,18,20,25                        48  51         171
Pig-tailed macaque  MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKEDHAE…  NSQAIKRRLERI…
Rhesus macaque      MALLTAKTFSLQFNNKRRVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAE…  NSQAIKRRLERI…
Sooty mangabey      MALLTAETFSLQFNNKRRVNKPHYPRKALLCYQLTPQNGSTPTRGHLKNKK-NHAE…  NSRAIKRRLERI…
De Brazza's Monkey  MALLTAKTFSLQFNNKR------YPRKALLCYQLTPQNGSTPTRGHLKNKKKDHAE…  NSRAIKRRLERI…
Talapoin            MALLTAKTFSLQFNNKRHVNKPYYPRKALLCYQLTPQNGSTPTRGHLKNKKKHHAE…  NSRAIKRRLERI…
                    ******:*********:       *  *********************:**: .***  **:*********
```

B

% Infectivity

WT H18R S20N Q48K Q171R H18R S20N H18R Q48K S20N Q48K WT H17R R25P E51K H17R R25P Node 1

AGM   Patas

HA

Vinculin

C

% Infectivity

AGM   Node 1   Node 1 Q48K

HA

Vinculin

Figure 6

A

# Species	M8a v M8 F3x4 2lnL	p-value	M8a vs M8 F61 2lnL	p-value	M8 BEB P.P. > 0.98
All primates: Hominoids, Old World monkeys, New World Monkeys					
27	69.625	p < 10⁻¹⁷	57.926	p < 10⁻¹⁴	**7**, **18**, **19**, **20**, **21**, **48**, **53**, 62, 90, **117**, **118**, **127**, **139**, **142**, **152**, **169**, **171**, **172**, **208**
Old World Monkeys					
15	30.051	p < 10⁻⁸	25.243	p < 10⁻⁷	**7**, **18**, **53**, **152**, **171**
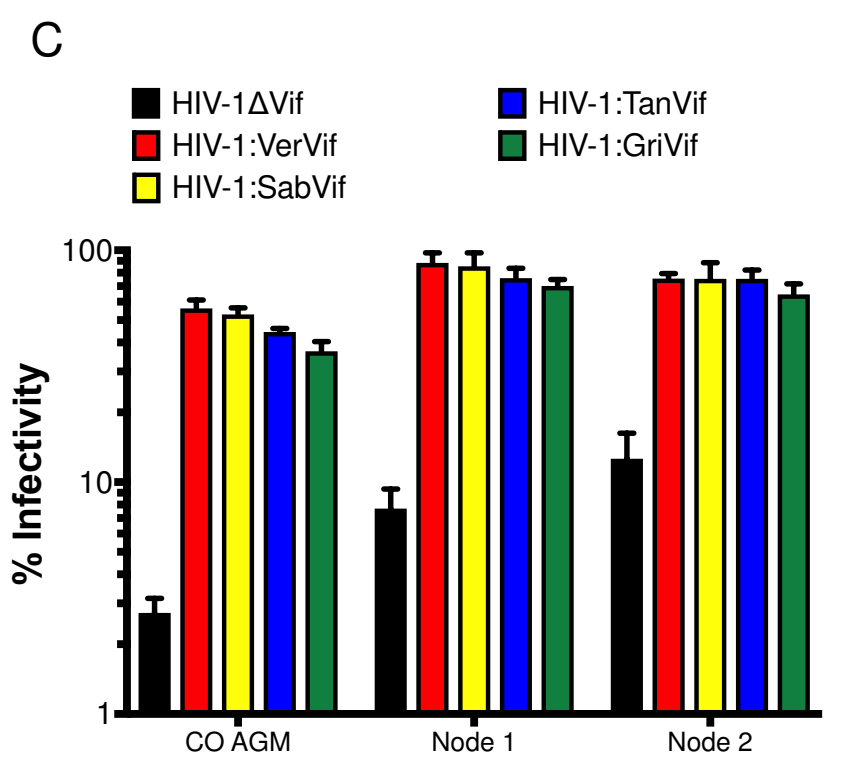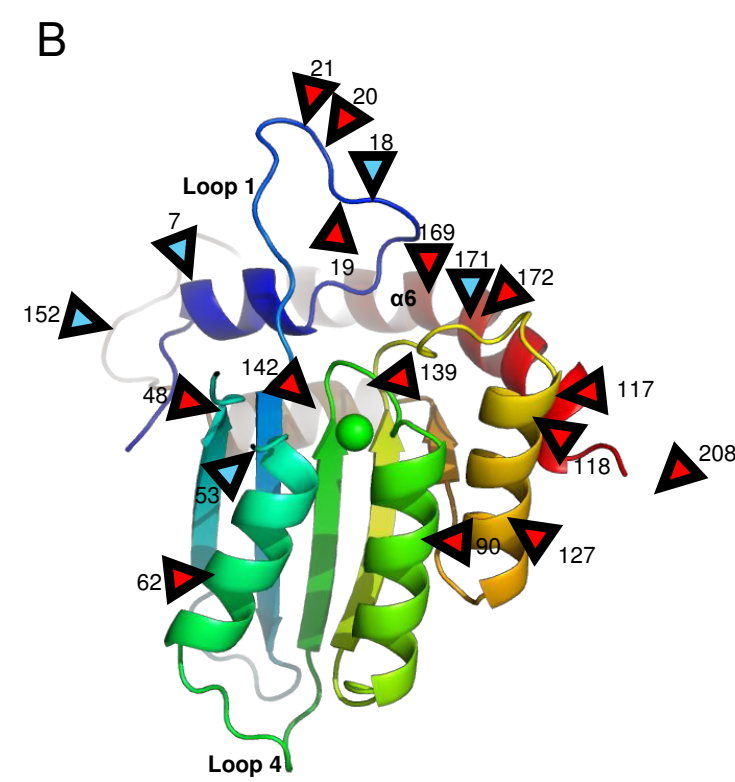
Table 1. Polymorphisms identified in AGM *A3H* sequences.

Site	Observed amino acids	Predicted location in structure[s]
17	R/H	
18	Y/R/H/L*	Loop 1
20	S/N	
25	P/R	
41	T/M*	Loop 2
44	R/K*	β-sheet
46	H/Q	
53	H/D	Loop 3
73	C/S*	Loop 4
75	R/Q	β-sheet
113	Y/C	
116	R/H	Loop 7
117	R/P*	
127	C/S	α4
128	G/R	
130	R/Q	Loop 8
134	E/K	β-sheet
138	L/R	α5
139	P/Q	
152	K/E	Loop 10
164	D/E	
171	Q/R	α6
181	K/E	
195	N/S	Not resolved in structure
204	S/A	

*Amino acid residues only identified in one animal

[a]Location in AGM A3H modeled onto a previously described pig-tailed macaque A3H (Bohn et al, PDB 5W3V)

Table 2. Polymorphic sites in tested A3H haplotypes.

Haplotype[a]	17[b]	18	20	25	53	75	113	116	127	128	130	134	138	139	152	164	171	182	195	204
													Amino acid							
1	R	H	S	P	D	Q	Y	R	C	G	R	E	L	P	K	Q	Q	K	S	S
2	R	H	S	P	D	Q	Y	R	C	G	R	E	L	P	K	Q	Q	K	N	A
3	H	H	S	R	H	R	Y	R	S	R	Q	K	R	Q	E	R	R	E	S	S
4	R	H	S	P	H	R	Y	R	S	R	Q	K	R	Q	E	R	R	E	S	S
5	R	H	S	P	D	Q	Y	R	S	R	Q	K	R	Q	E	R	R	E	N	A
6	R	R	N	R	D	Q	C	H	C	G	Q	K	L	Q	E	R	R	E	S	S
7	H	H	N	R	D	Q	Y	R	C	G	R	E	L	P	K	Q	Q	K	N	A
8	R	H	N	P	D	Q	Y	R	C	G	R	E	L	P	K	Q	Q	K	S	S
9	R	H	S	P	H	R	Y	R	S	R	Q	K	R	Q	E	R	R	E	N	A
10	R	H	S	R	D	Q	Y	R	C	G	R	E	L	P	K	Q	Q	K	S	S
11	H	H	S	R	D	Q	Y	R	C	G	R	E	L	P	K	Q	Q	K	S	S

[a]Haplotype number of A3H allele

[b]Amino acid position with nonsynonymous mutations across the tested haplotypes. Boxes colored in red have a different amino acid residue relative to AGM A3H haplotype 1.