Title: Functional rectification of the newly described African henipavirus fusion glycoprotein (Gh-M74a).

Running title (< 54 characters): Rectification of the African henipavirus F protein

Authors: Olivier Pernet¹§, Shannon Beaty¹§, Benhur Lee¹#.

Affiliations: ¹Department of Microbiology, Immunology and Molecular Genetics, University of California Los Angeles, Los Angeles, California, United States of America.

§ contributed equally

Correspondent footnote:

# To whom correspondence should be addressed. E-mail: benhur.lee@mssm.edu, phone: 310-794-2132, Fax: 310-267-2580

Word count

Abstract: 74 words

Text: 1,264 words
Recent evidence identified multiple *Henipavirus* species in Africa distinct from those in Southeast Asia and Australia. The reported fusion glycoprotein (F) sequence of the African Gh-M74a strain (GhV-F) is likely incorrect: a single base-pair deletion near the N-terminus results in multiple aberrancies. Rectifying single nucleotide insertions result in a GhV-F that now possesses a signal peptide, is efficiently cell-surface expressed, exhibits syncytia formation when co-expressed with GhV-G protein, and mediates pseudotyped viral particle entry.
One recent global survey of ~ 5,000 bat specimens identified 66 new species of paramyxoviruses—more species than were present in the Genbank database prior to the publication (2). This is worrisome, as paramyxoviruses are unusually susceptible to cross-species transfers, exhibiting one of the highest rates of host switching amongst RNA viruses (5). In particular, Drexler et al. reported >20 distinct viral clades within the Henipavirus genus with Nipah (NiV), Hendra (HeV) and Cedar (CedPV) viruses from Southeast Asia and Australia representing only three of those clades (2, 8). The phylogenetic diversity of henipaviruses (HNVs) is thus far greater than was previously thought. NiV and HeV cause lethal zoonoses in humans; infections can result in mortality rates exceeding 90% (3, 7, 9).

HNVs also exhibit a broader species tropism than other paramyxoviruses, in part due to the highly conserved ephrin-B2 and -B3 receptors that these viruses use (15). Although documented HNV spillover events are currently limited to Southeast Asia and Australia, the diversity of HNVs in endemic bat species (Eidolon helvum) throughout Africa (2, 12) raises the possibility of incipient spillover events. Indeed, a recent analysis of almost 500 human sera from Cameroon detected the significant presence of cross-neutralizing anti-NiV antibodies in groups considered to be at high risk for zoonotic exposure (e.g., bat bushmeat butchers) (13, 14). The multiple clades of African HNVs were determined from sequence analysis of an L gene segment. However, the almost full-length genomic sequence of one African HNV from Ghana (clone GH-M74a) was reported (Genbank accession number: HQ660129.1), which we will term GhV, and the fusion and attachment envelope glycoproteins of GhV as GhV-F’ and GhV-G, respectively. GhV-F’/G share only 26 – 56% sequence identity with their NiV/HeV counterparts.
Functional properties of GhV envelope glycoproteins. To examine the biological properties associated with GhV-F’ and GhV-G, we first codon-optimized and expressed the GhV-F’ and -G genes from their predicted Genbank reference sequences (AFH96010.1 and AFH96011.1, respectively). Consistent with recently published data (6), cell-surface GhV-G bound to soluble ephrinB2 (Fig. 1A). Next, we examined the fusogenicity of GhV-F’/G on human U87 glioblastoma cells, which express the HNV receptors ephrinB2 and -B3 (10) and are highly permissive to HNV envelope (F/G) mediated entry and syncytia formation (11). Curiously, while GhV-G cross-complemented NiV-F and HeV-F to form syncytia in permissive U87 cells (Fig. 1B, first column), GhV-F’ exhibited little to no fusogenic activity when complemented with heterotypic NiV-G and HeV-G or even its homotypic GhV-G (Fig. 1B, first row). Neither GhV-F’ (data not shown) nor GhV-G alone (Fig. 5A, “No F” control) induced any syncytia formation. Thus, while GhV-G was fully functional for receptor binding and could trigger heterotypic HNV-F proteins, GhV-F’ appeared to be functionally compromised compared to NiV-F or HeV-F (Fig. 1C). For clarity, we will designate this original GhV-F sequence as GhV-F’ (GhV-F prime).

N-terminus of GhV-F’ lacks a canonical signal peptide sequence. Sequence alignment and homology modelling shows that GhV-F shares the major structural features associated with other HNV-F proteins such as a cytoplasmic tail with the YxxL endocytic motif, HR1 and HR2 heptad repeat regions, a hydrophobic fusion peptide immediately following a basic cleavage site, and the conserved cysteines that are thought to link the F1 to F2 segments after protease cleavage (Fig. 2A). However, the predicted N-terminus of the GhV-F’ gene (Genbank: AFH96010.1) contains ~100 amino acids that are not conserved in the N-terminus of other HNV-F proteins (Fig. 2A and 2B),
and bioinformatic analysis indicates the absence of any recognizable signal peptide motif (Fig. 3A, 3B and Table I).

Signal peptides (SPs) are usually located within the first 20-30 amino acids of the N-terminus of plasma membrane-associated or secreted proteins, and they are characterized by a preponderance of aliphatic and hydrophobic amino acids such as leucines and isoleucines (1). Signal peptidases cleave SPs to generate the mature form of the associated proteins. Although there is no canonical motif in SPs that predicts cleavage by signal peptidases, SP prediction programs can identify potential SP cleavage sites with high confidence via a combination of contextual cues (see (16) and references therein). The median length of eukaryotic signal peptides is 22 aa (range 11-55) (1) and SPs were clearly present in the first ~ 25 amino acids of the F proteins from NiV_Mak, NiV_Bang, HeV, and CedPV, but not in GhV-F’ (Fig. 3B). Notably, four independent SP prediction programs failed to identify a SP within the first 70 aa of GhV-F’ (Table I).

Closer inspection reveals an isoleucine-/leucine-rich region in GhV-F’ (105-115 aa) that aligns well with the predicted SPs in the other HNV-F proteins (Fig. 3C). Remarkably, the genomic sequence encoding residues 95-115 of GhV-F’ is 81% AT-rich, and includes an out-of-frame ATG start codon motif that is upstream of the isoleucine-/leucine-rich region. Poly A/T sequences are well-known to cause stuttering during PCR amplification (4), which can result in inadvertent insertions or deletions. Insertion of one nucleotide in silico, which resulted in an in-frame start codon from the nearby ATG, rectified GhV-F’ such that the N-terminus of GhV-F’(+) aligned much better to the extant HNV-F proteins, both in terms of length and sequence similarity (Fig. 3A and 3C). Importantly, the rectified GhV-F’(+) was now predicted to have a bona fide SP (Fig. 3C and Table I).
Expression and cleavage of rectified GhV-F\(^{(\text{+1})}\). Depending on the identity of the inserted nucleotide, there are four possible versions of GhV-F\(^{(\text{+1})}\), which each have a different amino acid at position +4 from the new initiator methionine (Fig. 3C). We generated all four versions of GhV-F\(^{(\text{+1})}\) and evaluated their expression and their \(F_0 \rightarrow F_1+F_2\) cleavage levels, both in whole cell lysates and in pseudotyped viral particles. Interestingly, while GhV-F' is expressed in whole cell lysate (Fig. 4A and 4B), the protein is not efficiently cleaved (Fig. 4C) or incorporated into pseudotyped virions (Fig. 4D and 4E). The minimal levels that are cleaved or incorporated suggest that an internal hydrophobic domain might be used inefficiently as a surrogate SP. The diminished incorporation of GhV-F' into virions is consistent with inefficient trafficking through the secretory pathway due to the lack of any obvious SP sequence. GhV-F\(^{(\text{+T})}\) was poorly expressed in cell lysates (Fig. 4A and 4B) and on VSV-\(\Delta\)G-rLuc pseudotyped particles (Fig. 4D and 4E), however, cleavage was restored in the GhV-F\(^{(+A, +C, or +G)}\) proteins (Fig. 4C and 4F), and they were all more highly incorporated into pseudotyped virions than the fusion-compromised GhV-F' (Fig. 4D and 4E).

GhV-G levels of expression in cell lysates and incorporation into pseudotyped virions appeared to be inversely correlated with the expression levels of the GhV-F construct that was transfected \textit{in cis} (Fig. 4A and 4D). The significance of this is unclear, as pseudotyped virions bearing GhV-F\(^{(\text{+A}, +C, or +G)}\) were all more infectious than those bearing GhV-F' or \(^{(\text{+T})}\), despite the lower expression of GhV-G on the former (see Fig. 5C below).

Functionality of rectified GhV-F\(^{(\text{+1})}\). We evaluated the functionality of all four versions of GhV-F\(^{(\text{+1})}\) with a homotypic fusion assay. When co-expressed with GhV-G, GhV-F\(^{(\text{+A})}\) was the most fusogenic and generated the largest and most numerous syncytia.
F\(^{+C}\) and GhV-F\(^{+G}\) were also fusogenic but less so than GhV-F\(^{+A}\). GhV-F\(^{+T}\) was the least fusogenic and appeared no better than GhV-F\(^{-}\) (Fig. 5A and 5B). These fusogenic phenotypes were reflected in VSV-based pseudotyped virus infection assays, where GhV-F\(^{+T}\)/G and GhV-F\(^{-}\)/G pseudotypes were also markedly less infectious than GhV-F\(^{+A, C, or G}\)/G pseudotypes (Fig. 5C).

The functional defect of GhV-F\(^{-}\), but not its cause, was also recently recognized by Maisner and colleagues (17). Our results implicate a single nucleotide deletion in the reference sequence for GhV-F\(^{-}\) as the cause of its dysfunction. Given the A/T rich context (Fig. 3A) of the putative deletion, and that polymerase stuttering is most likely to result in an A or T deletion/insertion, we suggest that GhV-F\(^{+A}\) be tentatively designated as the working “wild-type” sequence for functional studies into this novel African henipavirus.
Acknowledgments.

This work was supported by NIH grants R01 AI069317 and a subproject grant from U54 AI065359 (PSWRCE) (to BL). S.B. was supported by the UCLA Molecular Pathogenesis Training Grant T32-AI07323. We thank members of the Lee lab for their critical comments on this manuscript.


Figure 1 | Functional properties of GhV envelope glycoproteins. (A) HNV-G proteins bind ephrinB2. HEK293T cells were transfected with GhV-G, NiV-G, or HeV-G expression plasmids. Empty pcDNA3.1-vector was also transfected to serve as a negative control. 27 hours post-transfection, cells were processed for FACS analysis using soluble ephrinB2-(hu)Fc as the primary staining agent, and Alexa Fluor 488 goat-anti-human (Fc) as the secondary antibody. (B) Fusogenicity of GhV-F'/G proteins. Human U87 glioblastoma cells were transfected with plasmids encoding the indicated F (GhV-F', NiV-F and HeV-F) and G (GhV-G, NiV-G, and HeV-G) proteins at a 1:1 ratio such that all homotypic and heterotypic combinations are represented. Cells were fixed and stained with Giemsa at 30 hours post-transfection. Images were taken from 5 random fields for each condition. Representative images are shown. (C) Quantification of fusogenicity of GhV-F'/G proteins. The images were quantified by manually counting nuclei in syncytia per field using ImageJ counting software. Results are presented as mean + S.D. from counting 5 fields. Significant differences were observed between GhV-F' and NiV-F (p < 0.01; one-tailed Student t-test).

Figure 2 | Comparison of conserved features in NiV-F and GhV-F'. (A) Comparison of conserved features across the full length of NiV-F and GhV-F'. The software PSIPRED was used to predict the secondary structure of the NiV-F and GhV-F' protein sequences. A schematic of the two proteins highlights major features and conserved regions. The unconserved region of the N-terminus in GhV-F' consists of about 100 amino acids. (B) Amino acid sequence of the full N-terminal region of GhV-F'. An extended version of Figure 2A is shown, which includes the complete amino acid sequence of the unconserved N-terminal region of GhV-F'.
Figure 3 | Sequence correction in the N-terminus of GhV-F' creates a conserved signal peptide sequence. (A) Addition of one nucleotide in GhV-F' creates a new start codon. Analysis of the protein and cDNA sequence in the unconserved region of GhV-F' revealed an out-of-frame ATG (highlighted yellow box) close to an A/T-rich region (outlined red box). Insertion of one nucleotide (+N) at the indicated position (right after nucleotide 7,166 in the HQ660129.1 reference sequence) shifts the reading frame of F', creates a new start codon from the upstream ATG (now position +1), and generates an amino acid insertion at position +4 (red letter, X). (B) Signal peptide prediction for GhV-F' and other HNV-F proteins. The first 70 amino acids of the indicated HNV-F proteins were bioinformatically analysed for the presence of signal peptide sequences using four independent signal peptide prediction programs (Table I). Results from the SignalP program are graphically presented, as described in Petersen et al. (16). SignalP is a neural network trained to predict signal peptides and provides three output scores for any query input sequence: the C-score (raw cleavage site score, red), S-score (signal peptide score, green) and Y-score (combined cleavage site score, blue). The C-, S- and Y-scores (y-axis) for each amino acid (x-axis) in the HNV-F query sequence are shown. In essence, the S score (green line) is high when the SignalP networks predict the sequence to be part of the signal peptide while the C-score (red vertical tics/lines) is trained to recognize the SP cleavage site and is therefore highest at the first residue after the cleavage site. The Y score (blue line) is a weighted score that helps identify the most likely cleavage site(s) when there are multiple C-score peaks. See http://www.cbs.dtu.dk/services/SignalP/output.php for details. (C) Alignment of HNV-F amino acid sequences. Insertion of a single nucleotide in GhV-F' creates an isoleucine- / leucine-rich N-terminus (dashed box) with a conserved signal peptide (green shaded box). (D) Signal peptide prediction for the rectified GhV-F(+1) proteins: GhV-F(+A), GhV-
F(+C), GhV-F(+G), and GhV-F(+T). Signal peptide motifs were analyzed using the SignalP 265 software, as described in (B).

**Table I | Bioinformatic analysis of HNV-F proteins to determine the presence of a signal peptide.** The first 70 amino acids of the indicated HNV-F proteins were analyzed by four different signal peptide prediction programs: SignalP, SPEPLip, PreDisi, and Signal-BLAST. Except for GhV-F', at least three, and often all four programs found signal peptide sequences at the N-terminus of all HNV-F proteins examined, including the rectified GhV-F(+1). Table shows the predicted signal peptide cleavage sites (between 21 and 28) in all the HNV-F sequences examined. PreDisi gives a predicted cleavage site of 36 for GhV-F', but the associated probability score that GhV-F' has a signal peptide is zero.

**Figure 4 | Expression of different GhV-F(+1) in association with GhV-G and resulting effect on pseudotyped virus formation and infectivity.** Viral pseudotypes were generated on HEK293T cells transfected with plasmids encoding GhV-G and the GhV-F(+1) variants in a 1:1 ratio. Cells were simultaneously infected with a VSV-ΔG-rLuc virus complemented with VSV-G. At 24 hours post infection (hpi), pseudotype virus-containing media were clarified by low-speed centrifugation, and then purified via ultracentrifugation through a 20% sucrose cushion. Purified GhV pseudotype particles (GhVpp) were resuspended in Opti-MEM, aliquoted and stored at -80°C. (A, D)

Expression of GhV-G and GhV-F' or GhV-F(+1) variants in whole cell lysates at 24 hpi (A) or in purified GhVpp (D). Primary anti-AU1 and anti-HA antibodies were used to detect AU1-tagged GhV-F and HA-tagged GhV-G, respectively. Fluorescent secondary antibodies were used for quantitative western blotting, and the blots imaged on a LI-
COR Odyssey scanner. β-tubulin is shown as a loading control. (B, E) The F₀ and F₁ bands from (A) and (D) were quantified using LI-COR Odyssey software and expression levels for the GhV-F(+1) variants in whole cell lysates (B) or GhVpp (E) are presented normalized to the expression of GhV-F' set at 100%. (C, F) Cleavage efficiency of GhV-F(+1) variants in whole cell lysates (C) or purified GhVpp (F) based on quantitative western blot values. Data is presented as % F cleavage, defined as F₁/(F₀ + F₁) *100. GhV-F(+1) was barely expressed in cell lysates and purified GhVpp, and thus reporting cleavage efficiency values would be meaningless (asterisk). Results are representatives of 3 different experiments.

Figure 5 | Functionality of GhV-F(+1) variants in syncytia formation and virus infectivity assays. Human U87 glioblastoma cells were transfected in duplicate with plasmids encoding the GhV-F(+1) variants and GhV-G proteins at a 1:1 ratio. (A) Cells were fixed and stained with Giemsa at 30 hours post transfection. Representative images are shown. (B) Quantification of fusogenicity of GhV-F/G proteins. The images were quantified by manually counting nuclei in syncytia per field using ImageJ software. Results are presented as mean ± S.D. from counting 10 fields per condition. (C) GhVpp infectivity. Purified GhVpp were produced as described for Fig. 4, and 5µL were used to infect U87 cells. Infections were performed in triplicate. Infected cells were lysed and assayed for renilla luciferase activity at 24 hpi. Activity was measured in Relative Luciferase Units (RLU), and data is presented as mean ± S.D. Statistical significance between the GhV-F' pseudotypes and the other groups was determined by a two-tailed Student t-test followed by the Holm step-down procedure for multiple comparisons (*p < 0.01).
Figure 1

(A) Histograms showing the ephrin-B2-AlexaFluor 488 counts for GhV-G, NiV-G, and HeV-G with and without the expression of GhV, NiV, and HeV.

(B) Microscopic images of GhV-F', NiV-F, and HeV-F for GhV-G, NiV-G, and HeV-G expression.

(C) Bar graph showing the average number of nuclei in syncyta per field for GhV-G, NiV-G, HeV-G, GhV-F', NiV-F, and HeV-F. Significant differences are indicated with p<0.01.
Figure 3

A

GhV-F

<table>
<thead>
<tr>
<th>Protein</th>
<th>DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>G3AAGA</td>
<td>G3AAGA</td>
</tr>
</tbody>
</table>

B

GhV-F

C

GhV-F

D

GhV-F

NIV-F

HeV-F

CedPV-F

Amino acid position

C-score

S-score

Y-score
Table I | Bioinformatic analysis of HNV-F proteins to determine the presence of a signal peptide. The first 70 amino acids of the indicated HNV-F proteins were analyzed by four different signal peptide prediction programs: SignalP, SPEPLip, PreDisi, and Signal-BLAST. Except for GhV-F', at least three, and often all four programs found signal peptide sequences at the N-terminus of all HNV-F proteins examined, including the rectified GhV-F(C). Table shows the predicted signal peptide cleavage sites (between 21 and 28) in all the HNV-F sequences examined. PreDisi gives a predicted cleavage site of 36 for GhV-F', but the associated probability score that GhV-F' has a signal peptide is zero.

<table>
<thead>
<tr>
<th></th>
<th>SignalP</th>
<th>SPEPLip</th>
<th>PreDisi</th>
<th>Signal-BLAST</th>
</tr>
</thead>
<tbody>
<tr>
<td>GhV-F'</td>
<td>Not Found</td>
<td>Not Found</td>
<td>36*</td>
<td>Not Found</td>
</tr>
<tr>
<td>GhV-F(U)</td>
<td>22-25</td>
<td>22</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>GhV-F(T)</td>
<td>22-25</td>
<td>22</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>GhV-F(D)</td>
<td>22-25</td>
<td>22</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>NiV-F_Malaysia</td>
<td>24-28</td>
<td>24</td>
<td>26</td>
<td>23</td>
</tr>
<tr>
<td>NiV-F_Bangladesh</td>
<td>24-28</td>
<td>Not Found</td>
<td>26</td>
<td>25</td>
</tr>
<tr>
<td>HeV-F</td>
<td>21-27</td>
<td>21</td>
<td>26</td>
<td>30</td>
</tr>
<tr>
<td>CedPV-F</td>
<td>23</td>
<td>Not Found</td>
<td>22</td>
<td>21</td>
</tr>
</tbody>
</table>

* Associated score = 0
Figure 5

A

GhV-G

GhV-F'  GhV-F(44A)  GhV-F(44C)

GhV-F(44G)  GhV-F(44T)

B

Average nuclei in syncytiotrophoblast per field

C

Pseudotype infectivity (10^6 RLU)

No-F  GhV-F  GhV-F(FM1)  GhV-F(FM9)  GhV-F(FM11)

GhV-G  GhV-G

*