Identification of *cis*-acting packaging signals in the coding regions of the influenza B virus HA gene segment

Lee Sherry, Karolina Punovuori, Louisa E. Wallace, Eliza Prangley, Sophie DeFries and David Jackson[#].

Biomolecular Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, KY16 9ST, United Kingdom.

[#] Corresponding author: Biomolecular Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, KY16 9ST, United Kingdom.

Phone: +44 13334 463 422

Email: dj10@st-andrews.ac.uk

Abstract: 149 words

Main text word count: 4563 words

Figures: 3

Running title: Influenza B virus genome packaging signals

Keywords: Influenza B virus, genome packaging, packaging signals

Abstract

For influenza A and B viruses to be infectious they require eight viral RNA (vRNA) genome segments to be packaged into virions. For efficient packaging influenza A viruses utilise *cis*-acting vRNA sequences, containing both non-coding and protein coding regions of each segment. Whether influenza B viruses have similar packaging signals is unknown. Here we show that coding regions at the 3' and 5' ends of the influenza B virus vRNA segment four are required for genome packaging, with the first 30 nucleotides at each end essential for this process. Synonymous mutation of these regions led to virus attenuation, an increase in defective particle production and a reduction in packaging of multiple vRNAs. Overall our data suggest that the influenza B virus vRNA gene segments likely interact with each other during the packaging process, which is driven by *cis*-acting packaging signals that extend into protein coding regions of the vRNA.

Introduction

Influenza A and B viruses are members of the *Orthomyxoviridae* family, containing an eight-segmented negative sense RNA genome. Each segment encodes one or more proteins that show high levels of similarity in structure and function between the two types of viruses, despite significant differences in amino acid identity (Palese & Shaw, 2007). Influenza A and B virions also show high levels of structural similarity, however a significant difference has been observed in genome arrangement within the core. The genome of both viruses is in the form of eight viral ribonucleoprotein complexes (vRNP) consisting of the viral RNA (vRNA) encapsidated by nucleoprotein and associated with a heterotrimeric polymerase complex. Electron microscopic analysis of influenza A virus particles demonstrated that the vRNPs are arranged in a highly ordered "7+1" orientation in which all segments align in a circular fashion with one segment in the centre (Noda *et al.*, 2006). However recent evidence suggests that this does not occur in influenza B viruses, as the vRNPs appear to twist around each other (Katz *et al.*, 2014). This suggests that influenza A and B viruses potentially adopt different mechanisms for genome packaging.

It is now widely accepted that influenza A viruses utilise a selective packaging process by which a single copy of each of the vRNPs is packaged into progeny virions in a process likely driven by bipartite *cis*-acting sequences within the vRNAs (Gerber *et al.*, 2014). Although early work suggested that the non-coding regions at the extreme termini of each vRNA were the minimal determinants of genome packaging (Luytjes *et al.*, 1989), more recent studies have shown that the extreme 5' and 3' ends of the protein coding regions are required for optimal genome packaging (reviewed in Hutchinson *et al.*, 2010). These *cis*-acting sequences are hypothesized to be responsible not only for packaging each genome segment into the virion but also for forming interactions between segments (Fournier *et al.*, 2012) to create a complex of vRNPs, thereby acting as bundling signals to allow a single copy of each vRNP to be packaged (Goto *et al.*, 2013). It is tempting to speculate that this may be responsible for the "7+1" genome arrangement, which questions whether influenza B viruses contain similar *cis*-acting packaging sequences due to the

85 significantly different appearance of the vRNPs within the virions. Although mutagenesis of the
86 non-coding regions of the influenza B virus vRNA segment four suggested these regions are
87 necessary for genome packaging and may therefore act as *cis*-acting packaging signals (Barclay &
88 Palese, 1995), whether these signals include protein coding regions similar to their influenza A
89 virus counterparts, is unknown.
90
91 Here we address this by analysing the role of the terminal protein coding regions of vRNA segment
92 four of influenza B viruses in genome packaging. We show that the first and last 150 nt of the
93 influenza B virus haemagglutinin (BHA) open reading frame are required for efficient packaging of
94 a viral-like RNA into virus-like particles and that the first 30 nt of the BHA ORF at the 3' end of the
95 negative sense vRNA are essential for this function. Analysis of mutant viruses containing
96 synonymous mutations in these regions demonstrated that mutations in the terminal 30 nt of the
97 protein coding region at both the 3' and 5' ends of vRNA segment four led to virus attenuation, an
98 increase in defective particle production and a reduction in packaging of multiple vRNA segments.
99 Our findings give the first evidence of the role of influenza B virus protein coding regions in
100 ensuring optimal genome packaging.
101
102 **Results and discussion**
103 A series of constructs were created for use in a mini-genome reporter assay (Sherry *et al.*, 2014)
104 to assess the importance of the terminal ends of the coding regions of influenza B virus vRNA
105 segment four (which encodes the haemagglutinin protein; BHA) in genome packaging. To create
106 the constructs the GFP coding sequence was flanked by the vRNA segment four non-coding
107 regions and the terminal 150 nucleotides at both termini of the BHA coding region, followed by
108 insertion into the pHH-21 plasmid in a 3'-5' negative sense orientation, thereby allowing expression
109 of a negative sense mini-genome vRNA segment (Fig. 1a). A series of truncations or deletions
110 were then introduced into the BHA coding regions at either terminus. The mini-genome assay
111 demonstrated the viral-like RNAs encoded on each construct expressed similar levels of GFP (Fig.
112 1b). The ability of each of these viral-like RNAs to be packaged into virus particles was assessed
113 by transfecting the constructs into 293T cells followed by infection with B/Yamanashi/98 virus and

the resultant virus-containing supernatant used to infect MDCK cells for FACS analysis. To determine the level of infectious virus present in the 293T cell supernatants the samples were serially 2-fold diluted and used to infect MDCK cells. At 10 hours post-infection (h.p.i.) cells were subjected to immunofluorescence analysis using an anti-BNP antibody to determine the number of virus antigen-positive cells. The dilution of supernatant that resulted in infection of 80-90% of cells was then placed onto fresh MDCK cells and at 10 h.p.i. cells were subjected to immunofluorescence and FACS analysis. Figure 1(c) shows representative GFP and viral antigen staining for cells infected with four of the supernatant samples. FACS analysis was performed to quantify the percentage of infected cells expressing GFP. 56% of virus-infected MDCK cells expressed the full-length HA(150)GFP(150) mini-genome vRNA segment, demonstrating the maximum packaging efficiency when using this approach, which is analogous to the level of packaging observed in a similar study using influenza A viruses (Fujii *et al.*, 2005). The packaging levels of all other segments were expressed as a percentage of the full-length mini-genome segment (Fig. 1d). Removal of the 150 nucleotide BHA coding region at the 3' end of the vRNA reduced the packaging efficiency to less than 1%, which could only be increased if the first 30 nucleotides of the this region were retained. When the 150 nucleotide BHA coding region at the 5' end of the vRNA was removed the packaging efficiency dropped to 9%, indicating that truncation at the 3' end of the vRNA had a larger effect on packaging. However when the terminal 30 nucleotides of the BHA coding region at the 5' end of the vRNA were replaced (HA(150)GFP(30)) packaging efficiency was increased to 73%. Overall the data indicate that although the entire 150 nucleotide regions at both ends of the BHA coding region are required for efficient genome packaging, the terminal 30 nucleotides at each end have a significant effect on this process, with those at the 3' end having the greatest effect on packaging efficiency. These findings are similar to those observed for packaging of various influenza A virus segments (Fujii *et al.*, 2005; Fujii *et al.*, 2003), suggesting that influenza B viruses may adopt a similar mechanism of genome packaging to their influenza A virus counterparts.

To assess the effects of the BHA coding regions on genome packaging in the context of infectious virus a series of mutant viruses were created in which synonymous mutations were introduced into

143 the first 30 nucleotides of the BHA coding region at the 3' end of the negative sense vRNA (3m30
144 mutations) or between nucleotides 31-60 (3m60 mutations) (Fig. 2a). Similar mutations were
145 introduced into the BHA coding region at the 5' end of the negative sense vRNA. Viruses were
146 generated containing mutations at either one or both termini. The rBHA-3m30/5m30 virus took
147 significantly longer to recover due to a propensity to generate defective particles, therefore this
148 virus was plaque purified to remove these particles prior to experimentation. The only virus
149 containing mutations at a single end of the gene that demonstrated attenuated replication kinetics
150 compared to the rBHA wt (wild type) virus was the rBHA-3m30 virus (Fig. 2b). Further attenuation
151 was observed when combining the 3m30 mutation with mutations in the BHA coding region at the
152 5' end of the vRNA, with the rBHA-3m30/5m30 virus demonstrating the largest degree of
153 attenuation (Fig. 2c). This confirmed that the terminal 30 nucleotides at both ends of the BHA
154 coding region have a significant effect on virus replication. This is remarkably similar to the results
155 of a previous study in which the packaging signals of the influenza A virus segment eight were
156 characterised (Fujii *et al.*, 2005). In this study synonymous mutations introduced into the 5' end of
157 the coding region did not alter viral replication, whereas mutations in the 3' end of the coding
158 region resulted in a slight attenuation, which was enhanced by introducing the mutations into both
159 the 3' and 5' coding regions. The phenotype of the double mutant reported by Fujii *et al.* was
160 therefore remarkably similar to the rBHA-3m30/5m30 virus, suggesting these sequences are
161 essential for efficient viral replication of both influenza A and B viruses, possibly indicating similar
162 functional requirements.
163
164 As the 3m30/5m30 mutations may have reduced the efficiency of genome packaging, resulting in
165 higher levels of defective virus production, the particle to infectivity ratio of all viruses stocks was
166 analysed. The only virus that demonstrated a marked increase in the number of defective particles
167 was the rBHA-3m30/5m30 virus (Fig. 2d). The 36 and 48 hour time point samples (Fig. 2a and b)
168 of viruses containing the 3m30 mutations were analysed for particle to infectivity ratios. The only
169 virus that demonstrated a significant increase in defective particles at both time points was the
170 rBHA-3m30/5m30 virus (Fig 2e). The level of defective virus particles increased for all viruses
171 between the two time points, however while most viruses demonstrated a 2- to 7-fold increase,

there was a 39-fold increase in rBHA-3m30/5m30 defective particles. The propensity for this virus to generate defective particles was likely responsible for its attenuation in the replication analysis.

Sequence analysis of vRNA segment four of multiple influenza B viruses isolated between 1940 and 2014 shows that the 3m30, 3m60, 5m30 and 5m60 regions of the genome are highly conserved, with the 3m30 region completely conserved in all analysed sequences (Fig. 3a). The lack of synonymous mutations in the 3m30 region over a 74-year time period may suggest that the sequence integrity of the RNA in this region is essential for virus viability, potentially through contributing to efficient genome packaging. This is not surprising given that previous studies have shown that the areas of various influenza A virus vRNA segments that contain the highest levels of sequence conservation are found in the terminal ends of the coding regions that have been implicated as packaging signals (Gog *et al.*, 2007; Marsh *et al.*, 2007). As the mutated regions of the rBHA-3m30/5m30 virus potentially represent packaging signals, changes in RNA sequence or secondary structure may have reduced the efficiency of segment four packaging. However it was also possible that such changes had affected the efficiency of mRNA transcription from the mutant vRNA, thereby leading to viral attenuation. To address this vRNA segment four from the wt and rBHA-3m30/5m30 viruses was cloned into a plasmid under the control of the human RNA polymerase I promoter allowing its expression in a mini-genome assay. The results showed similar levels of BHA for both wt and mutant segments (Fig. 3b and c). This demonstrated that the 3m30/5m30 mutations did not affect the ability of the polymerase to transcribe BHA mRNA and therefore did not affect BHA protein levels.

To analyse whether the BHA-3m30/5m30 mutations reduced vRNA content in released virions, resulting in increased levels of defective particles shown in Fig. 2(d), the BNP and vRNA levels in wt and mutant virions were assessed. After purification of the viruses an equal number of wt and mutant virions (determined by haemagglutination assay) were analysed for protein content by SDS-PAGE and coomassie staining. BHA was present in the trypsin-cleaved form of BHA_1 and BHA_2 (Fig. 3d). A slight reduction in overall viral protein levels were observed in the mutant virions compared to those of rBHA wt, suggesting the presence of fewer particles, however a greater

reduction was observed in BNP levels compared to other proteins. To accurately identify the viral proteins and to quantify their levels in both viruses the viral proteins were analysed by immunoblotting. While BM1 and BHA levels were similar for both viruses the BNP level in the mutant was only 63% that of wt (Fig. 3e). As BNP is associated with the viral genome this reduction likely indicated reduced vRNA content in the virions. This was confirmed by qRT-PCR analysis (Fig. 3f). Levels of three vRNA segments were present in the mutant virions at only 56 to 74% those of wt virus, which is a similar reduction to the levels of BNP observed in Fig. 3(e). This demonstrated a genome packaging deficiency in the mutant, and suggested that the deficiency in packaging segment four also reduced packaging of other segments. It has been suggested that intermolecular interactions form between influenza A virus vRNPs such that the full complement of vRNPs are present as a complex prior to packaging into virions (Goto *et al.*, 2013). A reduction in packaging efficiency of a single segment could therefore reduce or even prevent packaging of other segments. It is possible that this is also true of influenza B virus vRNPs and the mutation of packaging signals in the rBHA-3m30/5m30 virus segment four reduced the packaging of multiple segments. This would explain the increased level of defective particles and virus attenuation observed in Fig. 2, however whether the defective particles were semi-infectious as they lacked one or more vRNA segment, or lacked genome altogether is unknown. The level of vRNA segment eight was reduced to a larger extent than the other segments. Similar variations in packaging efficiencies between influenza A virus vRNPs have been described (Marsh *et al.*, 2007), further suggesting that influenza B virus genome packaging may adopt similar mechanisms to that of influenza A viruses, potentially indicating the likelihood of direct interactions between specific vRNPs during influenza B virus packaging.

It is not surprising that studies in recent years have identified coding regions in the influenza A virus vRNA as being important for efficient genome packaging, as prior work had shown that defective viruses containing truncated gene segments retained a full complement of eight vRNPs, with the defective segment retaining both the non-coding regions and a portion of the coding regions at either terminus (Duhaut & Dimmock, 1998; Duhaut & McCauley, 1996; Nayak *et al.*, 1982), with these coding regions increasing the stability of the defective particles (Duhaut &

Dimmock, 2000). It is plausible that short coding regions were retained as they contained the *cis*-acting packaging signals. Similar findings have been observed in defective influenza B viruses containing truncations in vRNA segment four, with the first and last 150-200bp of the BHA ORF retained (DJ unpublished results). This along with the data shown above suggests that influenza B viruses do contain similar genome packaging signals to those observed for influenza A viruses, and that these signals extend into the coding regions. Our results show that the extreme terminal 3' and 5' ends of the coding regions of the influenza B virus vRNA segment four are required for efficient genome packaging, with the first 30 nucleotides at the 3' end having the greatest influence on packaging efficiency, similar to findings in the context of influenza A virus segments six and eight packaging (Fujii *et al.*, 2005; Fujii *et al.*, 2003). It was previously shown through truncation analysis that the terminal 3' and 5' ends of the coding region of the influenza A virus segment four are required for efficient packaging to occur (Marsh *et al.*, 2007; Watanabe *et al.*, 2003). However unlike the findings for influenza B virus segment four, systematic synonymous mutagenesis in the influenza A virus segment revealed that a 15 nucleotide stretch in the terminal 80 nucleotides at the 5' end of the coding region had the greatest influence on packaging efficiency (Marsh *et al.*, 2007). Unfortunately the studies of Marsh *et al.* did not include analysis of viruses containing synonymous mutations at both the 3' and 5' ends of the same segment, therefore it is unknown whether this would have enhanced the attenuation of a recombinant virus, similar to the rBHA-3m30/5m30 virus in the current study. It is likely that these differences in segment four packaging sequence requirements between the two viruses are influenced by the mechanism of vRNA interaction between the individual segments, especially as it has been reported that vRNPs interact during transport to the site of assembly (Chou *et al.*, 2013; Gavazzi *et al.*, 2013; Lakdawala *et al.*, 2014). Although our results suggest that the mechanism of genome packaging utilised by influenza A and B viruses is similar, the fact that the 3' end of the coding region was more important in packaging of influenza B virus vRNA segment four compared to the 5' end in influenza A viruses may suggest that there is a difference in the mechanism behind vRNA-vRNA interaction formation between the two viruses, or that the influenza B virus vRNA segment four interacts with different vRNPs compared to its influenza A virus counterpart. Further experiments using labelled vRNPs (Lakdawala *et al.*, 2014) could address this. Furthermore while EM comparisons of vRNP structural

arrangement in influenza A and B virions might suggest different mechanisms of genome packaging between the two viruses (Katz *et al.*, 2014; Noda *et al.*, 2006), our results suggest a similar mechanism of *cis*-acting RNA signal-mediated packaging is employed by both viruses. The difference in vRNP arrangement is likely due to vRNA-vRNA interactions differing between the two viruses, leading to the more twisted vRNP complex observed in influenza B viruses.

Overall we propose that the 3' and 5' ends of the coding region of the influenza B virus vRNA segment four contains *cis*-acting packaging signals and that mutations in the first 30 nucleotides of these regions reduce the efficiency of genome packaging into progeny virions by approximately 40%, similar to that observed for mutagenesis of packaging signals in the same segment of influenza A virus (Marsh *et al.*, 2007). This reduction in segment four packaging subsequently reduces the packaging of other segments, resulting in an increase in the release of defective particles, thereby attenuating the virus. It is important to fully characterise the *cis*-acting packaging signals in influenza B viruses as this could have significant consequences for vaccine production. Contemporary influenza B viruses replicate poorly in eggs and therefore introducing segments four and six (encoding the BHA and BNA proteins respectively) into the background of a virus adapted for efficient growth in eggs by reverse genetics could offer significant advantages for vaccine production purposes (Hoffmann *et al.*, 2002). However natural mutations in the coding regions of these segments may reduce the efficiency of vRNA packaging, which has previously been shown to be a limiting factor in influenza A virus reassortment (Essere *et al.*, 2013). Therefore ensuring these signals are optimal for efficient recombinant virus generation could enhance vaccine seed production.

Materials and methods

Cells and Viruses

293T and MDCK cells were maintained in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) supplemented with 10% foetal calf serum (FCS) at 37 °C with 5% CO_2. B/Yamanashi/98 wild-type (rBHA wt) and mutant viruses were generated using plasmid-based reverse genetics as previously described (Hoffmann *et al.,* 2002). Briefly, 293T cells were

transfected with eight genome-encoding bi-directional (pAB) plasmids using FuGENE 6 transfection reagent (Promega) and at 16 h post-transfection the cells were co-cultured with MDCK cells in serum-free DMEM containing 2.5 µg/mL N-acetyl trypsin (Sigma). Supernatants were harvested four days post-transfection, viruses propagated twice through MDCK cells followed by plaque assay titration on MDCK cells. Viral RNA was extracted using the QIAamp viral RNA kit (QIAGEN), vRNA segment four of each virus was amplified by reverse-transcriptase PCR using genome specific primers and the resultant DNA sequenced to confirm presence of the desired mutations.

Plasmids

To create the pHH-HA(150)GFP(150) mini-genome plasmid the 3' and 5' non-coding regions of B/Yamanashi/98 vRNA segment four flanked by 150nt of the BHA coding region at either end of the gene segment were amplified by PCR using the parental pAB-HA plasmid as template. The eGFP coding sequence was amplified by PCR and inserted between the 3' and 5' end BHA PCR products by overlapping PCR. A translation termination codon was inserted after the GFP ORF. The entire HA(150)GFP(150) PCR product was inserted into the pHH-21 reverse genetics vector between *Bsm*BI restriction sites in a negative sense orientation such that expression of the negative sense viral-like RNA was under the control of the human polymerase I promoter. To create the series of truncation mutant mini-genome constructs truncated versions of the 150 nt BHA coding regions at either or both termini were amplified using specific primer sets, these PCR products were fused to the 3' or 5' non-coding region by overlapping PCR followed by insertion of the eGFP coding sequence by overlapping PCR. The resultant products were inserted into pHH-21 as above. All primer sequences are available upon request. For creation of mutant viruses synonymous mutations were introduced into the BHA open reading frame within the pAB-HA reverse genetics plasmid by site-directed mutagenesis and the resultant plasmids substituted into the reverse genetics system. The presence of the desired mutations were confirmed by DNA sequencing. To generate the pHH-BHA wt and pHH-BHA-3m30/5m30 plasmids total vRNA was extracted from rBHA wt or rBHA-3m30/5m30 virus and vRNA segment four was amplified by reverse transcriptase PCR. The resultant cDNA was inserted into the pHH-21 plasmid in a

317 negative sense orientation between the human RNA polymerase I promoter and terminator
318 sequences using *Bsm*BI restriction sites. Segment four sequences were confirmed by DNA
319 sequencing.
320
321 **Mini-genome assay**
322 To test expression of GFP from the mini-genome plasmids, 293T cells in 12-well plates were
323 transfected with 500 ng of pAB expression plasmids encoding PB1, PB2, PA and NP of
324 B/Yamanashi/98 virus and 1 µg of each mini-genome construct. 24 hours post-transfection cells
325 were fixed in 5% formaldehyde and GFP expression was observed at 20x magnification using a
326 Nikon Microphot-FXA fluorescence microscope. Cell nuclei were stained with 1 µg/ml 4',6-
327 diamidino-2-phenylindole (DAPI) and the total number of cells in each panel was determined by
328 manual counting using ImageJ. GFP-positive cells were manually counted and the percentage of
329 the total cell number positive for GFP was calculated. All samples were analysed in triplicate
330 experiments.
331
332 **Virus-like particle (VLP) assay**
333 293T cells were transfected with 1 µg of each mini-genome construct and at 24 h.p.t. infected with
334 B/Yamanashi/98 virus at an MOI of 10. At 16 h.p.i supernatants were harvested, N-acetyl trypsin-
335 treated at 37 °C for 30 mins, serially 2-fold diluted and placed on MDCK cells. At 10 h.p.i cells were
336 either analysed by immunofluorescence or FACS. For immunofluorescence analysis cells were
337 fixed in 5% formaldehyde, permeabilised in PBS/0.5% Triton X-100/0.5% NP40, blocked in PBN
338 blocking buffer (PBS/1% BSA/0.02% sodium azide) for 1 h at room temperature and stained with
339 an anti-BNP monoclonal antibody (Abcam) followed by an anti-mouse Texas-red conjugated
340 secondary antibody. Cell nuclei were stained with DAPI and samples analysed at 20x
341 magnification using a Nikon Microphot-FXA fluorescence microscope. For FACS analysis cells
342 were treated with PBS/EDTA to gain single cell suspensions, fixed in 5% formaldehyde, stained
343 with the anti-BNP monoclonal antibody and anti-mouse Texas-red conjugated secondary antibody
344 in suspension and subjected to FACS analysis to determine percentages of antigen-positive cells

(red) that also expressed GFP. 5000 cells were counted per sample and each sample was performed in duplicate.

Virus replication kinetics

MDCK cells were infected with either rBHA wt (wild-type) or mutant viruses at an MOI of 0.001 and supernatant samples harvested every 12 h until 72 hours post-infection. The infectivity of the samples was determined by plaque assay titration on MDCK cells. Results represent the average of three independent experiments ± S.D.

Particle to infectivity ratio analysis

The approximate number of virus particles in virus stocks was estimated by heamagglutination assay. Virus samples were serially two-fold diluted in PBS in rows of a V-bottomed 96-well plate in 50 µl volumes per well. 50 µl of chicken red blood cells (1 % diluted in PBS) were added to each well, mixed and incubated at 4 °C until haemagglutination was observed. The reciprocal of the final dilution of virus to display haemagglutination was used to determine the HA titre per ml of each virus. As approximately $1x10^6$ virus particles are required to achieve an HA titre of 1 (Donald & Isaacs, 1954), the HA titres of the virus stocks were multiplied by $1x10^6$ to determine the number of virus particles per ml in each sample. The infectious titre of each sample (pfu/ml) was determined by plaque assay titration on MDCK cells. The particle to infectivity ratio (particles/ml : plaque forming units/ml) was determined for each sample.

Viral protein analysis by SDS-PAGE

To analyse BHA mRNA transcription and protein expression levels from rBHA wt or rBHA-3m30/5m30 vRNA segment four, 1 µg of the pHH-BHA or pHH-BHA3m30/5m30 plasmids (described above) were transfected into 293T cells in a 12 well plate alongside 500 ng of pAB expression plasmids encoding PB1, PB2, PA and NP of B/Yamanashi/98 virus. 24 hours post-transfection cells were lysed in 2x disruption buffer (6 M urea, 2 M β-mercaptoethanol, 4% sodium dodecyl sulphate), proteins separated by SDS-PAGE and transferred to Immobilon-FL polyvinylidene difluoride membranes (Millipore). Membranes were blocked in blocking buffer (PBS,

0.1% Tween 20, 5% dried milk) and incubated with an anti-B/Hong Kong/73 polyclonal antisera (to detect viral BHA and BNP proteins) or an anti-actin monoclonal antibody. Protein detection was performed using IRDye 680- or IRDye 800-conjugated secondary antibodies (Licor) on an Odyssey CLx near infrared scanner (Licor), images were collected and protein band intensities were quantified using ImageStudio (Licor).

To analyse virion protein content 500 000 HA units of rBHA wt and rBHA-3m30/5m30 viruses were purified by sucrose gradient ultracentrifugation. Virus samples were placed onto a 30% sucrose cushion and subjected to ultracentrifugation at 112 000 x g for 2.5 h at 4 °C. Pelleted viruses were resuspended in 500 µl NTE buffer (150mM NaCl, 10mM Tris-HCl pH 7.5, 1mM EDTA) and placed onto a continuous 30-60% sucrose gradient, followed by ultracentrifugation at 112 000 x g for 2.5 h at 4 °C. Virus bands were extracted from the gradient and pelleted by ultracentrifugation at 112 000 x g for 2 h at 4 °C. Purified virus pellets were resuspended in 50 µl NTE buffer. All virus purifications were performed in triplicate. Viral proteins in 5 µl of each virus sample were then separated by SDS-PAGE and analysed by coomassie staining or immunoblotting using the anti-B/Hong Kong/73 polyclonal antisera as described above. Protein bands were detected and quantified for BHA, BNP and BM1.

Analysis of virion vRNA content by quantitative reverse transcriptase-PCR

Total vRNA was extracted from three separate samples of rBHA wt or rBHA-3m30/5m30 virus (5.12 x10^9 particles for each virus as determined by haemagglutination assay). vRNA segments four, five and eight were reverse transcribed using gene-specific primers and RevertAid Premium Reverse Transcriptase (Thermo Scientific). For qPCR viral gene-specific primers (sequences available on request) were designed to amplify a 150 nt fragment of DNA and various concentrations of primers were optimised against each other by qPCR using various concentrations of pAB-HA, pAB-NP or pAB-NS as standardized templates. cDNAs generated by reverse transcription were then assayed by qPCR using serial four-fold dilutions of cDNA and appropriate concentrations of gene-specific primers using Precision Mastermix (Primer Design) on a Stratagene Mx3005P real-time PCR thermocycler. A standard curve was generated using serial

10-fold dilutions of pAB-HA, pAB-NP or pAB-NS and used to convert ct values into cDNA concentrations.

Acknowledgements

The pAB plasmids for virus recovery were kindly provided by Dr Robert Webster (St Jude Children's Research Hospital, Memphis TN, USA). Lee Sherry is indebted to the University of St Andrews for a Ph.D. studentship through an MRC Doctoral Training Grant. We also gratefully acknowledge support by the University of St Andrews (KP, LEW, EP, SD and DJ), which is a charity registered in Scotland (No.SC013532).

References

Barclay, W. S. & Palese, P. (1995). Influenza B viruses with site-specific mutations introduced into the HA gene. *J Virol* **69**, 1275-1279.

Chou, Y. Y., Heaton, N. S., Gao, Q., Palese, P., Singer, R. H. & Lionnet, T. (2013). Colocalization of different influenza viral RNA segments in the cytoplasm before viral budding as shown by single-molecule sensitivity FISH analysis. *PLoS Pathog* **9**, e1003358.

Donald, H. B. & Isaacs, A. (1954). Counts of influenza virus particles. *J Gen Microbiol* **10**, 457-464.

Duhaut, S. & Dimmock, N. J. (2000). Approximately 150 nucleotides from the 5' end of an influenza A segment 1 defective virion RNA are needed for genome stability during passage of defective virus in infected cells. *Virology* **275**, 278-285.

Duhaut, S. D. & Dimmock, N. J. (1998). Heterologous protection of mice from a lethal human H1N1 influenza A virus infection by H3N8 equine defective interfering virus: comparison of defective RNA sequences isolated from the DI inoculum and mouse lung. *Virology* **248**, 241-253.

Duhaut, S. D. & McCauley, J. W. (1996). Defective RNAs inhibit the assembly of influenza virus genome segments in a segment-specific manner. *Virology* **216**, 326-337.

Essere, B., Yver, M., Gavazzi, C., Terrier, O., Isel, C., Fournier, E., Giroux, F., Textoris, J., Julien, T., Socratous, C., Rosa-Calatrava, M., Lina, B., Marquet, R. & Moules, V. (2013). Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc Natl Acad Sci U S A* **110**, E3840-3848.

Fournier, E., Moules, V., Essere, B., Paillart, J. C., Sirbat, J. D., Isel, C., Cavalier, A., Rolland, J. P., Thomas, D., Lina, B. & Marquet, R. (2012). A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res* **40**, 2197-2209.

Fujii, K., Fujii, Y., Noda, T., Muramoto, Y., Watanabe, T., Takada, A., Goto, H., Horimoto, T. & Kawaoka, Y. (2005). Importance of both the coding and the segment-specific noncoding regions of the influenza A virus NS segment for its efficient incorporation into virions. *J Virol* **79**, 3766-3774.

Fujii, Y., Goto, H., Watanabe, T., Yoshida, T. & Kawaoka, Y. (2003). Selective incorporation of influenza virus RNA segments into virions. *Proc Natl Acad Sci U S A* **100**, 2002-2007.

Gavazzi, C., Yver, M., Isel, C., Smyth, R. P., Rosa-Calatrava, M., Lina, B., Moules, V. & Marquet, R. (2013). A functional sequence-specific interaction between influenza A virus genomic RNA segments. *Proc Natl Acad Sci U S A* **110**, 16604-16609.

Gerber, M., Isel, C., Moules, V. & Marquet, R. (2014). Selective packaging of the influenza A genome and consequences for genetic reassortment. *Trends Microbiol* **22**, 446-455.

Gog, J. R., Afonso Edos, S., Dalton, R. M., Leclercq, I., Tiley, L., Elton, D., von Kirchbach, J. C., Naffakh, N., Escriou, N. & Digard, P. (2007). Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic Acids Res* **35**, 1897-1907.

Goto, H., Muramoto, Y., Noda, T. & Kawaoka, Y. (2013). The genome-packaging signal of the influenza A virus genome comprises a genome incorporation signal and a genome-bundling signal. *J Virol* **87**, 11316-11322.

Hoffmann, E., Mahmood, K., Yang, C. F., Webster, R. G., Greenberg, H. B. & Kemble, G. (2002). Rescue of influenza B virus from eight plasmids. *Proc Natl Acad Sci U S A* **99**, 11411-11416.

Hutchinson, E. C., von Kirchbach, J. C., Gog, J. R. & Digard, P. (2010). Genome packaging in influenza A virus. *J Gen Virol* **91**, 313-328.

Katz, G., Benkarroum, Y., Wei, H., Rice, W. J., Bucher, D., Alimova, A., Katz, A., Klukowska, J., Herman, G. T. & Gottlieb, P. (2014). Morphology of influenza B/Lee/40 determined by cryo-electron microscopy. *PLoS One* **9**, e88288.

Lakdawala, S. S., Wu, Y., Wawrzusin, P., Kabat, J., Broadbent, A. J., Lamirande, E. W., Fodor, E., Altan-Bonnet, N., Shroff, H. & Subbarao, K. (2014). Influenza a virus assembly intermediates fuse in the cytoplasm. *PLoS Pathog* **10**, e1003971.

Luytjes, W., Krystal, M., Enami, M., Parvin, J. D. & Palese, P. (1989). Amplification, expression, and packaging of foreign gene by influenza virus. *Cell* **59**, 1107-1113.

Marsh, G. A., Hatami, R. & Palese, P. (2007). Specific residues of the influenza A virus hemagglutinin viral RNA are important for efficient packaging into budding virions. *J Virol* **81**, 9727-9736.

Nayak, D. P., Sivasubramanian, N., Davis, A. R., Cortini, R. & Sung, J. (1982). Complete sequence analyses show that two defective interfering influenza viral RNAs contain a single internal deletion of a polymerase gene. *Proc Natl Acad Sci U S A* **79**, 2216-2220.

Noda, T., Sagara, H., Yen, A., Takada, A., Kida, H., Cheng, R. H. & Kawaoka, Y. (2006). Architecture of ribonucleoprotein complexes in influenza A virus particles. *Nature* **439**, 490-492.

Palese, P. & Shaw, M. L. (2007). Orthomyxoviridae: The viruses and their replication. In *Fields Virology*, 5th edn, pp. 1647-1689. Edited by D. M. Knipe & P. M. Howley. Philadelphia: Lippincott Williams and Wilkins.

Sherry, L., Smith, M., Davidson, S. & Jackson, D. (2014). The N terminus of the influenza B virus nucleoprotein is essential for virus viability, nuclear localization, and optimal transcription and replication of the viral genome. *J Virol* **88**, 12326-12338.

Watanabe, T., Watanabe, S., Noda, T., Fujii, Y. & Kawaoka, Y. (2003). Exploitation of nucleic acid packaging signals to generate a novel influenza virus-based vector stably expressing two foreign genes. *J Virol* **77**, 10575-10583.

Figure legends

Figure 1

The first and last 150 nucleotides of the BHA coding region of the influenza B virus vRNA segment four are required for efficient genome packaging. (a) Schematic diagram of the HA-GFP mini-genome constructs in a vRNA-like 3'-5' negative sense orientation. HA(150)GFP(150) contains the first and last 150 nucleotides of the BHA coding regions of B/Yamanashi/98 virus vRNA segment four (white) flanked by the 3' and 5' non-coding regions (grey) with the GFP open reading frame (green) inserted in frame between the BHA coding regions. A translation termination codon was inserted after the GFP ORF. The mutant constructs contained truncations in the BHA

coding regions at either the 3' or 5' end of the vRNA segment. All constructs were inserted into the pHH-21 vector under the control of the human RNA polymerase I promoter and terminator resulting in the expression of negative sense mini-genome viral-like RNAs. (b) GFP expression from each HA-GFP construct in a mini-genome replication assay. 293T cells were transfected with 1 µg of each mini-genome construct along with plasmids expressing the B/Yamanashi/98 virus polymerase complex and BNP proteins. At 24 hours post-transfection (h.p.t.) GFP expression was observed at 20x magnification. Cell nuclei were stained with DAPI and the total number of cells in each panel was determined by manual counting using ImageJ. GFP-positive cells were manually counted and the percentage of the total cell number positive for GFP was calculated. The percentages were calculated from three independent experiments and are shown in the white boxes in each panel ± S.D. (c and d) The packaging efficiency of each HA-GFP segment into virus-like particles (VLPs). 293T cells were transfected with 1 µg of each mini-genome construct and at 24 h.p.t. infected with B/Yamanashi/98 virus at an MOI of 10. Cells were monitored for GFP expression and at 16 hours post-infection (h.p.i.) all samples demonstrated GFP expression in 70-80% of cells. Supernatants were then harvested and N-acetyl trypsin-treated. Samples were serially 2-fold diluted and placed on MDCK cells for 10 hours. At 10 h.p.i. cells were fixed, permeabilised and stained with an anti-BNP antibody followed by a Texas-red conjugated secondary antibody. Representative images of GFP and viral antigen staining by immunofluorescence at 20x magnification are shown in (c) for four of the VLP-containing supernatants. Nuclei were stained with DAPI. (d) FACS analysis was performed to quantify the number of antigen positive cells that were also expressing GFP. 5000 cells were analysed per sample. The percentage of antigen-positive HA(150)GFP(150) VLP-infected cells that also expressed GFP was set at 100% packaging efficiency and the efficiency of all other constructs was determined as a percentage of this. Error bars represent the average of two independent experiments ± S.D.

Figure 2

Synonymous mutations in both the 3' and 5' coding regions of segment four vRNA lead to virus attenuation. (a) Schematic diagram of mutations inserted into B/Yamanashi/98 virus vRNA

segment four by reverse genetics. The RNA sequence is shown in positive sense mRNA orientation to highlight the lack of protein coding changes. The first and last 60 nucleotides of the BHA coding region are shown (nucleotides 1-60 and 1706-1766 respectively) in the mRNA-sense orientation with the corresponding amino acid sequence indicated. Synonymous mutations were inserted into the 3' end of the vRNA (ie the 5' end of the mRNA) between nucleotides 1-30 (3m30 mutations) or 31-60 (3m60 mutations) of the BHA coding region, and into the 5' end of the vRNA between nucleotides 1706-1736 (5m60 mutations) or 1737-1766 (5m30 mutations). Mutations are capitalised, highlighted and underlined. Wt = wild-type segment four sequence. (b) Multistep growth curve analysis of single site mutant viruses. MDCK cells were infected with wt and the four mutant viruses containing single site mutations (eg 3m30 mutations) at an MOI of 0.001, samples harvested every 12 hours until 72 h.p.i. and infectious titres determined by plaque assay. Results show the average of three independent experiments ± S.D. (c) Multistep growth curve analysis of double site mutant viruses. MDCK cells were infected with wt and mutant viruses containing double site mutations and replication kinetics determined as in (b). (d) Particle to infectivity analysis of the mutant virus panel. Stocks of wt and mutant viruses were titrated by plaque assay and haemagglutination assay. Particle number was calculated based on 1 HA unit requiring approximately 10^6 virus particles (Donald & Isaacs, 1954). The particle to infectivity ratio (particles/ml : plaque forming units/ml) was determined for each virus. (e) Particle to infectivity analysis of 3m30 mutation-containing viruses from growth curve samples in (b) and (c). Particle to infectivity ratios were determined for the 36 h and 48 h growth curve samples for wt and mutant viruses containing the 3m30 mutations. Results show the average of three independent experiments ± S.D.

Figure 3

The 3m30/5m30 mutations do not affect mRNA synthesis/protein production but reduce the packaging efficiency of vRNA segment four. (a) Sequence alignment of influenza B virus vRNA segment four sequences. The 3m30, 3m60, 5m30 and 5m60 regions of the influenza B virus vRNA segment four indicated in Fig. 2(a) (shown in positive sense orientation) were analysed in the published sequences of multiple viruses isolated between 1940 and 2014. Sequences were

554 extracted from the NCBI Influenza Virus Sequence Database and aligned using Geneious
555 software. Nucleotide differences from the consensus sequence are indicated with identical
556 nucleotides shown as dots. (b) BHA protein expression is unaffected by the 3m30/5m30 mutations.
557 Wt and mutant vRNA segment four were cloned into the pHH-21 plasmid under the control of the
558 human RNA polymerase I promoter, and transfected into 293T cells alongside plasmids
559 expressing the viral polymerase complex and BNP proteins. 24 h.p.t. cells were lysed and proteins
560 detected by immunoblotting using an anti-B/Hong Kong/73 polyclonal antibody (to detect BHA and
561 BNP proteins) and an anti-actin antibody (as a loading control). Images were obtained using an
562 Odyssey CLx NIR scanner. Results show three independent transfections for each construct. (c)
563 Quantification of protein levels from (b). Image Studio software was used to quantify the protein
564 bands in (b) in arbitrary units. The BHA and BNP values were normalised against the actin value
565 for each sample. Results show the average of three independent transfections ± S.D. * = $p > 0.05$.
566 (d) Analysis of viral proteins within purified virus particles. 500 000 HA units of rBHA wt and rBHA-
567 3m30/5m30 viruses were purified by sucrose gradient ultracentrifugation. Proteins within purified
568 viruses were separated by SDS-PAGE and analysed by coomassie staining. Three individual
569 preparations are shown for each virus. (e) Quantification of viral proteins within purified virus
570 particles. Viral proteins in the three preparations of rBHA wt and rBHA-3m30/5m30 viruses from (d)
571 were analysed by immunoblotting as in (b). Protein bands were detected and quantified for BHA,
572 BNP and BM1. As BM1 is the major structural component of the virus all viral proteins were
573 normalised against BM1 levels (set at 100% for each virus). Results show the average of the three
574 virus preparations ± S.D. * = $p > 0.05$, ** = $p < 0.0005$. (f) Quantification of vRNA within virus
575 particles. vRNA was extracted from 5.12 $x10^9$ rBHA wt and rBHA-3m30/5m30 virus particles
576 (determined by haemagglutination assay) and segments 4, 5 and 8 amplified by quantitative
577 reverse transcriptase PCR. Levels of each rBHA-3m30/5m30 vRNA segment are expressed as a
578 percentage of rBHA wt levels and represent three independent experiments ± S.D. * = $p > 0.05$, **
579 = $p < 0.0005$, *** = $p = 0.05$.
580
581

Fig. 1

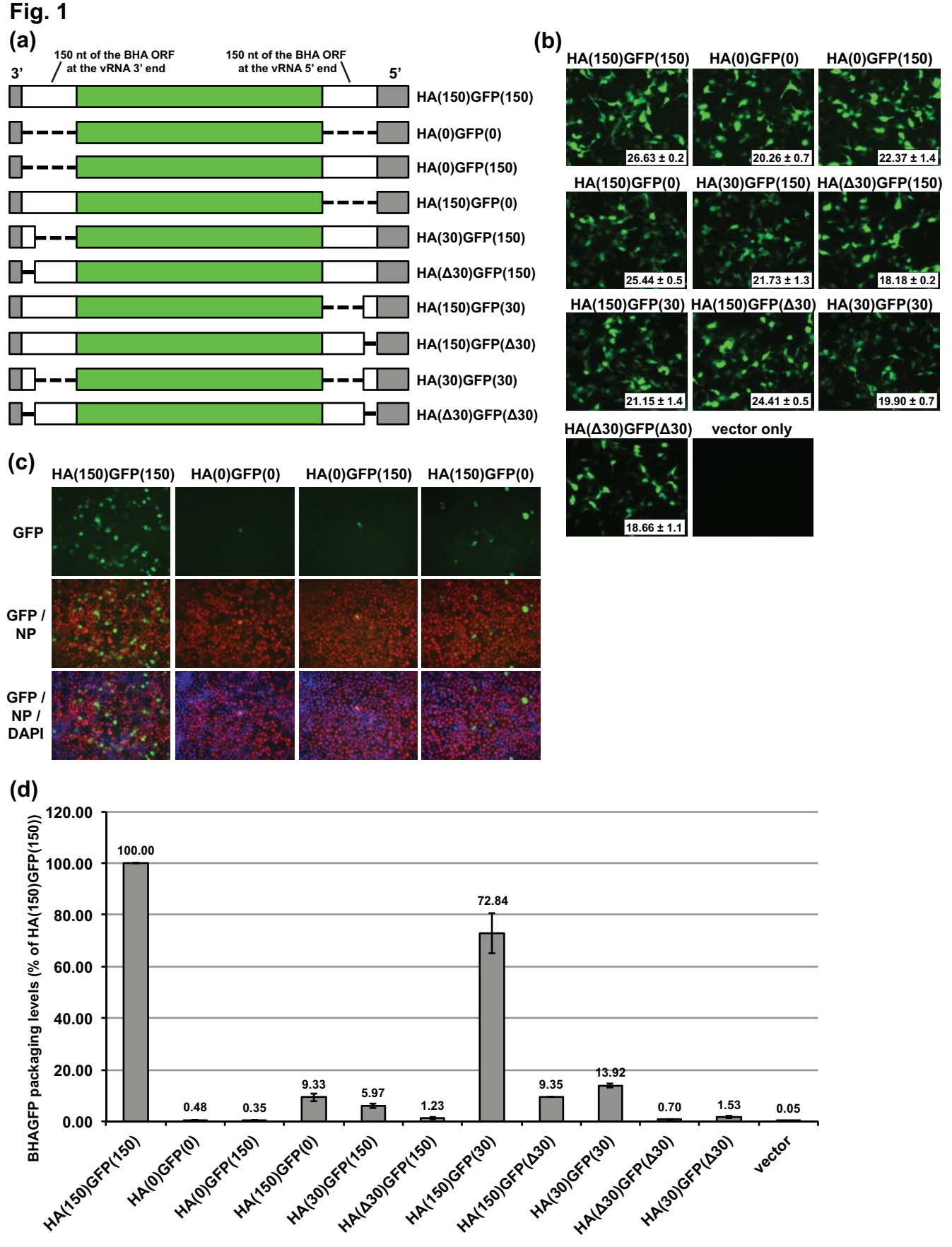
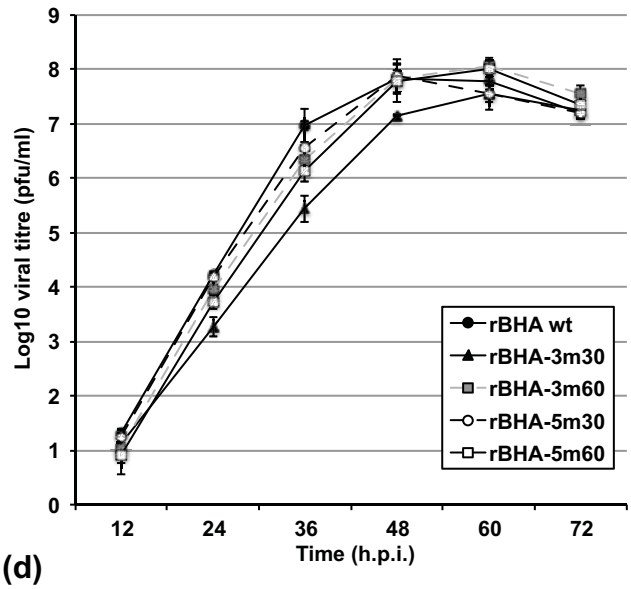

(a)
150 nt of the BHA ORF at the vRNA 3' end
150 nt of the BHA ORF at the vRNA 5' end
3'
5'
HA(150)GFP(150)
HA(0)GFP(0)
HA(0)GFP(150)
HA(150)GFP(0)
HA(30)GFP(150)
HA(Δ30)GFP(150)
HA(150)GFP(30)
HA(150)GFP(Δ30)
HA(30)GFP(30)
HA(Δ30)GFP(Δ30)
(b)
HA(150)GFP(150)
26.63 ± 0.2
HA(0)GFP(0)
20.26 ± 0.7
HA(0)GFP(150)
22.37 ± 1.4
HA(150)GFP(0)
25.44 ± 0.5
HA(30)GFP(150)
21.73 ± 1.3
HA(Δ30)GFP(150)
18.18 ± 0.2
HA(150)GFP(30)
21.15 ± 1.4
HA(150)GFP(Δ30)
24.41 ± 0.5
HA(30)GFP(30)
19.90 ± 0.7
HA(Δ30)GFP(Δ30)
18.66 ± 1.1
vector only
(c)
HA(150)GFP(150)
HA(0)GFP(0)
HA(0)GFP(150)
HA(150)GFP(0)
GFP
GFP / NP
GFP / NP / DAPI
(d)
BHAGFP packaging levels (% of HA(150)GFP(150))
120.00
100.00
80.00
60.00
40.00
20.00
0.00
100.00
0.48
0.35
9.33
5.97
1.23
72.84
9.35
13.92
0.70
1.53
0.05
HA(150)GFP(150)
HA(0)GFP(0)
HA(0)GFP(150)
HA(150)GFP(0)
HA(30)GFP(150)
HA(Δ30)GFP(150)
HA(150)GFP(30)
HA(150)GFP(Δ30)
HA(30)GFP(30)
HA(Δ30)GFP(Δ30)
HA(30)GFP(Δ30)
vector

Fig. 2

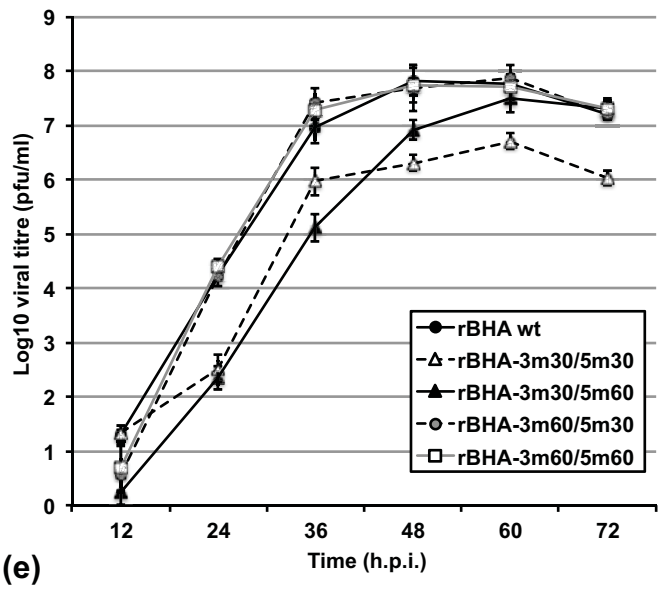
(a)

```
1                                   30                                  60
atg aag gca ata att gta cta ctc atg gta gta aca tcc aat gca gat cga atc tgc act  = wt
 M   K   A   I   I   V   L   L   M   V   V   T   S   N   A   D   R   I   C   T

atg aaA gcT atT atA gtT ctT ctG atg gtT gta aca tcc aat gca gat cga atc tgc act  = 3m30
 M   K   A   I   I   V   L   L   M   V   V   T   S   N   A   D   R   I   C   T

atg aag gca ata att gta cta ctc atg gta gtT acT tcG aaC gcT gaC cgT atT tgT acA  = 3m60
 M   K   A   I   I   V   L   L   M   V   V   T   S   N   A   D   R   I   C   T

1706                                1736                                1766
gct att ttt att gtt tat atg atc tcc aga gac aat gtt tct tgc tcc atc tgt cta tag  = wt
 A   I   F   I   V   Y   M   I   S   R   D   N   V   S   C   S   I   C   L   -

gcA atA ttC atA gtA taC atg atT tcG agG gac aat gtt tct tgc tcc atc tgt cta tag  = 5m60
 A   I   F   I   V   Y   M   I   S   R   D   N   V   S   C   S   I   C   L   -

gct att ttt att gtt tat atg atc tcc aga gaT aaC gtA tcA tgT tcT atT tgC ctT tag  = 5m30
 A   I   F   I   V   Y   M   I   S   R   D   N   V   S   C   S   I   C   L   -
```

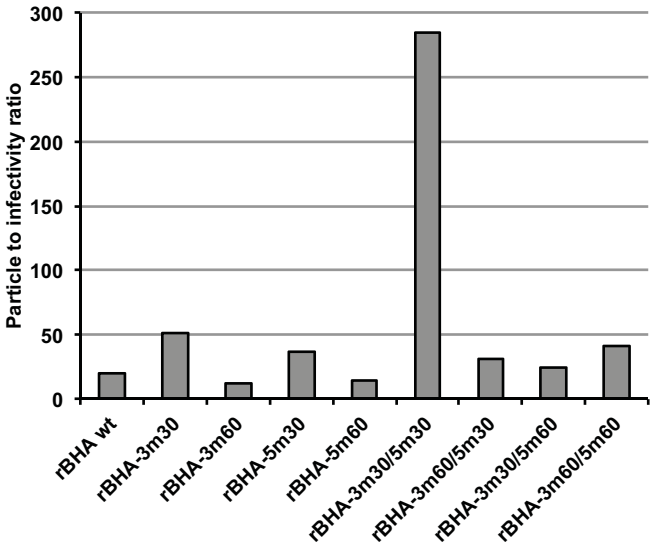
(b)



(c)



(d)



(e)

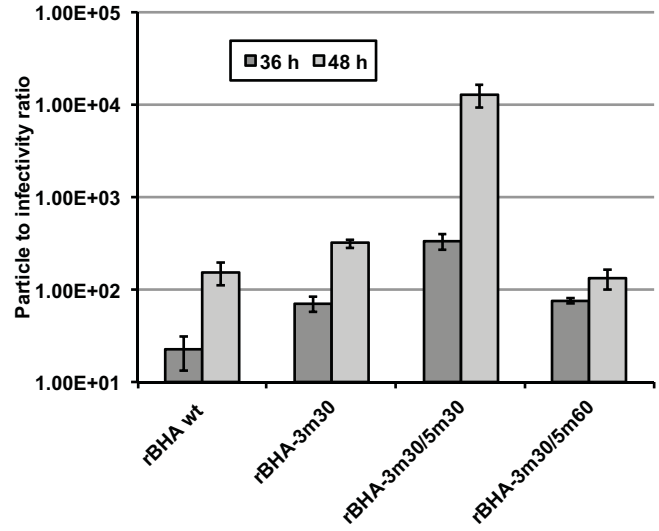
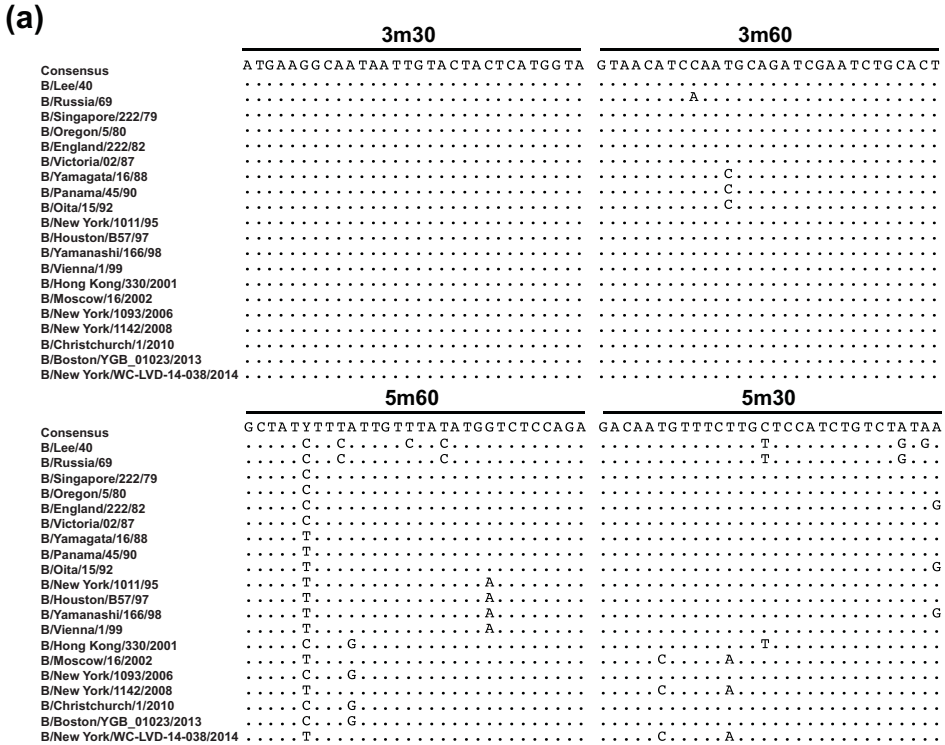
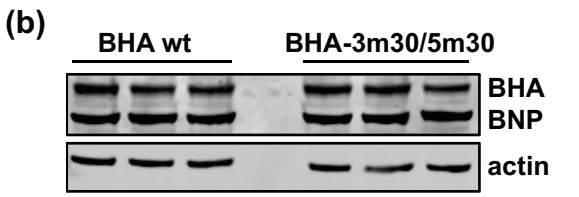
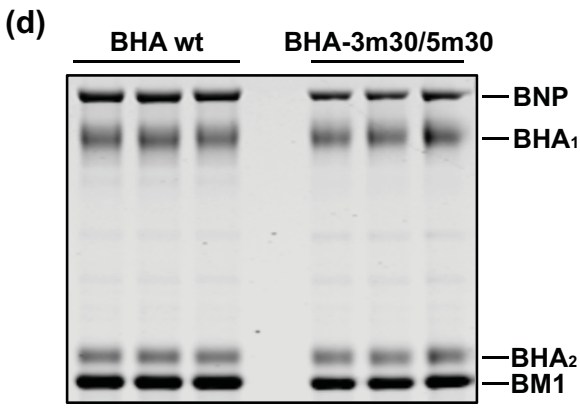
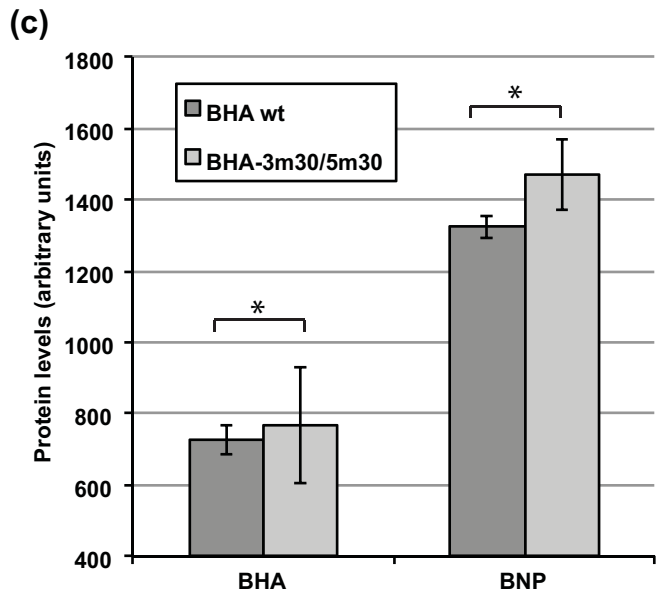


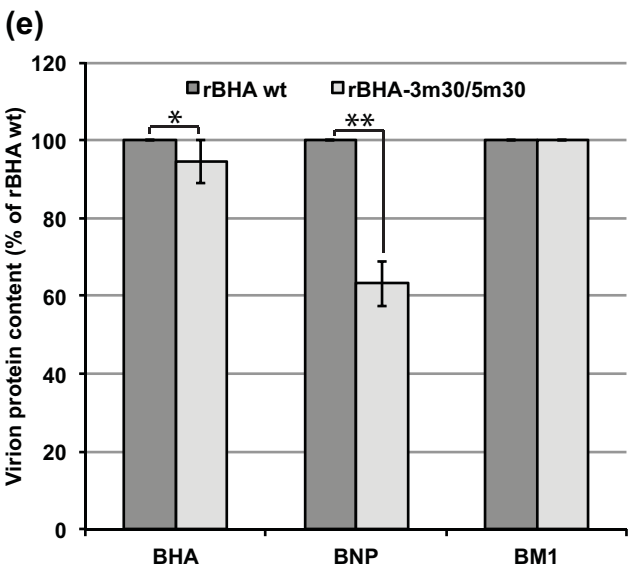
Fig. 3

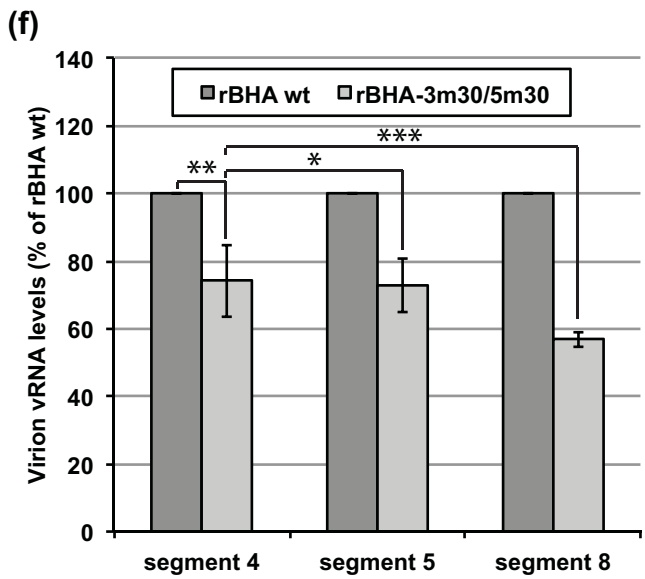

(a)
3m30
3m60
Consensus
ATGAAGGCAATAATTGTACTACTCATGGTA
GTAACATCCAATGCAGATCGAATCTGCACT
B/Lee/40
B/Russia/69
B/Singapore/222/79
B/Oregon/5/80
B/England/222/82
B/Victoria/02/87
B/Yamagata/16/88
B/Panama/45/90
B/Oita/15/92
B/New York/1011/95
B/Houston/B57/97
B/Yamanashi/166/98
B/Vienna/1/99
B/Hong Kong/330/2001
B/Moscow/16/2002
B/New York/1093/2006
B/New York/1142/2008
B/Christchurch/1/2010
B/Boston/YGB_01023/2013
B/New York/WC-LVD-14-038/2014
5m60
5m30
Consensus
GCTATYTTTATTGTTTATATGGTCTCCAGA
GACAATGTTTCTTGCTCCATCTGTCTATAA
B/Lee/40
B/Russia/69
B/Singapore/222/79
B/Oregon/5/80
B/England/222/82
B/Victoria/02/87
B/Yamagata/16/88
B/Panama/45/90
B/Oita/15/92
B/New York/1011/95
B/Houston/B57/97
B/Yamanashi/166/98
B/Vienna/1/99
B/Hong Kong/330/2001
B/Moscow/16/2002
B/New York/1093/2006
B/New York/1142/2008
B/Christchurch/1/2010
B/Boston/YGB_01023/2013
B/New York/WC-LVD-14-038/2014


(b)
BHA wt
BHA-3m30/5m30
BHA
BNP
actin


(d)
BHA wt
BHA-3m30/5m30
BNP
BHA1
BHA2
BM1


(c)
BHA wt
BHA-3m30/5m30
Protein levels (arbitrary units)
1800
1600
1400
1200
1000
800
600
400
*
*
BHA
BNP


(e)
rBHA wt
rBHA-3m30/5m30
Virion protein content (% of rBHA wt)
120
100
80
60
40
20
0
*
**
BHA
BNP
BM1


(f)
rBHA wt
rBHA-3m30/5m30
Virion vRNA levels (% of rBHA wt)
140
120
100
80
60
40
20
0
**
*
***
segment 4
segment 5
segment 8