The pseudoknot region and poly-(C) tract comprise an essential RNA packaging signal for assembly of foot-and-mouth disease virus

Authors

Chris Neil[1], Joseph Newman[1], Nicola J. Stonehouse[2], David J. Rowlands[2], Graham J. Belsham[4,5], Tobias J. Tuthill[1*]

[1] The Pirbright Institute, Ash Road, Pirbright, Surrey, United Kingdom

[2] The Faculty of Biological Sciences, University of Leeds, Leeds, United Kingdom

[4] DTU National Veterinary Institute, Technical University of Denmark, Lindholm, Kalvehave, Denmark (former affiliation)

[5] Department of Veterinary and Animal Sciences, University of Copenhagen, Copenhagen, Denmark (current affiliation)

Abstract

Virus assembly is a crucial step for the completion of the viral replication cycle. In addition to ensuring efficient incorporation of viral genomes into nascent virions, high specificity is required to prevent incorporation of host nucleic acids. For picornaviruses, including FMDV, the mechanisms required to fulfil these requirements are not well understood. However, recent evidence has suggested that specific RNA sequences dispersed throughout picornavirus genomes are involved in packaging. Here, we have shown that such sequences are essential for FMDV RNA packaging and have demonstrated roles for both the pseudoknot (PK) region and the poly-(C) tract in this process, where the length of the poly-(C) tract was found to influence the efficiency of RNA encapsidation. Sub-genomic replicons containing

longer poly-(C) tracts were packaged with greater efficiency *in trans*, and viruses recovered from transcripts containing short poly-(C) tracts were found to have greatly extended poly-(C) tracts after only a single passage in cells, suggesting that maintaining a long poly-(C) tract provides a selective advantage. We also characterised a critical packaging signal (PS) located in the pseudoknot (PK) region, adjacent to the poly-(C) tract, as well as several other non-essential but beneficial PSs elsewhere in the genome. Collectively, these PSs greatly enhanced encapsidation efficiency, with the poly-(C) tract possibly facilitating nearby PSs to adopt the correct conformation.

Using these data, we have proposed a model where interactions with capsid precursors control a transition between two RNA conformations, directing the fate of nascent genomes to either be packaged or alternatively to act as templates for replication and/or for protein translation.

Author summary

Genome packaging, whereby viral RNA is incorporated into protective protein capsids to produce more virus particles, is a crucial step in RNA virus life cycles. It is a stringent process as only viral RNA is encapsidated, while cellular RNA is excluded.

This study reveals the essential role of packaging signals in FMDV RNA packaging, specifically those in the pseudoknot region and in a region that can contain >100 cytosines, termed the poly-(C) tract. We demonstrate that the length of the poly-(C) tract significantly affects packaging efficiency; genomes containing longer poly-(C) tracts are favoured. This is the first role that has been identified for the poly-(C) tract in FMDV. We have also found an essential packaging signal in the pseudoknot region, which is assisted by other packaging signals located throughout the genome, that together facilitate encapsidation of FMDV RNA. These

results provide compelling evidence for the involvement of RNA packaging signals in FMDV assembly. Based on this, we propose a simple model for FMDV RNA packaging, which involves a transition from genome replication to genome packaging and is controlled by packaging signals. This knowledge could pave the way for future research and development of novel antiviral strategies targeting FMDV and other picornaviruses.

Introduction

Foot-and-mouth disease virus (FMDV), a species within the *Aphthovirus* genus of the family *Picornaviridae*, is the aetiological agent of foot-and-mouth disease (FMD) in cloven-hooved animals (1). FMD is extremely contagious and is endemic in parts of Africa and Asia. It is controlled through vaccination in regions at risk of infection, while countries including the USA, UK, Australia and much of mainland Europe maintain FMD-free status without vaccination. However, outbreaks of the disease in countries that are normally FMD-free can cause severe economic losses and the ongoing cost to combat the virus in endemic countries is significant (2).

FMDV is a small, non-enveloped, single-stranded, positive-sense RNA virus with a genome approximately 8.4 kb in length that is packaged into a pseudo *T* = 3 icosahedral capsid (3). The genome contains a single large open reading frame, encoding both the structural and non-structural proteins. The first protein in this polyprotein is the Leader protease, followed by the capsid precursor P1-2A, and the non-structural proteins 2B, 2C, 3A, $3B_{(1-3)}$, $3C^{pro}$ and 3D (4). The capsid precursor P1-2A is cleaved by the viral protease $3C^{pro}$ into VP0, VP1, VP3 and 2A, of which VP0, VP1 and VP3 remain associated in the protomer (5). Five protomers then assemble into a pentamer and 12 pentamers associate with a molecule of genomic RNA to form the complete infectious virus particle. Finally, there is an assembly dependent

68 maturation cleavage of VP0 into VP2 and VP4 which is more efficient in the presence of
69 packaged RNA, but can still occur independently (6). Although this maturation cleavage is
70 common to most picornaviruses, it is not ubiquitous. For example, viruses in the *Parechovirus*
71 genus do not undergo this process and VP0 remains intact in the mature virus (7). In addition
72 to the formation of mature virions, assembly of capsids lacking RNA can also occur to
73 produce non-infectious empty particles (8).

74 The FMDV genome is organised into a highly structured 5'-UTR, with a virus-encoded peptide
75 (VPg) covalently bound to the 5' terminus of the RNA (9, 10), the open reading frame; and a
76 3'-UTR terminating in a poly-A tail. The key elements of the 5'-UTR are the: S-fragment, poly-
77 (C) tract, pseudoknot (PK) region, the *cis*-acting replication element (*cre)* and the internal
78 ribosome entry site (IRES). Of particular interest for this study are the poly-(C) tract and the
79 adjacent PK region, both of which are not present in some better studied picornaviruses, e.g.
80 enteroviruses.

81 The poly-(C) tract typically consists of a stretch of about 50-200 cytosines (11). The
82 requirement for the poly-(C) tract to be over a certain length has been linked to virulence in
83 both cardioviruses and aphthoviruses (12, 13). However, conflicting results have also been
84 reported (14). This is, in part, due to the difficulty of cloning such a long homopolymer and a
85 mechanism allowing truncated poly-(C) tracts to revert to a *wt* length during passage, thus
86 obscuring potential differences (15, 16). However, it has been shown that this feature is not
87 essential for replication.

88 The PK region is made up of 2-4 PKs with high sequence similarity, which directly follow the
89 poly-(C) tract. While at least one PK must be present for *wt* levels of replication, RNA
90 transcripts lacking all four PKs are still able to replicate. Despite this, we have shown that RNA

transcripts without the PKs cannot be recovered as infectious virus, which we hypothesised may be due to defects in packaging (17).

Picornavirus RNA is thought to be encapsidated during capsid assembly and, although no clear mechanism has been established for how this is achieved, recent work has highlighted the role of RNA packaging signals (PSs) in this process. PSs are used by other RNA viruses to incorporate viral RNA into capsids, with the bacteriophage MS2 being seen as a model system for PS-mediated packaging (18). Additionally, PSs have been identified in other picornaviruses – most notably aichi virus, human parechovirus 1 and enterovirus-E (19-25). In these systems, multiple small, low affinity PSs are dispersed across the viral genome, which function in concert to package the genome more efficiently into the virus capsid (26). These PSs may each recruit a capsid subunit, for example a picornavirus pentamer, sequentially in Hamiltonian pathways (27). The arrangement of PSs throughout the genome, each with different affinities for the capsid subunits, may coordinate the assembly process by initially binding to PSs with the strongest affinities, followed by recruitment of the lower-affinity PSs to complete the capsid structure. For some viruses, the presence of PSs is essential for capsid formation, for example MS2 PS-coat protein interactions are required for a conformational change in the coat protein which enables efficient assembly (28). Although FMDV pentamers can assemble into empty capsids in the absence of RNA (and PSs are therefore not essential for capsid formation) these PSs may facilitate and stabilise the process, making assembly more efficient and preventing the capsid from dissociating prematurely. Additionally, other studies on picornaviruses have described alternative mechanisms for encapsidation. For the enterovirus poliovirus (PV), for example, all non-essential sections of the genome have been either deleted or extensively mutated while maintaining the encoded protein sequence, or swapped with those of other picornaviruses without affecting virus viability, suggesting that

PV RNA encapsidation does not rely solely on PSs (29-32). As an alternative to RNA PSs, an interaction between 2C, a viral protein involved in RNA replication, and the capsid protein VP3 was proposed as the link between replication and encapsidation (33). Recent work has identified PS motifs across the PV genome (25), however, which suggests that picornaviruses may employ both protein-protein interactions and RNA-protein interactions during encapsidation.

FMDV packaging appears to be extremely stringent since only genomic RNA (or FMDV replicon RNA) is packaged into virus capsids (34). Additionally, we have previously identified functional PSs dispersed across the FMDV genome (35). The 5'-most PS identified is located in the PK region, and deletion of sequences within this region prevented virus recovery (17).

A number of factors influence the assembly of infectious virus, including the relative rates of protein and RNA synthesis, and these complicate the study of genome encapsidation in picornaviruses generally and in FMDV specifically. This is particularly relevant when the final assay of virus assembly is the measurement of virus viability. Approaches to identify picornavirus packaging signals have helped to establish the role of RNA-capsid interactions during virus assembly (24, 25, 35), but even these are limited by the reliance on virus viability for validation and more direct methods to assess assembly and encapsidation are needed. To this end, an assay designed to look at the relative encapsidation efficiency of GFP-expressing replicons *in trans* was developed based on the assays carried out by Barclay et al. and McInerney et al. (30, 34). This assay demonstrated that poliovirus replicons could be packaged into poliovirus capsids *in trans*, but the assay was significantly less efficient when FMDV was used.

In the work presented here, we have further optimised the *trans*-encapsidation assay to enable the study of FMDV RNA encapsidation, solving the problem of inefficient replicon *trans*-encapsidation. We then used this assay to characterise key features of the genome required for packaging, including the 5'-most PS, the PK region and the poly-(C) tract.

Methods

Cell lines

The baby hamster kidney-21 (BHK-21) cell line was obtained from the central services unit (CSU) at The Pirbright Institute. This cell line was used for all cell culture experiments. The cells were grown in Glasgow's minimum essential media (GMEM, Merck) with 10% foetal bovine serum (FBS; Gibco), 2 mM L-glutamine (Merck), 100 units/L penicillin, 100 µg/mL streptomycin (Merck) and 5% tryptose phosphate broth (Gibco). For virus infections, virus growth media (VGM) was used, which contained the same components but with either 1% or 5% FBS.

Cells used for transfection and/or infection were seeded the previous day using a suitable dilution to reach 80% confluence at the start of the experiment.

Plasmids

Three plasmids were used as the basis for the work described here: an O1 Kaufbeuren (O1K) FMDV ΔP1 ptGFP replicon, containing the Leader protease coding region but with a ptGFP coding region replacing the capsid coding region (36); an O1K aqGFP Leaderless (ΔLb) infectious copy plasmid (ΔLb.GFP.FMDV), which does not contain the Leader protein coding region but does contain the aqGFP coding region (37); and the full-length infectious copy of O1K FMDV, pT7S3 (38). These are shown in Fig 1.

Four variants of the ΔP1 GFP replicon were used here: C11 ΔPK1234, ΔPK234, ΔPK34 and C11, which we have previously described (17). An additional variant, C29, was made for this study by ligating synthetic DNA from GeneArt containing the alternative-length poly-(C) tract.

The ΔLb.GFP.FMDV infectious copy plasmid was modified to make a replicon by changing the encoded 3Cpro cleavage sites to prevent P1-2A processing. This replicon was called the Leaderless defective-capsid GFP replicon (ΔLbdcap GFP replicon). The sequence encoding the VP2/VP3 junction was changed from GAGGGA to GCTGCT (PSKE/GIFP to PSKA/AIFP), the sequence encoding the VP3/VP1 junction was changed from GAAACC to GCCGCA (ARAE/TTSA to ARAA/ATSA) and a codon in the VP0 myristoylation site sequence was changed from GGG to GCC (VP0 G1A) by the introduction of a synthetic fragment of DNA from GeneArt. Variations were made to the ΔLbdcap GFP replicon, again by incorporating synthetic DNA from GeneArt, to make the: C11, C29, C40 ΔPK1 and C39 ΔPK2 ΔLbdcap GFP replicons.

The pT7S3 plasmids containing truncated poly-(C) tracts were made by digesting the corresponding ΔLbdcap GFP replicons with suitable restriction enzymes and ligating in the fragment containing the poly-(C) tract to replace the equivalent region of pT7S3.

All restriction enzyme digests were performed using enzymes from New England Biolabs (NEB) based on the product protocols. The vector portions were dephosphorylated using rAPid alkaline phosphatase (Roche) following digestion with the appropriate restriction enzymes. Fragments were then separated using DNA electrophoresis, purified using the Monarch DNA Gel Extraction kit (NEB) and ligated using T4 DNA ligase (Invitrogen).

In vitro transcription

Plasmids were linearised using appropriate restriction enzymes (*AscI* for the ΔP1 ptGFP based replicons, *HpaI* for all others). The DNA was then purified using the Monarch PCR & DNA

Cleanup Kit (NEB), and RNA was synthesised from the purified DNA using the MEGAscript T7 Transcription Kit (Invitrogen) according to the manufacturer guidelines. Input DNA was digested using the Turbo included in the kit. The RNA was purified using a MEGAclear kit (Invitrogen) and quantified using the Qubit RNA BR Assay Kit (Invitrogen) with a Qubit Fluorometer 4 according to the manufacturer's protocol.

BHK cell transfections

BHK cells were transfected with RNA using Lipofectamine 2000, based on the manufacturer's protocol (Invitrogen). The RNA was mixed with Opti-Mem (Gibco), with 1 µg (suitable for a 24-well transfection) made up to 50 µL. The Lipofectamine 2000 was diluted 1 in 10 in Opti-Mem to make an equal volume of Lipofectamine 2000-Opti-Mem mix (per RNA sample). Each was incubated at room temperature for 5 minutes, before the RNA/Opti-Mem solution was mixed with an equal volume of the Lipofectamine 2000/Opti-Mem solution and incubated for a further 15 minutes at room temperature. The medium was aspirated from cell monolayers and replaced with the RNA transfection mix, along with sufficient Opti-Mem to cover the cells and prevent them from drying out. The medium was replaced after two hours with an appropriate volume of 1% VGM.

Trans-encapsidation assay

In the first round of the assay, BHK-21 cells were seeded in multi-well plates (6-, 12- or 24-well, Scientific Laboratory Supplies) and transfected with an appropriate amount of the replicon RNA transcript (4, 2 and 1 µg respectively). If an infectious clone transcript such as the C11 transcript was used as the helper genome, it was co-transfected with the replicon RNA by including it with the RNA-Opti-Mem mixture (no change to final volume or amount of Lipofectamine 2000 used). If a helper virus was used, it was added to the cells at 1-hour post-

transfection. At 2-hours post-transfection the medium was replaced with an appropriate volume of VGM containing 1% serum to cover the cells. Cells transfected with the ΔP1 GFP replicon were incubated until 5 hours post-transfection, while cells transfected with the ΔLbdcap GFP replicon were incubated until 7 hours post-transfection, both at 37°C, before they were frozen at -20°C overnight. Duplicate plates were prepared in parallel and were analysed using the Incucyte S3 Live-Cell Analysis System, which imaged the cells hourly using phase light and detected fluorescence (emission wavelength 524 nm, excitation wavelength 460 nm). Images were analysed to calculate the number of GFP expressing cells present and the mean fluorescence intensity (MFI) was determined as a measure of replicon replication.

For the second round of the assay, the cells were thawed and pelleted, and the supernatant, containing encapsidated RNA, was treated with 1 µL RNase A per 100 µL of supernatant for 10 minutes at room temperature to remove input transcripts. Fresh cells were then inoculated with 50, 100 or 200 µL of supernatant (24-, 12- and 6-well plates respectively) and loaded into the Incucyte S3 for hourly imaging using phase light and fluorescence detection (as above). Several images were taken per well which were analysed in real time for the proportion of the well covered with cells (% confluence), the number of fluorescent objects per well (GFP objects per well) and/or the MFI of the well where relevant. The parameters of the analysis were adjusted based on the characteristics of genuine GFP-expressing cells to optimise detection of GFP-expressing cells and to minimise background auto-fluorescence, which included setting minimum thresholds for both the intensity of the GFP and the area of the foci. Typical thresholds included: a threshold of 1.0 GCU (green calibrated unit); a maximum area of 1000 μm^2; a maximum eccentricity of 0.7; and a minimum integrated intensity (GCU x μm^2) of 2000. All other parameters were typically left at the default setting. Plates were typically read for 24 hours or until the GFP object count started decreasing.

CPE development assay

The CPE (cytopathic effect) development assay was performed by transfecting a 96-well plate of BHK-21 cells with RNA transcripts made from the virus infectious clone plasmids. Plates were loaded into the Incucyte S3 Live-Cell Analysis System and imaged using phase light every four hours until either complete CPE was reached, or the cell only negative control became unhealthy. Multiple images were taken per well, which were analysed in real time to determine the proportion of the well covered in cells to calculate the confluence of the cell monolayers. The parameters of the analysis were adjusted to optimise detection of healthy cells compared to unhealthy cells to better observe the spread of the rescued virus through the culture.

RNA extraction and digestion

RNA transcripts were transfected into BHK21 cells, and cells were left to develop CPE. At the point of complete CPE, the cells were lysed by freeze/thawing and the lysate was clarified by centrifugation. A portion of the clarified lysate was then added to fresh BHK21 cells with VGM, and the cells were again left until complete CPE was reached. At this point, cells were processed using a Qiagen RNeasy Kit (74004) to extract the RNA. The purified RNA samples were then digested with RNase T1 (ThermoFisher Scientific) to release the poly(C) tract based on the manufacturer guidelines.

Automated electrophoresis

DNA and RNA samples were analysed using either the Tapestation (Tapestation D1000 High Sensitivity DNA kit and Tapestation RNA ScreenTape kit, Agilent) or the Bioanalyzer (DNA 12000 kit, Agilent) according to the manufacturer's guidelines.

Results

A novel *trans*-encapsidation assay to investigate FMDV RNA packaging

The *trans*-encapsidation assay involves co-transfecting two RNA transcripts into cells: a GFP-expressing replicon which cannot produce a functional capsid and acts as the genome-donor; and an infectious copy transcript which acts as the capsid-donor. When these RNAs replicate in the same cell, the infectious copy produces functional capsid subunits, which can assemble together with either viral RNA derived from the infectious copy *in cis*, or with replicon RNA *in trans*. Progeny particles incorporating infectious genomes or replicon RNAs can then infect new cells, with cells infected by the *trans*-encapsidated replicon RNA being quantifiable by GFP expression. The relative t*rans*-encapsidation efficiencies of replicons containing mutations, which may affect encapsidation, can therefore be measured by comparing the GFP expression in monolayers of cells infected with lysates from the transfected cells of the first-round with that from a *trans*-encapsidated control replicon (Fig 2). The number of cells expressing GFP in the second round reflects the packaging efficiency of the *trans*-encapsidated replicon. The transfection efficiencies and replication kinetics of the replicons used can be assessed from the GFP counts and mean fluorescent intensities (MFI) produced in the first round of the assay to ensure they are comparable. Where there were significant differences between the replicons, these were factored into the conclusions drawn from the experiments and their potential to influence *trans*-encapsidation efficiency assessed.

We hypothesised that the efficiency of viral RNA packaging would correlate with the number of PSs. To test this, two types of replicon RNA, with either normal or reduced numbers of PSs, were used in the *trans*-encapsidation assay: a GFP replicon with fewer PSs (due to replacement of the majority of the P1 capsid-coding region with the GFP reporter sequence,

termed ΔP1 GFP replicon); and a Leaderless (ΔLb) GFP replicon including the capsid coding region and the associated PSs, albeit with substitutions at the $3C^{pro}$ cleavage sites to prevent capsid processing and assembly, termed the Leaderless defective capsid (ΔLbdcap) GFP replicon. Schematics of these constructs are shown in Fig 1.

Since the ΔP1 GFP replicon contains a GFP coding region in place of the capsid coding region, it is unable to produce infectious virions. Although this does not affect RNA replication and translation, the deletion of the capsid coding region also removes any PSs contained within this region, including four PSs identified by Logan et. *al*.: PS3, PS4, PS5 along with the majority of PS2 (35).

An alternative replicon was therefore designed with the maximum number of PSs possible, with the hypothesis that it would be packaged more efficiently in the *trans*-encapsidation assay. This alternative replicon retains the capsid coding region and the associated PSs. However, inclusion of both the capsid with the GFP reporter rendered the genome too large to be packaged (39). Therefore, a ΔLb GFP virus was used where the deletion of the Lb portion of Leader coding region provides capacity in the genome for the GFP reporter. This region does not contain any currently identified PSs and has little effect on viability in BHK-21 cells if the remainder of Lab is still present (37). Substitutions in the 3C cleavage sites within the capsid sequences were introduced to prevent the capsid proteins from being processed, thus preventing capsid assembly in the absence of helper viruses.

In the first round of the trans-encapsidation assay, the GFP signal from the ΔLbdcap GFP replicon was expressed less rapidly compared to the ΔP1 GFP replicon (Fig 3a). This indicates that the ΔLbdcap replicon replicates with slower kinetics, but it reached the same peak signal for both mean fluorescence intensity (Fig 3a) and GFP object count (which represents the

297 number of GFP positive cells) (Fig 3b). Both samples were harvested at the point of peak GFP
298 expression (5 and 7 hours respectively for the ΔP1 and ΔLbdcap GFP replicons) to ensure that
299 these were harvested at equivalent points in the lifecycle.

300 The second round of the *trans*-encapsidation assay revealed that the ΔLbdcap GFP replicon
301 was packaged 25% more efficiently compared to the ΔP1 GFP replicon (Fig 3c). This indicated
302 that the benefits for *trans*-encapsidation provided by including additional PSs outweigh the
303 fitness costs incurred by deleting the Leader protein coding sequence.

304 ## The length of the poly-(C) tract affects encapsidation

305 The PK region and PS1 are located immediately adjacent to the poly-(C) tract (Fig 1), and for
306 practical reasons the DNA fragments used to alter the PK region contained a synthetic poly-
307 (C) tract which was shorter than the poly-(C) tract in the parental replicons. To investigate the
308 possible influence of poly-(C) tract length on packaging, two variations of the ΔLbdcap GFP
309 replicon were generated using synthetic DNA sequences containing poly-(C) tract lengths of
310 11 and 29 nt respectively for comparison to the *wt* ΔLbdcap replicon, which has
311 approximately 35 cytosines. Furthermore, it should be noted that these poly-(C) tracts are all
312 shorter than the ca. 50-200 nt present in the rescued viruses due to the difficulty of cloning
313 and maintaining a long poly-(C) tract in the plasmid.

314 The poly-(C) tract length of each construct was verified by digesting the plasmids at
315 restriction sites to excise a fragment containing 316 base pairs (bp) plus the poly-(C) tract and
316 analysing the size of each DNA fragment using electrophoresis (Fig 4). Although the
317 resolution was not sufficient to determine the exact lengths of the poly-(C) tracts consistently
318 across multiple experiments, it was possible to approximate the lengths in each analysis. The
319 fragment sizes in each experiment were also compared to determine the relative differences

in poly-(C) tract lengths between the three samples to confirm that the *wt* was largest in each experiment, as expected, followed by the C29 replicon and then the C11 replicon. The C11 poly-(C) tract was consistently shorter than that from both other replicons, and was estimated to contain between 9 and 12 cytosines. The C29 and *wt* replicons appeared to be of a similar size, with the *wt* replicon appearing either slightly larger or of the same size, consistent with the small difference in expected lengths. Both were estimated to be between 30 and 36 cytosines in length.

The RNA transcripts from these plasmids were then tested in the *trans*-encapsidation assay described above. No differences in replication of these replicons were observed in the first round, either by GFP counts or the MFI of the GFP objects (Fig 5a-b), indicating that the poly-(C) tract truncations had no measurable effect on RNA replication. In the second round, however, *trans*-encapsidation efficiency correlated with the length of the poly-(C) tract; the *wt* C35 replicon with the longest poly-(C) tract had the greatest peak GFP count of 5700 at 16 hours, while the C11 and C29 replicons had peak counts reaching only 1500 +/- 50 (26% of *wt*) and 4000 +/- 150 (71% of *wt*) respectively at 16 hours (Fig 5c). All differences were significant ($p < 0.0001$).

The effect of truncations to the poly-(C) tract on encapsidation efficiency has not been previously reported and no differences in virus viability have been observed, provided that the poly-(C) tract is at least 6 nucleotides in length (16). We also saw no effect on RNA replication or ability to rescue virus from an infectious whole-genome RNA transcript with 11 cytosines in the poly-(C) tract (17). To extend these observations, we investigated the effect of poly-(C) tract truncation in the context of initial rates of grow*t*h of infectious virus. For this,

we generated an additional infectious copy plasmid containing 29 cytosines for comparison against the C11 and C35 (*wt)* versions.

RNA transcripts were made from these plasmids and transfected into cells in a CPE development assay to investigate the effect of these truncations on virus rescue, CPE development and cell-to-cell spread (Fig 5d). However, no clear differences were observed, and although the C11 transcript appeared to induce slightly less CPE than the other two transcripts initially, this was not statistically significant. The impaired encapsidation observed in the *trans*-encapsidation assay was therefore not reflected in the recovery of infectious virus.

These infectious copy RNA transcripts were subsequently used as capsid-donors in a 'reverse' *trans*-encapsidation assays, using the *wt* ΔLbdcap (replicon lacking functional capsid) as the genome-donor in each case. In this scenario, it was expected that the more competitive the capsid-donor RNA is at packaging, the fewer the resources available for replicon RNA packaging, resulting in reduced *trans*-encapsidation efficiency. In support of this prediction, the *trans*-encapsidation efficiency of the genome donor replicon was reduced when using capsid donor infectious copy transcripts with longer (C29, C35) poly-(C) tracts, and replicon *trans*-encapsidation increased when using the capsid donor infectious copy transcript with a shorter (C11) poly-(C) tract. Although there was also a slight increase in replicon *trans*-encapsidation when using the C29 capsid donor transcript over the *wt* capsid donor transcript, this was not statistically significant (Fig 5e). This was consistent with the results described earlier with the standard *trans*-encapsidation assay using the poly-(C) tract-modified replicons.

In summary, these findings demonstrate that longer poly-(C) tracts confer better packaging efficiency. This represents the first clear role that has been identified for this feature in the FMDV lifecycle.

The poly-(C) tract extends rapidly due to selection pressure from encapsidation

The FMDV poly-(C) tract is typically >100 cytosines in length in *wt* virus genomes (11), which is considerably longer than the poly-(C) tracts present in the RNA transcripts used in this study. It is only possible to consistently maintain a poly-(C) tract length of around 35 nt in the infectious copy plasmid, which is why the poly-(C) tract lengths of our constructs are considerably shorter than the poly-(C) tracts found in *wt* virus genomes. However, providing they are greater than a minimal length, poly-(C) tracts shorter than *wt* have not been reported to effect replication, translation, virus viability or genome packaging. Even when virus was recovered from the C11 transcript alongside the C35 transcript, there was no noticeable difference in terms of the speed of rescue (Fig 5d). However, short poly-(C) tracts have been reported to revert to a *wt* length of 75-100 nt. during virus growth, thus obscuring effects the truncations may have on the virus (15, 16). To determine if this extension of the poly-(C) tract is rapid enough to explain the apparent *wt* viability of transcripts with severely truncated poly-(C) tracts, we compared the lengths of the poly-(C) tracts in RNA transcripts derived from the three infectious constructs, C11, C29, C35 (*wt*), with those present in the recovered viruses. Viruses were recovered by transfection of cells with transcript RNA, followed by a single passage on new cells until development of complete CPE (approximately 24 hours). This ensured that the majority of the viral RNA taken forward for analysis was derived from virions produced during the initial transfection, thus having passed through the packaging bottleneck at least once.

Transcripts or recovered RNA were digested with RNase T1 and resulting RNA fragments were analysed by automated electrophoresis to assess the lengths of the RNase T1 resistant poly-(C) tracts in each case (40). A peak at ~25 nt was seen in digests of the transcripts derived from the original modified plasmids (Fig 6a) whereas a major peak at ~70 nt was seen in digests of the RNAs from virus recovered after a single passage following transfection (Fig 6b), as expected for a newly recovered poly-(C) tract (16). This suggests that the rapid restoration of the poly-(C) tract length obscures differences between the three viruses in the CPE development assay and hinders the study of encapsidation.

At least one pseudoknot is required for encapsidation

Our previous work on the PK region showed that PK deletions did not prevent RNA replication, but appeared to be detrimental to RNA encapsidation, as inferred from the lack of virus recovery (17). To investigate the hypothesised loss of packaging associated with PK deletions, we used modified replicons in the *trans*-encapsidation assay.

These replicons encompassed distinct deletions within the PK region, each within the ΔP1 GFP replicon backbone, i.e.: ΔPK34, with PKs 3 and 4 deleted; ΔPK234, with PKs 2, 3 and 4 deleted; and ΔPK1234, with all four PKs deleted (Fig 1). Due to cloning difficulties arising from the proximity of the PK region to the poly-(C) tract, the ΔPK1234 replicon contained a poly-(C) tract of only 11 cytosines as opposed to the approximately 35 cytosines present in the poly-(C) tracts of the other replicon plasmids. To differentiate effects on replication of the truncated poly-(C) tract from the effects of the complete PK region deletion, we previously created a C11 replicon with all four PKs present but containing a truncated poly-(C) tract with 11 cytosines. This C11 replicon was used here in a *trans*-encapsidation experiment to clarify

409 whether defects on packaging were due to the deletion of the PK region, the truncation of
410 the poly-(C) tract or a combination of both.

411 In this *trans*-encapsidation experiment (as in Fig 2) the replication kinetics of the C11, ΔPK34
412 and ΔPK234 replicons were similar in terms of GFP expression in the first round of the assay,
413 both from GFP object count (Fig 7a), which represents the number of cells in which the GFP
414 replicon is replicating based on the parameters defined in the analysis, and from the MFI (Fig
415 7b). This allowed direct comparison of these constructs in the second round of the assay
416 which assesses encapsidation efficiency. However, the C11 ΔPK1234 replicon produced
417 significantly fewer and less intense GFP foci (Fig 7a-b), consistent with our previous
418 observation (17). Despite this reduction in replication kinetics, it was still tested in the second
419 round of the assay, but considerations were made when drawing conclusions to account for
420 the effects that the reduced replication of the C11 ΔPK1234 replicon would have on
421 encapsidation.

422 In the second round, it was apparent that all three ΔPK mutant transcripts had been
423 packaged less efficiently *in trans* compared to their parent replicons (*wt* GFP replicon
424 compared to the ΔPK34 and ΔPK234 replicons and C11 GFP replicon compared to the
425 ΔPK1234 replicon) (Fig 7c). *Trans*-encapsidation efficiency for the C11 replicon was also
426 reduced, consistent with the previous findings in this study (Fig 4).

427 Most notably, packaging of the C11 ΔPK1234 transcripts was severely reduced with a 12-fold
428 reduction in the number of GFP positive cells per well compared to the C11 replicon. In
429 contrast, the peak levels for the ΔPK234 and ΔPK34 replicons were only reduced by 2.5- and
430 3.3-fold respectively compared to the *wt* replicon (Fig 7c). Although the C11 ΔPK1234
431 replicates less well than the other transcripts (as described above and in Fig 7a and 7b), this

reduction was thought to be insufficient to account for the near complete loss of encapsidation and the complete loss of virus recovery seen in our previous study (17). A caveat for the interpretation of these results is that the reduced level of replication seen with the C11 ΔPK1234 construct suggests that it may produce less replicon RNA for packaging compared to the other replicons.

Therefore, from these data alone, it was uncertain exactly why the C11 ΔPK1234 replicon was packaged so poorly. It may be because PK1 (or at least one PK if not specifically PK1), is essential; or it may be a compound outcome resulting from the effects of two mutations that are detrimental for packaging in addition to the poor replication of the replicon. Although a PK appears to be required for packaging, it is not clear whether PK1 specifically is necessary for efficient *trans*-encapsidation or if any of the PKs would suffice.

PK 1 is required for *wt* levels of encapsidation

To test if the PK1 deletion alone was attributable for the loss of *trans*-encapsidation efficiency, or if it was the result of multiple detrimental mutations, individual PK deletions were introduced into the replicon and tested in the assay. All four PKs have extremely similar sequences, suggesting that each may be able to functionally replace PK1 to facilitate packaging. PK1 and PK2 are particularly similar, differing only within a region of 11 nt at the 5' end (Fig 8a). Therefore, PK1 and PK2 were deleted individually in the ΔLbdcap GFP replicon (Fig 1) and tested in the *trans*-encapsidation assay (Fig 2). Due to limitations in the accuracy of the DNA synthesis, the poly-(C) tract length of the two constructs differed by one cytosine; the ΔPK1 replicon had 39 nt (C39 ΔPK1 GFP ΔLbdcap) and the ΔPK2 replicon had 40 nt (C40 ΔPK2 GFP ΔLbdcap) in this tract according to the manufacturer. Both poly-(C) tracts were greater in length compared to the estimated 35 cytosines in the *wt* ΔLbdcap replicon poly-(C)

tract. These constructs were deemed to be sufficiently close in poly-(C) tract length to enable reliable comparisons of their *trans*-encapsidation efficiencies.

In the first round of the *trans*-encapsidation assay, deletion of either PK1 or PK2 did not significantly diminish the levels of replication of each replicon; although both had higher GFP object counts compared to the *wt* replicon, these differences were not significant (Fig 8b). Additionally, the C39 ΔPK1 ΔLbdcap GFP replicon expressed GFP to a similar level of intensity as the *wt* replicon while the C40 ΔPK2 ΔLbdcap GFP replicon produced significantly more intense GFP fluorescence than either of the other two (Fig 8c). The similarity of the replication of these replicons made direct comparisons in the second round of the *trans*-encapsidation assay possible.

In the second round, the peak GFP count for both replicons was reached at equivalent times but with significantly different values; ΔPK1 was encapsidated less well than the *wt* replicon (at 43%), while the ΔPK2 replicon was encapsidated significantly more efficiently (at 127%) than the *wt* replicon (Fig 8d).

To reinforce the results of the *trans*-encapsidation assays by the analysis of virus viability, the individual PK deletions were also introduced into infectious copy plasmids, and RNA transcripts were then transfected into cells for the CPE development assay. The recovery of infectious virus and development of CPE after these transfections was based on the monolayer confluence; decreases in confluence represent progression of CPE through cell rounding and cell lysis/death, which both result in decreases in the area occupied by infected cells. CPE development was thus used to determine relative rates of progression of infection through cell monolayers, following transfection.

Following transfection, the ΔPK1 virus spread through the cells and caused CPE more slowly than either the *wt* or the ΔPK2 virus, which were comparable to each other (Fig 8e). This delay in virus spread could be attributed to the defects in encapsidation observed in the equivalent constructs in the *trans*-encapsidation assay (Fig 8d), with the reduced rates of encapsidation delaying the production of progeny virions.

Discussion

In earlier published work, *trans*-encapsidation in FMDV was reported to be inefficient, especially when compared to PV (34). It was speculated that this could be due to either a preference for FMDV genomes to be packaged into capsid proteins provided *in cis*, or a preference for full sized genomes to be packaged over the smaller replicons, but neither of these hypotheses was verified.

The current study indicates that a major issue with the previously published FMDV *trans*-encapsidation assays was that packaging of the replicon RNA was outcompeted by the *wt* helper virus RNA. This is in part due to the lack of the capsid coding region in the replicon, and consequent absence of PSs present in that region (35). However, the earlier *trans*-encapsidation studies with PV replicons showed relatively efficient *trans*-encapsidation despite equivalent deletion of the PV capsid region (25). Enteroviruses likewise contain PSs dispersed across the entire genome, including in the capsid coding region (25), suggesting that the deletion of the capsid coding region in the FMDV replicon is unlikely to be the sole reason for inefficient *trans*-encapsidation.

This inefficiency could be overcome in part by reducing the competitiveness of the capsid-donor. Crucially, by competing with virus RNA transcripts modified to contain shorter poly-(C) tracts than those in the replicon, the replicon was encapsidated with much greater efficiency.

Considering that previous iterations of the assay have used *wt* virus as the capsid-donor, and that the *wt* virus typically contains a poly-(C) tract with over 100 cytosines (11), it is likely that the reduced *trans*-encapsidation efficiency previously observed can be directly correlated to the length of the poly-(C) tract in the capsid-donor used. PV does not contain a poly-(C) tract and so this feature cannot explain differences in *trans*-encapsidation of PV and FMDV replicons. To counter this problem in FMDV, we used a transcript with a truncated poly-(C) tract as the capsid-donor. This reduced the competition faced by the FMDV replicon, making the *trans*-encapsidation assay a sensitive, specific and robust method for determining the effects of various mutations on the packaging of FMDV genomes.

The *trans*-encapsidation assay allowed us to confirm that PSs play a significant role in FMDV encapsidation. Not all PSs in the genome have equal effects, however. The PSs tested functionally by Logan et *al*. (35), namely PS2, PS3 and PS4, cumulatively resulted in a drop in virus viability compared to the *wt* virus, but the viruses were still able to package and propagate. Similarly, re-introducing PS3-5 to the GFP replicon by using a ΔLb GFP virus as the recipient resulted in an increase in encapsidation efficiency compared to the ΔP1 GFP replicon, although the ΔP1 GFP replicon was still packaged, indicating that these PSs alone are clearly not essential.

In contrast, the PS(s) in the PK region appear to be far more important for packaging compared to those present in the capsid coding region. Deleting the PK region almost abolished replicon *trans*-encapsidation, and while the deletions in C11 ΔPK1234 also affected the replicon replication kinetics, this is not sufficient to explain the reduction in *trans*-encapsidation.

Deleting only parts of the PK region also severely hindered encapsidation, but without affecting replication or translation. PK1 specifically appeared to be of particular importance, as deleting it severely reduced encapsidation whereas the equivalent deletion of PK2 seemingly increased encapsidation, although this may be attributable to the increased length of the poly-(C) tract in this transcript compared (C40) to the *wt* replicon (C35). PK1 also corresponds to the putative PS with the greatest constraint that Logan et *al*. (35) identified, supporting the hypothesis that the PS in this position is particularly important for FMDV encapsidation. However, a considerable amount of encapsidation still occurred when PK1 alone was deleted, indicating that PK1 itself is not essential. An explanation for this can be seen from the alignment of PK1 and PK2 (Fig 8a); aside from a stretch of nucleotides near the 5' end, the sequences are almost identical and PK2 is thus likely able to adopt the role of the PS. This similarity does include PK3 and PK4, although the sequences do diverge slightly more for the latter two PKs. When larger portions of the PK region were deleted, in ΔPK234 and ΔPK34, *trans*-encapsidation was similarly reduced. This reduction could be attributed to either "off-site" effects disrupting the PS in PK1 or to the deletion of active PSs. If the latter, the redundancy of the PKs could be since multiple PSs may act in concert to strengthen the PS interaction, with the 5'-most PS (PK1) being particularly important.

Due to the importance of the PK region and the PS(s) that it contains for encapsidation, we hypothesise that the influence the poly-(C) tract has on encapsidation is due to its proximity to PK1 and PS1. Rather than being a PS itself, the poly-(C) tract may instead act as a spacer between the PK region and the highly structured S-fragment, allowing the PS to fold independently of upstream components. A minimal number of cytosines may be necessary to achieve this, with an optimal number likely to be about 100 as seen in the *wt* virus (11). This requirement could explain why viruses with truncated poly-(C) tracts rapidly recover their

poly-(C) tracts and why all strains of FMDV contain this unusual feature, as they are selected for during the encapsidation bottleneck.

The role for the poly-(C) tract in encapsidation has likely remained undiscovered for two reasons. Firstly, the poly-(C) tract can rapidly recover from truncations. Truncations leaving at least six nucleotides in the poly-(C) tract have been reported to revert to *wt* length (16), and here we have demonstrated the same for an eleven nucleotide poly-(C) tract after one passage. This makes investigations into its role difficult since the length is so variable. Constructs must therefore be assessed directly following transfection instead of using recovered virus, because the process of rescuing virus allows the poly-(C) tract to extend and potentially nullify the effects of truncations within the original plasmid constructs. Secondly, results may have previously been obscured by the lack of competition from more packaging-competent constructs. Here, we have shown that although genomes containing longer poly-(C) tracts are packaged preferentially, the rates of virus recovery from constructs with poly-(C) tracts of different lengths were indistinguishable in the absence of competition.

It should be noted that a virus containing a poly-(C) tract with only two nucleotides (C2) has been shown to infect mice and to grow in cell culture, but to significantly lower titres than *wt* virus (16), possibly due to compromised encapsidation. Virus recovered from this C2 construct often had a 42 nt deletion, corresponding to PK1, which we have highlighted as an important PS (35). We have shown that constructs lacking only PK1 are still able to be encapsidated *in trans*, albeit inefficiently, suggesting that similar sequences in the other PKs are able to act as substitutes albeit at reduced efficiency. However, this reduced efficiency might only be true when PK1 is in the presence of a poly-(C) tract that can help it fold into the correct PS confirmation. It is therefore possible that when combined with a severely

569 truncated poly-(C) tract, which prevents PK1 from folding correctly, deletion of PK1 may
570 instead be beneficial and selected for during encapsidation.

571 A potential model for packaging based on these results proposes that the PK region can
572 adopt two alternative conformations; the first being the typical PK region, comprising four
573 pseudoknots, and the second being a packaging competent conformation in which a PS is
574 exposed and able to bind to a capsid precursor and initiate encapsidation (Fig 9). This would
575 enable the RNA to follow one of two pathways depending on the conformation adopted; a
576 packaging focussed pathway, and a replication focussed pathway. This transition between
577 two RNA conformations to control separate aspects of the virus lifecycle is a mechanism
578 which has been observed in other RNA virus families (41-43).

579 In this proposed model, as a newly replicated positive strand RNA molecule emerges from
580 the replication complex, the RNA of the PK1/PS1 region initially adopts a transient
581 conformation displaying the 'primary' PS, PS1. In this nascent state the PS1 binds to a
582 pentamer subunit to stabilise PS1 in the packaging conformation and initiate encapsidation.
583 This RNA-pentamer complex may then act as the nucleation site for capsid assembly with the
584 emerging replicated RNA engaging additional pentamers through 'secondary' PSs and
585 arranging them strategically within the assembling capsid. This may assist folding of the
586 FMDV genome into a compact conformation with a small hydrodynamic radius based on the
587 PS interactions with the nearby pentamers. The resulting particle is converted to infectious
588 virus by maturation cleavage of VP0 into VP2 and VP4.

589 Alternatively, if the RNA emerging from the replication complex fails to interact with a capsid
590 precursor, the transient PS conformation collapses into a more energetically favourable PK
591 conformation, and the RNA serves as a template for further replication and/or for translation.

In this model, we have divided the PSs into two categories – a singular primary PS, which we hypothesise to be essential for genome encapsidation and to act as the encapsidation primer; and secondary PSs, which individually are not essential, but each contributes to the cumulative effect on encapsidation. This model, involving the interplay between PSs and pentamer subunit concentration, may explain how picornavirus encapsidation is highly specific and why it only involves actively replicating RNA (44). It also provides a mechanism for optimising the timing of the transition from replication and translation to packaging.

Furthermore, this model is not incompatible with encapsidation models proposed previously. In PV, for example, earlier work has shown that protein-protein interactions are key for encapsidation (4, 33). This protein-focused approach could act in concert with RNA PSs, where the protein interactions facilitate the recruitment of capsid precursors to the replication complex in picornaviruses, providing opportunities for the RNA and the capsid precursors to interact. Following this localisation, a primary PS initiates encapsidation and secondary PSs dispersed throughout the genome facilitate the RNA encapsidation process, as suggested by our model.

Acknowledgements

The authors would also like to acknowledge Prof. Stephen Curry for his support as a supervisor to CN during his PhD.

Author Contributions

Conceptualisation: Tobias J. Tuthill, Graham J. Belsham, Chris Neil

Data curation: Chris Neil

Formal Analysis: Chris Neil

Funding acquisition: Tobias J. Tuthill, Graham J. Belsham

Investigation: Chris Neil

Methodology: Chris Neil, Joseph Newman, Tobias J. Tuthill

Project Administration: Chris Neil, Joseph Newman, Tobias J. Tuthill

Supervision: Tobias J. Tuthill, Joseph Newman, Graham Belsham

Validation: Chris Neil

Visualisation: Chris Neil

Writing – Original Draft Preparation: Chris Neil, Tobias J. Tuthill

Writing – Review & Editing: Chris Neil, Joseph Newman, Nicola J. Stonehouse, David J. Rowlands, Graham J. Belsham, Tobias J. Tuthill

References

1. Alexandersen S, Mowat N. Foot-and-mouth disease: host range and pathogenesis. Curr Top Microbiol Immunol. 2005;288:9-42.

2. James AD, Rushton J. The economics of foot and mouth disease. Rev Sci Tech. 2002;21(3):637-44.

3. Belsham GJ, Kristensen T, Jackson T. Foot-and-mouth disease virus: Prospects for using knowledge of virus biology to improve control of this continuing global threat. Virus Res. 2020;281:197909.

4. Jiang P, Liu Y, Ma H-C, Paul AV, Wimmer E. Picornavirus Morphogenesis. Microbiol Mol Biol Rev. 2014;78(3):418-37.

5. Newman J, Asfor AS, Berryman S, Jackson T, Curry S, Tuthill TJ. The Cellular Chaperone Heat Shock Protein 90 Is Required for Foot-and-Mouth Disease Virus Capsid Precursor Processing and Assembly of Capsid Pentamers. J Virol. 2018;92(5):10.1128/jvi.01415-17.

6. Gullberg M, Muszynski B, Organtini LJ, Ashley RE, Hafenstein SL, Belsham GJ, Polacek C. Assembly and characterization of foot-and-mouth disease virus empty capsid particles expressed within mammalian cells. J Gen Virol. 2013;94(8):1769-79.

7. Stanway G, Kalkkinen N, Roivainen M, Ghazi F, Khan M, Smyth M, et al. Molecular and biological characteristics of echovirus 22, a representative of a new picornavirus group. J Virol. 1994;68(12):8232-8.

8. Curry S, Fry E, Blakemore W, Abu-Ghazaleh R, Jackson T, King A, et al. Dissecting the roles of VP0 cleavage and RNA packaging in picornavirus capsid stabilization: the structure of empty capsids of foot-and-mouth disease virus. J Virol. 1997;71(12):9743-52.

9. Nomoto A, Kitamura N, Golini F, Wimmer E. The 5'-terminal structures of poliovirion RNA and poliovirus mRNA differ only in the genome-linked protein VPg. Proc Natl Acad Sci U S A. 1977;74(12):5345-9.

10. Sangar DV, Rowlands DJ, Harris TJ, Brown F. Protein covalently linked to foot-and-mouth disease virus RNA. Nature. 1977;268(5621):648-50.

11. Mellor EJ, Brown F, Harris TJ. Analysis of the secondary structure of the poly(C) tract in foot-and-mouth disease virus RNAs. J Gen Virol. 1985;66 (Pt 9):1919-29.

12. Duke GM, Osorio JE, Palmenberg AC. Attenuation of Mengo virus through genetic engineering of the 5' noncoding poly(C) tract. Nature. 1990;343(6257):474-6.

13. Harris TJ, Brown F. Biochemical analysis of a virulent and an avirulent strain of foot-and-mouth disease virus. J Gen Virol. 1977;34(1):87-105.

14. Costa Giomi MP, Bergmann IE, Scodeller EA, Auge de Mello P, Gomez I, La Torre JL. Heterogeneity of the polyribocytidylic acid tract in aphthovirus: biochemical and biological studies of viruses carrying polyribocytidylic acid tracts of different lengths. J Virol. 1984;51(3):799-805.

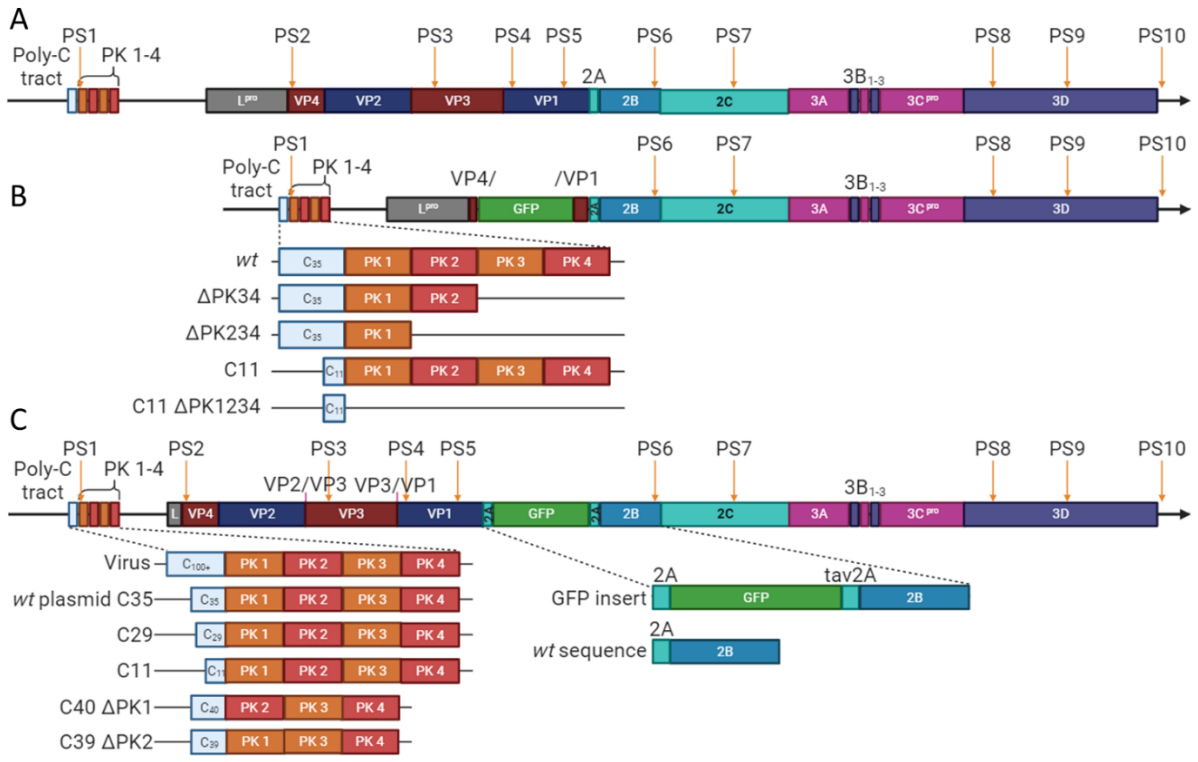
660 15. Zibert A, Maass G, Strebel K, Falk MM, Beck E. Infectious foot-and-mouth disease virus
661 derived from a cloned full-length cDNA. J Virol. 1990;64(6):2467-73.
662 16. Rieder E, Bunch T, Brown F, Mason PW. Genetically engineered foot-and-mouth disease
663 viruses with poly(C) tracts of two nucleotides are virulent in mice. J Virol. 1993;67(9):5139-45.
664 17. Ward JC, Lasecka-Dykes L, Neil C, Adeyemi OO, Gold S, McLean-Pell N, et al. The RNA
665 pseudoknots in foot-and-mouth disease virus are dispensable for genome replication, but essential
666 for the production of infectious virus. PLoS Pathog. 2022;18(6):e1010589.
667 18. Rolfsson Ó, Middleton S, Manfield IW, White SJ, Fan B, Vaughan R, et al. Direct Evidence for
668 Packaging Signal-Mediated Assembly of Bacteriophage MS2. J Mol Biol. 2016;428(2 Pt B):431-48.
669 19. Sasaki J, Nagashima S, Taniguchi K. Aichi virus leader protein is involved in viral RNA
670 replication and encapsidation. J Virol. 2003;77(20):10799-807.
671 20. Sasaki J, Taniguchi K. The 5'-end sequence of the genome of Aichi virus, a picornavirus,
672 contains an element critical for viral RNA encapsidation. J Virol. 2003;77(6):3542-8.
673 21. Zhu L, Wang X, Ren J, Porta C, Wenham H, Ekstrom JO, et al. Structure of Ljungan virus
674 provides insight into genome packaging of this picornavirus. Nature communications. 2015;6:8316.
675 22. Kalynych S, Pálková L, Plevka P. The Structure of Human Parechovirus 1 Reveals an
676 Association of the RNA Genome with the Capsid. J Virol. 2016;90(3):1377-86.
677 23. Shakeel S, Westerhuis BM, Domanska A, Koning RI, Matadeen R, Koster AJ, et al. Multiple
678 capsid-stabilizing interactions revealed in a high-resolution structure of an emerging picornavirus
679 causing neonatal sepsis. Nature communications. 2016;7:11387.
680 24. Shakeel S, Dykeman EC, White SJ, Ora A, Cockburn JJB, Butcher SJ, et al. Genomic RNA
681 folding mediates assembly of human parechovirus. Nature communications. 2017;8(1):5.
682 25. Chandler-Bostock R, Mata CP, Bingham RJ, Dykeman EC, Meng B, Tuthill TJ, et al. Assembly
683 of infectious enteroviruses depends on multiple, conserved genomic RNA-coat protein contacts.
684 PLoS Pathog. 2020;16(12):e1009146.

685 26. Twarock R, Leonov G, Stockley PG. Hamiltonian path analysis of viral genomes. Nature
686 communications. 2018;9(1):2021.
687 27. Dykeman EC, Grayson NE, Toropova K, Ranson NA, Stockley PG, Twarock R. Simple rules for
688 efficient assembly predict the layout of a packaged viral RNA. J Mol Biol. 2011;408(3):399-407.
689 28. Stockley PG, Rolfsson O, Thompson GS, Basnak G, Francese S, Stonehouse NJ, et al. A simple,
690 RNA-mediated allosteric switch controls the pathway to formation of a T=3 viral capsid. J Mol Biol.
691 2007;369(2):541-52.
692 29. Johnson VH, Semler BL. Defined recombinants of poliovirus and coxsackievirus: sequence-
693 specific deletions and functional substitutions in the 5'-noncoding regions of viral RNAs. Virology.
694 1988;162(1):47-57.
695 30. Barclay W, Li Q, Hutchinson G, Moon D, Richardson A, Percy N, et al. Encapsidation studies
696 of poliovirus subgenomic replicons. J Gen Virol. 1998;79 (Pt 7):1725-34.
697 31. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. Reduction of the rate of
698 poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral
699 virulence by lowering specific infectivity. J Virol. 2006;80(19):9687-96.
700 32. Todd S, Towner JS, Brown DM, Semler BL. Replication-competent picornaviruses with
701 complete genomic RNA 3' noncoding region deletions. J Virol. 1997;71(11):8868-74.
702 33. Liu Y, Wang C, Mueller S, Paul AV, Wimmer E, Jiang P. Direct interaction between two viral
703 proteins, the nonstructural protein 2C and the capsid protein VP3, is required for enterovirus
704 morphogenesis. PLoS Pathog. 2010;6(8).
705 34. McInerney GM, King AM, Ross-Smith N, Belsham GJ. Replication-competent foot-and-mouth
706 disease virus RNAs lacking capsid coding sequences. J Gen Virol. 2000;81(Pt 7):1699-702.
707 35. Logan G, Newman J, Wright CF, Lasecka-Dykes L, Haydon DT, Cottam EM, Tuthill TJ. Deep
708 sequencing of foot-and-mouth disease virus reveals RNA sequences involved in genome packaging. J
709 Virol. 2017.

710 36. Tulloch F, Pathania U, Luke GA, Nicholson J, Stonehouse NJ, Rowlands DJ, et al. FMDV
711 replicons encoding green fluorescent protein are replication competent. Journal of Virological
712 Methods. 2014;209(0):35-40.
713 37. Belsham GJ. Influence of the Leader protein coding region of foot-and-mouth disease virus
714 on virus replication. J Gen Virol. 2013;94(7):1486-95.
715 38. Ellard FM, Drew J, Blakemore WE, Stuart DI, King AMQ. Evidence for the role of His-142 of
716 protein 1C in the acid-induced disassembly of foot-and-mouth disease virus capsids. J Gen Virol.
717 1999;80(8):1911-8.
718 39. Seago J, Juleff N, Moffat K, Berryman S, Christie JM, Charleston B, Jackson T. An infectious
719 recombinant foot-and-mouth disease virus expressing a fluorescent marker protein. J Gen Virol.
720 2013;94(Pt 7):1517-27.
721 40. Black DN, Stephenson P, Rowlands DJ, Brown F. Sequence and location of the poly C tract in
722 aphtho- and cardiovirus RNA. Nucleic Acids Res. 1979;6(7):2381-90.
723 41. Villordo SM, Gamarnik AV. Genome cyclization as strategy for flavivirus RNA replication.
724 Virus Res. 2009;139(2):230-9.
725 42. Larman BC, Dethoff EA, Weeks KM. Packaged and Free Satellite Tobacco Mosaic Virus
726 (STMV) RNA Genomes Adopt Distinct Conformational States. Biochemistry. 2017;56(16):2175-83.
727 43. Baranovskaya I, Sergeeva M, Fadeev A, Kadirova R, Ivanova A, Ramsay E, Vasin A. Changes in
728 RNA secondary structure affect NS1 protein expression during early stage influenza virus infection.
729 Virol J. 2019;16(1):162.
730 44. Nugent CI, Johnson KL, Sarnow P, Kirkegaard K. Functional coupling between replication and
731 packaging of poliovirus replicon RNA. J Virol. 1999;73(1):427-35.

732

733



734

735 **Fig 1: Schematic of selected features within the virus, ΔP1 GFP replicon and ΔLbdcap GFP**

736 **replicon.** Representations of the (A) virus, (B) ΔP1 GFP replicon (including the ΔPK34, ΔPK234,

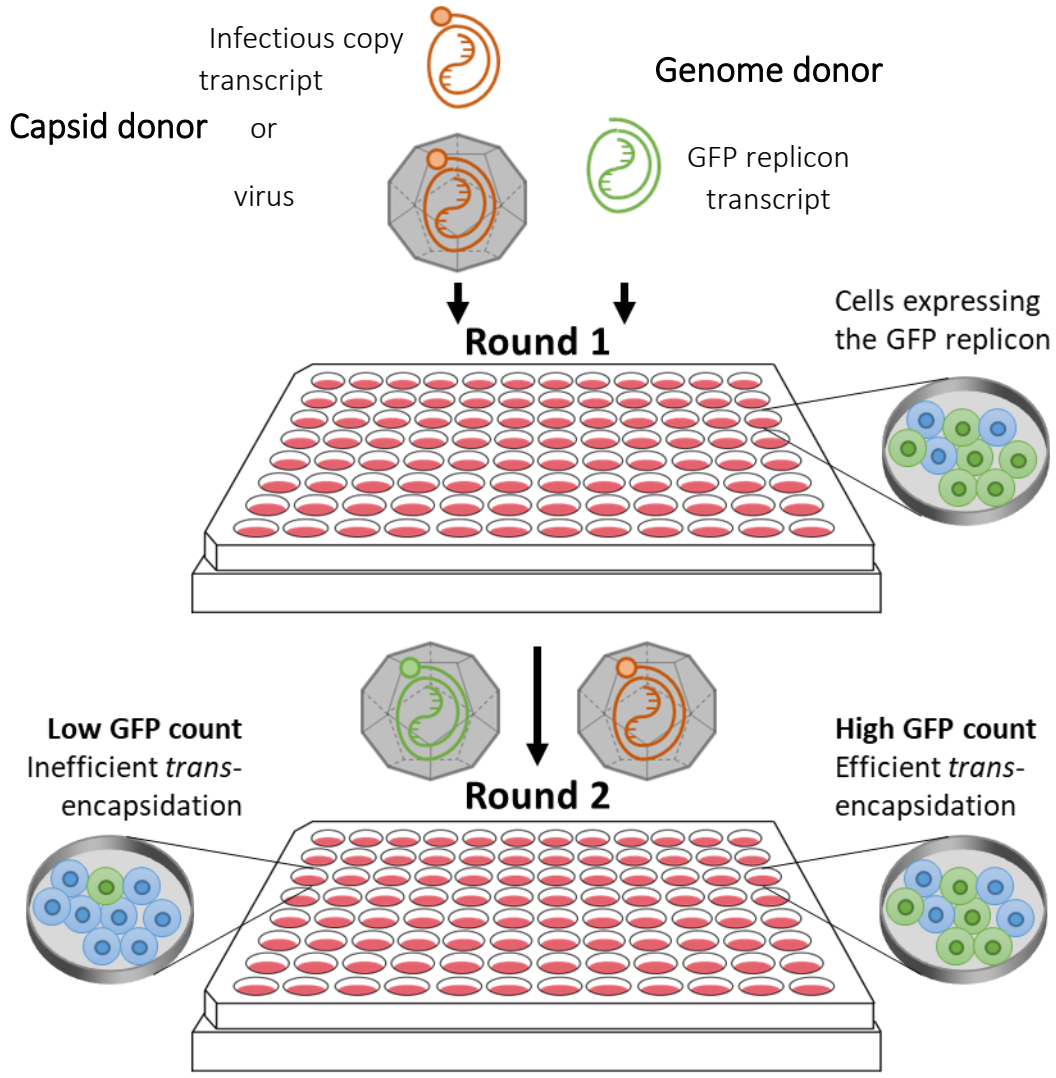
737 C11 and C11 ΔPK1234 versions) and (C) ΔLbdcap GFP replicon (including the *wt* C35, C29,

738 C11, C40 ΔPK1 and C39 ΔPK2 versions). Substitutions are labelled above their respective

739 positions (VP2/VP3 and VP3/1 $3C^{pro}$ cleavage sites), and packaging signals (described by

740 Logan et al.) are labelled above with orange arrows.

741



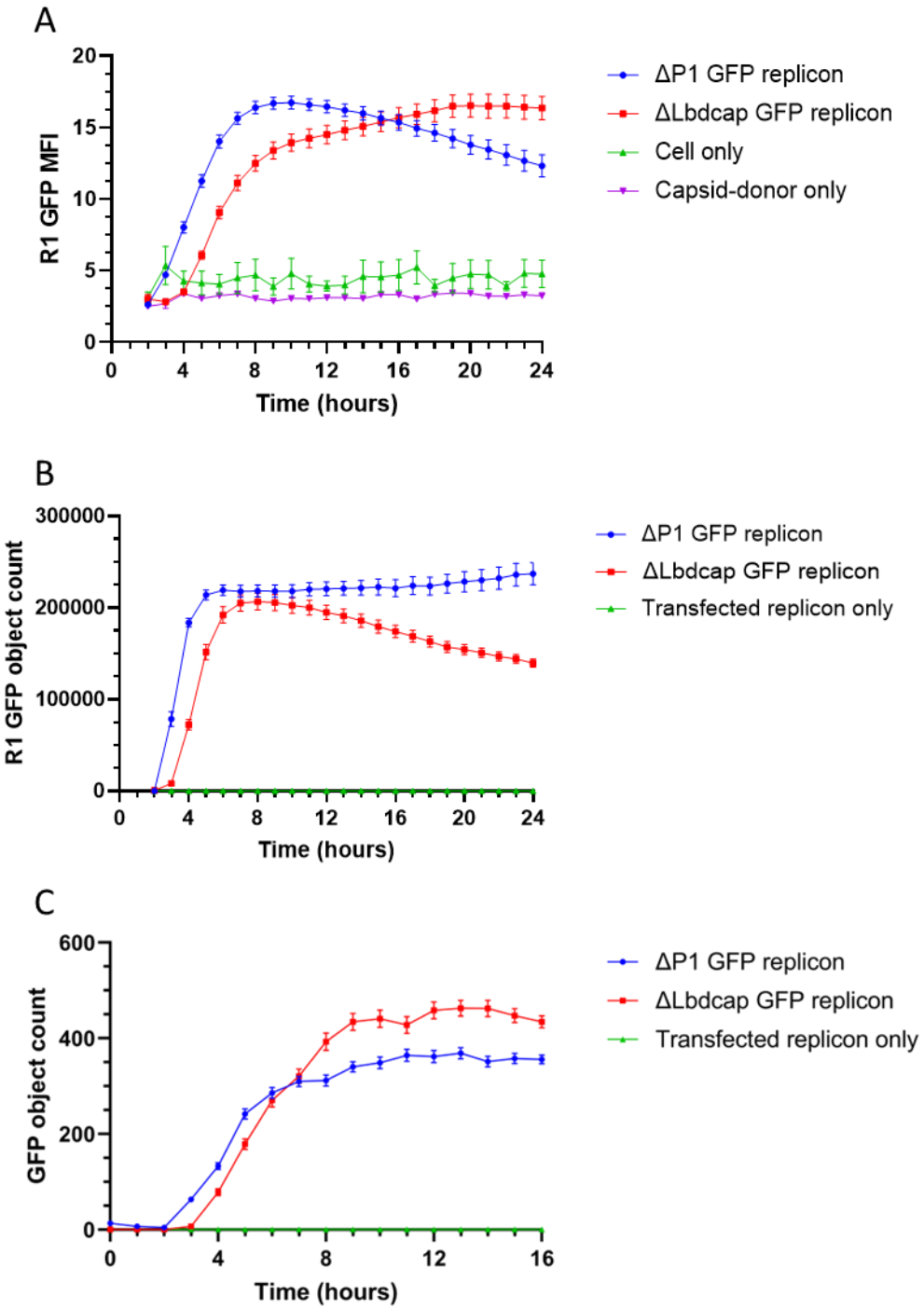
742

743 **Fig 2. *Trans*-encapsidation assay overview** (A) In the first round of the assay, cells were
744 transfected with a GFP replicon transcript and either infected with *wt* virus or co-transfected
745 with an infectious copy transcript, so the replicon and virus/infectious copy transcripts were
746 replicating in the same cells. This resulted in the production of both progeny *wt* virus and
747 progeny trans-encapsidated virus particles, with the GFP replicon packaged into the capsid
748 provided by the *wt* virus. The proportion of cells expressing GFP corresponds to the efficiency
749 of transfection. (B) The cell lysate containing the progeny *wt* virus and trans-encapsidated
750 virus particles was then used to infect a second round of fresh cells. Cells infected with the
751 trans-encapsidated GFP replicon expressed GFP, and the proportion of GFP expressing cells

752 compared to the standard GFP replicon sample represented the relative trans-encapsidation

753 efficiency of mutant GFP replicons.

754



755

756 **Fig 3. The presence of packaging signals enhances *trans*-encapsidation efficiency.** Cells were
757 transfected with the ΔP1 GFP replicon (blue) and ΔLbdcap GFP replicon (red) with virus co-
758 infection, along with an uninfected ΔLbdcap GFP replicon transfection only control (green).
759 (A) The MFI for the first round and the GFP counts for the (B) first and (C) second rounds of
760 the trans-encapsidation assays are shown. 'GFP objects' are the GFP positive foci which are in

761 the range of parameters defined in the analysis setup. Data shown are from a single

762 experiment, representative of multiple experiments, and error bars represent the SEM

763 calculated from (A-B) 5 and (C) 12 images.

764

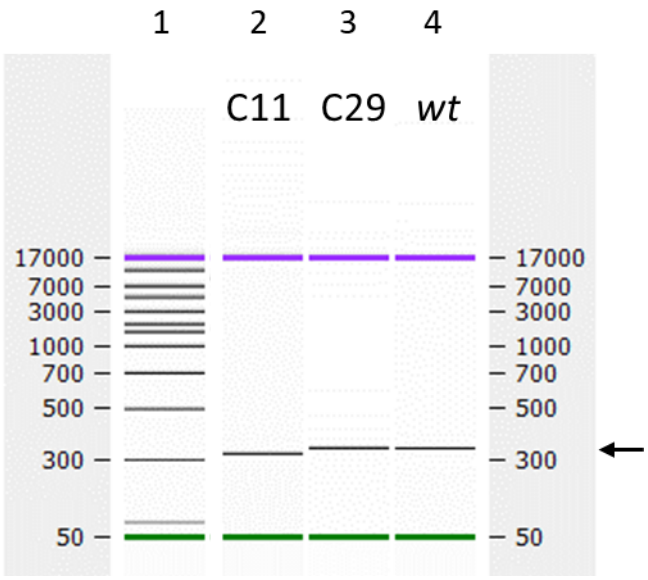
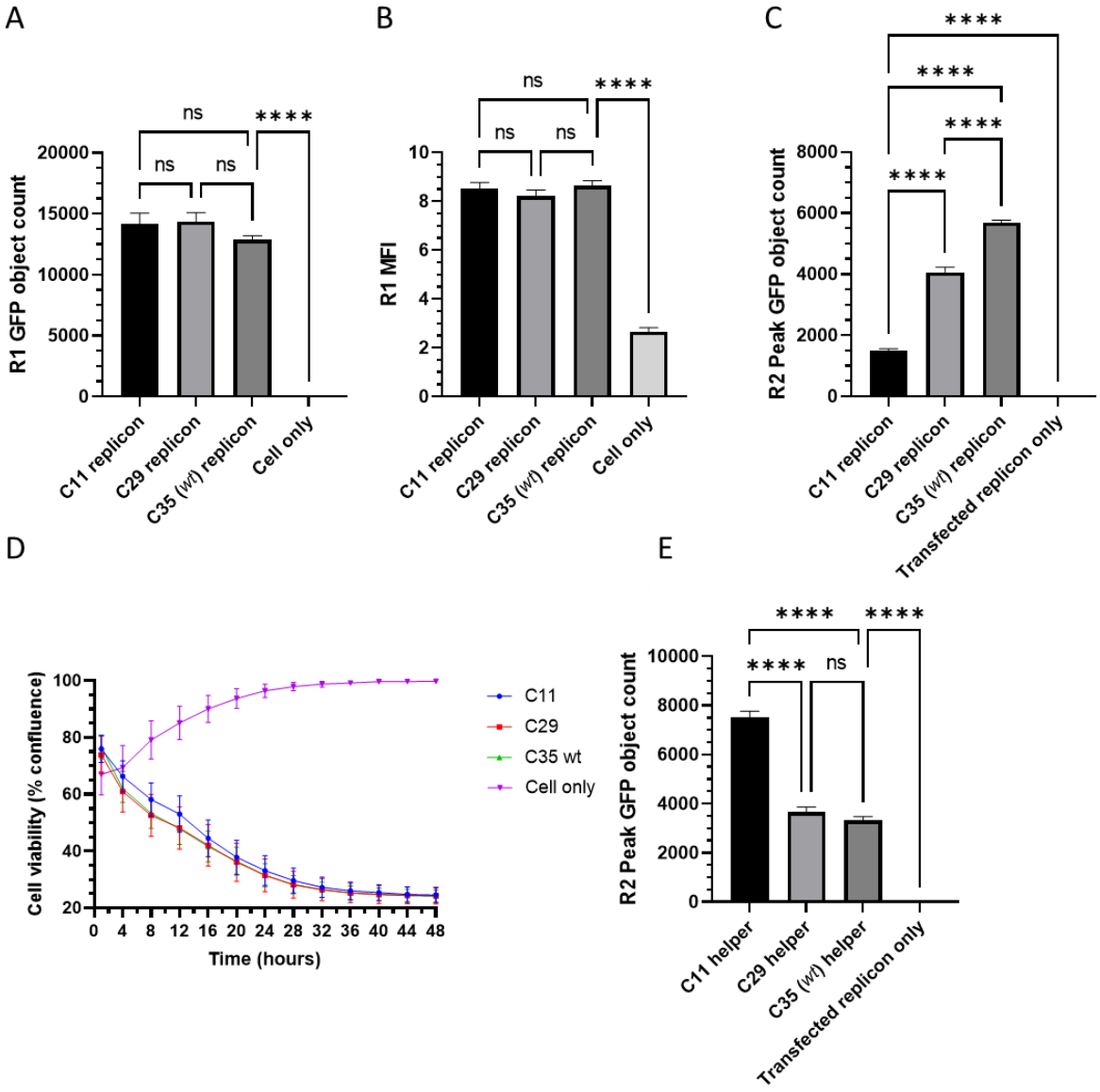


Fig 4. The poly-(C) tract in the C11 ΔLbdcap GFP replicon is shorter than the poly-(C) tracts in the C29 and C35 ΔLbdcap GFP replicons. Plasmids were digested using *XbaI* and *NheI*, gel purified to obtain the 300 bp fragment containing the poly-(C) tract and analysed using the Bioanalyser. The ladder (in bp) is shown in Lane 1, the C11 replicon in Lane 2, the C29 replicon in Lane 3 and the *wt* C35 replicon in Lane 4. The single analysis shown is representative of three experiments. The fragment of interest is indicated on the right by the arrow. The fragment from the C11 replicon was estimated to be 325-328 bp in size (C9-12), while the C29 and C35 *wt* replicon fragments were both estimated to be 347-352 bp in size (C30-36). Expected sizes were 327, 345 and 351 bp respectively.



776

777 **Fig 5. Transcripts with shorter poly-(C) tracts are less competitive during encapsidation.**

778 Replicons with differing length poly-(C) tracts were compared in the *trans*-encapsidation
779 assay and were assessed according to the (A) first round GFP object count, (B) first round
780 green MFI and (C) second round GFP object count. (D) Virus transcripts containing the
781 equivalent poly-(C) tracts were compared in terms of speed of CPE development and (E) their
782 ability to compete with a standard replicon for packaging resources, when used as the capsid-
783 donor in the *trans*-encapsidation assay; high replicon *trans*-encapsidation efficiency equates
784 to the capsid-donor being poorer at competing with the replicon. Data shown represent the

mean from triplicate wells at either the point of harvest (A-B) or the time point with peak GFP expression (C and E), and error bars represent the SEM calculated from 5 (A-B), 12 (C and E), and 15 (D) images. Significance is shown for A-C and E comparing the samples to each other and between the *wt* replicon and the cell only/transfected replicon only controls using a one-way ANOVA (**** $p < 0.0001$). Significance in D was calculated using Wilcoxon tests between each sample, but none were significant and are not shown ($p < 0.001$).

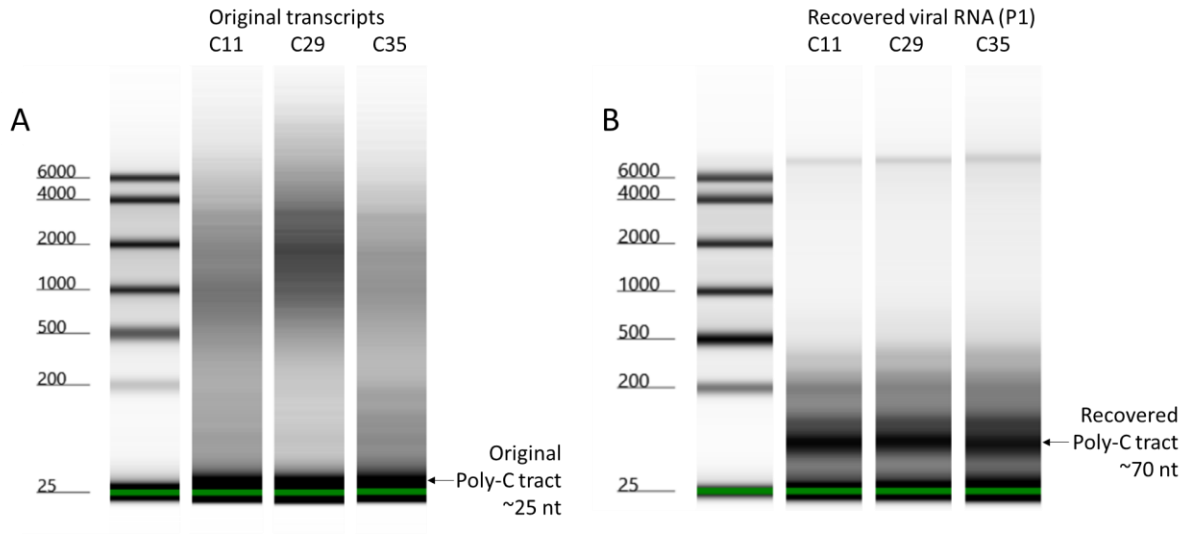


Fig 6. Truncated poly-(C) tracts extend to near *wt* length during the first passage. Automated electrophoresis analysis of RNase T1 digests of; (A) RNA samples from the *in vitro* transcription reactions and (B) the corresponding extracted RNAs after transfecting the transcripts and passaging the recovered viruses. Lane 1 in each contains the RNA ladder; Lane 2 the C11 sample, Lane 3 the C29 sample; and Lane 4 the *wt* C35 sample. The lower marker is at 25 nt. The fragments containing the poly-(C) tracts are indicated by arrows on the right-hand side.

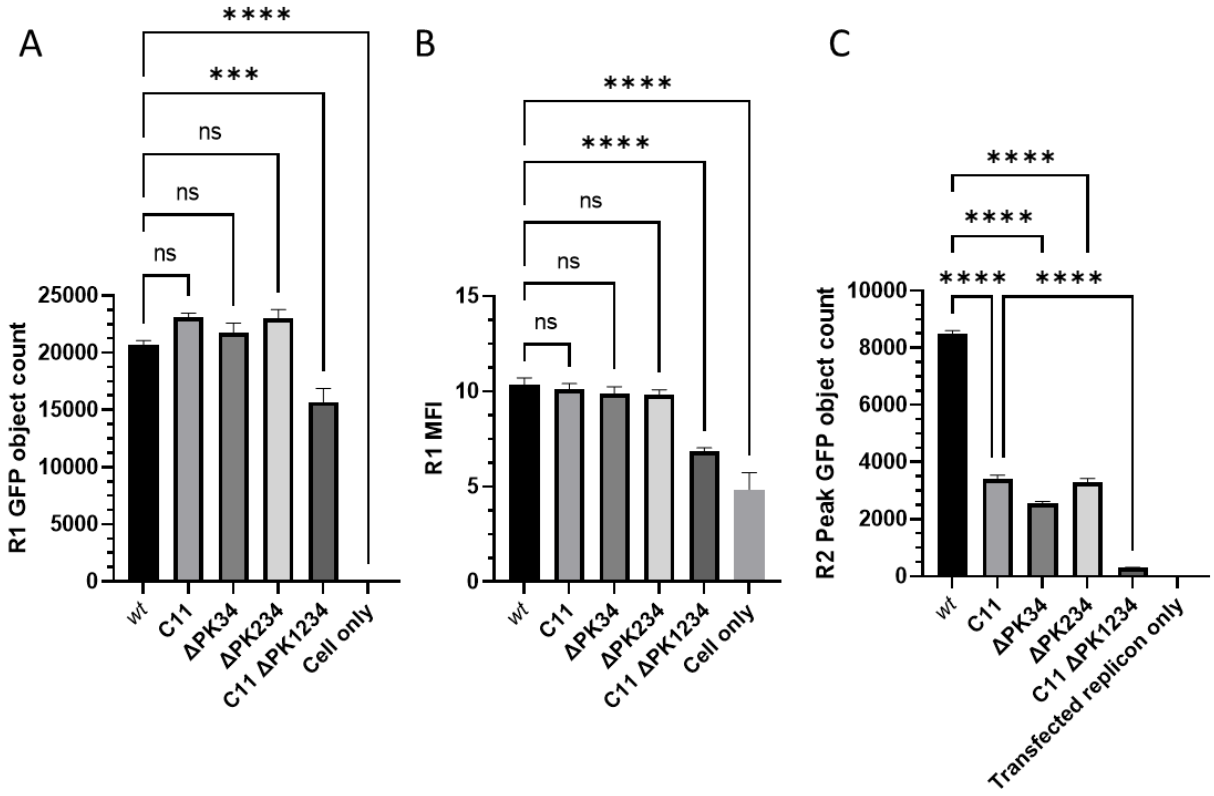
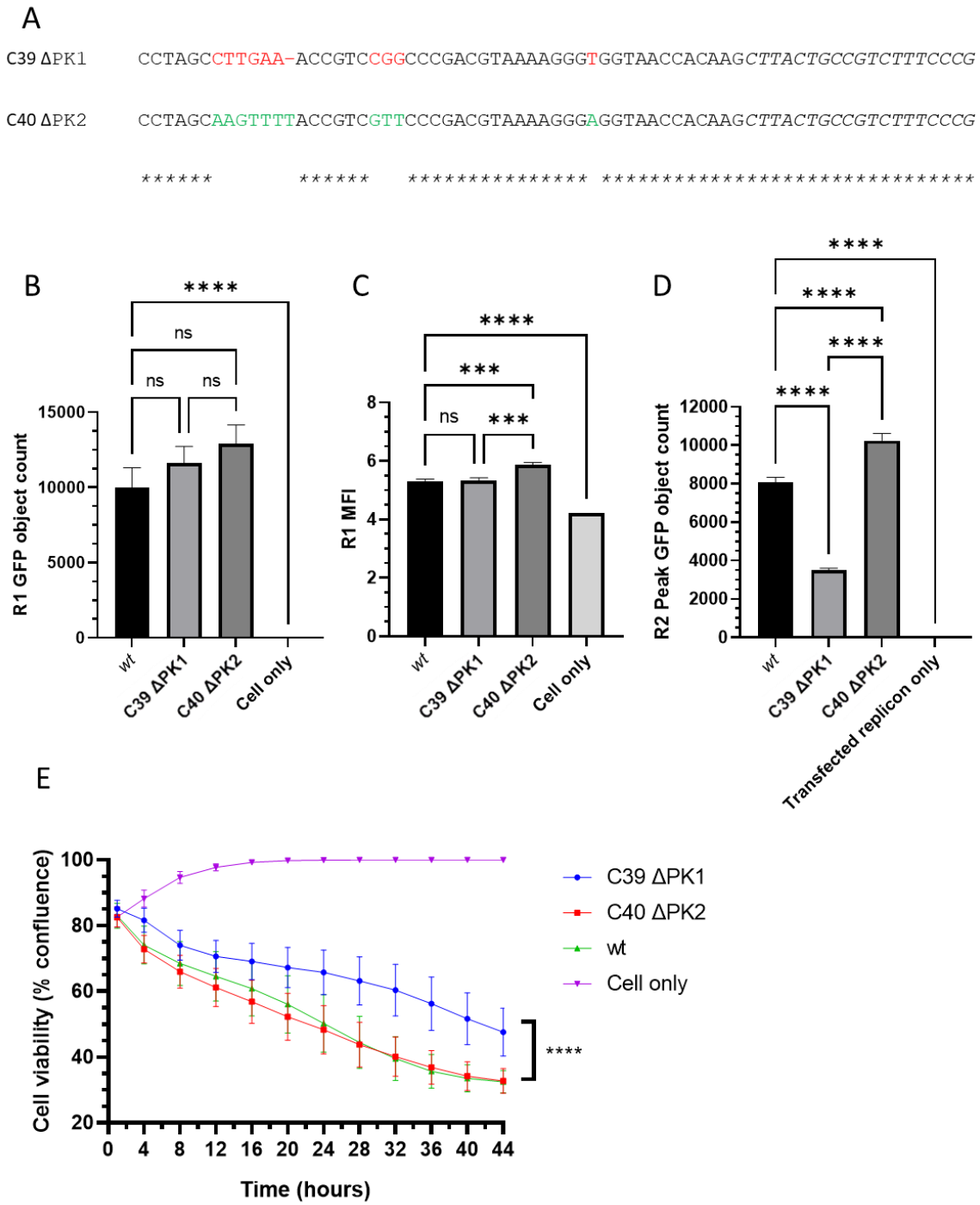


Fig 7. Replicons lacking the pseudoknots and poly-(C) tract are not *trans*-encapsidated.

Replicons containing deletions in the PK region and/or poly-(C) tract were used in the *trans*-encapsidation assay. The (A) GFP object count and (B) green MFI of the first round are shown, along with the (C) GFP object count of the second round. The data shown represent the mean from triplicate wells, and the error bars represent the SEM from 12 images. Significance is shown compared to the *wt* GFP replicon using a one-way ANOVA (**** $p < 0.0001$).



809

810 **Fig 8. Replicons lacking PK1 have reduced *trans*-encapsidation efficiencies.** (A) An alignment
811 of the 5'-end of the PK region is shown for the ΔPK1 and ΔPK2 deletion mutants, using red to
812 denote nucleotides unique to PK2 (in ΔPK1, top) or green to denote those unique to PK1 (in
813 ΔPK2, bottom). Replicons containing these mutants were tested in the *trans*-encapsidation
814 assay, and the (B) first round GFP object count, (C) first round green MFI and (D) second

round GFP object count are reported. (E) CPE development assay using transfected virus transcripts containing the deletions. The data shown represent the mean from triplicate wells, and the error bars represent the SEM from 12 images. Significance is shown compared to the *wt* GFP replicon using a one-way ANOVA (B-D) or by using a Wilcoxin test between the samples and the *wt* transcript (E) (**** $p < 0.0001$).

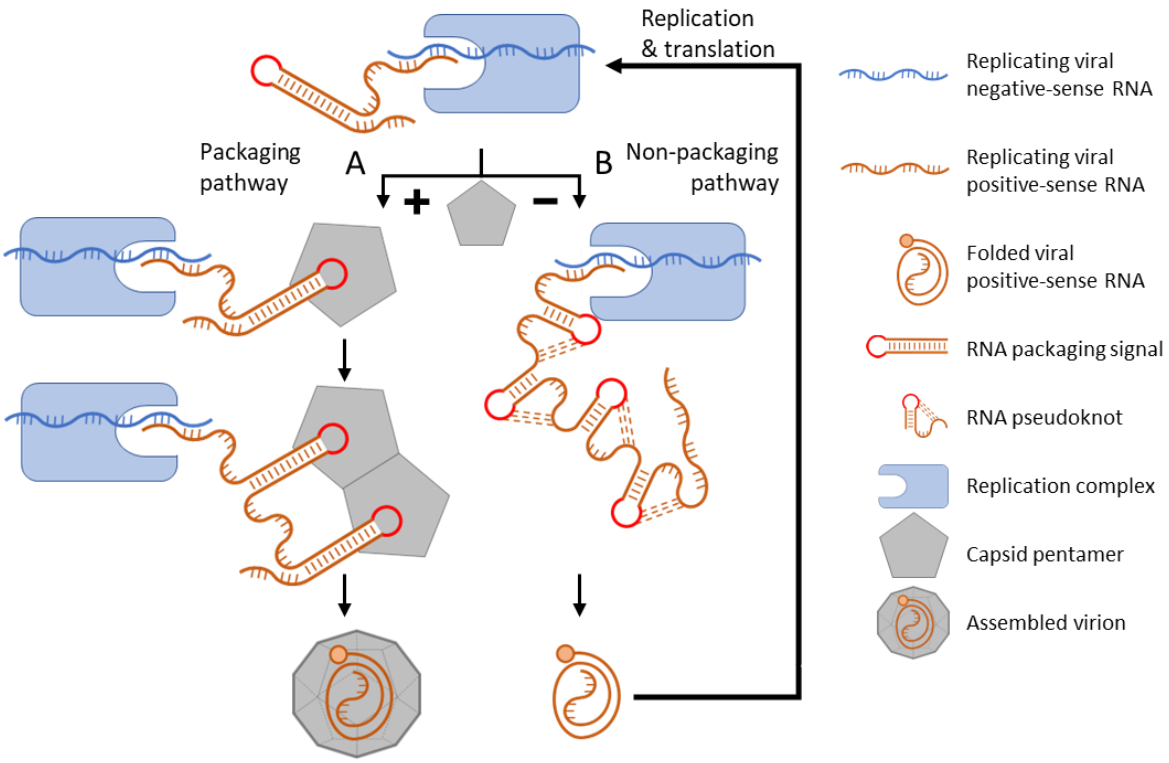


Fig 9. Model for capsid assembly. Negative-sense RNA (blue) acts as the template for the formation of corresponding positive-sense strand RNA molecules (brown), and PS1 forms a stem-loop as the RNA emerges from the replication complex. (A) If nascent PS1 encounters a pentamer, the interaction stabilises the RNA in the stem-loop conformation. Subsequent packaging signals then form as the RNA continues to emerge from the replication complex and these interact with additional pentamers to form the 'nucleus' of capsid assembly. Once completed, the genome is completely encapsidated within the new virion. (B) If no pentamers are present to stabilise PS1, the RNA stem-loop collapses into the pseudoknots conformation and the completed the positive-sense strand is used for translation and/or further RNA replication.