Article

Not peer-reviewed version

Picornavirus Evolution: Genomes Encoding Multiple 2A^{NPGP} Sequences – Biomedical and Biotechnological Utility

Garry Luke , Lauren Ross , Yi-Ting Lo , Hsing-Chieh Wu , Martin Ryan *

Posted Date: 25 September 2024

doi: 10.20944/preprints202409.1973.v1

Keywords: Picornaviruses; Polyprotein 2A region; 2A^{NPGP} sequences; ribosome skipping; translation; biotechnology

Article

Picornavirus Evolution: Genomes Encoding Multiple 2A^{NPGP} Sequences - Biomedical and Biotechnological Utility

Garry A. Luke [1], Lauren S. Ross [1], Yi-Ting Lo [2], Hsing-Chieh Wu [2] and Martin D. Ryan [1,*]

[1] School of Biology, University of St. Andrews, Biomolecular Sciences Research Complex, North Haugh, St. Andrews, KY16 9ST, Scotland, U.K

[2] International College, National Pingtung University of Science and Technology, 1, Shuefu Rd., Neipu, Pingtung 91201, Taiwan

* Correspondence: martin.ryan@st-andrews.ac.uk; Tel.: +01334-463403; Fax: +01334-462595

Abstract: Alignment of picornavirus proteinase/polymerase sequences reveals this family evolved into five 'supergroups'. Interestingly, the nature of the 2A region of the picornavirus polyprotein is highly correlated with this phylogeny. Viruses within supergroup 4, the *Paavivirinae*, have complex 2A regions with many viruses encoding multiple 2A^{NPGP} sequences. *In vitro* transcription/translation analyses of a synthetic polyprotein comprising green fluorescent protein (GFP) linked to β-glucuronidase (GUS) *via* individual 2A^{NPGP}s showed two main phenotypes: highly active 2A^{NPGP} sequences - similar to Foot-and-Mouth Disease Virus 2A^{NPGP} and, surprisingly, a novel phenotype of some 2A^{NPGP} sequences which apparently terminate translation at the C-terminus of 2A^{NPGP} - without detectable re-initiation of downstream sequences (GUS). Probing databases with the short sequences between 2A^{NPGP}s did not reveal any potential 'accessory' functions. The novel, highly active, 2A-like sequences we identified substantially expands the toolbox for biomedical / biotechnological co-expression applications.

Keywords: picornaviruses; polyprotein 2A region; 2A^{NPGP} sequences; ribosome skipping; translation; biotechnology

1. Introduction

Picornavirus genomes are single-stranded, positive sense, RNA of some 7.5 to 8.5 kb in length. Their genomes have a common architecture: (i) a long 5′ untranslated region (UTR) comprising an internal ribosome entry site (IRES), conferring a m^{7}G cap-independent mode of the initiation of translation, (ii) a long open reading frame (ORF; ~2,200aa 'polyprotein') and (iii) a short 3′ UTR preceding a poly(A) tail. In many genera the N-terminal region of the polyprotein comprises a 'leader' sequence. The P1 domain of the polyprotein comprises 4 different proteins (1A-1D), 60 copies each of which assemble to form the capsid. Proteins within the P2 (2A, 2B, 2C) and P3 (3A-3D) domains are replication proteins. Early studies showed the 3C protein to be a virus-encoded proteinase (3C^{pro}), responsible for a 'primary' (co-translational) cleavage between the P2 and P3 regions, but also subsequent ('secondary') polyprotein processing: 3C^{pro} being conserved amongst all picornaviruses (for review see [1]).

In the case of entero- and human rhinoviruses, a second virus-encoded proteinase, 2A^{pro} (~17kDa), was identified and shown to be responsible for a single primary cleavage between the capsid proteins (P1) and P2 domains of the polyprotein – and, shown latterly, also to degrade key cellular proteins thereby enhancing virus replication [2-5]. Simple inspection of the relatively modest number of picornavirus genome sequences available at that time - such as entero- and rhinoviruses, Foot-and-Mouth Disease Virus (FMDV; aphthovirus), Encephalomyocarditis Virus (EMCV; cardiovirus) and Hepatitis A Virus (HAV; hepatovirus) showed their 2A regions to be quite different from one another: indeed, the 2A protein of FMDV was only 18aa long. Research into polyprotein

 doi:10.20944/preprints202409.1973.v

processing mechanisms of these other, non-enterovirus, genera showed that the capsid proteins domain was separated from the replication protein domains by either (i) the 3C proteinase cleaving the polyprotein in this region [6, 7], or, (ii) a proposed 'ribosome skipping' mechanism mediated by the 2A oligopeptide sequence proposed to interact with the ribosome exit tunnel, bringing about a discontinuity in the polypeptide backbone at the C-terminus of 2A: not a proteolytic 'cleavage' but by 'skipping' the synthesis of a specific peptide backbone bond [7-9]. A completely conserved motif at the C-terminus of these 2A sequences (-NPG↓P-; position of discontinuity of the peptide backbone shown) gives rise to this type of picornavirus 2A being referred-to in the literature as 2A^{NPGP}. It should be noted, however, that 2A^{NPGP} ('2A-like') sequences are also found in in a wide range of non-picornavirus virus families [10-12] - plus non-LTR retrotransposons [13-15] and some cellular sequences (NOD-like receptor proteins – NLRs) [16, 17]. Our previous work has shown that the 'ribosome skipping' activity resides within a ~25aa tract (the 2A C-terminal delimiter being -NPG↓P) [13, 18]. Cardiovirus 2A proteins are ~143aa long, however, the C-terminal region of the cardiovirus-like 2As shares sequence similarity with the short aphthovirus-like 2As: both possessing the same proposed ribosome-skipping function [19], outlined in Figure 1, Panel A.

Figure 1

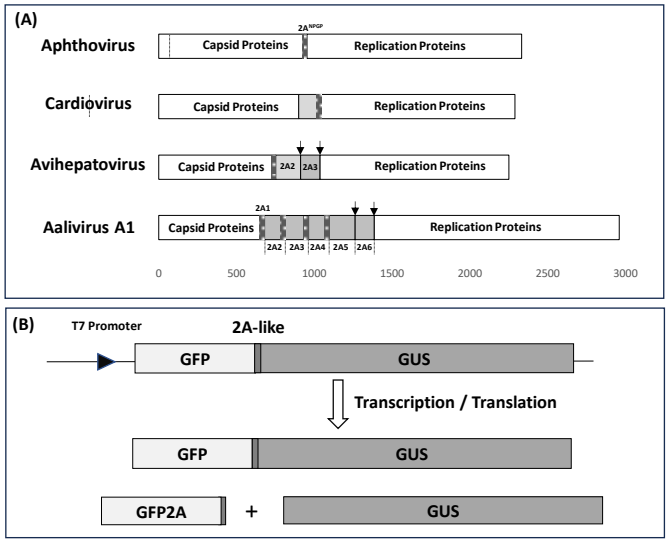


Figure 1. Schematic (drawn to scale) showing the positions of 2A^{NPGP} sequences (25aa) within the genomes of Aphtho-, Cardio-, Avihepato- and Aalivirus polyproteins (Panel A). The [GFP2AGUS] artificial polyprotein (GFP stop codon removed) used to test the various 2A^{NPGP} sequences. In the case of FMDV 2A^{NPGP}, three translation products are observed: 'uncleaved' [GFP2AGUS] together with the 'cleavage' products [GFP2A] and GUS (Panel B).

It has been shown recently, however, that the cardioviruses EMCV and Theilers Murine Encephalitis Virus (TMEV) 2A proteins are bifunctional in that they also stimulate a programmed ribosomal (-1) frameshift by binding an RNA secondary structure formed closely downstream – the first demonstration of protein-stimulated regulation of programmed ribosomal frameshifting. Here,

 doi:10.20944/preprints202409.1973.v

a shift site occurs just 5′ of a stem-loop structure within the region encoding the 2B protein, producing a truncated 2B′ protein - effectively terminating translation prior to the translation of downstream replication proteins – which comprises some 50% of the entire polyprotein [20-22]. In the case of one genus, the dicipiviruses, however, the P1 region is encoded by an ORF separated from the second ORF (encoding the P2-P3 region) by a second, intergenic, IRES that initiates translation of the downstream ORF [23].

Presently, the family *Picornaviridae* comprises almost 70 genera, (Knowles www. picornaviridae. com; [24]). Alignment and phylogenetic analyses of the nucleotide sequences corresponding to the uncleaved form ([3CD]) of $3C^{pro}$ and the 3D RNA-dependent RNA polymerase ($3D^{pol}$) shows five major lineages, or 'supergroups' (SG), within the family: the *Caphthovirinae* (SG1), the *Kodimesavirinae* (SG2), the *Ensavirinae* (SG3), the *Paavivirinae* (SG4), and the *Heptrevirinae* (SG5) [25].

In the case of supergroups 2 (*Kodimesavirinae* - 22 genera) and 5 (*Heptrevirinae* - 7 genera), the separation between capsid and replication proteins is brought about by $3C^{pro}$ cleavage. In the case of supergroup 3 (*Ensavirinae* - 8 genera) this separation is brought about by $2A^{pro}$, a proteinase unique to this supergroup. In supergroup 1 (*Caphthovirinae* – 16 genera) this separation is brought about, in the vast majority cases, by a single copy of $2A^{NPGP}$. In the case of supergroup 4 (*Paavivirinae* – 13 genera), different species encode either a single $2A^{NPGP}$, or, multiple copies of $2A^{NPGP}$s.

In the case of the avihepatoviruses, typified by Duck Hepatitis Virus type 1 (DHV-1), the 2A region comprises three proteins (Figure 1, Panel A); (i) 2A1 (20aa) is highly similar to the short aphthovirus 2A sequence – a highly efficient ribosome skipping sequence separating the capsid from replication proteins domains [26], (ii) 2A2 (161aa) has been shown to possess GTPase activity inducing apoptosis [27] and (iii) 2A3 (124aa) possessing similarity with parechovirus 2A, comprising an H-box/NC motif and related to the host-cell *H-rev* protein family [28], recently shown to promote cell proliferation [29]. In the case of Duck Picornavirus GL/12 (Aalivirus A1) the complexity of the 2A region is increased since it comprises six 2A proteins [30] (Figure 1, Panel A). Here, the 2A1 protein (19aa) is aphthovirus-like, 2A2 is 133aa, 2A3 is 150aa and 2A4 is 131aa: all have C-terminal regions (~25aa) similarity with aphthovirus-like $2A^{NPGP}$ sequences. Protein 2A5 has similarity with the avihepatovirus 2A2 protein, whilst 2A6 has similarity with the avihepatovirus 2A3 (parechovirus-like). Recently, the genome of Duck Egg-Reducing Syndrome Virus (DERSV) has been determined [31]. This virus encodes seven 2A proteins with six, highly conserved, $2A^{NPGP}$s.

Since a single iteration of $2A^{NPGP}$ is sufficient to separate encapsidation from replication proteins (e.g. SG1), the question arises what function(s) do these 'additional' $2A^{NPGP}$- type 2As serve? The polypeptide tracts between successive $2A^{NPGP}$s are between ~50 and ~150aa long and could represent the acquisition of 'accessory' polypeptides. Multiple $2A^{NPGP}$s between the capsid and replication polyprotein domains could affect protein biogenesis: a key question being what is the ribosome skipping activity of each of these sequences? Our previous work showed the activity of 2A/2A-like sequences resides within a ~25aa tract - including the N-terminal proline of the 2B downstream protein. In this study we inserted a 2A/2A-like sequence (25aa, in-frame) between green fluorescent protein (GFP: ~27kDa – stop codon deleted) and β-glucuronidase (GUS: ~70kDa), creating a single, long, ORF [9, 13].

It should be noted that certainly not all 2A-like sequences containing the conserved [D-V/I-E-X-NPGP] motif are active in ribosome skipping: the sequence immediately upstream, although not highly conserved, plays an essential role in 'cleavage'. It is essential, therefore, to perform an assay to determine the activity of each sequence. Here we used an *in vitro* transcription / translation system - the 'skipping' activity of each test sequence was determined by the incorporation of ^{35}S-methionine into each translation product (Figure 1, Panel B). For viruses in supergroup 4, however, amino acid tracts between the $2A^{NPGP}$s are generally short: analysing the 2A region as a single tract could produce a significantly more complex mixture of translation products with either (i) poor resolution on SDS gels, (ii) with low methionine content proving very difficult to detect or (iii) produce translation products too small to be detected. We chose, therefore, to use our GFP/GUS system determine the 'skipping' activity of these $2A^{NPGP}$s individually, rather than in their native, concatenated, forms.

 doi:10.20944/preprints202409.1973.v

2. Materials and Methods

2.1. Bioinformatic Analyses

Picornavirus genome sequences were downloaded from the National Center for Biotechnology Information (NCBI) *via* the links provided on https://www.picornaviridae.com. The uncleaved form of the proteinase and polymerase ([3CD]) protein sequences were aligned using CLUSTALX [32]. Phylogenetic trees were visualised using FIGTREE (http://tree.bio.ed.ac.uk/software/figtree/). Polyprotein sequences were searched for occurrences of the 'signature' -NPGP- motif, completely conserved amongst all 2A-like sequences, to identify 2A^{NPGP}s. Arbitrarily, 25aa tracts were chosen for further analyses (Table 1). Nucleotide and amino acid sequences of tracts between picornavirus 2A^{NPGP}s were submitted to NCBI BlastN / BlastP (https://blast.ncbi.nlm.nih.gov/Blast.cgi) for database similarity searches.

Table 1. 2A^{NPGP} amino acid sequences and accession codes.

Genus	Species	2A^{NPGP}	Amino Acid Sequence	Acc.#
Aphthovirus	FMDV O1K	2A	VAPVKQTLNFDLLKLAGDVESNPGP	GNNYF
Aalivirus (SG4)	AalV- A1	2A1	LLTSEGATNSSLLKLAGDVEENPGP	KJ000696
		2A2	FEMPYDDPEWDRLLQAGDIEQNPGP	
		2A3	PIPARPDPQWNNLQQAGDVEMNPGP	
		2A4	EHFNQTGGWVPDLTQCGDVESNPGP	
	AalV-B1	2A1	ATTLQVSEYLKDLTIDGDVESNPGP	MH453803
		2A2	LKVKKLEGDYVRDLTQEGVEPNPGP	
		2A3	SVRVTDAGWVRDLTVDGDVESNPGP	
		2A4	VFKCHDKCWVDDLTNCGDVEPNPGP	
		2A5	IFKCHEGCWVEDLTVDGDVESNPGP	
DERSV-AH204		2A1	TSTAQATSYVKDLTIDGDVESNPGP	UYL81882
		2A2	KTCREVEGSYVKDLTEEGIEPNPGP	
		2A3	LLKIGNAAWVRDLTEDGDVEENPGP	
		2A4	VYNCHESCWNRDLTIDGDVELNPGP	
		2A5	VFKCHEKCWQKDPTQDGDVEQNPGP	
		2A6	EFKCHEHCWVRDLTMDGDVEENPGP	
Avisivirus (SG4)	AsV-A1	2A1	EVGAYDEVDHRDILMGGDIEENPGP	KC465954
		2A2	EMGVFDETDHRDILLGGDIEENPGP	
	AsV-B1	2A1	PQFEKERSAHEDVLLGGDVESNPGP	KF979333
		2A2	SESVQYLEPQIDICVCGDVERNPGP	
Grusopivirus (SG4)	GrV-A1	2A1	FEKHVKPWRSQEDLSKEGIEPNPGP	KY312544
		2A2	ITDNRYKETDAKWLSRYGVEMNPGP	
		2A3	VTQDLYAATNQDQLSNQGIESNPGP	
	GrV-C	2A1	YFEERSPHPTQKELGQFGVETNPGP	MK443503
		2A2	ENNSNYSERDAKHLSRYGIEMNPGP	
		2A3	CVCTRWSPTMQSELGKYGIEKNPGP	
	YC-4	2A1	PERQYFSPKAKEELSKYGIEPNPGP	KY312543
Kunsagivirus (SG4)	Kuv-C1	2A1	IAAASAQGWQRDLTQDGDVESNPGP	KY670597
		2A2	LGIVISDSVWQRDLPREGVEENPGP	
		2A3	SYDPLAPSQWCRDLTCEGIEPNPGP	
Limnipivirus (SG4)	A1	2A1	CKEFVRESDNQELLKCGDVESNPGP	JX134222
		2A2	WDLSTGWFHFFRLLRSGDVEQNPGP	
	B1	2A1	MDVVDDYPFKRDLTRDGDVESNPGP	KF306267
		2A2	IDLVQAAYSRMRLLLSGDVEQNPGP	
	C1	2A1	KLLEQILAYKRDLTACGDVESNPGP	KF874490

 doi:10.20944/preprints202409.1973.v

		2A2	SRWIHARFARLRLLLSGDVEQNPGP	
	D1	2A1	EEEVDWGVGRMRLKMSGDVEENPGP	MG600094
		2A2	AVHLLVTWMRRRLTLSGDIESNPGP	
		2A3	DLRAVKSFIESQLMRAGDVERNPGP	
Mosavirus (SG1)	B1	2A1	ESRGTGNCDATTISQCGDVETNPGP	KY855435
		2A2	YVRRSANRTAADISQDGDVETNPGP	
Parechovirus (SG4)	E	2A1	WFDARTGFKTPLMNPCGDVEENPGP	KY645497
		2A2	QIEKRYGYRFWLLMLCGDVELNPGP	
	RtPV	2A1	MLDRRMGYRSRILCQCGDVEENPGP	MF352429
		2A2	WFNKRSGYRSRLLSQCGDVEENPGP	
Potamipivirus (SG4)	B1	2A1	LMEKTEEAGWLRDLTREGVEENPGP	MK189163
		2A2	FDDYHQEGGWIRDLTAEGVEPNPGP	
Unassigned (SG4)	WCP	2A1	MKEDEAGGWKEDLTEDGDVESNPGP	MG600066
		2A2	EQAIPETTWRRDLTQSGDVESNPGP	
		2A3	PGAIPASVWVHDLTTDGDVESNPGP	
Unassigned	WP-LV 48	2A1	GPSCYDRNNHCNILLSGDIEENPGP	NC_032820
		2A2	VFNASYLDCFISLLSCGDIESNPGP	
		2A3	PIQGLTQRFESTLLLGGDIEENPGP	

2.2. Cloning of 2A^{NPGP} Sequences into pSTA1

All plasmids were constructed using standard methods and confirmed by automated nucleotide sequencing (Eurofins Genomics, Ebersberg, Germany). Restriction enzymes were purchased from Promega (Southampton, UK) and New England Biolabs (Hitchin, UK), whilst oligonucleotides were obtained from, and automated DNA sequencing by, Eurofins Genomics (Ebersberg, Germany). PCR products encoding each 2A-like sequence were amplified from our pcDNA™ 3.1 mammalian expression vector encoding ([pGFPF2AGUS]; pSTA1) [9] using the T7 "forward" primer (5′-TAATACGACTCACTATAGGG-3′) and "reverse" oligonucleotide primers listed in Table 2, such that a panel of GFP/2A-like PCR products was generated. Each PCR product was restricted with *BamHI* and *ApaI*, gel purified, and cloned into pSTA1, similarly restricted. Sequence identities were confirmed by automated DNA sequencing using an oligonucleotide primer corresponding to that encoding a C-terminal region of GFP (5′-CTGTCCACACAATCTGCCC-3′).

Table 2. Reverse oligonucleotide primers used to amplify GFP thereby adding each 2A^{NPGP} 3′ extension: restriction sites used in cloning are indicated in bold typeface.

Genus	Species	2A^{NPGP}	Reverse Primer Sequence (5′-3′)
Aalivirus	AalV- A1	2A1	GCGCGC**GGGCCC**TGGATTCTCTTCCACATCTCCAGCTAACTTTAACAGAC TTGAATTTGTGGCTCCCTCTGATGTGAGCAA**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**AGGATTCTGTTCTATGTCTCCAGCCTGGAGCAGCCTGT CCCATTCTGGGTCATCATATGGCATTTCGAA**TCTAGA**CCCGGACTTGTAT
		2A3	GCGCGC**GGGCCC**TGGATTCATCTCAACATCACCAGCTTGCTGCAAATTAT TCCATTGTGGGTCAGGCCTGGCTGGAATTGG**TCTAGA**CCCGGACTTGTAT
		2A4	GCGCGC**GGGCCC**GGGATTGGACTCTACATCACCACACTGCGTCAGATCGG GGACCCATCCCCCTGTCTGGTTGAAGTGCTC**TCTAGA**CCCGGACTTGTAT
	AalV-B1	2A1	GCGCGC**GGGCCC**AGGATTTGATTCAACATCTCCGTCAATGGTTAAATCTT TCAGATACTCAGACACTTGCAAAGTAGTTGC**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**AGGGTTAGGTTCCACACCCTCTTGAGTTAAATCTCTAA CATAATCTCCCTCAAGTTTCTTAACTTTCAA**TCTAGA**CCCGGACTTGTAT
		2A3	GCGCGC**GGGCCC**AGGGTTTGATTCCACATCTCCATCAACTGTGAGGTCTC TCACCCACCCAGCATCTGTTACTCTAACCGA**TCTAGA**CCCGGACTTGTAT
		2A4	GCGCGC**GGGCCC**TGGATTTGGCTCAACATCCCCACAATTCGTCAGGTCGT CAACCCAACATTTATCGTGGCACTTAAAAAC**TCTAGA**CCCGGACTTGTAT
		2A5	GCGCGC**GGGCCC**AGGGTTCGACTCCACATCACCATCAACAGTTAGATCCT CAACCCAACAGCCCTCATGACACTTAAAAAT**TCTAGA**CCCGGACTTGTAT

 doi:10.20944/preprints202409.1973.v

Avisivirus	AsV-A1	2A1	GCGCGC**GGGCCC**AGGGTTTTCTTCAATGTCACCCCCCATGAGAATGTCTC TGTGGTCCACTTCATCATAAGCTCCAACTTC**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**TGGATTCTCTTCAATGTCACCTCCAAGTAGTATGTCTC TGTGGTCAGTCTCATCAAAGACTCCCATCTC**TCTAGA**CCCGGACTTGTAT
	AsV-B1	2A1	GCGCGC**GGGCCC**AGGGTTTGATTCTACATCTCCACCTAGCAGAACATCCT CATGGGCTGAGCGCTCCTTTTCAAACTGTGG**TCTAGAC**CCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**GGGATTCCTCTCTACATCACCACAAACACAGATATCAA TCTGGGGCTCCAAATATTGAACAGACTCACT**TCTAGA**CCCGGACTTGTAT
Grusopivirus	GrV-A1	2A1	**GGGCCC**AGGGTTTGGTTCAATTCCCTCCTTAGATAGATCTTCTTGTGATT TCCAAGGTTTCCCATGTTTTTCAAA**TCTAGA**CCCGGACTTGTATAGTTC
		2A2	**GGGCCC**TGGGTTCATTTCCACTCCATATCGGCTCAACCATTTAGCGTCGG TTTCCTTATAACGATTGTCCGTAAT**TCTAGA**CCCGGACTTGTATAGTTC
		2A3	**GGGCCC**AGGATTTGATTCAATGCCTTGATTTGATAACTGATCTTGATTAG TAGCAGCATAAAGATCCTGAGTGAC**TCTAGA**CCCGGACTTGTATAGTTC
	GrV-C	2A1	GCGCGC**GGGCCC**AGGATTAGTTTCTACTCCAAATTGCCCCAATTCCTTCT GAGTTGGATGTGGAGATCTTTCTTCAAAATA**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**AGGATTCATCTCTATGCCATATCGTGATAAGTGTTTGG CATCTCTCTCAGAATAATTTGAGTTGTTCTC**TCTAGA**CCCGGACTTGTAT
		2A3	GCGCGC**GGGCCC**TGGATTCTTCTCAATTCCATACTTACCTAATTCAGACT GCATGGTGGGACTCCACCTAGTGCAAACACA**TCTAGA**CCCGGACTTGTAT
	YC-4	2A1	GCGCGC**GGGCCC**AGGATTAGGCTCGATACCATATTTAGACAGTTCTTCCT TCGCCTTTGGAGAGAAATATTGACGTTCTGG**TCTAGA**CCCGGACTTGTAT
Kunsagivirus	Kuv-C	2A1	GCGCGC**GGGCCC**AGGATTGCTCTCAACATCACCATCTTGAGTAAGGTCTC TTTGCCAGCCCTGTGCACTAGCCGCGGCAAT**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**TGGATTTTCCTCAACACCTTCGCGGGGTAGATCCCGCT GCCACACAGAGTCGGAGATGACAATACCTAA**TCTAGA**CCCGGACTTGTAT
		2A3	GCGCGC**GGGCCC**TGGATTAGGCTCGATACCCTCACAAGTCAAATCCCTAC ACCACTGGCTGGGGGCCAGAGGGTCGTAGCT**TCTAGA**CCCGGACTTGTAT
Limnipivirus	A1	2A1	GCGCGC**GGGCCC**TGGGTTAGACTCCACATCTCCACACTTGAGTAGCTCCT GGTTGTCTGATTCTCTTACAAATTCTTTGCA**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**AGGGTTCTGTTCAACATCTCCGCTCCTCAACAACCGGA AAAAGTGAAACCATCCTGTTGAAAGGTCCCA**TCTAGA**CCCGGACTTGTAT
	B1	2A1	GCGCGC**GGGCCC**TGGGTTGCTCTCAACATCTCCATCACGTGTTAAGTCAC GTTTGAAAGGGTAATCATCAACGACATCCAT**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**GGGATTTTGTTCAACGTCCCCCGATAACAACAACCTCA TGCGTGAGTAGGCAGCTTGCACCAAGTCGAT**TCTAGA**CCCGGACTTGTAT
	C1	2A1	GCGCGC**GGGCCC**AGGGTTGGACTCCACATCGCCACAAGCAGTCAAATCT C GCTTGTATGCCAGAATTTGTTCAAGCAGTTT**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**AGGGTTGGACTCCACATCGCCACAAGCAGTCAAATCT C GCTTGTATGCCAGAATTTGTTCAAGCAGTTT**TCTAGA**CCCGGACTTGTAT
	D1	2A1	GCGCGC**GGGCCC**TGGGTTCTCCTCAACATCACCAGACATCTTCAGCCGCA TCCTGCCCACGCCCCAGTCGACTTCCTCCTC**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**TGGGTTGGATTCAATGTCTCCAGAAAGCGTCAATCGTC TGCGCATCCAAGTAACCAGTAAATGAACAGC**TCTAGA**CCCGGACTTGTAT
		2A3	GCGCGC**GGGCCC**TGGGTTTCTCTCCACGTCACCAGCGCGCATCAATTGAC TTTCAATGAATGACTTCACTGCTCTTAAATC**TCTAGA**CCCGGACTTGTAT
Mosavirus	B1	2A1	GCGCGC**GGGCCC**AGGATTGGTTTCAACATCCCCGCACTGACTGATAGTC G TCGCATCACAGTTTCCTGTGCCACGAGATTC**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**GGGGTTGGTCTCAACATCTCCATCCTGACTGATATCA G CGGCAGTACGGTTTGCGGACCGCCTGACGTA**TCTAGA**CCCGGACTTGTAT
Parechovirus	E	2A1	GCGCGC**GGGCCC**TGGGTTTTCTTCCACATCACCACAGGGGTTCATTAGGG GTGTTTTAAACCCCGTGCGTGCATCAAACCA**TCTAGA**CCCGGACTTGTAT

 doi:10.20944/preprints202409.1973.v

		2A2	GCGCGC**GGGCCC**AGGATTTAACTCAACATCTCCACAGAGCATTAGCAAC C AGAAACGATAGCCATATCGCTTCTCTATCTG**TCTAGA**CCCGGACTTGTAT
	RtPV	2A1	GCGCGC**GGGCCC**TGGATTTTCCTCGACATCTCCACATTGACAGAGGATTC TGCTCCGATAGCCCATTCTCCTGTCAAGCAT**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**AGGATTTTCTTCTACATCACCACATTGAGACAACAAT C TTGACCTGTATCCTGATCTTTTGTTGAACCA**TCTAGA**CCCGGACTTGTAT
Potamipi virus	A	2A1	GCGCGC**GGGCCC**TGGATTTGGTTCAACTCCTTCTTGTGTCAGATCTCTGA TCCACATCTTGTCCGTGAGTATTTCGCCAAT**TCTAGA**CCCGGACTTGTAT
	B	2A1	GCGCGC**GGGCC**CGGGGTTCTCCTCAACTCCCTCTCTTGTCAAATCTCTTA GCCATCCTGCTTCTTCTGTTTTCTCCATCAA**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**CGGGTTGGGCTCCACACCCTCAGCAGTGAGGTCCCGT A TCCAACCACCTTCCTGGTGGTAATCATCAAA**TCTAGA**CCCGGACTTGTAT
Unassig ned	WCP	2A1	CGCGCG**GGGCCC**AGGGTTACTCTCCACATCACCGTCCTCAGTGAGGTCTT CTTTCCACCCACCAGCTTCATCCTCCTTCAT**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**TGGATTGGATTCCACATCACCAGATTGTGTGAGATCT C GACGCCATGTGGTTTCAGGAATTGCTTGCTC**TCTAGA**CCCGGACTTGTAT
		2A3	GCGCGC**GGGCCC**AGGATTGGATTCAACATCACCATCTGTTGTGAGGTCA T GAACCCAGACACTTGCTGGTATGGCACCAGG**TCTAGA**CCCGGACTTGTA T
Unassig ned	WP-LV 48	2A1	GCGCGC**GGGCCC**TGGATTCTCTTCAATATCTCCTGAAAGTAAGATGTTGC AATGATTATTCCTGTCGTAGCAAGATGGACC**TCTAGA**CCCGGACTTGTAT
		2A2	GCGCGC**GGGCCC**TGGATTTGACTCGATATCCCCACAAGATAATAAGCTG A TGAAACAATCTAAATAACTGGCATTAAAAAC**TCTAGA**CCCGGACTTGTA T
		2A3	GCGCGC**GGGCCC**TGGATTTTCTTCAATATCGCCCCCCAAAAGAAGAGTT G ACTCAAAACGTTGTGTAAGACCTTGTATTGG**TCTAGA**CCCGGACTTGTAT

2.3. In Vitro Transcription/Translation

Plasmid DNA was linearised with *PstI* and purified using the Wizard SV system (Promega). Purified, linearised, DNA (200ng) was used to program a Wheat Germ Extract coupled transcription/translation system (Promega) supplemented with L-[^{35}S]-methionine (EasyTag™ Perkin Elmer; 1µl = 10µCi) and the amino acid mixture (minus Met). The final reaction volume was adjusted to 25µl using nuclease-free water. Reactions were incubated at 30^{o}C for 90 minutes before the addition of 2x SDS-PAGE loading buffer (Jena Bioscience). Translation products were analysed by 4-12 % gradient SDS-PAGE (NuPAGE, Invitrogen) and the distribution of the radiolabel visualised by autoradiography.

3. Results

3.1. Bioinformatic Analyses

NOTE: Alignments and virus polyprotein sequences showing the positions of 2A^{NPGP}s discussed below are shown in the Supplementary Data. Our alignment of [3CD] amino acid sequences produced the same phylogenetic relationships as the nucleotide sequence alignments reported by Zell and co-workers [25]. Our phylogenetic analyses show that picornaviruses encoding 2A^{NPGP}s fall into two, distinct, supergroups – the *Caphovirinae* (SG1) and the *Paavivirinae* (SG4): viruses within the latter encoding either a single or multiple 2A^{NPGP}s: indeed, sometimes this being the case for different species within the same genus (Table 1). Interestingly, there is a high correlation between each super-group and the nature of the 2A region - the most plastic amongst picornavirus polyproteins (Figure

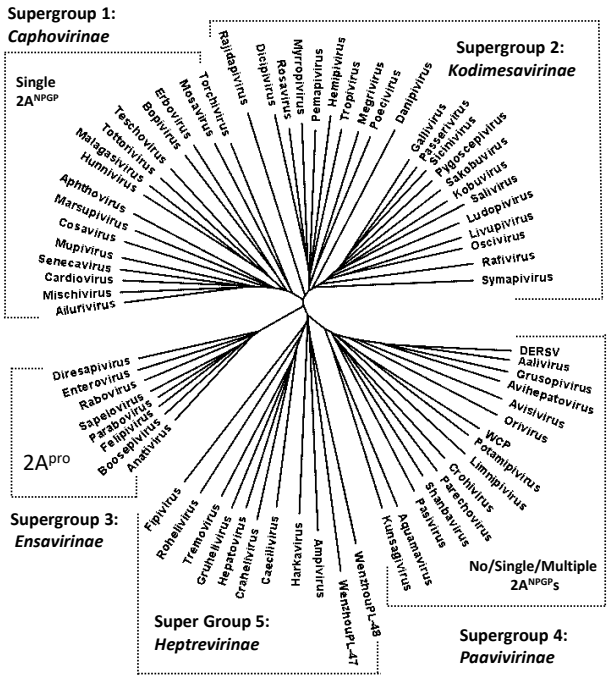
 doi:10.20944/preprints202409.1973.v

2). In the case of supergroup 1 (*Caphthovirinae* – 16 genera) viruses encode a single copy of 2A^{NPGP}. The first exception here is Mosavirus B1 which encodes two 2A^{NPGP}s. In this virus, the first copy of 2A^{NPGP} appears to occur within the P1 region, whilst the second copy aligns with the single copy of Mosavirus A 2A^{NPGP} – which lies between the P1 and P2 polyprotein domains. The second exception is Mupivirus (both A1 and A2), neither of which encode a 2A^{NPGP}.

In the case of supergroup 4 (*Paavivirinae* – 13 genera), variability is observed between different species within a genus. Here, a genus may comprise viruses which encode either: (i) no 2A^{NPGP} (Orivirus A1 and A2: alignment with Avisiviruses shows a relative deletion in this region, see Supplementary data), (ii) a single 2A^{NPGP} (Aquamavirus, Avihepatovirus, Crohivirus, Pasivirus, Shanbavirus), (iii) multiple copies of 2A^{NPGP}s - Aalivirus, Limnipivirus, Wuhan Carp Picornavirus (WCP - unassigned, but clusters within SG4) NOTE: DERSV is most closely related to Aaliviruses and clusters within the *Paavivirinae* (Figure 2), (iv) genera comprising viruses encoding either a single or multiple 2A^{NPGP}s (Avisivirus, Kunsagivirus, Potamipivirus), or, (v) genera comprising viruses encoding either none, or multiple 2A^{NPGP}s, (Grusopivirus, Parechovirus: see Supplementary data for alignments).

Wenzhou picorna-like virus 48 (WPLV-48; NC_032820) encodes three 2A^{NPGP} sequences, but is unique in that 2A1 appears to lie between the P1 capsid and P2 replication protein domains, whilst 2A2 and 2A3 occur in the C-terminal region of the polyprotein – downstream of the WPL-48 sequences that align with the C-terminal 3CD region of all other picornaviruses (see Supplementary Data). In our alignment / phylogenetic analyses this region was deleted. Although unassigned, WPLV-48 clustered within SG5 (Figure 2), the *Heptrevirinae,* which do not encode any 2A^{NPGP}s (Figure 2). Interestingly, our 3CD alignment shows Wenzhou picorna-like virus 47 (WPLV-47; NC_033150) also clusters within SG5, although does not encode a 2A-like sequence. Like many other viruses in this supergroup, WPLV-47 diverged at an early stage in the evolution of this supergroup.

Figure 2



 doi:10.20944/preprints202409.1973.v

Figure 2. Dendrogram of aligned 3CD amino acid sequences rendered using FigTree. The supergroups indicated are consistent with that of Zell and co-workers [25]. Viruses encoding a 2A proteinase only within supergroup 3, whilst 2A^{NPGP} sequences are found only within supergroups 1 and 4.

For viruses with multiple 2A^{NPGP}s, peptide sequences lying between each 2A^{NPGP} were used to probe the sequence database: other than similarities with related viruses, no significant matches were detected.

3.2'. Cleavage' Activities of 2A-like Sequences

For all cases, *in vitro* translation profiles are shown in Figure 3 with estimated 'cleavage' activities shown in Table 3.

Figure 3

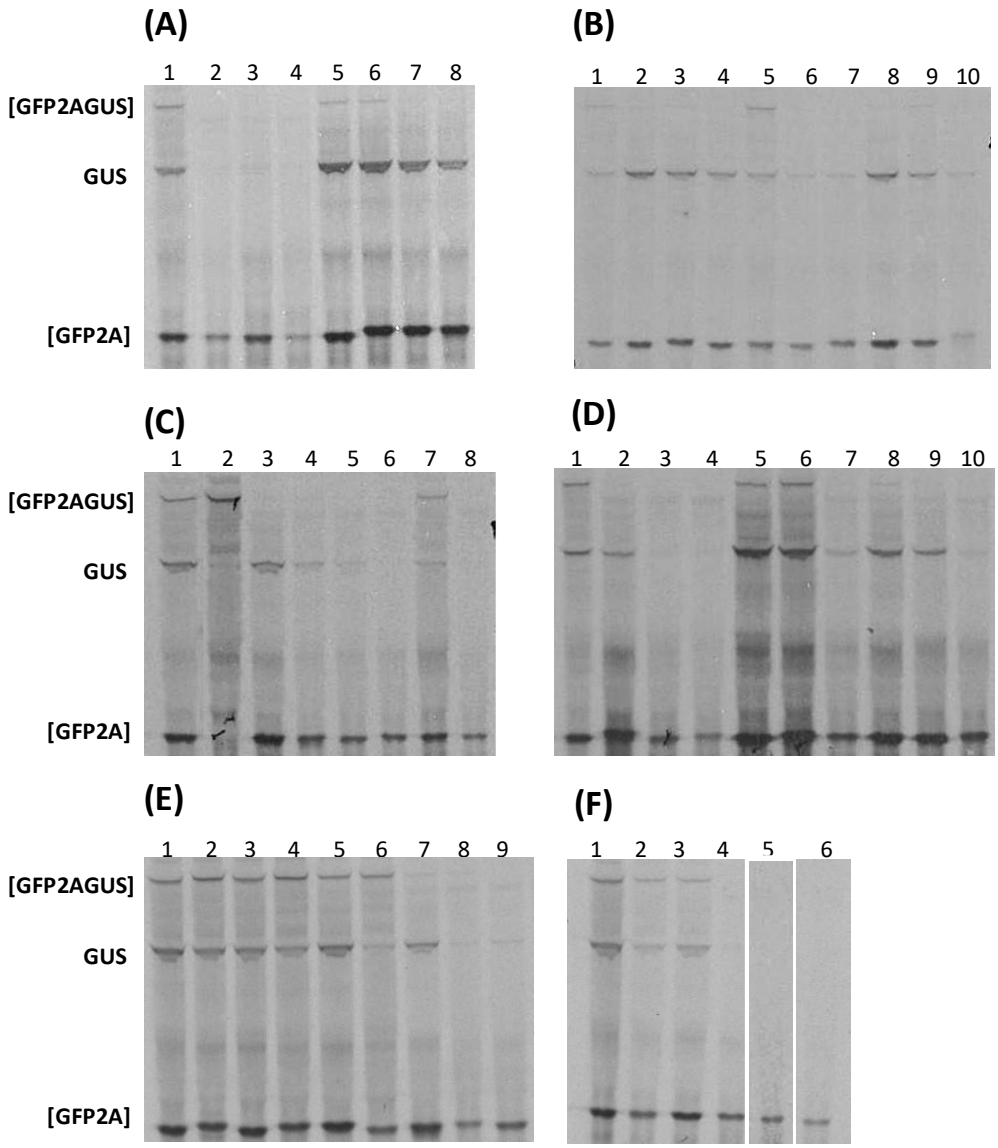


Figure 3. Translation *in vitro*. Coupled transcription/translation wheat germ extracts were programmed with the plasmid constructs indicated. Translation products were labelled with [^{35}S] methionine, separated on 4-12% gradient SDS polyacrylamide gels, and detected by autoradiography. Lane 1 panels A-F, control construct FMDV 25 encoding [GFP-2A(25aa)-GUS]; positions of the

 doi:10.20944/preprints202409.1973.v

uncleaved [GFP-2A-GUS] and cleavage products [GFP-2A] and GUS are indicated. (A) Lanes 2-4, WP-LV48: 2A1-2A3; lanes 5-8, Aalivirus-A1: 2A1-2A4. (B) Lanes 2-6, Aalivirus-B1: 2A1-2A5; lanes 7 and 8, Avisivirus-A1: 2A1/2A2; lanes 9 and 10, Avisivirus-B1: 2A1/2A2. (C) Lanes 2 and 3, Parechovirus E: 2A1/2A2; lanes 4 and 5, RtPV: 2A1/2A2; lane 6, Potamipivirus A1: 2A1; lanes 7 and 8, Potamipivirus B1: 2A1/2A2. (D) Lanes 2 and 3, Limnipivirus A1: 2A2 and 2A2; lanes 4 and 5, Limnipivirus B1: 2A1 and 2A2; lanes 6 and 7, Limnipivirus C1: 2A1 and 2A2; lanes 8-10, Limnipivirus D1: 2A1, 2A2 and 2A3. (E) Lanes 2-4, Grusopivirus-A1: 2A1, 2A2 and 2A3; lanes 5-7, Grusopivirus-C: 2A1-2A3; lane 8, YC-4; lane 9, Kungsagivirus-C1: 2A1. (F) Lanes 2 and 3, Mosavirus B1: 2A1/2A2; lanes 4-6, WCP: 2A1-2A3.

Table 3. Estimated 'cleavage' activities of 2A-like sequences.

Genus	Species	2A^{NPGP}	[GFP2AGUS]	[GFP2A]	GUS
Aphthovirus	FMDV O1K	2A	+	++++	++
Aalivirus (SG4)	AalV- A1	2A1	+	+++++	+++
		2A2	+	+++++	+++
		2A3	-	+++++	+++
		2A4	-	+++++	+++
	AalV-B1	2A1	-	++++	++
		2A2	-	++++	++
		2A3	-	++++	++
		2A4	+	++++	++
		2A5	-	++++	+
Unassigned	DERSV	2A1-6	**ND**	**ND**	**ND**
Avisivirus (SG4)	AsV-A1	2A1	-	++++	+
		2A2	-	++++	++
	AsV-B1	2A1	-	++++	++
		2A2	-	++	+
Grusopivirus (SG4)	GrV-A1	2A1	+	++++	++
		2A2	+	++++	++
		2A3	+	++++	++
	GrV-C	2A1	+	++++	++
		2A2	+	++++	(+)
		2A3	(+)	++++	++
	YC-4	2A1	-	++++	(+)
Kunsagivirus (SG4)	Kuv-C1	2A1	-	++++	-
		2A2	**ND**	**ND**	**ND**
		2A3	**ND**	**ND**	**ND**
Limnipivirus (SG4)	A1	2A1	-	++++	++
		2A2	-	++++	-
	B1	2A1	-	+++	-
		2A2	++	+++	++++
	C1	2A1	++	++++	++++
		2A2	-	++++	+
	D1	2A1	-	++++	++
		2A2	-	++++	++
		2A3	-	++++	+
Mosavirus (SG1)	B1	2A1	+	++++	++
		2A2	+	++++	++
Parechovirus (SG4)	E	2A1	+	-	-
		2A2	-	++++	+
	RtPV	2A1	-	++++	+

 doi:10.20944/preprints202409.1973.v

		2A2	-	++++	+
Potamipivirus (SG4)	B1	2A1	(+)	++++	+
		2A2	-	+++	-
Unassigned (SG4?)	WCP	2A1	-	+++	-
		2A2	-	+++	-
		2A3	-	+++	-
Unassigned (SG5?)	WP-LV 48	2A1	-	+++	-
		2A2	-	++++	-
		2A3	-	+++	-

Alignments showing relative insertions/deletions in polyproteins within a genus - plus highlights of the position of 2A^{NPGP}s are provided in the Supplementary Data.

In the single instance of a virus clustering within SG1, Mosavirus B1 - encoding two 2A^{NPGP}s, the activity of both sequences was directly comparable to FMDV 2A. The first Mosavirus B1 2A^{NPGP} occurs within P1, whilst the second aligns with the single 2A^{NPGP} present within other Mosavirus sequences and corresponds to the junction between P1 and P2 domains of the polyprotein.

In the case of viruses encoding 2A^{NPGP}s within SG4, however, a more complex pattern of activity was observed. Aalivirus A1/B1 2A^{NPGP}s show activities comparable to FMDV 2A, although Aalivirus B1 sequences yielded slightly less of the GUS translation product. For both Aalivirus A1 and B1 2A^{NPGP}s, 2A1-2A4 (A1) and 2A1-2A5 (B1) produced similar translation profiles. Similarly, in the related Grusopivirues different viruses encode different numbers of 2A^{NPGP}s. Grusopiviruses encode a sequence resembling 2A within the (predicted) P1 capsid protein domain of the polyprotein, although point mutations within the conserved C-terminal consensus of active 2A/2A-like sequences ([D(V/I)ExNPGP]) strongly suggest these are inactive. Alignment of Grusopivirus polyproteins shows relative insertion/deletions within the 2A region is the cause of variable numbers of 2A^{NPGP}s (see supplementary data). 2A1-2A3 of Grusopivirus A1 and C viruses produced a similar translation profile as FMDV 2A, although GrV-C 2A2 with lower translation of the downstream protein, GUS. Again, relative insertions/deletions within the 2A region caused variable numbers of 2A^{NPGP}s – notably virus YC-4, which only encodes one 2A^{NPGP} (see supplementary data). In the case of Avihepatoviruses, related to Aali- and Grusopiviruses, all viruses encode a single 2A^{NPGP}. Avisivirus A1 and B1 viruses encode two 2A^{NPGP}s: both active but with no detectable 'uncleaved' [GFP2AGUS], and somewhat lower levels of GUS in comparison with FMDV 2A. Avisivirus A1 strain USA-IN1 and Avisivirus C1, however, encode a single 2A^{NPGP} whilst Oriviruses do not encode a 2A^{NPGP} - again due to a relative insertions/deletions (see supplementary data). Wuhan Carp Picornavirus (WCP) encodes three 2A^{NPGP}s: the translation profile produced by each sequence showed the [GFP2A] translation product but, surprisingly, no detectable 'uncleaved' [GFP2AGUS] or downstream GUS translation product. A similar translation profile was observed for the two Potamipivirus B1 2A^{NPGP}s, high [GFP2A] with low/no GUS translation, whilst Potamipivirus A1 encodes a single 2A^{NPGP} - due to a relative insertions/deletions. The 2A^{NPGP}s encoded by members of the Limnipivirus genus showed a perhaps surprisingly wide range of translation profiles: whilst Limnipivirus A1 2A1, type B1 2A2, type C1 2A1 and type D1 2A1-3 gave essentially the same translation profile as FMDV 2A, type A1 2A2 and type B1 2A1 produced the upstream translation product [GFP2A], but the downstream product, GUS, was not detected. Both Crohivirus A and B encode a single 2A^{NPGP}, whilst in the Parechoviruses relative insertions/deletions in this region has produced a variable number of 2A^{NPGP}s: type A has none, types B, C, D and F have one, whilst types E and RtPV have two. Type E 2A1 only produced a full-length ([GFP2AGUS)] translation product, but type E 2A2 and RtPV 2A1/2A2 produced [GFP2A] as the major translation product with a low level of GUS. Shanba-, Pasi-, Aquama- and Kunsagiviruses encode a single 2A^{NPGP}, with the exception of Kunsagivirus C, which encodes three 2A^{NPGP}s – again arising from relative insertions/deletions in this genus. Kunsagivirus C 2A3 aligns with the single 2A^{NPGP} within types A and B.

Analyses of the ribosome skipping activities showed a mixed pattern (Table 3). Certain genera showed activities similar to the control FMDV 2A^{NPGP}: Aaliviruses, Avisiviruses, Grusopiviruses,

 doi:10.20944/preprints202409.1973.v

Mosavirus B1 and Parechovirus 2A2. In a number of cases only [GFP2A] could be detected: Kunsagivirus C1 2A1, Limnipivirus A1 2A2, Limnipivirus B1 2A1, Potamivirus B1 2A2, WCP 2A1/2A2/2A3 and WP-LV48 2A1/2A2 and 2A3. In the case of Parechovirus E, 2A1 was inactive whilst Parechovirus E 2A2, along with RtPV 2A1 and 2A2, was similar to FMDV 2A.

4. Discussion

4.1.

Our phylogenetic analyses were based upon alignment of 3CD amino-acid sequences and showed a surprisingly high, but not complete, correlation with the nature of the 2A region of the polyprotein. It should be noted, however, that in the interpretation of these data picornavirus mixed infections may show high levels of recombination (reviewed in [33]). Whilst SG1 comprises almost exclusively mammalian viruses, SG4 comprises 'sub-lineages' of avian (DERSV, Aali-, Avihepato-, Avisi-, Grusopi-, Kunsagi-, Ori-), fish (WCP, Limnipi-, Potamipi-) and mammalian (Crohi-, Parecho-, Shanba-, Pasi-, Aquama-, Kunsagi-) viruses. The picornavirus 2A region is highly plastic; viruses within SG4 may encode no 2A^{NPGP}s (e.g. Oriviruses), a single 2A^{NPGP} (e.g. Aquama- and Pasiviruses), or, multiple 2A^{NPGP}s. Indeed, viruses within the same genus may encode different numbers of 2A^{NPGP}s (e.g. Aali-, Avisi-, Kunsagi-, Limnipi-, Parecho- and Potamipiviruses). SG4 polyprotein alignments show relative insertions/deletions within the 2A region determines the variable numbers of 2A^{NPGP}s (see Supplementary Data).

In the case of SG1 viruses, the single 2A^{NPGP} sequence is responsible for a highly efficient 'primary', co-translational, cleavage between capsid and replication proteins. Our *in vitro* transcription/translation analyses showed many 2A^{NPGP} sequences in SG4 to be as active as the FMDV 2A control: Aalivirus A1/B1 2A1-2A4, Grusopivirus A1/C 2A1-2A3, Limnipivirus A1 2A1, B1 2A2, C1 2A1 and D1 2A1-3, and Mosavirus B1 2A1, 2A2. We have shown that somewhat less efficient 2A^{NPGP} sequences can also be used for other virus genomes to acquire additional functional 'modules'. For example, different RNA segments/proteins of certain rota- and totiviruses have acquired a [2A-like/dsRNA-binding protein] 'module' [10, 34, 35]. In the case of SG4 viruses with multiple iterations of 2A^{NPGP}, could sequences between 2A^{NPGP}s comprise 'accessory' protein(s)? Protein BLAST was used to probe databases using such sequences. These analyses only produced, however, significant matches against corresponding regions of closely related viruses, providing no indications of accessory functions. As outlined above, it has been shown that cardiovirus 2A proteins, ~143aa long, alongside the ribosome skipping activity, mediate programmed ribosomal frame-shifting at a proximal site within protein 2B, such that the frame-shifted ribosome quickly encounters a stop codon. Therefore, as the infectious cycle progresses and the level of the 2A protein increases, the expression of (downstream) replication proteins is progressively diminished. Remaining aminoacyl-tRNAs can be devoted, therefore, to the production of capsid proteins thereby increasing the yield of particles.

4.2. Implications for our Model of Ribosome Skipping

In our proposed model of 2A^{NPGP} ribosome skipping, the interaction between 2A^{NPGP} and the ribosome exit tunnel leads to the C-terminal portion of 2A^{NPGP} adopting a conformation within the P-site of the peptidyl transferase centre (PTC) such that tRNA^{Pro} in the A-site cannot form a peptide bond since the imino group is sterically restrained being within a ring structure: the phi angle is constrained. We proposed that the tRNA^{Pro} exits the A-site allowing release factors 1 and 3 (eRF1/3) to enter this site to hydrolyse the bond between the nascent polypeptide and tRNA^{Gly} [7-9]. eRF1 exit from the A-site is promoted by eRF3 [36]. For the 'pseudo-reinitiation' necessary to allow the elongation cycle to continue: (i) the vacant A-site must be re-occupied by tRNA^{Pro} and (ii) the tRNA^{Pro} must be translocated from the A- to P-site by elongation factor 2 (eEF2), thereby allowing (iii) the next aminoacyl-tRNA to enter the A-site to recommence polypeptide elongation – in this case the downstream replication proteins. Should any of these steps be inhibited translation would be terminated at this point without 'pseudo-reinitiation' – effectively the phenotype we observed using

 doi:10.20944/preprints202409.1973.v

our reporter system in the case of the Kunsagivirus C1 2A1, Limnipivirus A1 2A2, Limnipivirus B1 2A1, Potamivirus B1 2A2, WCP 2A1/2A2/2A3 and WP-LV48 2A1/2A2 and 2A3 2A-like sequences: we observed synthesis of [GFP2A], but no GUS.

Our model proposes that the C-terminal residues of 2A^{NPGP} adopt a conformation within the PTC such that a peptide bond with prolyl-tRNA (A-site) cannot be formed: it is possible that certain subset of 2A^{NPGP} sequences adopt a conformation that also precludes peptidyl-tRNA hydrolysis by eRF1 – halting, rather than stalling, elongation. The translation of sequences upstream of 2A (GFP) without detectable translation of sequences downstream of 2A (GUS) has, perhaps, a parallel with 'No-go decay'. It has been shown that stalled translation complexes can be 'rescued' by the activities of mammalian Pelota /Hbs1/ABCE1 [37]. Pelota is a molecular mimic of eRF1, although (i) lacks the -GGQ- motif of eRF1 – the peptidyl-tRNA ester linkage is, therefore, not hydrolysed by Pelota, and (ii) binds into a vacant A-site in a stop codon independent manner: Hbs1 is a structural homologue of eRF3 [reviewed in 38-40]. ABCE1 promotes dissociation of the ribosome subunits and nascent peptidyl-tRNA is released – peptidyl-tRNA hydrolase (Pth) subsequently releasing tRNA from peptidyl-tRNA by cleaving the peptide/tRNA ester linkage. Consistent with our observations this would produce [GFP2A] alone although, following this analogy, the subsequent decay of cellular mRNA in the no-go decay pathway implies degradation of the virus RNA!

The evidence that 2A^{NPGP} sequences function within the ribosome arises from two main sources. Firstly, construction of artificial polyproteins comprising two proteins each of which bear N-terminal signal sequences. Here, sequences encoding the p40 and p35 subunits of IL12 were linked *via* FMDV 2A^{NPGP} to encode [p402Ap35] (p40 stop codon removed). Both subunits were secreted from the cell to form active IL12: the (formerly) N-terminal signal sequence of p35 was recognised by signal recognition particle (SRP) as a nascent N-terminal feature and p35 secreted from the cell, along with p40 [41]. This property has been observed for many other artificial polyprotein systems e.g. heavy and light chains of monoclonal antibodies, T-cell receptor proteins etc. [see 42]. Secondly, the 2A^{NPGP} 'cleavage' has been mapped to a 20-30aa tract that can be accommodated within the ribosome exit tunnel – one type amongst a family of ribosome arrest peptides (RAPs) [43]. Indeed, the C-terminal portion of a stalled nascent peptide in the ribosome exit tunnel has been shown to modulate selectivity of the A-site [44].

Studies of 2A^{NPGP} activity using strains of *S. cerevisiae* with compromised levels of release factor activity showed in strains with eRF1 depleted, a greater proportion of ribosomes translated through the 2A coding sequence, whilst in strains with impaired eRF3 GTPase activity, many ribosomes failing to 'pseudo-reinitiate' and translate sequences downstream of 2A^{NPGP} [45]: these data lead us to develop our model of 2A^{NPGP} 'cleavage' activity. In contrast, however, studies using reconstituted translation systems *in vitro* did not show any involvement of eRF1/3 [46]. In conclusion, this sub-set of 2A-like sequences that appear to terminate translation at 2A^{NPGP} poses a conundrum for both (i) our current model of 2A^{NPGP}-mediated ribosome skipping, and (ii) the effect of such sequences on the replication of Kunsagi-, Limnipi-, Potami-, WCP and WP-LV48 viruses.

4.3. Biotechnological and Biomedical Applications

2A/2A-like sequences have been used in a huge range of biotechnological and biomedical applications: these sequences have been shown to be active in all eukaryotic cell types (amoeba, yeast, fungi, algae, plant, animal) - but not in prokaryotic cells [see 42]. In this paper we report a series of 2A-like sequences that are directly comparable – in some cases superior - to those highly active sequences already in use: those encoded by Aalivirus A1 2A1-2A4, Aalivirus B1 2A1-2A4, Avisivirus A1 2A2, Avisivirus B1 2A1, Grusopivirus A1 2A1-2A3, Grusopivirus C 2A1/2A3, Limnipivirus A1 2A1, Limnipivirus B1 2A2, Limnipivirus C1 2A1, Limnipivirus D1 2A1/2A2 and Mosavirus B1 2A1/2A2.

In the majority of cases, 2A^{NPGP}s from different viruses have been used to co-express multiple, different, proteins (e.g. components of macromolecular structures, biochemical pathways etc.) from a single ORF: for example, six different genes were co-expressed using five different 2A-like sequences to create autonomous bioluminescent human cells [47]. Alternatively, multiple iterations

 doi:10.20944/preprints202409.1973.v

of the T2A peptide sequence, with different codon usages, were used to co-express nine different proteins comprising the carotenoid and violacein pathways in *Pichia pastoris* [48]. In both cases, however, constructs were designed to minimise the possibility of gene deletion *via* homologous recombination. In many cases this co-expression technology has replaced the need for time-consuming, costly, sequential transformations by the ability to link multiple genes into a single, self-processing, construct thereby introducing a 'trait' by a single transformation: e.g. production of pluripotent stem cells, cancer gene therapies, CAR T-cell therapies, golden rice development, co-expression of glyphosate tolerance / BT toxins, introduction of novel biosynthetic pathways etc. [42]. Here we describe a series of 2A-like sequences which substantially expand the toolbox for biotechnologists.

Author Contributions: Conceptualisation M.D.R and G.A.L; Methodology (clone construction, *in vitro* transcription / translation etc.) L.S.R. & G.A.L; Bioinformatic analyses Y-T.L., H-C.W. & M.D.R. .

Funding. This work was supported by the University of St. Andrews.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of interest: The author(s) declare that there are no conflicts of interest.

References

1. Martĩnez-Salas, E.; Ryan, M.D. Translation and protein processing. In *The Picornaviruses* Ehrenfeld, E., Domingo, E., Roos, R.P., Eds.; ASM Press: Washington, DC, USA, 2010; pp. 141-161.
2. Yu, S.F.; Benton, P.; Bovee, M.; Sessions, J.; Lloyd, R.E. Defective RNA replication by Poliovirus mutants deficient in 2A protease cleavage activity. *J Virol* **1995**, 69, 247-252.
3. Toyoda, H.; Nicklin, M.J.; Murray, M.G.; Anderson, C.W.; Dunn, J.J.; Studier, F.W.; Wimmer, E. A second virus-encoded proteinase involved in proteolytic processing of poliovirus polyprotein. *Cell* **1986**, 45, 761-770.
4. Hsu, Y-Y.; Liu, Y-N.; Wang, W.; Kao, F-J.; Kung, S-H. *In vivo* dynamics of enterovirus protease revealed by fluorescence resonance emission transfer (FRET) based on a novel FRET pair. *Biochem Biophys Res Commun* **2007**, 353, 939–945.
5. Lu, J.; Lina, Y.; Zhao, J.; Yu, J.; Chen, Y.; Lin, M.C.; Kung, H-F.; He, M-L. Enterovirus 71 disrupts Interferon signaling by reducing the level of Interferon Receptor 1. *J Virol* **2012**, 86, 3767-3776.
6. Martin, A.; Escriou, N.; Chao, S-F.; Girard, M.; Lemon, S.M.; Wychowski, C. Identification and site-directed mutagenesis of the primary (2A/2B) cleavage site of the Hepatitis A Virus polyprotein: functional impact on the infectivity of HAV RNA transcripts. *Virology* **1995**, 213, 213-222.
7. Palmenberg, A.; Neubauer, D.; Skern, T. Genome organization and encoded proteins. In *The Picornaviruses* Ehrenfeld, E., Domingo, E., Roos, R.P., Eds.; ASM Press: Washington, DC, USA, 2010; pp. 3-17.
8. Ryan, M.D.; Donnelly, M.L.L.; Lewis, A.; Mehrotra, A.P.; Wilkie, J.; Gani. D. A model for nonstoichiometric, co-translational protein scission in eukaroytic ribosomes. *Bioorg Chem* **1999**, 27, 55-79.
9. Donnelly, M.L.L.; Luke, G.A.; Mehotra, A.; Li, X.; Hughes, L.E.; Gani, D.; Ryan, M.D. Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal "skip". *J Gen Virol* **2001**, 82, 1013-1025.
10. Luke, G.A.; de Felipe, P.; Lukashev, A.; Kallioinen, S.E.; Bruno, E.A.; Ryan, M.D. Occurrence, function and evolutionary origins of "2A-like" sequences in virus genomes. *J Gen Virol* **2008**, 89, 1036-1042.
11. de Lima, J.G.S.; Teixeira, D.G.; Freitas, T.T.; Lima, J.P.M.S. Evolutionary origin of 2A-like sequences in *Totiviridae* genomes. *Virus Res* **2019**, 259, 1-9.
12. de Lima, J.G.S.; Lanza, D.C.F. 2A and 2A-like sequences: Distribution in different virus species and applications in biotechnology. *Viruses* **2021**, 13, 2160.
13. Donnelly, M.L.L.; Hughes, L.E.; Luke, G.A.; Mendoza, H.; ten Dam, E.; Gani, D.; Ryan, M.D. The "cleavage" activities of foot-and-mouth disease virus 2A site-directed mutants and naturally occurring "2A-like" sequences. *J Gen Virol* **2001**, 82, 1027-1041.
14. Heras, S.R.; Thomas, M.C.; García-Canadas, M.; de Felipe, P.; García-Perez, J.L.; Ryan, M.D.; Lopez, M.C. L1Tc non-LTR retrotransposons from Trypanosoma cruzi contain functional viral-like self-cleaving 2A sequence in frame with active proteins they encode. *Cell Mol Life Sci* **2006**, 63, 1449-1460.
15. Odon, V.; Luke, G.A.; Roulston, C.; de Felipe, P.; Ruan, L.; Escuin-Ordinas, H.; Brown, J.D.; Ryan, M.D.; Sukhodub, A. APE-type non-LTR retrotransposons of multicellular organisms encode virus-like 2A oligopeptide sequences, which mediate translational recoding during protein synthesis. *Mol Biol Evol* **2013**, 30(8), 1955-1965.

16. Brown, J.D.; Ryan, M.D. Ribosome "Skipping": "Stop-Carry On" or "Stop-Go" Translation. In *Recoding: Expansion of Decoding Rules Enriches Gene Expression* Atkins, J.F., Gesteland, R.F., Eds.; Springer: New York, NY, USA, 2010; pp. 101-121.
17. Roulston, C.; Luke, G.A.; de Felipe, P., Ruan, L.; Cope, J.; Nicholson, J.; Sukhodub, A.; Tilsner, J.; Ryan, M.D. "2A-Like" signal sequences mediating translational recoding: A novel form of dual protein targeting. *Traffic* **2016**, 17, 923-939.
18. Minskaia, E.; Nicholson, J.; Ryan, M.D. Optimisation of the foot-and-mouth disease virus 2A co-expression system for biomedical applications. *BMC Biotechnology* **2013**, 13, 67.
19. Donnelly, M.L.L.; Gani, D.; Flint, M.; Monaghan, S.; Ryan, M.D. The cleavage activities of aphthovirus and cardiovirus 2A proteins. *J Gen Virol* **1997**, 78, 13-21.
20. Loughran, G.; Firth, A.E.; Atkins, J.F. Ribosomal frameshifting into an overlapping gene in the 2B-encoding region of the cardiovirus genome. *Proc. Natl Acad. Sci. USA* **2011**, 108, E1111-E1119.
21. Napthine, S.; Ling, R.; Finch, L.K.; Jones, J.D.; Bell, S.; Brierley, I.; Firth, A.E. Protein-directed ribosomal frameshifting temporally regulates gene expression. *Nat Commun.* **2017**, 8, 15582.
22. Napthine, S.; Bell, S.; Hill, C.H.; Brierley, I.; Firth, A.E. Characterization of the stimulators of protein-directed ribosomal frameshifting in Theiler's murine encephalomyelitis virus. *Nucleic Acids Res* **2019**, 47, 8207–8223.
23. Woo, P.C.Y.; Lau, S.K.P.; Choi, G.K.Y.; Huang, Y.; Teng, J.L.L.; Tsoi, H-W.; Tse, H.; Yeung, M.L.; Chan, K-H.; Jin, D-Y.; Yuen, K-Y. Natural occurrence and characterization of two internal ribosome entry site elements in a novel virus, canine Picodicistrovirus, in the picornavirus-like superfamily. *J Virol* **2012**, 86(5), 2797-2808.
24. Zell, R. *Picornaviridae*—the ever-growing virus family. *Arch Virol.* 2018, 163, 299–317.
25. Zell, R.; Knowles, N.J.; Simmonds, P. A proposed division of the family *Picornaviridae* into subfamilies based on phylogenetic relationships and functional genomic organization. *Arch Virol.* **2021**, 166, 2927–2935.
26. Yang, X.; Zeng, Q.; Wang, M.; Cheng, A.; Pan, K.; Zhu, D.; Liu, M.; Jia, R.; Yang, Q.; Wu, Y.; Chen, S.; Zhao, X.; Zhang, S.; Liu, Y.; Yu, Y.; Zhang, L. DHAV-1 2A1 peptide – A newly discovered co-expression tool that mediates the ribosomal "Skipping" function. *Front. Microbiol* **2018**, 9, 2727.
27. Cao, J.; Ou, X.; Zhu, D.; Ma, G.; Cheng, A.; Wang, M.; Chen, S.; Jia, R.; Liu, M.; Sun, K.; Yang, Q.; Wu, Y.; Chen, X. The 2A2 protein of Duck hepatitis A virus type 1 induces apoptosis in primary cell culture. *Virus Genes* **2016**, 52(6), 780-788.
28. Hughes, P.J.; Stanway, G. The 2A proteins of three diverse picornaviruses are related to each other and to the H-rev107 family of proteins involved in the control of cell proliferation. *J Gen Virol.* **2000**, 81, 201–207.
29. Yang, X.; Cheng, A.; Wang, M.; Jia, R.; Sun, K.; Pan, K.; Yang, Q.; Wu, Y.; Zhu, D.; Chen, S.; Liu, M.; Zhao, X-X.; Chen, X. Structures and corresponding functions of five types of picornaviral 2A proteins. *Front Microbiol.* **2017**, 8, 1373.
30. Wang, X.; Liu, N.; Wang, F.; Ning, K.; Li, Y.; Zhang, D. Genetic characterization of a novel duck-origin picornavirus with six 2A proteins. *J Gen Virol.* **2014**, 95, 1289-1296.
31. Su, X.; Shuo, D.; Luo, Y.; Pan, X.; Yan, D.; Li, X.; Lin, W.; Huang, D.; Yang, J.; Yuan, C.; Liu, Q.; Teng, Q.; Li, Z. An emerging duck egg-reducing syndrome caused by a novel picornavirus containing seven putative 2A peptides. *Viruses* **2022**, 14(5), 932.
32. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Wilm, A.; Lopez, R.; Thompson, J.D.; Gibson, T.J.; Higgins, D.G. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, 23, 2947-2948.
33. Lukashev, A.N. Recombination among picornaviruses. *Rev Med Virol* **2010**, 20(5), 327-37.
34. Dantas, M.D.A.; Cavalcante, G.H.O.; Oliveira, R.A.C.; Lanza, D.C.F. New insights about ORF1 coding regions support the proposition of a new genus comprising arthropod viruses in the family *Totiviridae*. *Virus Res* 2016, 211, 159-164.
35. Shao, Q.; Jia, X.; Gao, Y.; Liu, Z.; Zhang, H.; Tan, Q.; Zhang, X.; Zhou, H.; Li, Y.; Wu, D.; Zhang, Q. Cryo-EM reveals a previously unrecognized structural protein of a dsRNA virus implicated in its extracellular transmission. *PLOS Pathog* **2021** https://doi.org/10.1371/journal.ppat.1009396
36. Eyler, D.E.; Wehner, K.A.; Green, R. Eukaryotic release factor 3 is required for multiple turnovers of peptide release catalysis by eukaryotic release factor 1. *J Biol Chem.* **2013**, 288, 29530-8.
37. Pisareva, V.P.; Skabkin, M.A.; Hellen, C.U.; Pestova, T.V.; Pisarev, A.V. Dissociation by Pelota, Hbs1 and ABCE1 of mammalian vacant 80S ribosomes and stalled elongation complexes. *EMBO J.* **2011**, 30, 1804-1817.
38. Yip, M.C.J.; Shao, S. Detecting and rescuing stalled ribosomes. *Trends Biochem Sci.* **2021**, 46, 731-743.
39. Powers, K.T.; Szeto, J.A.; Schaffitzel, C. New insights into no-go, non-stop and nonsense-mediated mRNA decay complexes. *Curr Opin Struct Biol.* **2020**, 65, 110-118.
40. Harigaya, Y.; Parker, R. No-go decay: a quality control mechanism for RNA in translation. *Wiley Interdiscip Rev RNA. 2010*, 1, 132-141.

 doi:10.20944/preprints202409.1973.v

41. Chaplin, P.J.; Camon, E.B.; Villarreal-Ramos, B.; Flint, M.; Ryan, M.D.; Collins, R.A. Production of interleukin-12 as a self-processing 2A polypeptide. *J Interferon Cytokine Res.* **1999**, 19, 235-241.
42. https://www.st-andrews.ac.uk/ryanlab/page10.htm
43. [43] Ito, K.; Chiba, S. Arrest peptides: cis-acting modulators of translation. *Annu Rev Biochem*. **2013**, 82, 171-202.
44. Ramu, H.; Vázquez-Laslop, N.; Klepacki, D.; Dai, Q.; Piccirilli, J.; Micura, R.; Mankin AS. Nascent peptide in the ribosome exit tunnel affects functional properties of the A-site of the peptidyl transferase center. *Mol Cell.* **2011**, 41, 321-330.
45. Doronina, V.A.; Wu, C.; de Felipe, P.; Sachs, M.S.; Ryan, M.D.; Brown, J.D. Site-specific release of nascent chains from ribosomes at a sense codon. *Mol Cell Biol.* **2008**, 28, 4227-4239.
46. Machida, K.; Mikami, S.; Masutani, M.; Mishima, K.; Kobayashi, T.; Imataka. H. A translation system reconstituted with human factors proves that processing of encephalomyocarditis virus proteins 2A and 2B occurs in the elongation phase of translation without eukaryotic release factors. *J Biol Chem.* **2014**, 289, 31960-31971.
47. [47] Xu, T.; Ripp, S.; Sayler, G.S.; Close, D.M. Expression of a humanized viral 2A-mediated lux operon efficiently generates autonomous bioluminescence in human cells. *PLoS One* **2014**, 9, e96347.
48. [48] Geier, M.; Fauland, P.; Vogl, T.; Glieder, A. Compact multi-enzyme pathways in P. pastoris. *Chem Commun (Camb)***. 2015**, 51, 1643-1646.