

Addendum to “The Pan-SL-CoV/GD sequences may be from contamination.”

Daoyu Zhang

The Pan-SL-CoV/GD sequences are likely originated from a synthetic construct. Recently, a dataset containing sanger sequencing amplicons of the Pan-SL-CoV/GD sequence was uploaded under the accession number SRX9503273. These sequences are entirely viral in origin and contains 240 Amplicon sequences from the GD_1 sequence made by the authors of PRJNA607174.

Distribution of the top 243 Blast Hits on 240 subject sequences

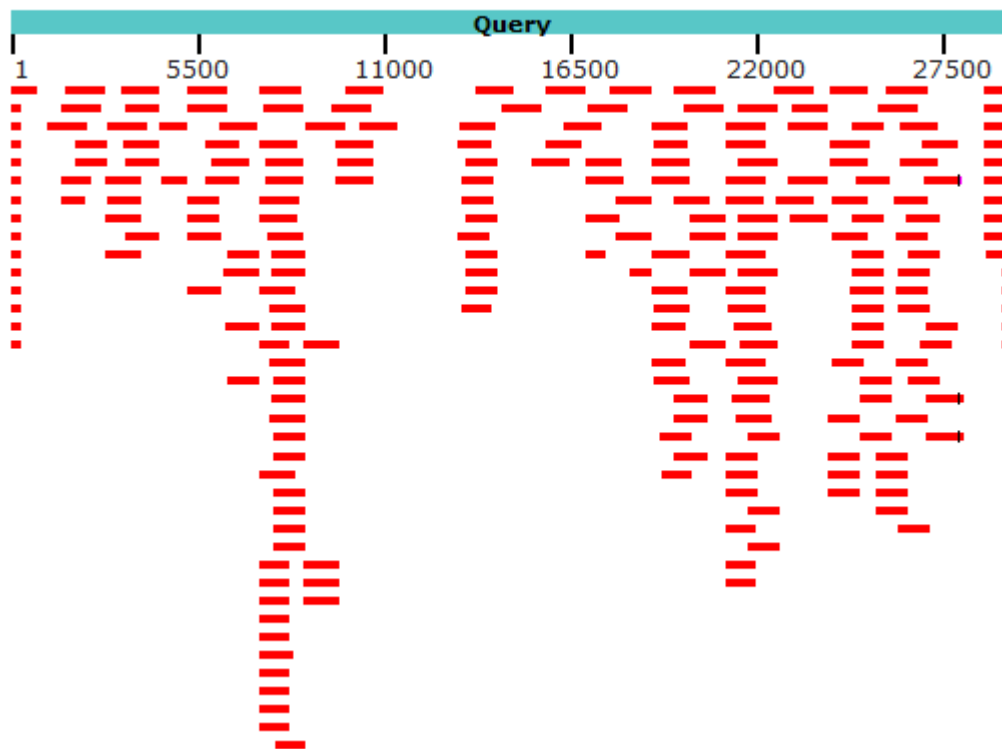


Fig. A1: All 240 sequences are mapped onto the MP789/MT121216.1 sequence.

147 of these sequences contains the tag “M13”, which corresponding to sequencing primers found on popular phagemid cloning vectors.

These sequences appear to cover much of the genome, and contained the backbone of the virus including ORF1a, ORF1b, S1, S2 and the 3'-end of the genome that includes the E, M and N.

Distribution of the top 147 Blast Hits on 147 subject sequences

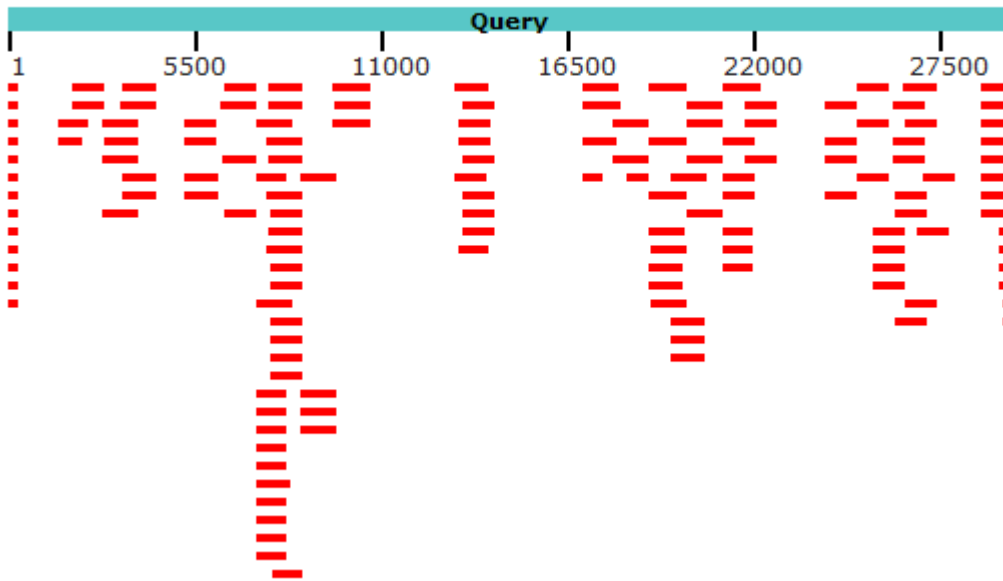


Fig. A2: Mapping of the 147 sequences containing “M13” onto the MT121216.1 genome.

We successfully isolated a Multiple Cloning Site sequence, matching a known pEASY-T1 vector, from these 147 sequences that contained the tag “M13”.



Fig. A3: The isolated Multiple Cloning Site sequence.

We then hypothesized that the trace amount of viral reads found in [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#) May have also been the result from contamination by these clone sequences. In order to confirm this hypothesis, the Multiple Cloning Site sequence was BLASTed against in [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#), and as expected, sequences resembling the Multiple Cloning Site was obtained from the BLAST result.

>gnl|SRA|SRR11119762.25655142.1 25655142 (Biological)
ACACATGGGATGGCACCTCCAGGCTCAAGGAAGGACCGGGGAGGGCCAAACACAGAT
GGCTGGCAATAGAAAGGACAGTCCAGGCTGATCAGCGGGTTAAACGGGCCCTAGAC
TGTGCATCGTCTGTAGTCTCGAGCG

Cloning vector pcDNA3.1_+, complete sequence

Sequence ID: [MN996867.1](#) Length: 5428 Number of Matches: 1

Range 1: 952 to 1051 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
133 bits(147)	4e-27	100/124(81%)	24/124(19%)	Plus/Plus
Query 28	GAAATTCGAGATATCCAGCAGTGGCGCCCTCGAG	-----	-----	-----
Sbjct 952	GAAATTCGAGATATCCAGCAGTGGCGCCCTCGAG	-----	-----	-----
Query 88	AAGTCTAGAGGGCCGTTTAAACCCGCTGATCAGCCTCGACTGCTCTAGTTGCCAG	-----	-----	-----
Sbjct 991	---TCTAGAGGGCCGTTTAAACCCGCTGATCAGCCTCGACTGCTCTAGTTGCCAG	-----	-----	-----
Query 148	CCAT 151	gac tac aag gac gac gat gac aag	-----	-----
Sbjct 1048	CCAT 1051	D Y K D D D D K	-----	-----

>gnl|SRA|SRR11119762.25655142.2 25655142 (Biological)
AFTAAAGGAGAGACCTTCCAGGACCGAATTCCTCAGATATCCAGCAGTGGCGCCG
CTCCAGCTCAGCAGGACGATGAGAACTCTAGAGGGCCGTTTAAACCCGCTGATCA
GCCCTCGACTGTCCTCTAGTTGCCAGCCAT "B"
IKKEFTTETFCRYPAQWRPLEDYKDDDDKSRGPV-TR-SASTVPSSCP
MCP-p65TA-HSF1TA, partial [Expression vector pKK44]

Sequence ID: [QH94762.1](#) Length: 481 Number of Matches: 2

Range 1: 456 to 481 [GenPept](#) [Graphics](#) [Next Match](#)

Score	Expect	Method	Identities	Positives	Gaps
43.1 bits(100)	0.004	Composition-based stats.	21/34(62%)	23/34(67%)	8/34(23%)
Query 2	KEEFTTETFCRYPAQWRPLEDYKDDDDKSRGPV	SRGPV	35		
Sbjct 456	KDTVSCTEFCRYPAQWRPLEDYKDDDDKSRGPV	SRGPV	481		

>gnl|SRA|SRR11119763.138114743.1 138114743 (Biological)
TCCCTCCATCAGCAGAGCTCCAGCCCTAAGAGGAGACCCAGCTTTAAGGAAGAGA
CCTTCCAGCAGACGAAATTCGAGATATCCAGCAGTGGCGCCGCTCGAGATAC
AGGACGAGATGACAGCTCTAGAGGGCCGCT

Cloning vector pcDNA3.1_+, complete sequence

Sequence ID: [MN996867.1](#) Length: 5428 Number of Matches: 1

Range 1: 952 to 1046 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
124 bits(137)	2e-24	95/119(80%)	24/119(20%)	Plus/Minus
Query 1	TGGCACTAGAAAGGACAGTCCAGGCTGATCAGCGGTTTAAACGGGCCCTTAGACTTG	-----	-----	-----
Sbjct 1046	TGGCACTAGAAAGGACAGTCCAGGCTGATCAGCGGTTTAAACGGGCCCTTAGACTTG	-----	-----	-----
Query 61	TCATCGTCGCTTGTAGTCTCGAGGGCCGCACTGCTGATATCCAGATTC	-----	-----	-----
Sbjct 990	---TCCAGGGCCGCACTGCTGATATCCAGATTC	-----	-----	-----
Query	gac tac aag gac gac gat gac aag	-----	-----	-----
Sbjct	D Y K D D D D K	-----	-----	-----

>gnl|SRA|SRR11119763.138114743.2 138114743 (Biological)
TCATCGCTCGCTTGTAGTCTCGAGGGCCGCACTGCTGATATCCAGATTC
GTCCTGGTGAAGTCTCTTCTTAAATAGGGC
PKKEFTTETFCRYPAQWRPLEDYKDDDDKSRGPV-TR-SASTVPSSC
MCP-p65TA-HSF1TA, partial [Expression vector pKK44]

Sequence ID: [QH94762.1](#) Length: 481 Number of Matches: 1

Range 1: 456 to 481 [GenPept](#) [Graphics](#) [Next Match](#)

Score	Expect	Method	Identities	Positives	Gaps
41.6 bits(96)	0.013	Composition-based stats.	21/34(62%)	23/34(67%)	8/34(23%)
Query 3	KEEFTTETFCRYPAQWRPLEDYKDDDDKSRGPV	SRGPV	36		
Sbjct 456	KDTVSCTEFCRYPAQWRPLEDYKDDDDKSRGPV	SRGPV	481		

>gnl|SRA|SRR11119766.149640833.1 149640833 (Biological)
AACTAGAGGACAGCTCGAGCTGATCAGCGGTTTAAACGGGCCCTTAGACTTGTCA
CTCGCTCTGTAGTCTCGAGGGCCGCACTGCTGATATCCAGATTCGAGATTCGCT
CGGTGAAGTCTCTCTTAAATAGGGCTGGT

Cloning vector pcDNA3.1_+, complete sequence

Sequence ID: [MN996867.1](#) Length: 5428 Number of Matches: 1

Range 1: 952 to 1036 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
106 bits(117)	5e-19	85/109(78%)	24/109(22%)	Plus/Plus
Query 43	GAAATTCGAGATATCCAGCAGTGGCGCCCTCGAG	-----	-----	-----
Sbjct 952	GAAATTCGAGATATCCAGCAGTGGCGCCCTCGAG	-----	-----	-----
Query 103	AAGTCTAGAGGGCCGTTTAAACCCGCTGATCAGCCTCGACTGCTCTAGTTGCCAG	-----	-----	-----
Sbjct 991	---TCTAGAGGGCCGTTTAAACCCGCTGATCAGCCTCGACTGCTCTAGTTGCCAG	-----	-----	-----
Query	gac tac aag gac gac gat gac aag	-----	-----	-----
Sbjct	D Y K D D D D K	-----	-----	-----

>gnl|SRA|SRR11119766.149640833.2 149640833 (Biological)
GAGGAGACCCGCTTAAAGGAAGAGCCTTCCAGCAGCCGAAATTCGAGATTCCTCA
GCACAGTGGCGCCGCTCGAGGACACTGAGGACGATGAGAACTCTAGAGGGCCGCT
TAAACCCGCTGATCAGCTCGACTGCTCT

EETSPIKEETFTETFCRYPAQWRPLEDYKDDDDKSRGPV-TR-SASTVP
MCP-p65TA-HSF1TA, partial [Expression vector pKK44]

Sequence ID: [QH94762.1](#) Length: 481 Number of Matches: 1

Range 1: 451 to 481 [GenPept](#) [Graphics](#) [Next Match](#)

Score	Expect	Method	Identities	Positives	Gaps
41.2 bits(95)	0.023	Composition-based stats.	22/39(56%)	24/39(61%)	8/39(20%)
Query 2	EETSPIKEETFTETFCRYPAQWRPLEDYKDDDDKSRGPV	SRGPV	40		
Sbjct 451	EPFAADFTVSCTEFCRYPAQWRPLEDYKDDDDKSRGPV	SRGPV	481		

>ACH1550.1 FLAG, partial [Expression vector pDM320]
MDYKDDDDK

Mammalian expression vector pcDNA3mB7-2, complete sequence

Sequence ID: [LT727375.1](#) Length: 6407 Number of Matches: 1

Range 1: 861 to 948 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
163 bits(88)	2e-36	88/88(100%)	0/88(0%)	Plus/Plus
Query 1	ATTAATACGACTCACTATAGGAGAGCCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAA	-----	-----	-----
Sbjct 861	ATTAATACGACTCACTATAGGAGAGCCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAA	-----	-----	-----
Query 61	CGGGCCAGCTGTGCTGGAATCGGCTT 88	-----	-----	-----
Sbjct 921	CGGGCCAGCTGTGCTGGAATCGGCTT 948	-----	-----	-----

>gnl|SRA|SRR11119766.121997761.1 121997761 (Biological)
CGCTGTAGAGCTGCTTCAAGTCTTCCCTCGAGATCCGCTTGAAGGGCTCCCTCAGCAGG
TCTTGAAGCAGCTCGTCAATATTAAGTCCGCTGGTGTGGTGTATGATGATCGCTCCGGGG
ATCCCGGAGATCTCGCGGGCTGGCCCGAG

>gnl|SRA|SRR11119766.121997761.2 121997761 (Biological)
ATTAATACGACTCACTATAGGAGAGCCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAA
CGGGCCAGCTGTGCTGGAATCGGCTTGGGATATCCACCATTGGAGCAGACACACTCC
TGCATGGGTACTGCTGCTCGGGTCCAGG

"A"

Mammalian expression vector pcDNA3mB7-2, complete sequence

Sequence ID: [LT727375.1](#) Length: 6407 Number of Matches: 1

Range 1: 845 to 948 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
193 bits(104)	3e-45	104/104(100%)	0/104(0%)	Plus/Plus
Query 1	TTACTGGCTTATCGAAATTAATACGACTCACTATAGGAGAGCCCAAGCTTGGTACCGAGG	-----	-----	-----
Sbjct 845	TTACTGGCTTATCGAAATTAATACGACTCACTATAGGAGAGCCCAAGCTTGGTACCGAGG	-----	-----	-----
Query 61	TCCGATCCACTAGTAAAGGGCCGCACTGCTGGAATTCGGCTT 104	-----	-----	-----
Sbjct 905	TCCGATCCACTAGTAAAGGGCCGCACTGCTGGAATTCGGCTT 948	-----	-----	-----

>gnl|SRA|SRR11119763.163404363.1 163404363 (Biological)
AGCAGGAGTGTCTGTCTCCAGTGGTGGATATCCCAAGCCGAAATCCAGCAGCAGCTGGCG
CGCCTTATAGTGGATCCAGGCTCGGTACCAAGCTTGGGTCTCCCTATAGTACTCGTAT
TAATTTGATAGCCGCTAAAGATCGGAAG

>gnl|SRA|SRR11119763.163404363.2 163404363 (Biological)
TACTGGCTTATCGAAATTAATACGACTCACTATAGGAGAGCCCAAGCTTGGTACCGAGG
TCCGATCCACTAGTAAAGGGCCGCACTGCTGGAATTCGGCTTGGGATATCCACCATTG
GAGCAGACACTCTCTGTAGTCCGAAGA

Fig.A5: Presence of pcDNA3.1 vectors within in [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#). The single large gap on the alignment with other available GenBank was identified as encoding an 1xFLAG peptide, DYKDDDDK, indicating a modification to the standard pcDNA3.1 vector.

A:"TTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGGCTTGGGGATCCACCATGGAGACAGACACTCTGCTAGATCGGAAGA"

Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Mammalian expression vector pcDNA3mB7-2_complete sequence	Mammalian ex...	193	193	68%	3e-45	100.00%	6407	LT727375.1
<input checked="" type="checkbox"/> Expression vector pcDNA3-AQP4f_complete sequence	Expression vec...	185	185	66%	5e-43	100.00%	6510	EF437956.1
<input checked="" type="checkbox"/> Expression vector pcDNA3-AQP4e_complete sequence	Expression vec...	185	185	66%	5e-43	100.00%	6675	EF437953.1
<input checked="" type="checkbox"/> Expression vector pcDNA3-AQP4d_complete sequence	Expression vec...	185	185	66%	5e-43	100.00%	6349	EF437951.1
<input checked="" type="checkbox"/> Mammalian expression vector pcDNA-B-catenin_complete sequence	Mammalian ex...	183	183	65%	2e-42	100.00%	7806	LT727265.1
<input checked="" type="checkbox"/> Mammalian expression vector pcDNA-mCASP-8_complete sequence	Mammalian ex...	183	183	65%	2e-42	100.00%	6950	LT727263.1
<input checked="" type="checkbox"/> Mammalian vector pcDNA-HA_complete sequence	Mammalian ve...	183	183	65%	2e-42	100.00%	7173	LT727252.1
<input checked="" type="checkbox"/> Mammalian expression vector pcDNA1-hTNFR55-E_complete sequence	Mammalian ex...	183	183	67%	2e-42	99.02%	5563	LT727190.1
<input checked="" type="checkbox"/> Shuttle vector pMSK1-E-D565A_complete sequence	Shuttle vector...	183	183	65%	2e-42	100.00%	6575	LT727062.1
<input checked="" type="checkbox"/> Shuttle vector pMSK1-E_complete sequence	Shuttle vector...	183	183	65%	2e-42	100.00%	6575	LT727061.1
<input checked="" type="checkbox"/> Mammalian expression vector pcDNA3-hIRAK1-T66A_complete sequence	Mammalian ex...	183	183	65%	2e-42	100.00%	7476	LT726935.1
<input checked="" type="checkbox"/> Mammalian expression vector pcDNA3-hIRAK1_complete sequence	Mammalian ex...	183	183	65%	2e-42	100.00%	7476	LT726934.1
<input checked="" type="checkbox"/> Mammalian expression vector pcDNA3-hIRAK1-KD_complete sequence	Mammalian ex...	183	183	65%	2e-42	100.00%	7476	LT726931.1
<input checked="" type="checkbox"/> Expression vector pcDNA/HA-FLAG_complete sequence	Expression vec...	183	183	65%	2e-42	100.00%	5560	FJ624378.1
<input checked="" type="checkbox"/> Cloning vector pOnR6K-zeo-ie_complete sequence	Cloning vector...	183	183	65%	2e-42	100.00%	2477	AY700022.1
<input checked="" type="checkbox"/> Expression vector pCDPT	Expression vec...	183	183	65%	2e-42	100.00%	3840	AJ132038.1
<input checked="" type="checkbox"/> Cloning vector pIRES1hyg_complete plasmid sequence	unidentified clg...	182	182	64%	6e-42	100.00%	5726	U89672.1
<input checked="" type="checkbox"/> Cloning vector pRcCMV-luc luciferase gene_complete cds	Cloning vector...	180	180	64%	2e-41	100.00%	7290	U43958.1
<input checked="" type="checkbox"/> Mammalian expression vector pM6DraA6_complete sequence	Mammalian ex...	169	169	66%	5e-38	97.03%	7949	LT727632.1
<input checked="" type="checkbox"/> Mammalian vector pCR3-EYFPC-GW_complete sequence	Mammalian ve...	137	137	49%	1e-28	100.00%	7475	LT726804.1

B:"ATTAAGGAAGAGACCTTACCAGACCGAATTCTGCAGATATCCAGCACAGTGGCGCCGCTCGAGGACTACAAGGACGACGATGCAAGTCTAGAGGGCCGTTAAACCCGCTGATCAGCCTCGACTGTGCCTCTAGTTGCCAGCCAT"

select all 500 sequences selected

[GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Mammalian expression vector pcDNA3-N-HA-LIC_complete sequence	Mammalian ex...	113	113	82%	3e-21	84.00%	7397	KT359064.1
<input checked="" type="checkbox"/> Cloning vector pcDNA3.1-anti-VEGF-scFV_complete sequence	Cloning vector...	109	109	41%	4e-20	100.00%	6259	MN996868.1
<input checked="" type="checkbox"/> Cloning vector pcDNA3.1+_complete sequence	Cloning vector...	109	178	67%	4e-20	100.00%	5428	MN996867.1
<input checked="" type="checkbox"/> Cloning vector pcDNA3.1-AFP_complete sequence	Cloning vector...	109	109	41%	4e-20	100.00%	7258	MN996866.1
<input checked="" type="checkbox"/> Cloning vector pcDNA3.1-LATS1_complete sequence	Cloning vector...	109	109	41%	4e-20	100.00%	8821	MN996863.1
<input checked="" type="checkbox"/> Cloning vector pADNpcDNA3.1 KanR_complete sequence	Cloning vector...	106	106	40%	4e-19	100.00%	10000	KX176867.1
<input checked="" type="checkbox"/> Cloning vector pTCHS-Nluc/Orange-G/Zr_complete sequence	Cloning vector...	106	106	40%	4e-19	100.00%	7604	KM603067.1
<input checked="" type="checkbox"/> Mammalian expression vector pSA89_complete sequence	Mammalian ex...	106	106	40%	4e-19	100.00%	5594	JQ624674.1
<input checked="" type="checkbox"/> Mammalian expression vector pSA83_complete sequence	Mammalian ex...	106	106	40%	4e-19	100.00%	5599	JQ624673.1
<input checked="" type="checkbox"/> Cloning vector pTT-PB-SOKMLN-Cer_complete sequence	Cloning vector...	104	104	50%	1e-18	92.31%	13592	KM279352.1
<input checked="" type="checkbox"/> Rabies viral vector pHEP5.0-CVSG-mRFP DNA_complete sequence	Rabies viral ve...	104	104	41%	1e-18	98.41%	17394	AB855657.1
<input checked="" type="checkbox"/> Rabies viral vector pHEP5.0-delG-mRFP DNA_complete sequence	Rabies viral ve...	104	104	41%	1e-18	98.41%	15772	AB855650.1
<input checked="" type="checkbox"/> Cloning vector RS474_ErbB-RASER1C-dCas9VP64_complete sequence	Cloning vector...	102	102	40%	4e-18	98.39%	16350	MK801288.1

Fig.A6: These complete reads are not found in non-synthetic sequences, confirming their origin as from a synthetic cloning vector.

In order to validate the existence of the vector sequences within [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#), A BLAST analysis using the vector backbone, MN996868.1, was performed on all 3 datasets. All 3 datasets yielded near-complete sequences from the backbone confirming the existence of a synthetic eukaryotic expression vector within these datasets.



Fig.A7: Near-complete sequence mapping to MN996868.1 in [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#)

No evidence of the the Pan-SL-CoV/GD sequence in other datasets deposited by Liu et al. or Xiao et al.

In order to confirm that these datasets that we have already covered were in deed the only datasets of which Coronavirus-like reads were present, we performed a BLAST analysis on the rest of the PRJNA607174 and PRJNA573298 datasets. We did not obtain any evidence of Coronavirus-like reads from these datasets.

Job Title **gb|MT121216|**

RID [WVGXCM9K01R](#) Search expires on 12-08 12:54 pm
[Download All](#) ▾

Program [Citation](#) ▾

Database SRA [See details](#) ▾

SRA Blast search set information

SRX7756764	SRR11119764
------------	-------------

Query Length 29521

Other reports [?](#)

⚠ No significant similarity found. For reasons why [click here](#)

Job Title **gb|MT121216|**

RID [WVGXV0JY01R](#) Search expires on 12-08 12:55 pm
[Download All](#) ▾

Program [Citation](#) ▾

Database SRA [See details](#) ▾

SRA Blast search set information

SRX7756761	SRR11119767
------------	-------------

Query Length 29521

Other reports [?](#)

⚠ No significant similarity found. For reasons why [click here](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**

RID [WVGXNWGF01R](#) Search expires on 12-08 12:55 pm
[Download All](#) ▾

Program [Citation](#) ▾

Database SRA [See details](#) ▾

SRA Blast search set information

SRX7756763	SRR11119765
------------	-------------

Query Length 29521

Other reports [?](#)

⚠ No significant similarity found. For reasons why [click here](#)

Job Title **gb|MT121216|**

RID [WVR2588J01R](#) Search expires on 12-08 14:39 pm
[Download All](#) ▾

Program [Citation](#) ▾

Database SRA [See details](#) ▾

SRA Blast search set information

SRX6893158	SRR10168373
SRX6893152	SRR10168379
SRX6893151	SRR10168380
SRX6893150	SRR10168381
SRX6893149	SRR10168382
SRX6893148	SRR10168383
SRX6893147	SRR10168384
SRX6893146	SRR10168385
SRX6893145	SRR10168386
SRX6893144	SRR10168387
SRX6893143	SRR10168388
SRX6893142	SRR10168389
SRX6893140	SRR10168391
SRX6893141	SRR10168390
SRX6893138	SRR10168393

⚠ No significant similarity found. For reasons why [click here](#)

Query ID [MT121216.1](#)

Description Pangolin coronavirus isolate MP789, complete genome

Molecule type nucleic acid

Query Length 29521

Other reports [?](#)

⚠ No significant similarity found. For reasons why [click here](#)

Fig.A8: BLAST analysis of the other datasets within PRJNA573298 and PRJNA607174. No evidence of Coronavirus-like reads were found in these datasets.

A putative vector backbone sequence isolated from [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#)

Using De-novo sequence assembly with EGAssembler[1] and sequence alignment with BLAST2[2], contiguous sequences were assembled and identified as originated from Mammalian Expression vectors, and complete sequence backbones of two mammalian expression vectors were obtained from the datasets [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#) using paired-end information.

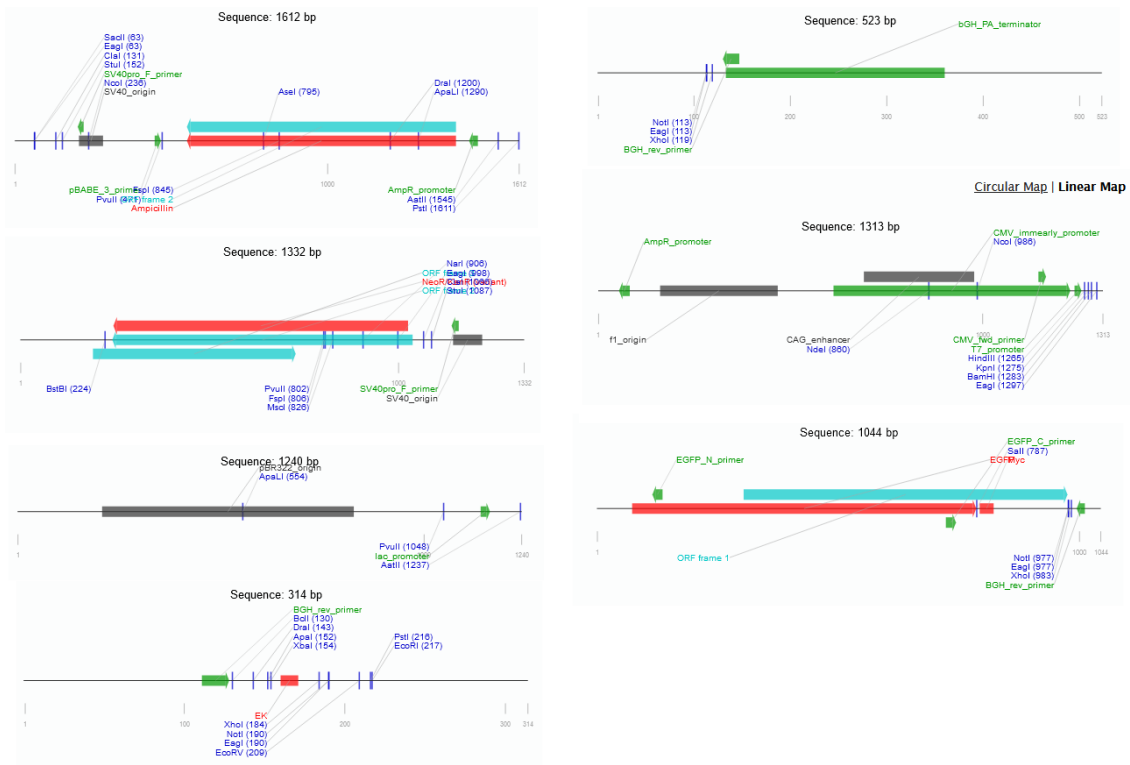


Fig.A9: Contigs identified from [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#) using BLAST results with pcDNA3.1 as the query sequence.

Using paired-end information and sequences from [SRX7756765](#) alone, we managed to obtain the full-length vector backbones from the obtained sequences using De-Novo scaffold assembly.

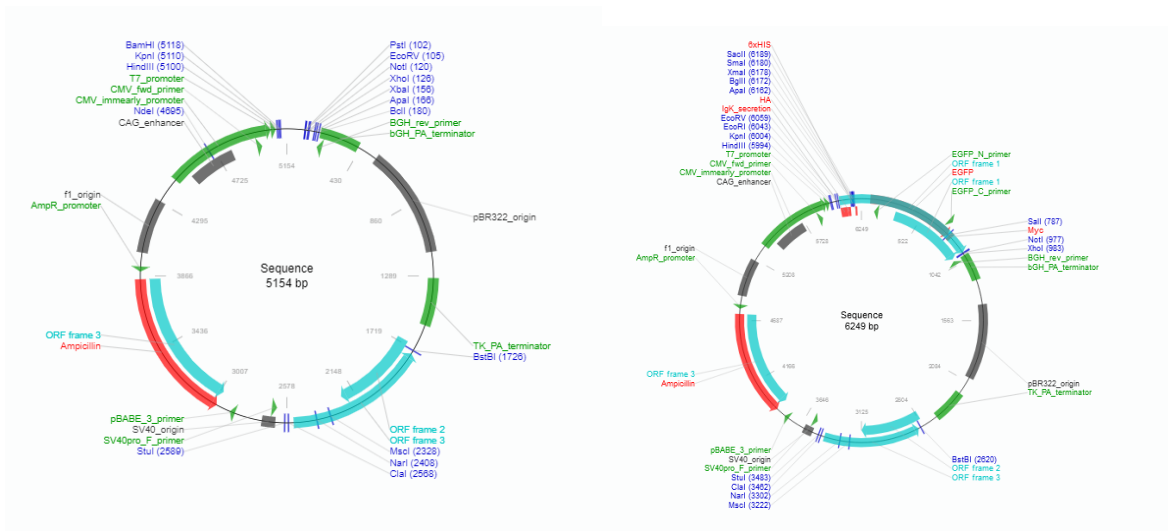


Fig.A10: Complete sequence backbone of two cloning vectors obtained from [SRX7756765](#).

These sequence backbones were then verified against [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#). Full coverage of the sequences by reads are obtained in all 3 datasets, which confirmed the existence of this vector in all 3 datasets.



Fig.A11: existence of the sequence backbone within [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#)

This sequence was then mapped using The Addgene sequence Analyzer[3] and identified as a modified pCR3.1 vector, with a f1(+) origin to prepare (+)ssDNA in the same direction as the mammalian expression cassette beginning with a CMV promoter and terminating with a bGH poly(A) signal.

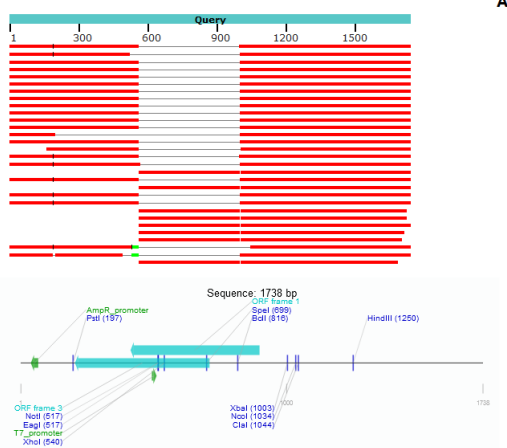
select all 100 sequences selected		GenBank	Graphics	Distance tree of results				
Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Eukaryotic expression vector pCR3.1mBCL-XL_complete sequence	Eukaryotic expression vector p...	7847	9478	97%	0.0	99.74%	5771	AF060226.1
<input checked="" type="checkbox"/> Mammalian expression vector pCR3Flag-E8-FLIP_complete sequence	Mammalian expression vector ...	7806	9339	97%	0.0	99.08%	5611	LT727397.1
<input checked="" type="checkbox"/> Mammalian expression vector pCR3FLAG-CED-3_complete sequence	Mammalian expression vector ...	7805	9317	95%	0.0	99.74%	6593	LT727396.1
<input checked="" type="checkbox"/> Mammalian expression vector pCR3VSV-CED-4_complete sequence	Mammalian expression vector ...	7805	9317	95%	0.0	99.74%	6734	LT727395.1
<input checked="" type="checkbox"/> Mammalian expression vector pCR3-FLAGhRIP1_complete sequence	Mammalian expression vector ...	7805	9317	95%	0.0	99.74%	7098	LT727049.1
<input checked="" type="checkbox"/> Mammalian expression vector pCR3-pro-IL-1B_complete sequence	Mammalian expression vector ...	7805	9317	95%	0.0	99.74%	5843	LT726864.1
<input checked="" type="checkbox"/> Mammalian expression vector pCR3-pro-IL-1B-D105A_complete sequence	Mammalian expression vector ...	7805	9317	95%	0.0	99.74%	5843	LT726863.1
<input checked="" type="checkbox"/> Mammalian expression vector pCR3-pro-IL-1B-D108A_complete sequence	Mammalian expression vector ...	7805	9317	95%	0.0	99.74%	5843	LT726862.1
<input checked="" type="checkbox"/> Mammalian expression vector pCR3-pro-IL-1B-D117A_complete sequence	Mammalian expression vector ...	7805	9317	95%	0.0	99.74%	5843	LT726861.1

Fig.A12 :BLAST alignment of the assembled vector sequence against the NCBI nr/nt database.

As sequences resembling mammalian expression vectors are unique to synthetic sequences, the identification of this mammalian expression vector backbone provides irrefutable proof of contamination of [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#) by customized, synthetic DNA sequences created with the intention of expressing foreign RNA and protein sequences within mammalian cells, which directly contradicts the claim of Xiao et al [4] that the samples were allegedly obtained from lung tissues of Malayan pangolins “confiscated by Customs and Department of Forestry of Guangdong Province in March and August 2019”, which would not have contained any synthetic sequences due to their supposed origin from the wild.

A synthetic recombinant African Swine Fever Virus was found within PRJNA607174. During the sequence assembly process, one specific contig that was found to be divergent from the rest of the vector backbones in the 5'-terminus was obtained, which revealed itself to be the result of direct fusion of an African Swine Fever Virus genome to a vector containing the pBR322 origin sequence. We attempt to investigate further this sequence, which revealed an intact insert sequence in a separate, distinct vector backbone, originating from the African Swine Fever Virus isolate Wuhan-2019-1.

Distribution of the top 194 Blast Hits on 100 subject sequences



African swine fever virus isolate ASFV Wuhan 2019-1, complete genome

Sequence ID: [MN393476.1](#) Length: 190576 Number of Matches: 2
[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

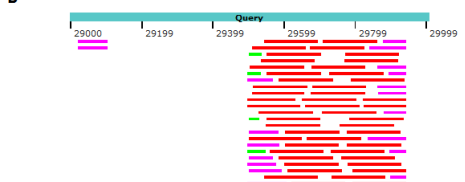
Range 1: 29498 to 29935 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
868 bits(438)	0.0	438/438(100%)	0/438(0%)	Plus/Minus
Query 560	CATACCTCAGATGCCATATATATTTTGGATTGTAAGTGGCAACATAAAGGCTATATTT	619		
Sbjct 29935				
Query 620	TTTGATGAGGTTGATGGAAAGGTTAGTGAGACAGTAAATTTCAAGGTTTCTCGGA	679		
Sbjct 29875				
Query 680	AGATTTTGGCTTTTATCGCCACTAGTAAACATTTGCTTATCTGGCGTGGGTTGATGAGG	739		
Sbjct 29815				
Query 740	AATGAAATAGATTAAAATTTTTCATAGGTTATAGGTTGACAGCTTGTGTTAGATAATATT	799		
Sbjct 29755				
Query 800	TATATATCTTTATATGATCAATCTTAAAGTCTTAGATATAAATCTAAGTATCCCTTT	859		
Sbjct 29695				
Query 860	GGCTGCCACTTATAGAGGTTTGGGCTGAAAGCATGCGAATCTAAATGATGATA	919		
Sbjct 29635				
Query 920	CTTATCGGAGACATCTGTTATATATAAATCTATACCCCAACATATAGATGATGATC	979		
Sbjct 29575				
Query 980	CTTATGGCAATAGCACA 997			
Sbjct 29515				

Range 2: 41127 to 41151 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#) [First Match](#)

Score	Expect	Identities	Gaps	Strand
42.1 bits(21)	123	24/23(96%)	0/23(0%)	Plus/Minus
Query 660	AGGTTCAAGGTTTCTCGGAGATT 684			
Sbjct 41151	AGGTTCAAGGTTTCTCGGAGATT 41127			

Distribution of the top 225 Blast Hits on 225 subject sequences



Cloning vector AGG1533, complete sequence

Sequence ID: [MN720705.1](#) Length: 6646 Number of Matches: 3

Range 1: 3043 to 3783 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
1453 bits(733)	0.0	739/741(99%)	0/741(0%)	Plus/Plus
Query 997	AATCTTTCTAGAGATCTCTACATATTTCTGAGTGGCATGAAATCGAGTCTTCT	1056		
Sbjct 3043				
Query 1057	TTTATCTCTCAAGATTTTCAGGCTGTATATTAACCTATATTAAGAACTATGCTAAC	1116		
Sbjct 3103				
Query 1117	ACCTCATCAGAACCGTTGAGTGGCGTGGTTTTCTGGCAATCGACTCATGAAAA	1176		
Sbjct 3163				
Query 1177	CTACGAGCTAAATATCAATATGTTCCCTTACCAACTTTATCTGCATTTTTTTTGA	1236		
Sbjct 3223				
Query 1237	CGAGGTTAGAGCAAGCTTCAGAACTGAGACAGAAATTTTAAATAATTAATTTTT	1296		
Sbjct 3283				
Query 1297	GAGAAAGTTCAGGTTAATAGCATCCATTTTTTGGTTTTCAGGTTTCTCAGCATCTTA	1356		
Sbjct 3343				
Query 1357	ACAAAAGAGCTCTCTTTGACATGTTTAAAGTTTAACTCCCTGCGAAATTTATATCC	1416		
Sbjct 3403				
Query 1417	GCTCATATTCACACATATACAGAGCCGAGCATAAAGTTAAAGCTGGGGTCCCTA	1476		
Sbjct 3463				
Query 1477	ATGATGAGCTAACCTACATTAATGCGTTGGCTCACTGCCAATGCTTCCAGTCGG	1536		
Sbjct 3523				
Query 1537	AAACCTGCTGCCACTGCTAATGAACTCGCCCAAGCGGGGAGAGGGCGGTTTCCG	1596		
Sbjct 3583				
Query 1597	TATTGGGCGCTCTCCGCTTCTCCCTCACTGACTCGCTGGCTGGCTGGCTGGCTGG	1656		
Sbjct 3643				
Query 1657	CGAGCGGTATCAGCTCACTAAGAGCGTAAATAGGTTATCCAGCAATCAGGGATAA	1716		
Sbjct 3703				
Query 1717	CGAGAAAGAACATGTGAC 1737			
Sbjct 3763				

Range 2: 1 to 378 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#) [First Match](#)

Score	Expect	Identities	Gaps	Strand
726 bits(366)	0.0	375/378(99%)	0/378(0%)	Plus/Plus
Query 189	GCCTTCGAGCGAATATATATTTTGGCAATAATTTTAAACAAAGCTCGAAGTC	248		
Sbjct 1				
Query 249	TTCTTCATTTAAATCTTAGATGATCTCATCGGAAATGCTCCCAATGATGACATC	308		
Sbjct 61				
Query 309	ACGCTGAGTAAAGTCTAAGCAATTTTTATGTTGATATCTCCAACTTACTACT	368		
Sbjct 121				
Query 369	CGATGATTTTGGTATATCTCTATTTTAACTGGATAGGTTCAATCATGTTTTT	428		
Sbjct 181				
Query 429	TTCACTAGTAAATAAATCACTGCTTTAAACCTGCTGCTGACACATATCACTCC	488		
Sbjct 241				
Query 489	GGCTAATACGACTACTATAGGAGAGCGCCGAGATCTCCGAGTGGCTCGAGTT	548		
Sbjct 301				
Query 549	TTGACCAAGTACTACTC 566			
Sbjct 361				

Range 3: 5459 to 5646 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#) [First Match](#)

Score	Expect	Identities	Gaps	Strand
373 bits(188)	3e-98	188/188(100%)	0/188(0%)	Plus/Plus
Query 1	TTCTTTTTCAATATTAATGAAGCATTTATCAGGGTTATGTTCTCATGAGCGGATACATA	60		
Sbjct 5459				
Query 61	TTTGAATGATTTAGAAAATAAACAATAGGGTTCCCGGCACATTTCCCGAAAAGTG	120		
Sbjct 5519				
Query 121	CCACCTGAGCTTAAGAAACCATATATATCATGACATTAACCTATAAAAATAGGCGTATC	180		
Sbjct 5579				
Query 181	ACGAGGCC 188			
Sbjct 5639				

Fig.A13: A: partial DNA cassette from [SRX7756765](#) harboring a fragment of DNA sequence from African Swine Fever Virus Wuhan-2019-1 inserted into a AGG1533-related backbone. B: One additional pair of reads just 5' of the insert sequence can be seen on the extreme 5' end of the alignment.

During assembly of the sequence, we obtained a single pair of reads from [SRX7756765](#) which displays gene fusion to both the vector backbone, a sequence fragment from nearby the insert sequence and a sequence fragment from far outside of the insert sequence, with a distance between the two fragments being more than 7500bp apart. Such a peculiar sequence fusion is only possible if these two sequences transcribed from two distinct positions on the African Swine Fever Virus backbone were fused together through mRNA splicing, which require the fusion of the vector DNA and the ASFV genomic DNA as a single contiguous strand of DNA. This indicates the presence of an African Swine Fever Virus within the sample, containing an exogenously inserted synthetic DNA cassette into it's DNA genome replacing the DNA sequence from position 29104 to position 29498, making it a synthetic recombinant strain of ASFV. As no evidence of pig DNA was found within PRJNA607174, the presence of such recombinant African Swine Fever Virus within these supposedly wild samples of lung tissue have important implications on the integrity of Xiao et al[4] and the real sequence of events that happened during March-August 2019, the alleged collection date of the samples used in this study.

```

>gnl[SRA]SRR11119763.160830215.1 160830215 (Biological)
ACCC TGGCCCTTTTCCCAATGAAGGAGGTCCCAGTATGGCAGCTTTTCTTTAAAGATGGC
TGCCATCAATGCTTCCCTTCGGCTACACTCCCATCCAGAGGAACCCATAAAGCAGCCCA
GGCTAATCTGCTGGAAAGACAGGTCACGAAG

>gnl[SRA]SRR11119763.160830215.2 160830215 (Biological)
ATGTTGGCATCTTTTATGGTCTTGTGGTTGACTGGGCTCAGCTACATGGTTCCTGCAGTG
TTAGATGATTCAGGCAGATGGTGGCTGGGGCTGGAGTTATCTGAAGGCTGGACATCCTG
GATCGCTTCCTTACCCTCTTGTCTGGCACGT

```

African swine fever virus isolate ASFV Wuhan 2019-1, complete genome

Sequence ID: [MN393476.1](#) Length: 190576 Number of Matches: 2

[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 29023 to 29104 [GenBank](#) [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
155 bits(78)	1e-33	81/82(99%)	0/82(0%)	Plus/Minus
Query 47	TGCTGCTGTTCTCAGGGCTTCATTTACGGTAATGAGGGTATGTTGCTGTATAAGCACACA	106		
Sbjct 29104	TGCTGCTGTTCTCAGGGCTTCATTTACGGGAATGAGGGTATGTTGCTGTATAAGCACACA	29045		
Query 107	GTTGTTTCTGCTCGTTGAATA	128		
Sbjct 29044	GTTGTTTCTGCTCGTTGAATA	29023		

Range 2: 36665 to 36710 [GenBank](#) [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#) [▲ First Match](#)

Score	Expect	Identities	Gaps	Strand
91.7 bits(46)	1e-14	46/46(100%)	0/46(0%)	Plus/Minus
Query 1	GGTTTTGATTACAAAGGGTAGAAGCCCTTTTATTGAAACGTCTCGC	46		
Sbjct 36710	GGTTTTGATTACAAAGGGTAGAAGCCCTTTTATTGAAACGTCTCGC	36665		

Cloning vector AGG1533, complete sequence

Sequence ID: [MN720705.1](#) Length: 5646 Number of Matches: 1

Range 1: 3042 to 3100 [GenBank](#) [Graphics](#) [▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Identities	Gaps	Strand
117 bits(59)	2e-22	59/59(100%)	0/59(0%)	Plus/Plus
Query 127	TAATCTTTCTAGAAGATCTCCTACAATATTCTCAGCTGCCATGGAAAATCGATGTTCTT	185		
Sbjct 3042	TAATCTTTCTAGAAGATCTCCTACAATATTCTCAGCTGCCATGGAAAATCGATGTTCTT	3100		

Fig.A14: A pair of reads found within [SRX7756765](#) Shows a triple fusion of two African Swine Fever Virus DNA fragments and a one pBR322 origin of replication.

Methods

Sequence extraction and alignment

The Multiple Cloning Site (MCS) sequence was obtained by BLAST analysis of reads with title “M13” found in SRX9503273. sequences aligned to the GD-1 reference genome MT121216 was masked, and the remaining end sequences of the reads were searched against the NCBI nr/nt database and

identified as a Multiple Cloning Site sequence. The 5' sequence and the 3' sequence was merged and used for downstream analysis.

This MCS sequence was then searched against [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#), which revealed matches to this sequence in all 3 datasets. This match was then searched against the NCBI nr/nt database and identified as a Multiple Cloning Site from the cloning vector pcDNA3.1(+), with an additional inserted sequence identified as the coding sequence for the FLAG octapeptide DYKDDDDK.

Mammalian Expression Vectors from the NCBI nr/nt database are used to obtain alignments from [SRX7756766](#), [SRX7756765](#) and [SRX7756762](#), the ILLUMINA sequencing adapter sequence Read 1: *AGATCGGAAGAGCACACGTCTGAACTCCAGTCA* and Read 2: *AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT* were trimmed away from the ends of the reads, and Contigs were de-novo assembled from the resulting alignments in [SRX7756765](#) using EGAssembler [2]. These contigs were identified as from cloning vectors and mammalian expression vectors in origin using the NCBI BLAST server and using the Addgene sequence analyser [3].

Assembly of full-length sequence of cloning vectors from [SRX7756765](#)

Contigs were analysed using the Addgene sequence analyser [3] and overlapping ends were identified. These overlapping ends are identified as being contiguous, and their order in the original sequence was deduced using Pair-end information from the [SRX7756765](#) dataset. Specifically, spots matching the overlapping regions were obtained from [SRX7756765](#) and the paired-end reads were aligned against each other in each spot and merged. Overlapping spots are obtained which fully bridges the assembly-related inconsistencies at the ends of each contig, which were then joined together into a single contiguous sequence using this information.

References

[1] Masoudi-Nejad A, Tonomura K, Kawashima S, Moriya Y, Suzuki M, Itoh M, Kanehisa M, Endo T, Goto S. EGAssembler: online bioinformatics service for large-scale processing, clustering and assembling ESTs and genomic DNA fragments. *Nucleic Acids Res.* 2006 Jul 1;34(Web Server issue):W459-62. doi: 10.1093/nar/gkl066. PMID: 16845049; PMCID: PMC1538775.

[2] Zheng Zhang, Scott Schwartz, Lukas Wagner, and Webb Miller (2000), "A greedy algorithm for aligning DNA sequences", *J Comput Biol* 2000; 7(1-2):203-14.

[3] <https://www.addgene.org/analyze-sequence/>

[4] Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou JJ, Li N, Guo Y, Li X, Shen X, Zhang Z, Shu F, Huang W, Li Y, Zhang Z, Chen RA, Wu YJ, Peng SM, Huang M, Xie WJ, Cai QH, Hou FH, Chen W, Xiao L, Shen Y. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature.* 2020 Jul;583(7815):286-289. doi: 10.1038/s41586-020-2313-x. Epub 2020 May 7. PMID: 32380510.

Appendix: FASTA sequence of the assembled vector backbone isolated from

PRJNA607174

>vector1

```
CTTGCGTCGTCAGCAATCTGCTCCCTATCACCGAGACCTCCAGCGCTAAGGAGGAGACCAGCCCTATTAAGGA
AGAGACCTTACCGAGACCGAATTCTGCAGATATCCAGCACAGTGGCGGCCGCTCGAGGACTACAAGGACGA
CGATGACAAGTCTAGAGGGCCGTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCT
GTTGTTTGGCCCTCCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTCTTAATAAAATGAG
```

GAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGG
GGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAGGCGGAAAGAA
CCAGTGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCA
GCAAAAGGCCAGGAACCGTAAAAAGGCCGCTTGGCTGGCGTTTTTCCATAGGCTCCGCCCCCTGACGAGCA
TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCCCT
GGAAGCTCCCTCGTGCCTCTCCTGTTCCGACCCTGCCGTTACCGGATACCTGTCCGCCTTTCTCCCTCGGG
AAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCCGCTCCAAGCTGGGC
TGTGTGCACGAACCCCCGTTACGCCGACCCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGCT
AAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGC
TACAGAGTTCCTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTG
AAGCCAGTTACCTTCGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGT
TTTTTGTTCGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGG
GTCTGACGCTCAGTGGAAACGAAACTCACGTTAAGGGATTTTGGTATGAGATTATCAAAAAGGATCTTCACC
TAGATCCTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACTGAGGCTATGGCAGG
GCCTGCCGCCCCGACGTTGGCTGCGAGCCCTGGGCCTTACCCGAACCTGGGGGGTGGGGTGGGGAAAAGG
AAGAAACGCGGGCGTATTGGCCCAATGGGGTCTCGGTGGGGTATCGACAGAGTGCCAGCCCTGGGACCGA
ACCCCGCTTTATGAACAAACGACCCAACACCGTGCCTTTTATTCTGTCTTTTATTGCCGTCATAGCGCGGGT
TCCTCCGGTATTGTCTCCTCCGTGTTTCAGTTAGCCTCCCCCTAGGGTGGGCGAAGAACTCCAGCATGAGAT
CCCCCGCTGGAGGATCATCCAGCCGGCGTCCCGAAAACGATTCCGAAGCCCAACCTTTCATAGAAGGCGG
CGGTGGAATCGAAATCTCGTATGGCAGGTTGGGCGTGCCTTGGTCCGTCATTTGAAACCCAGAGTCCCGCT
CAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCAC
GAGGAAGCGGTCAGCCATTGCGCGCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAG
CGGTCCGCCACACCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCA
AGCAGGCATCGCCATGGGTCACGACGAGATCCTCGCCGTCGGGCATGCTCGCCTTGAGCCTGGCGAACAGTT
CGGCTGGCGGAGCCCTGATGCTCTTCGTCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACG
TGCTCGCTCGATGCGATGTTTCGCTTGGTGGTGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCG
CATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCAC
TTCGCCAATAGCAGCCAGTCCCTCCCGCTTCAGTGACAACGTCGAGCACAGCTGCGCAAGGAACGCCCGTC
GTGGCCAGCCACGATAGCCGCGCTGCCTCGTCTTCGAGTTCATTCAGGGCACCGGACAGGTCCGGTCTTGACAA
AAAGAACCGGGCGCCCTGCGCTGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCC
AGTCATAGCCGAATAGCCTCTCCACCAAGCGGCCGAGAACCTGCGTGCAATCCATCTTGTTCAATCATGCG
AAACGATCCTCATCCTGTCTCTTGATCGATCTTTGCAAAAGCCTAGGCCTCAAAAAAGCCTCCTCACTACTTCT
GGAATAGCTCAGAGGCCGGGCTCGGCCTCTGCATAAATAAAAAAATTAGTCAGCCATGGGGCGGAGAAT
GGGCGAACTGGGCGGAGTTAGGGGCGGGATGGGCGGAGTTAGGGGCGGGACTATGGTTGCTGACTAATT
GAGATGCATGCTTTCGATACTTCTGCCTGCTGGGGAGCCTGGGGACTTTCCACACCTGGTTGCTGACTAATTG
AGATGCATGCTTTCGATACTTCTGCCTGCTGGGGAGCCTGGGGACTTTCCACACCCTAACTGACACACATTCCA
CAGCTGGTTCCTTCCGCTCAGGACTCTTCTTTTTCAATAAATCAATCTAAAGTATATATGAGTAACTTGGTC
TGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTTCGTTCCATAGTTGCCTG
ACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGA
GACCACGCTCACCGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGT
CCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCCCGAGTTAA
TAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACGCTCGTCTTGGTATGGCTTCATTCA
GCTCCGTTCCCAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGG
TCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTC
TACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGT
ATGCGGCGACCGAGTTGCTCTTCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAAGTTAAAA
GTGCTCATCATTGAAAACGTTCTTCGGGGCGAAAACCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGA
TGTAACCCACTCGTGCACCCAAGTATCTTCAGCATCTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAAACA

GGAAGGCAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTCCTTTTT
CAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGATTTAGAAAAATAA
ACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGCGCCCTGTAGCGGCGCATTAAAGCGC
GGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTGCTTTT
TTCCCTTCTTTCTGCCACGTTGCGCCGGCTTTCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGGTTCCGA
TTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCT
GATAGACGGTTTTTTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTCCAACTGGAACA
ACACTCAACCCTATCTCGGTCTATTCTTTGATTATAAGGGATTTTGGCGATTTGCGCCTATTGGTTAAAAAAT
GAGCTGATTTAACAAAAATTTAACGCGAATTTAACAAAAATTAACGCTTACAATTTACGCGCGCGTTGACAT
TGATTATTGACTAGTTAATAAGTAATCAATTACGGGGTTCATTAGTTCATAGCCCATATATGGAGTTCCGCGT
TACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCAACGACCCCGCCATTGACGTCAATAATGACG
TATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCC
ACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGC
CTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTAT
TACCATGGTGATGCGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGTTTTGACTCACGGGGATTTCCAAGT
CTCCACCCATTGACGTCAATGGGAGTTTTTTTTGGCACAAAATCAACGGGACTTTCCAAAATGTCGTAACAA
CTCCGCCCCATTGACGCAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCT
AACTAGAGAACCCTGCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTTGGTAC
CGAGCTCGGATCCACTAGTAACGGCCCGCAGTGTGCTGGAATTC

>vector2

CTCGTGGTCTCAGCCTCAGCAATCAACTTTCTGCTGCTCCAGAACGGGTGACCCGTGCTGAGATACTCAACAGT
GAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAAACGGCC
ACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCA
CCACCGCAAGCTGCCCGTGCCTGGCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTACGCCG
CTACCCCGACCACATGAAGCAGCACGACTTCTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC
ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAAC
CGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAATA
CAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACCTCAAGATCCGCCA
CAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGT
GCTGCTGCCCGACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCA
CATGGTCTGCTGGAGTTCGTGACCGCCCGGGGATCACTCTCGGCATGGACGAGCTGTACAAGGTGACGGA
ACAAAACTCATCTCAGAAGAGGATCTGAATGCTGTGGGCCAGGACACGCAGGAGGTCATCGTGGTGCCACA
CTCCTTGCCCTTTAAGGTGGTGGTGTCTCAGCCATCCTGGCCCTGGTGGTGTCTACCATCATCTCCCTTATCAT
CCTCATCATGCTTTGGCAGAAGAAGCCACGTTAGGCGGCCGCTCGAGATCAGCCTCGACTGTGCCTTCTAGTT
GCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCT
AATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTATTCTATTCTGGGGGGTGGGGTGGGGCAGG
ACAGCAAGGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGCTTCTGAG
GCGGAAAGAACCAGTGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAG
CAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGTCCACAGAATCAGGGGATAACGCA
GGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCCTTTTT
CCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGG
ACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGTTACCG
GATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCG
GTGTAGGTCGTTGCTCCAAGCTGGGCTGTGTGCACGAACCCCCGTTACGCCGACCGCTGCGCCTTATCCG
GTAACATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGAT
TAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
AACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGC

AAACAAACCACCGCTGGTAGCGGTGGTTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTC
AAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTAAATTAATAAATGAAGTTTTAAATCAATCTAAAGTA
TATATGAGTAACCTGAGGCTATGGCAGGGCTGCCGCCCGACGTTGGCTGCGAGCCCTGGGCCTCACCCG
AACTTGGGGGGTGGGGTGGGGAAAAGGAAGAAACGCGGGCGTATTGGCCCAATGGGGTCTCGGTGGGGT
ATCGACAGAGTGCCAGCCCTGGGACCGAACCCCGCTTTATGAACAAACGACCCAACACCGTGCCTTTTATTC
TGCTTTTTATTGCCGTCATAGCGCGGGTTCCTTCCGGTATTGTCTCCTTCCGTGTTTCAGTTAGCCTCCCCCTAG
GGTGGGCGAAGAACTCCAGCATGAGATCCCCGCGCTGGAGGATCATCCAGCCGGCGTCCCGGAAAACGATTC
CGAAGCCCAACCTTTCATAGAAGGCGGGCGGTGGAATCGAAATCTCGTGATGGCAGGTTGGGCGTGCCTTGGT
CGGTCATTTGAAACCCAGAGTCCCCTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGA
ATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTGAGCCATTGCGCCGAAGCTCTTACGAATATC
ACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAA
GCGGCCATTTCCACCATGATATTCGGCAAGCAGGCATCGCCATGGGTACACGACGAGATCCTCGCCGTCGGGC
ATGCTCGCCTTGAGCCTGGCGAACAGTTCCGGTGGCGCGAGCCCCTGATGCTCTTCCGTCCAGATCATCCTGAT
CGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCGATGCGATGTTTCGCTTGGTGGTCAATGGGCAGGT
AGCCGGATCAAGCGTATGCAGCCGCCGATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTG
AGATGACAGGAGATCCTGCCCCGGCACTTCGCCAATAGCAGCCAGTCCCTTCCCGCTCAGTGACAACGTCG
AGCACAGCTGCGCAAGGAACGCCGTCGTGGCCAGCCACGATAGCCGCGCTGCCTCGTCTTGACGTTCAATCA
GGCACCCGGACAGGTCGGTCTTGACAAAAGAACCGGGCGCCCCTGCGCTGACAGCCGGAACACGGCGGCA
TCAGAGCAGCCGATTGTCTGTTGTGCCAGTCATAGCCGAATAGCCTCTCCACCCAAGCGGCCGGAGAACCTG
CGTGCAATCCATCTTGTTCAATCATGCGAAACGATCCTCATCCTGTCTCTTGATCGATCTTTGAAAAGCCTAGG
CCTCCAAAAAGCCTCCTCACTACTTCTGGAATAGCTCAGAGGCCGGCCTCGGCCTCTGCATAAATAAAAAA
AATTAGTCAGCCATGGGGCGGAGAATGGGCGGAACTGGGCGGAGTTAGGGGCGGGATGGGCGGAGTTAGG
GGCGGGACTATGGTTGCTGACTAATTGAGATGCATGCTTGCATACTTCTGCCTGCTGGGGAGCCTGGGGACT
TTCCACACCTGGTTGCTGACTAATTGAGATGCATGCTTGCATACTTCTGCCTGCTGGGGAGCCTGGGGACTTT
CCACACCTAACTGACACACATTCCACAGCTGGTCTTTCCGCCCTCAGGACTCTTCTTTTTCAATAAATCAATCT
AAAGTATATATGAGTAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGT
CTATTTGTTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGG
CCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCC
GGAAGGGCCGAGCGCAGAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAATTGTTGCCGGGAAG
CTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCACAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTACG
CTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCTAACGATCAAGGCGAGTTACATGATCCCCATGTTGT
GCAAAAAGCGGTTAGCTCCTTCCGTCTCCGATCGTTGTGAGAAGTAAGTTGGCCGAGTGTATCACTCAT
GGTTATGGCAGCACTGCATAATTCTTACTGTGATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGTGACT
CAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGGCCGGCGTCAATACGGGATAATAC
CGCGCCACATAGCAGAACTTAAAGTGCTCATCATTGGAACGTTCTTCCGGGGCGAAAACCTCAAGGATC
TTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAATGATCTTCAGCATCTTTACTTTACC
AGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGCGACACGGAAAT
GTTGAATACTCATACTCTTCTTTTTCAATATTATTGAAGCATTATCAGGGTATTGTCTCATGAGCGGATACA
TATTTGAATGATTTAGAAAAATAAACAATAGGGGTTCCGCGCACATTTCCCGAAAAGTGCCACCTGACGC
GCCCTGTAGCGGCGCATTAAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGC
CCTAGCGCCCCTCTTTCTGCTTTCTTCCCTCCTTTCTCGCCACGTTCCCGGCTTTCCCGTCAAGCTCTAAAT
CGGGGGCTCCCTTAGGGTCCGATTTAGTGCTTACGGCACCTCGACCCCAAAAACCTTGATTAGGGTGATG
GTTACGTAAGTGGCCATCGCCCTGATAGACGGTTTTTCCGCCCTTGGAGTCCACGTTCTTTAATAGT
GGACTCTTGTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTGATTTATAAGGGATTTGGCC
ATTTCCGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAATTTAACGCGAATTTAACAAAATATTAACGCT
TACAATTTACGCGCGCGTTGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCA
TAGCCCATATATGGAGTTCGCGGTTACATAACTTACGGTAAATGGCCCGCTGGCTGACCGCCCAACGACCCC

CGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGT
GGAGTATTACGGTAAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCATTGAC
GTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGT
ACATCTACGTATTAGTCATCGCTATTACCATGGTGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCG
GTTTGACTCACGGGGATTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTGGCACCAAAATCAAC
GGGACTTTCAAAATGTCGTAACAACTCCGCCCATGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGG
TCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCTGCTTACTGGCTTATCGAAATTAATACGACTCAC
TATAGGGAGACCCAAGCTTGGTACCGAGCTCGGATCCACTAGTAACGGCCGCCAGTGTGCTGGAATTCGGCT
TGGGGATATCCACCATGGAGACAGACACTCTGCTATGGGTACTGCTGCTCTGGGTTCCAGGTTCCACTGG
TGACTATCCATATGATGTTCCAGATTATGCTGGGGCCAGCCGGCCAGATCTCCCGGGATCCGCGGCATGCAT
CATCACCATCACCACGCATCTGGGGGAGCATTCTGTCTGATCGCCAACGAC

>ASFV-like cassette

TTCCTTTTTCAATATTATTGAAGCATTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAG
AAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTA
TCATGACATTAACCTATAAAAAATAGGCGTATCACGAGGCCGCCCTGCAGCCGAATTATATTATTTTTGCCAAA
TAATTTTTAACAAAAGCTCTGAAGTCTTCTTCATTTAAATTCTTAGATGATACTTCATCTGGAAAATTGTCCCAA
TTAGTAGCATCACGCTGTGAGTAAGTCTAAACCATTTTTTATTGTTGTATTATCTCTAATCTTACTACTCGATG
AGTTTTCGGTATTATCTCTATTTTTAACTTGGATAAGGTTCAATTCATTGTTTTTTTCATCATAGTGAATAAAATC
AACTGCTTTAACACTTGTGCCTGAACACCATATCCATCCGGCGTAATACGACTCACTATAGGGAGAGCGGCCG
CCAGATCTTCCGGATGGCTCGAGTTTTTCAGCAAGATCATACTCAGAATGCCTATTATATTTGTTGAATTGTAA
GTGGCAACATAATGGCTTATTTTTTATGATGAGGTTGATGGAAAGGTTTAGGTGAGACAGTAGTAAGTTCAAGG
TTTTCTGGAAGATTTTGCTTTATCGCCACTAGTAAACATTGTTCTATCTGTGCGTGGGTTGATTGAGGAATG
AAATAGATTAATAATTTTTCTATGGTTATAGGTTGCAGCTTGTGTTAGAATAATATTTATATTATCTTTATTATGA
TCATACTTTAAGGTTTCTAGGATAAAAATCTAAGTATCCTTTGGCTGCCACCTTATGTAGGAGTTGTAGGCTTGA
AAGCATGCGAATATCTAAATGATGATACTTATCGGGTACGATTCTGTTATATATAAATCTATACCCCAAACAAT
ATAGATGTAGATCCTTATGGGCAATAGCACAATCTTTCTAGAAGATCTCCTACAATATTCTCAGCTGCCATGGA
AAATCGATGTTCTTCTTTTATTCTCTCAAGATTTTCAGGCTGTATATTAACCTTATATTAAGAACTATGCTAACC
ACCTCATCAGGAACCGTTGTAGGTGGCGTGGGTTTTCTTGGCAATCGACTCTCATGAAAACCTACGAGCTAAAT
ATTCAATATGTTCTCTTGACCAACTTTATTCTGCATTTTTTTGAACGAGGTTTAGAGCAAGCTTCAGGAAACT
GAGACAGGAATTTTATTAATAATTTAAATTTGAAGAAAGTTCAGGGTTAATAGCATCCATTTTTTCTTTGCA
AGTTCCTCAGCATTCTTAACAAAAGACGTCTCTTTTGACATGTTTAAAGTTTAAACCTCCTGTGTGAAATTATTA
TCCGCTCATAATTCCACACATTATACGAGCCGGAAGCATAAAGTGTAAGCCTGGGGTGCCTAATGAGTGAGC
TAACTCACATTAATTGCGTTGCGCTCACTGCCAATTGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTA
ATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCATTGGGCGCTCTCCGCTTCTCGCTCACTGACTCG
CTGCGCTCGGTGTTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGTTATCCACAGAAT
CAGGGGATAACGCAGGAAAGAACATGTGAGCC