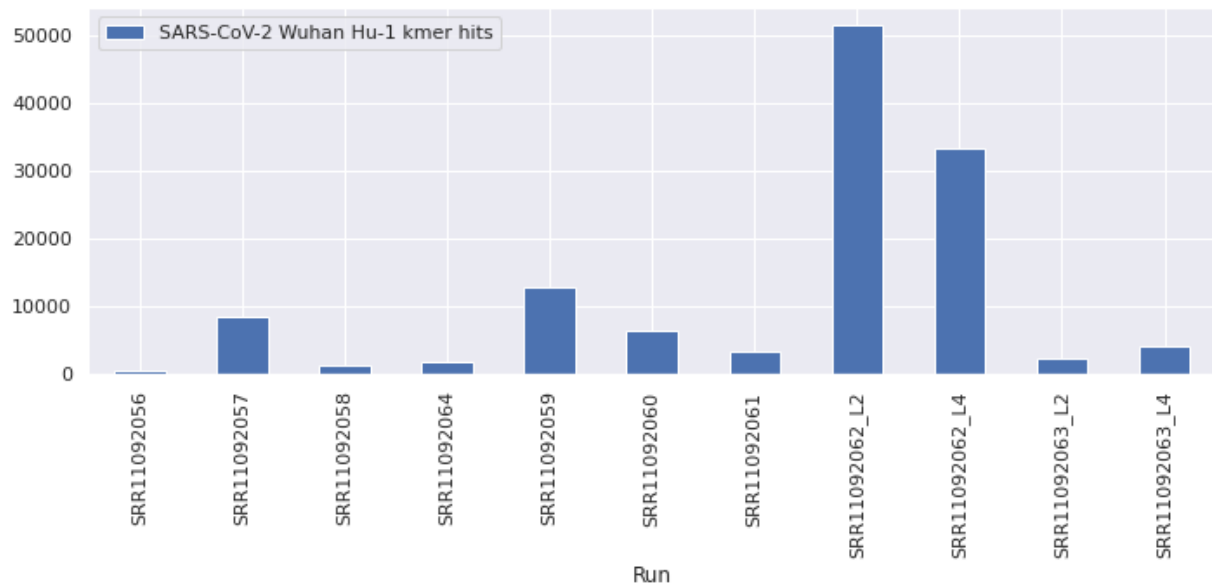


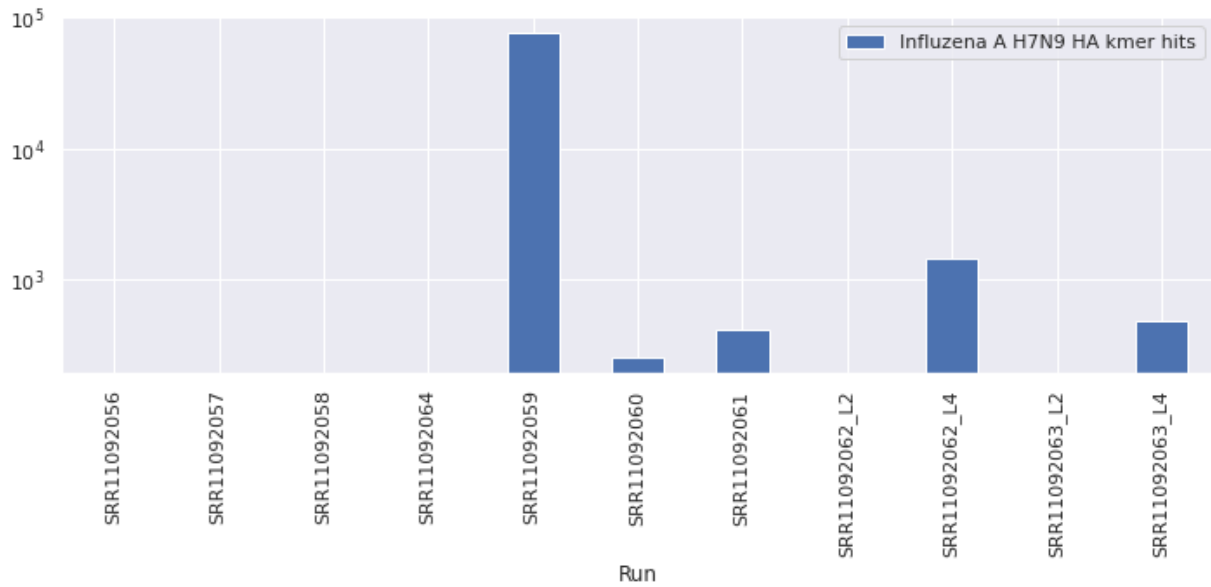
SUPPLEMENTAL ANALYSES

Fastv analysis

The number of kmer hits to SARS-CoV-2 and the Influenza A H7N9 HA gene identified using fastv for analysis of each SRA in PRJNA605983 are shown in Supp. Figs. 1 and 2 below. Note that for SRR11092062 and SRR11092063 Lanes 02 and 04 are plotted separately.



Supp. Fig. 1. SARS-CoV-2 Wuhan Hu-1 genome kmer hits identified in fastv analysis of SRA datasets in PRJNA605983.

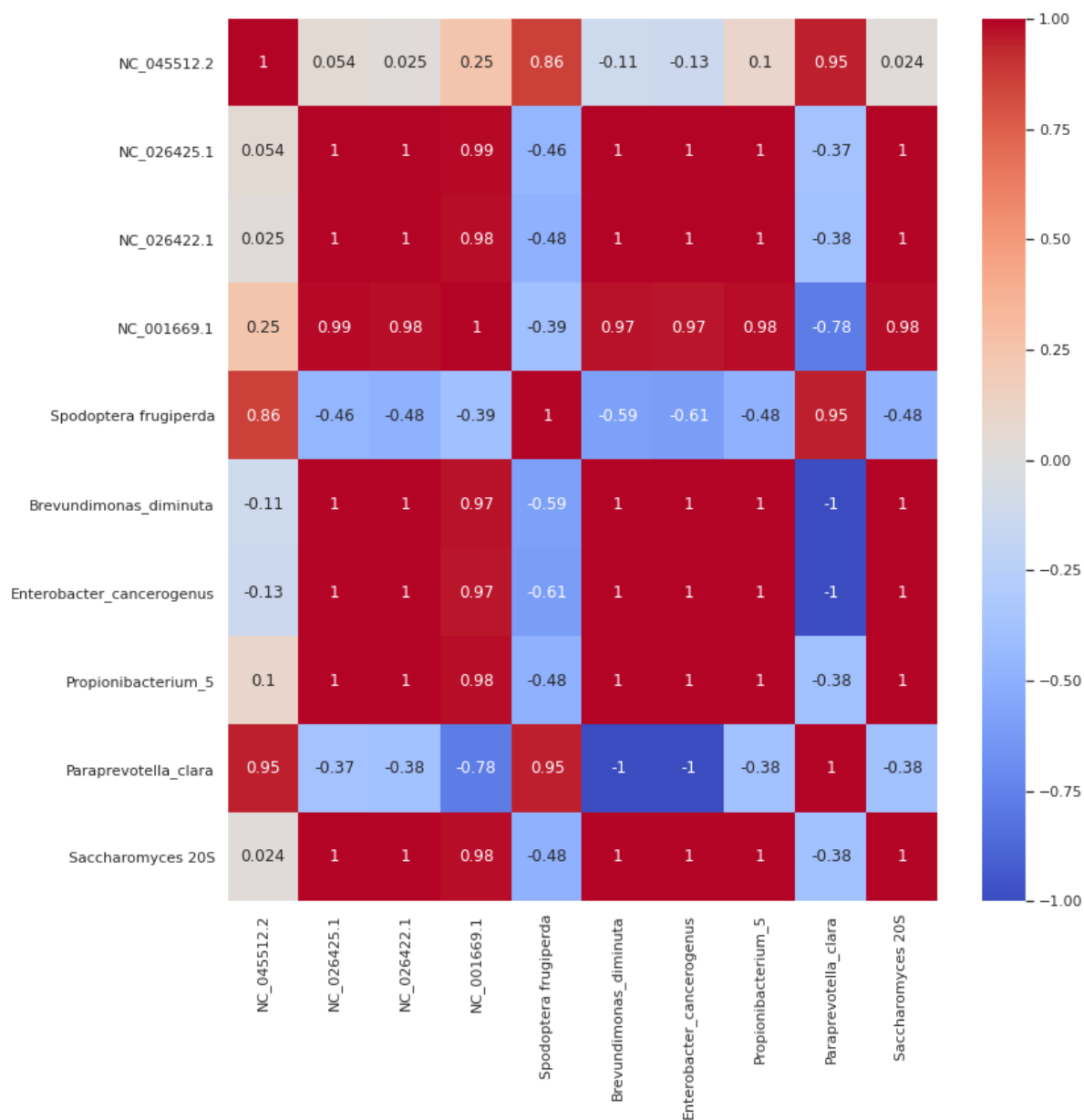


Supp. Fig. 2. Influzena A H7N9 HA gene K-mer hits identified in fastv analysis of SRA datasets in PRJNA605983 (log scale).

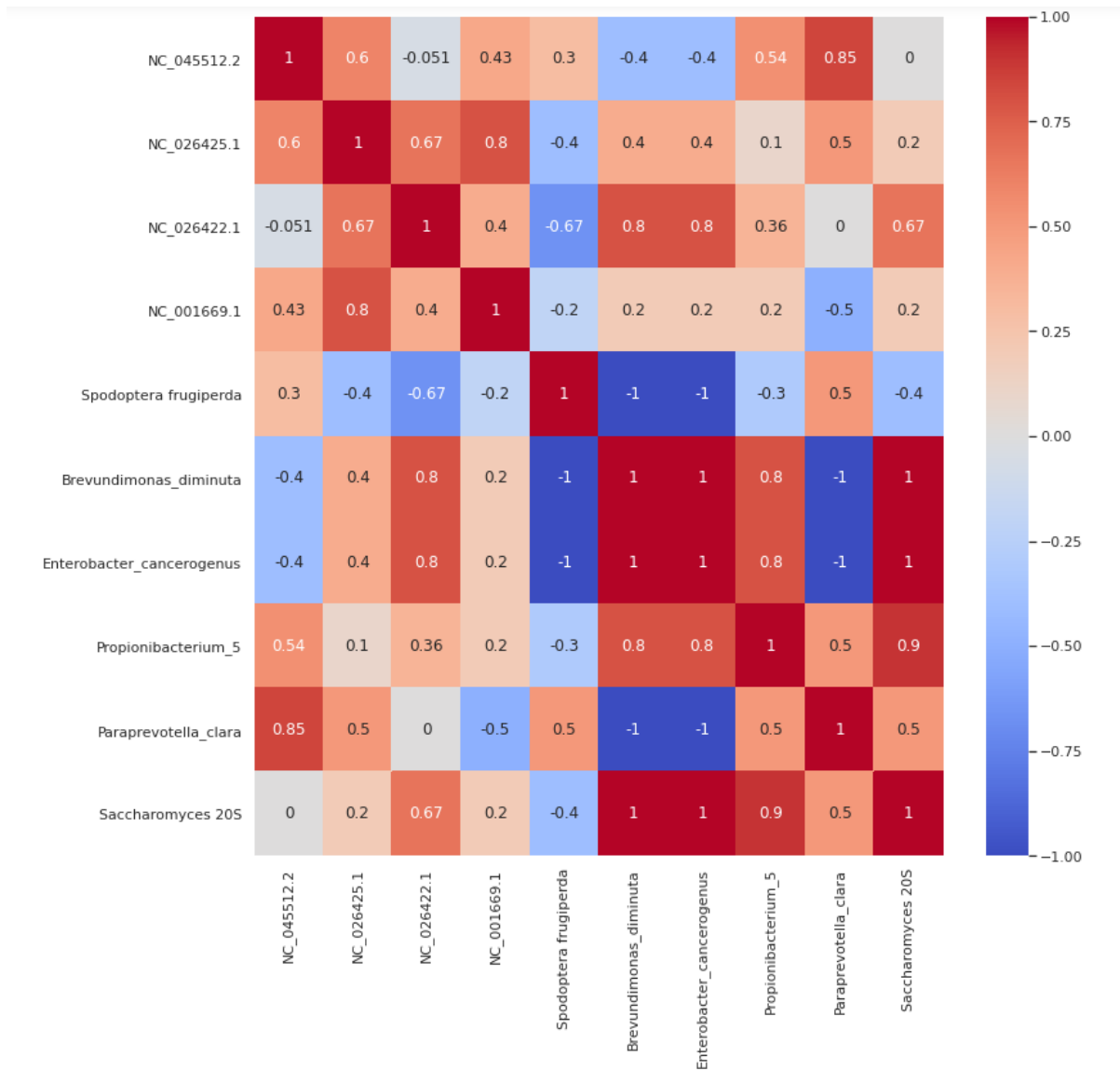
GenBank URL	GenBank Biosample URL	SRA Run Size	GISAID ID	CoV-2 Isolate
SRX7730879	SAMN14082200	3.6M spots, 1G bases, 548.1Mb	EPI_ISL_402130	WIV07; Lineage B; mutations NSP3 D1761A, NSP4 T327I; passage original
SRX7730880	SAMN14082196	67.1M spots, 20.1G bases, 12.6Gb	EPI_ISL_402127	WIV02; Lineage B; mutations NSP16 D220N; passage original
SRX7730881	SAMN14082197	61.3M spots, 18.4G bases, 11.4Gb	EPI_ISL_402124	WIV04; Lineage B; no mutations; passage original
SRX7730882	SAMN14082198	34.3M spots, 10.3G bases, 6.4Gb	EPI_ISL_402128	WIV05; Lineage B; NSP3 G1433S, NSP16 K160R; passage original

SRX7730883	SAMN14082199	29.7M spots, 8.9G bases, 5.6Gb	EPI_ISL_402129	WIV06; Lineage B; no mutations; original passage
SRX7730884	SAMN14082200	38.5M spots, 11.5G bases, 7.1Gb	EPI_ISL_402130	WIV07; Lineage B; mutations NSP3 D1761A, NSP4 T327I; passage original
SRX7730885	SAMN14082196	8.3M spots, 2.2G bases, 1.2Gb	EPI_ISL_402127	WIV02; Lineage B; mutations NSP16 D220N
SRX7730886	SAMN14082197	5.2M spots, 1.5G bases, 768.3Mb	EPI_ISL_402124	WIV04; Lineage B; no mutations; passage original
SRX7730887	SAMN14082199	5.2M spots, 1.5G bases, 772.9Mb	EPI_ISL_402129	WIV06; Lineage B; no mutations; original passage
SRX8032202	SAMN14479127	5.2M spots, 1.6G bases, 583.4Mb	EPI_ISL_412898	hCoV-19/Wuhan/HBCDC-HB-02/2019
SRX8032203	SAMN14479128	5.8M spots, 1.7G bases, 634.3Mb	EPI_ISL_402132	Wuhan HBCDC-HB-01/2019; Lineage B; mutation Spike F32I; original passage

Supplementary Table/Fig 3



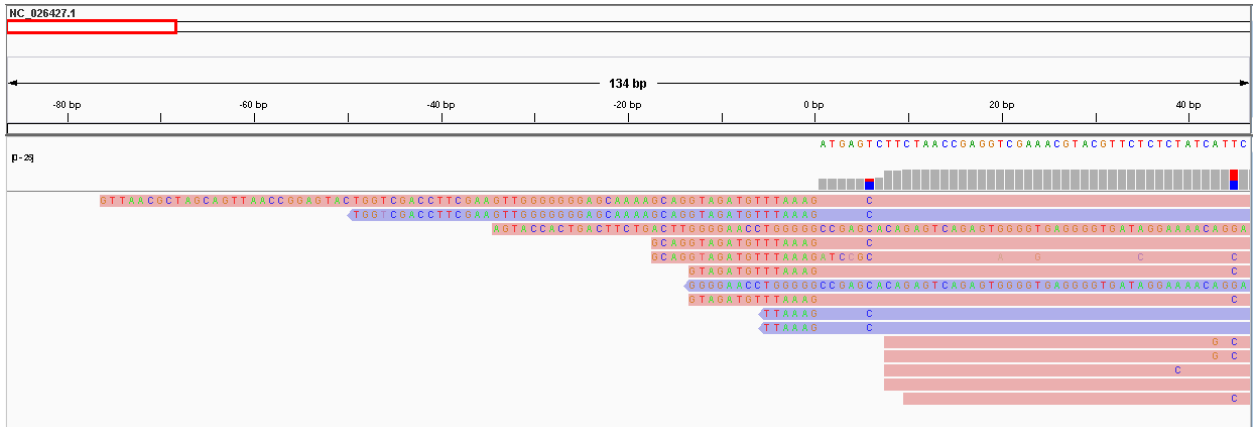
Supp. Fig. 4. Pearson correlation matrix plot for selected viruses and microbes from SRR11092059. Key: NC_045512.2: Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome; NC_026425.1: Influenza A virus (A/Shanghai/02/2013(H7N9)) segment 4 hemagglutinin (HA) gene, complete cds; NC_026422.1: Influenza A virus (A/Shanghai/02/2013(H7N9)) segment 1 polymerase PB2 (PB2) gene, complete cds; NC_001669.1: Simian virus 40 complete genome; Spodoptera: NC_025382.1 Spodoptera frugiperda rhabdovirus isolate Sf, complete genome; Brevundimonas: Brevundimonas_diminuta_470_4_uid67195; Enterobacter: Enterobacter_cancerogenus_ATCC_35316_uid28663; Propionibacterium: Propionibacterium_5_U_42AFAA_uid39413; Paraprevotella: Paraprevotella_clara_YIT_11840_uid48501; Saccharomyces: NC_004051.1 Saccharomyces 20S RNA narnavirus, complete genome



Supp. Fig. 5. Spearman correlation matrix plot for selected viruses and microbes from SRR11092059. Key: as per Supp. Fig. 4.

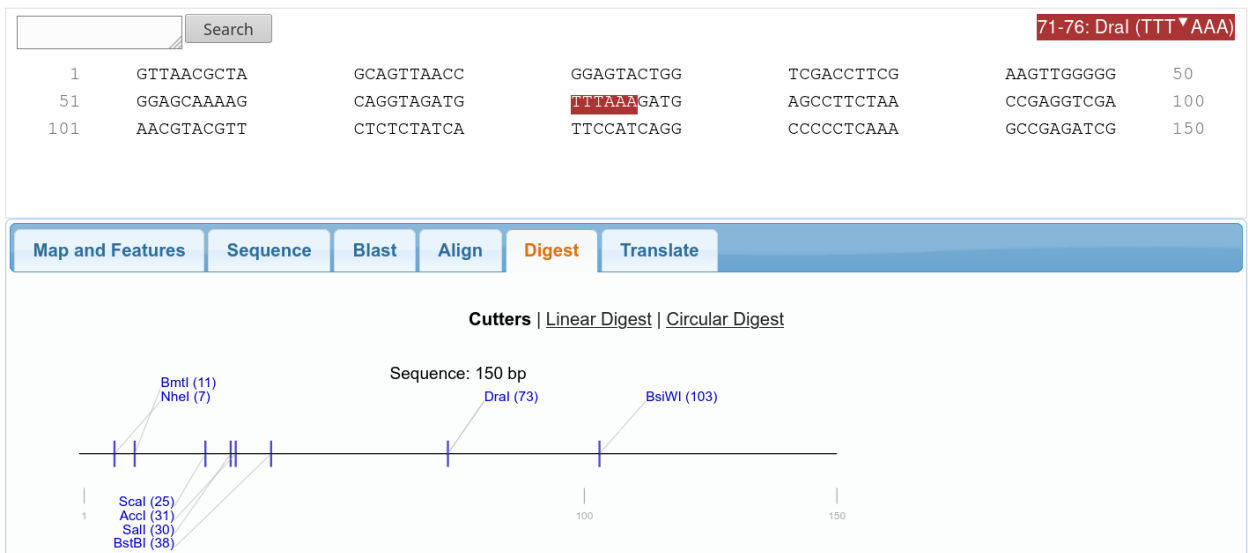
H7N9 M gene

Raw reads from SRR11092059 were aligned to NC_026427.1 (H7N9 M gene) using bwa mem with default parameters. As shown in the figure below, several reads extend upstream of the 3' end of the Influenza A H7N9 M gene (NC_026427.1).



Supp. Fig. 6. 3' most end of SRR11092059 reads aligned to NC_026427.1, plotted in IGV. The M gene sequence starts at 0bp. Mismatched bases only are plotted with letters.

Read v300043428L4C001R0371251734/2 (top read in figure above) was then plotted in Addgene sequence analyzer (<https://www.addgene.org/analyze-sequence/>). A HincII restriction enzyme site at the 3' end (GTTAAC) and a DraI restriction enzyme site (TTTAAA) was identified at positions -7 to -2bp relative to the 3' end of the M gene (Supp. Fig. 7). The entire 77nt sequence upstream of the M gene was then analysed using NCBI BLASTN against the nt database and was found to exhibit a high [similarity/homology](#) to H5N1 segment 8 with one nt insertion and one substitution (Supp. Figs 8 and 9)



Supp. Fig. 7. Read v300043428L4C001R0371251734/2 plotted in addgene. The DraI site at the intersection of the H5N1 sement8 section and H7N9 M gene highlighted in red.

Job Title **Nucleotide Sequence**

RID [B9KTJYKC01R](#) Search expires on 06-01 19:57 pm [Download All](#) ▼

Program **BLASTN** [Citation](#) ▼

Database **nt** [See details](#) ▼

Query ID **lc|Query_502797**

Description **None**

Molecule type **dna**

Query Length **77**

Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

Filter Results

exclude

Organism only top 20 will appear

[+ Add organism](#)

Percent Identity to

E value to

Query Coverage to

[Filter](#) [Reset](#)

Descriptions | [Graphic Summary](#) | [Alignments](#) | [Taxonomy](#)

Sequences producing significant alignments [Download](#) ▼ [New](#) [Select columns](#) ▼ Show [?](#)

select all 1 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [New](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Influenza A virus (A/chicken/Guangdong/1/2005(H5N1)) segment 8 nuclear export protein (NEP) and nonstructural...	Influenza A virus...	95.3	95.3	75%	4e-16	96.55%	899	EU874904.2

Supp. Fig. 8. NCBI BLASTN analysis results for the 77bp sequence upstream of the H7N9 gene queried against the nt database with default parameters.

Influenza A virus (A/chicken/Guangdong/1/2005(H5N1)) segment 8 nuclear export protein (NEP) and nonstructural protein 1 (NS1) genes, complete cds

Sequence ID: [EU874904.2](#) Length: 899 Number of Matches: 1

Range 1: 1 to 57 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps	Strand
95.3 bits(51)	4e-16	56/58(97%)	1/58(1%)	Plus/Plus

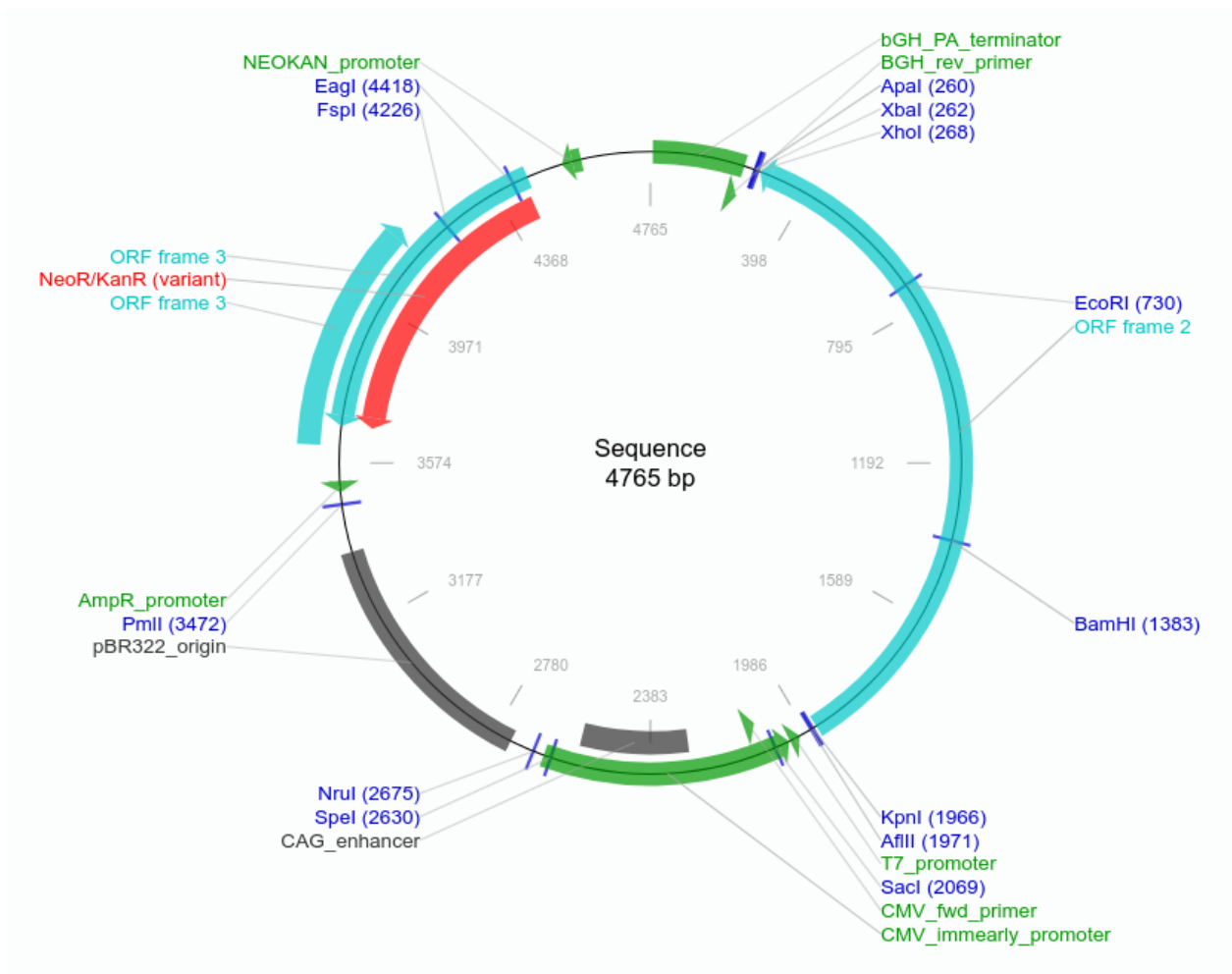
```

Query 7  GCTAGCAGTTAACCGGAGTACTGGTCGACCTTCGAAGTggggggAGCAAAAGCAGG 64
Sbjct 1  GCTAGCAGTTAACCGGAGTACTGGTCGACCTTCGAAGT-GGGGGAGCAAAAGCAGG 57
  
```

Supp. Fig. 9. NCBI BLASTN analysis results for the 77bp sequence upstream of the H7N9 gene queried against the nt database with default parameters showing alignment sequence.

pVAX1

A complete pVAX1 plasmid sequence was identified from WIV07-2 SRA dataset SRR11092059 as a single contig (Supp. Fig. 10), and was also identified in WIV04-4 (SRR11092062) and WIV02-2 (SRR11092063).



Supp. Fig. 10. pVAX1 plasmid containing Influenza A H7N9 HA gene from de novo assembly of SRR11092059 and plotted in addgene.

The ORF 2 region (274-1956bp) was analyzed using NCBI BLASTN for alignment to the nt database and 100% homology was found to Influenza A virus (A/Shanghai/02/2013(H7N9)) segment 4 hemagglutinin (HA) gene, complete cds KF021597.1 and Influenza A virus (A/chicken/Guangdong/G2/2013(H7N9)) segment 4 hemagglutinin (HA) gene, complete cds KJ395948.1 and synthetic constructs (Supp. Fig. 11).

Job Title Nucleotide Sequence

RID [CPN1C6UH013](#) Search expires on 06-18 21:54 pm [Download All](#) ▼

Program BLASTN [Citation](#) ▼

Database nt [See details](#) ▼

Query ID lcl|Query_62817

Description None

Molecule type dna

Query Length 1683

Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

Filter Results

Organism only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to **E value** to **Query Coverage** to

[Filter](#) [Reset](#)

Descriptions | [Graphic Summary](#) | [Alignments](#) | [Taxonomy](#)

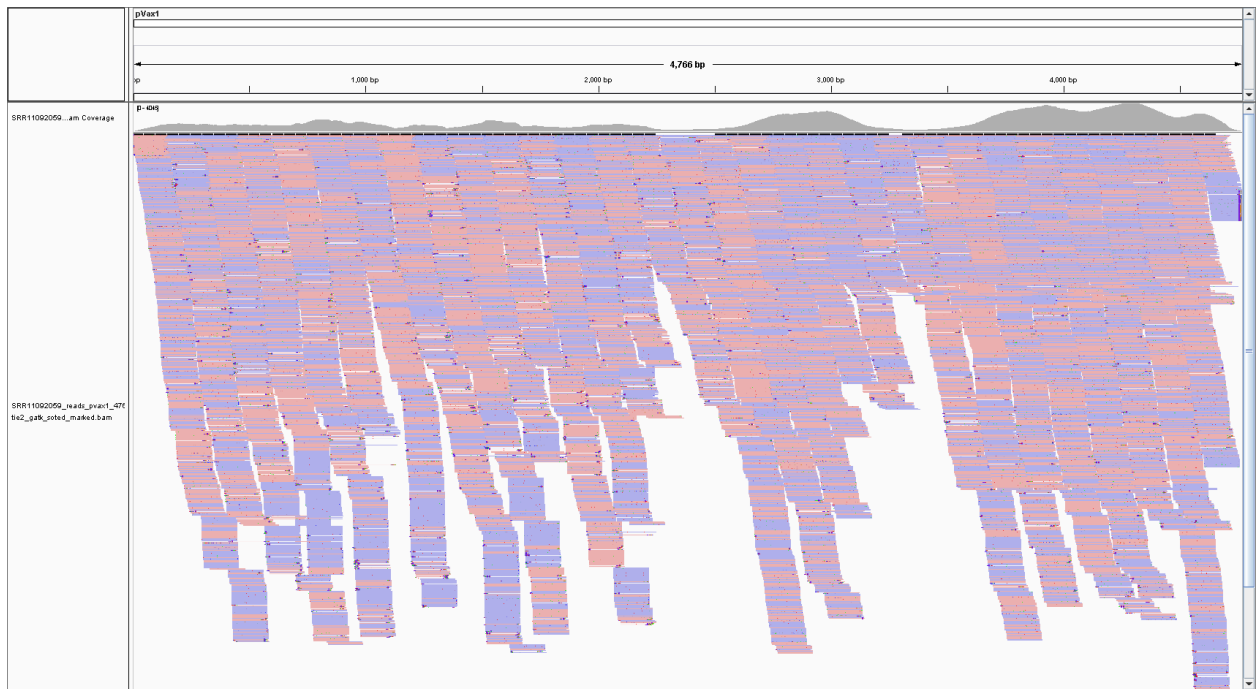
Sequences producing significant alignments Download ▼ [New](#) Select columns ▼ Show [?](#)

select all 100 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [New](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Synthetic construct H7N9 HA gene, complete cds	synthetic construct	3109	3109	100%	0.0	100.00%	9599	KY199425.1
<input checked="" type="checkbox"/>	Influenza A virus (A/chicken/Guangdong/G2/2013(H7N9)) segment 4 hemagglutinin (HA) gene, complete cds	Influenza A virus ...	3109	3109	100%	0.0	100.00%	1683	KJ395948.1
<input checked="" type="checkbox"/>	Influenza A virus (A/Shanghai/02/2013(H7N9)) segment 4 hemagglutinin (HA) gene, complete cds	Influenza A virus ...	3109	3109	100%	0.0	100.00%	1708	KF021597.1
<input checked="" type="checkbox"/>	Synthetic construct HA-A/Shanghai/2/2013_R3delHA_H7N9 gene, complete cds	synthetic construct	3109	3109	100%	0.0	100.00%	1683	MW298210.1
<input checked="" type="checkbox"/>	Influenza A virus (A/chicken/Guangdong/SD014/2014(H7N9)) segment 4 hemagglutinin (HA) gene, complete cds	Influenza A virus	3103	3103	100%	0.0	99.94%	1727	MN037516.1
<input checked="" type="checkbox"/>	Influenza A virus (A/Anhui/1-YK_RG01/2013(H7N9)) hemagglutinin (HA) gene, complete cds	Influenza A virus ...	3103	3103	100%	0.0	99.94%	1708	CY191659.1
<input checked="" type="checkbox"/>	Influenza A virus (A/Anhui/1-JCV11_RG1/2013(H7N9)) hemagglutinin (HA) gene, complete cds	Influenza A virus ...	3103	3103	100%	0.0	99.94%	1732	CY187410.1
<input checked="" type="checkbox"/>	Influenza A virus (A/Anhui/DEWH72-02/2013(H7N9)) hemagglutinin (HA) gene, complete cds	Influenza A virus ...	3103	3103	100%	0.0	99.94%	1732	CY181521.1
<input checked="" type="checkbox"/>	Influenza A virus (A/Zhejiang/DTID-ZJU01/2013(H7N9)) segment 4 hemagglutinin (HA) gene, complete cds	Influenza A virus ...	3103	3103	100%	0.0	99.94%	1732	KJ633809.1
<input checked="" type="checkbox"/>	Influenza A virus (A/chicken/Jiangsu/1021/2013(H7N9)) segment 4 hemagglutinin (HA) gene, complete cds	Influenza A virus ...	3103	3103	100%	0.0	99.94%	1696	KF938946.1
<input checked="" type="checkbox"/>	Influenza A virus (A/chicken/Guangdong/SD641/2013(H7N9)) hemagglutinin (HA) gene, complete cds	Influenza A virus ...	3103	3103	100%	0.0	99.94%	1683	CY146908.1

Supp. Fig. 11. NCBI BLASTN analysis on ORF 2 region of the pVax1 plasmid identified in contigs from de novo assembly of SRR11092059.

Reads from SRR11092059 were then aligned to the pVax1 plasmid (de novo assembled contig k141_31) to assess coverage and depth. 62,599 reads were found to be mapped to the pVax1 plasmid sequence as shown in Supp Fig. 12.



Supp. Fig. 12. 65299 reads from SRR11092059 were mapped to a pVax1 plasmid sequence identified from de novo assembly of SRR11092059 (contig k141_31), plotted in IGV.

Addendum

De novo assembled pVax1 vector sequence has been saved as pvax1_4765.fa in Supplementary Data

```
>pvax1 SRR11092059 k141_31
```

```
AGAAGCCATAGAGCCACCGCATCCCCAGCATGCCTGCTATTGTCTTCCAATCCTCCCCCTTGCTGTCCTGCCACCCACCCCAAGAATAGAATGACA
CCTACTCAGACAATGCGATGCAATTTCTCATTTTATTAGGAAAGGACAGTGGGAGTGGCACCTTCCAGGGTCAAGGAAGGCACGGGGAGGGGCAAAC
AACAGATGGCTGGCAACTAGAAGGCACAGTCGAGGCTGATCAGCGGGTTAAACGGGCCCTTAGACTCGAGTTATATACAATAGTGCACCCGATGTTT
CCATTTTACACATATGAAGACAAGGCCATTACAATGGCTAGAAGTATGAAACATGATGCCCGAAGCTAAACCAAAGTATCACATCTTTGTAGCCGCT
GCTTAGTTTGACTGGGCAATCTGTATTCTATTTTGCATTGCCTCTTCCCTGTATTGTGTGATCAGGTGTTATTCTAATACTGGCCATACAGTCATCAT
CACACTTGTGAAATATTTCAAAGCAACCGAGTCCATCTTCTCAGCATTCTCTCAGCTGTCTTTCACTCGTTCGTACAGTTTGTCCATTTCTGAATCAGCC
AGATCAATTGTATGCTGTTTCTCCATTGCTACCAAGAGTTCAGCATTGTATGACCACACTTCTGTTATAGAATCTCTGGTCCAATTTATCACATTACCGATT
GCTTCTACTCTCATTGAATTCATTGTCTATCAACTCAAATGTTGGTTGGTTTTTCTATAAGCCGTTTAATTTTCTGTTATTTGATCAATGCCGATTGA
GTGCTTTTGAATCTGCAGCAGTCCCTCTCCCTGTGCATTCTGGTGTCTGAAACCATACCAACCATCAATTAGGCCTTCCATCTTTCAATGAAACCCG
CTATAGCACCAAATAGGCCTCTTCTTTTGAATCTCAGGAACATTCTTATCCCTGTGCTAGCAGCAGACTCCTTTGCTTAACATATCTCGGACATTTTCC
AACTGCCCTGCTATCTATGTTCTGAAATGGCAAGTTACTTATTATTGTCCCTCACTATGATAGCAGTCCCTTCACAATTGGCATCAACCTGACTCCACTCT
GGATTCCCATAGATTTTCTCTCAGGAAGCTTGACAGGCTGGAGCTATGAAAGCCCAATTGAAACTGAAAGTGACTGTATCATTGGGATTTAGCATTAGC
CAATGAAAGTCAATTTCCAGATAGACCATTAACCTTGTGGTCTCGCTCCTGGACTCGGTACAAAAGATTGTTGATAATTAGAAGTCCCACTGTACCAGT
TTGTTTCCACTCCCATATAGCTTGGTTTGTCTGCGAGTTGATACGGAATGATGGATCCCCATACTATTAGAGCTGGGCTTTTCTGTATTTTATATGACTT
AGTCATCTGCGGGAATGCAGCATTATCTGTGTTTACAGGAGCCATTTTCTTCTGCATAGAATGAAGATCCTGATCTCTACATGCACTGTTGTCTCATT
AGTTCTTATCCACTGTATGTAATCCCATTTGCTTCTTGTCAATCCGCCTGATTCTCTGAGAATTTGCCTCAGAGCTTCTTATTACGAATTTCCAGGAT
AACAGACATCACTTCTTCTCGCCTCTCAATAAATTAATCGGCTGAAATTTAGGAATGGTACATTGAGGTGTTCCAGTGATTGTCCCAGGAGTCCAC
ATTGACCGAGGTCAACTGTCTTTTCCCTTTGAGCAGATCTGGGGATGTTTGTCTGTTCCACTGTTTCAAGTTGATTGACGACTTCCACTCTCTTTAGTT
AATGTGTTTACTTTGGTTCCGTTTACAGCGCATGATGTCGAGGCAGATTTTGTCTGCATTTGTTGGAATGATCGCAATCAGAGCGAATACCAGGATTTG
```

AGTGTTTCATGGTGGGGTACCAAGCTTAAGTTTAAACGCTAGCCAGCTTGGGTCTCCCTATAGTGAGTCGTATTAATTCGATAAGCCAGTAAGCAGTGGGT
TCTCTAGTTAGCCAGAGAGCTCTGCTTATATAGACCTCCACCGTACACGCCTACCGCCATTTCGTCATGGGGCGGAGTTGTTACGACATTTTGAAA
GTCCCGTTGATTTTGGTGCCAAAACAACTCCATTGACGTCAATGGGGTGGAGACTTGGAAATCCCGTGAGTCAAACCGCTATCCACGCCATTGATGT
ACTGCCAAAACCGCATCACCATGGTAATAGCGATGACTAATACGTAGATGTACTGCCAAGTAGGAAAGTCCATAAAGTCATGTACTGGGCATAATGCCA
GGCGGGCCATTTACCGTCATTGACGTCAATAGGGGGCGTACTTGGCATATGATACACTTGATGTACTGCCAAGTGGGCAGTTTACCGTAAATACTCCACCC
ATTGACGTCAATGGAAAGTCCCTATTGGCGTTACTATGGGAACATACGTCAATTATTGACGTCAATGGGCGGGGGTCTGTTGGCGGTGACCCAGGGGGC
CATTTACCGTAAGTTATGTAACGCGAACTCCATATATGGGCTATGAACTAATGACCCCGTAATTGATTACTATTAATAACTAGTCAATAATCAATGTCAAC
GCGTATATCTGGCCGTACATCGGAAGCAGCAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGCCGCTTGTGCGGTTTT
TCCATAGGCTCCGCCCCCTGACGAGCATCACAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTTCCCC
TGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCTGCCGTTACCGGATACCTGTCCGCTTTCTCCCTTCGGGAAGCGTGGCGCTTCTCATAGCTCACG
CTGTAGGTATCTCAGTTCGGTGTAGGTCGTTGCTCCAAGTGGGCTGTGTGCACGAACCCCCGTTCCAGCCGACCGCTGCGCTTATCCGGTAACTATC
GTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT
TCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCT
TGATCCGGCAAAACACCCTGAGCGGTGTTTTTTTTGTTTGAAGCAGCAGATTACCGCGCAAAAAAAGGATCTCAAGAAGATCCTTTGATCTT
TTCTACGGGGTCTGACGCTCAGTGAACGAAAACCTCACGTTAAGGGATTTTGGTATGAGATTATCAAAAAGGATCTTACCTAGATCCTTTAAATAAA
AATGAAGTTTTAGCACGTGCTATTATTGAAGCATTATCAGGGTATTGTTCTCATGAGCGGATACATATTTGAATGATTTAGAAAAATAACAATAGGG
GTTCCGCGCACATTTCCCGAAAAGTGCCACCTGATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGAAATTGTAAGCGTTA
ATAATTCAGAAGAAGTCTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAGCACGAGGAAGCGGTGAGCCATTC
GCCGCAAGCTCTTCAAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAGCGGTCCGCCACCCAGCCGCGCCACAGTCGATGAATCCAGAAAAGCGG
CCATTTCCACCATGATATTCGGCAAGCAGGCATCGCCATGGGTACGACGAGATCCTCGCCGTCGGGCATGCTCGCCTTGAAGCTGGCGAACAGTTCGG
CTGGCGGAGCCCTGATGCTCTTCTGTCAGATCATCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTCGATGCGATGTTTCGCTTGGTGG
TCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGATTGCATCAGCCATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGA
TCTGCCCCGGCACTTCGCCCAATAGCAGCCAGTCCCTTCCCGTTCAGTGACAACGTCGAGCACAGCTGCGCAAGGAACGCCCGTCTGGCCAGCCACG
ATAGCCGCGCTGCCTCGTCTTGCAGTTCATTCAGGGCACCGGACAGGTGCGTCTTGACAAAAAGAACCGGGCGCCCTGCGCTGACAGCCGGAACACGG
CGGCATCAGAGCAGCCGATTGCTGTTGTGCCAGTCATAGCCGAATAGCCTCTCCACCAAGCGGCGGAGAACCTGCGTGCAATCCATCTTGTCAATC
ATGCGAAACGATCCTCATCTGTCTTGTGATCAGAGCTTATCCCTGCGCCATCAGATCCTTGGCGCAAGAAAGCCATCCAGTTTACTTTGACAGGGCTTC
CCAACCTTACCAGAGGGCGCCAGCTGGCAATTCGGTTCGCTTGTCCATAAAAACCGCCAGTAGAAGCCATAGAGCCACCGCATCCCCAGCATGC
CTGCTATTGCTTCCCAATCCTCCCCCTGTGCTGCTGCCACCCACCCCCAGAAATAGAATGACACCTACTCAGACAATGCGATGCAATTTCTCATTTTA
TTA