Anomalies in BatCoV/RaTG13 sequencing and

provenance

Daoyu Zhang

ABSTRACT

To this date, the most critical piece of evidence on the purposed "natural origin" theory of SARS-CoV-2, was the sequence known as RaTG13, allegedly collected from a single fecal sample from Rhinolophus Affinis. Understanding the provenance of RaTG13 is critical on the ongoing debate of the Origins of SARS-CoV-2. However, this sample is allegedly "used up" and therefore can no longer be accessed nor sequenced independently [1], and the only available data was the 3 related Genbank accessions: MN996532.1, SRX7724752 and SRX8357956.

We report these datasets possessed multiple significant anomalies, and the provenence of the promised claims of RaTG13 or it's role in proving a "probable bat origin"[2] of SARS-CoV-2 can not be satisfied nor possibly be confirmed.

RESULTS

Anomalous enrichment of telomere-like repeat sequences in

the dataset SRX7724752

>gnl|SRA|SRR11085797.3.1 3 (Biological)

CTAACCCTAACCCTAGCACTATCCTGTTTCCAACCCCAACCCTAACCCTCACCCTAACCC TAACCCCAGCCTGTTTCATACCTTAACTCGCACCTCATCGCTAACCCCAGCCCTCACCCG ATCCTGTTTCCTCCCCGAACATAACCCCT

>gnl|SRA|SRR11085797.3.2 3 (Biological) GGTTAGGGTTAGGGTTAGGGTTGGAAACAGGATAGGGTTAGGGTTAGGGTGAGGGTTAGG GTTAGAGTTAGGGTGGGAAACAGGATAGGGGTAGGGGTAGGGCTAGGGCGAGGGCTAGGGATAGGG AGGGAAACAGGATAGTGGGAGGGCTAGGGGT

>gnl|SRA|SRR11085797.8.1 8 (Biological)

GTTAGGGTTAGGGTTAGGGTTAGGGTTGGGGTTGGATACAGGATATGGTTAGGGTTAGGG GTAGGGTCAGGGTTAGGATTGGAAACGAGATAGGTTACGTGATAGGGTTAGGGT TAGGTTTAGTAATCCGCAACGGCTTAGGGTT

>gnl|SRA|SRR11085797.8.2 8 (Biological)

CCTAACCCTAACCCTAACCCTAACCCTAACCCTATCCTGTTCCCAACCCTAACC CTAACCCTAACCCTAACCCTAACACAAAACATAACCCTAACCCCAACCCCAAACCCTAACC CCATCTTTACTCACACCCTAACCCCAAAACTC

>gnl|SRA|SRR11085797.10.1 10 (Biological)

>gnl|SRA|SRR11085797.10.2 10 (Biological)

GTTCCCAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTTTCCTTTTTCCAACCCTA ACACTAACCCTAACACTAAACCTAACCCCAACCCTACCACTATACTATATCCGACTCTCA CGCTAACACTAAACATAAGTAATCACAAATT

>gnl|SRA|SRR11085797.13.1 13 (Biological)

>gnl|SRA|SRR11085797.13.2 13 (Biological)

GGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTGGGAACAGGATAGG GTTAGGGTTAGGGTTGGGATCAGGATAGGGATAGGGATAGGGATAGGGTTAGGG TGGGGAACAGGAGAGCGTTAGGCAAGG

>gnl|SRA|SRR11085797.14.1 14 (Biological)

GGTTAGGGTTAGGGTGGAGAAGAGAGGATAGGTTTAGGTTTAGGGTTAGGGTTAGGGTGAGG GTTAGGGAGAGGGGTTAGCTACACGATAGGAGTAGGGTAACGATTAGGGTTAGGGTTAGGT TTGGAAAAAGCATAGGCTATGAGGTACGGT

>gnl|SRA|SRR11085797.14.2 14 (Biological)

CTGCTTCCAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACC CTATCCTGTTCCCAACCCTAACCCTAACCCCTAACCCCAACCCCTAACCCCAAACCCT AACCCTAACCCCAACCCATACCCCAACCAT

>gnl|SRA|SRR11085797.15.1 15 (Biological)

TGTTCCCAACCCTAAACCTAAACCTAAGCCGATCCTGTTCCCAACCCTAACCCTAACCCT ATCCTGTAAACAACCCCCACCCTAAAAACATCCTCGTACAAACCCTAACCCCAACCCCCAT CCCAAACCACATACCCCGTCACGAACCCCACCC

>gnl|SRA|SRR11085797.15.2 15 (Biological)

>gnl|SRA|SRR11085797.16.1 16 (Biological)

CTAACCCTAACCCTACCCCTATCCTGTTCCTAACCCGAACCCTAACCCTAACCCTAACCC TAACCCTCACCTGTTCCAGACCGTAATGCTAACCCTTAACACTATCCTGTGCGCTACCCCG ACCCTAACCCTCAGCCGACGCGTCACGCCCG

>gnl|SRA|SRR11085797.16.2 16 (Biological)

GGTAAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG GTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGATAGGGTTAGGGA TAGGGGTAGAGATAGGGTGAGGTGGTGGAA

>gnl|SRA|SRR11085797.17.1 17 (Biological)

AACCCTAACCCTAACCCTAACGCTATCATGATCCCATCCCTAACCCTAACCCTAACCCTA ACCCTAATACTAACCCTACCCTTTTCATCTCCCCTTACACTACCCCCAACACGCCACCCAT CCCCAACCACTATGCATGCACTGTCCTAAAC

>gnl|SRA|SRR11085797.17.2 17 (Biological)

GGTTAGGGTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGAACAGGATAGGG GTAGGGTTAGGGCTAGGGTGAGAAACAGGGTAGGGGTAGGGTGAGGATAAGGGATAGGGT TGGGGTTGGGAACAGAGAAGGGGAAGGGCA

>gnl|SRA|SRR11085797.18.1 18 (Biological)

CTAACCTGTTCCCAAACTTAAATCCAATCCTAACCCTATCCTGTTCCCAACCCTAACCCT AAACCTATACCTATCCTGCCCCACACACCGACCCTATACACCACCCTAAACGCAACCCTA ACCCCATCCTGTTATCGAAGCATACCCCCCAC

>gnl|SRA|SRR11085797.18.2 18 (Biological)

>gnl|SRA|SRR11085797.19.1 19 (Biological)

CCTAACCCTAAACCTAACCCTCTCCTGTTTCCAACCATAACCCTAACCCTAACCCTAACC CTAACCCACTCCTGTTCTTAACACTAACCTTAACTCTGAGCTCATCCCCAAACCTAACCA TAACCCCACCAGTTCCGATACCATCACCCCC

>gnl|SRA|SRR11085797.19.2 19 (Biological)

AGGGTTAGGGTTGGAAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTG GGGACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTGG GATTGGGTATGGGTAGTGGTCAGGGATAGTG

>gnl|SRA|SRR11085797.20.1 20 (Biological)

CCTGTTTCCAACCCTCACCCTGACACTGACCCTAACACTAACCCTAACCCTAACCCGATC CTGTTTCTGACCCTAACGACAAGCCTGGCACTAAACTGATCGCGTTTCCAATCGTTACCG CTTCCCTAACACCGTCTGTGAAGATACTCCG

>gnl|SRA|SRR11085797.20.2 20 (Biological)

CTTTAGGTTTAGTGTTAGGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG GGTTGGGAACAGGATAGGGTTAGGGTTAGGGATAGGGTTGGGGGTCTGGATAGGGTTGGGG GTAGGGTTAGAGTGAGTGTGGGGCAGCAGGCG

Figure 1: The reads that contained Telomere-like repeat sequences within the first 20 reads of SRX7724752.

Despite the theoretical presence of traces of Telomere-like repeats in total RNA of most cells, such repeats comprise only a tiny fraction of the total cellular RNA within real biological samples, and normally does not show up in the first 100 reads. RaTG13 contained an anomalous amount of such repeats, which comprises 63% of the dataset and exist in nearly any set of 10 reads within this dataset. In comparison, the next highest content of such repeats within any other sample of similar context on NCBI, contained merely 4% of these repeats, which does not show up in the first 20 reads of the dataset. Telomere-like repeats are not detected in the first 100 reads of any other datasets examined.

In comparison, the related SRX7724693 lacked such reads within the first 100 reads of the dataset.

>gnl|SRA|SRR11085736.100.1 100 (Biological)

CTACTGTGTCATCCCATTTCACAAACGCTTATTGGCGGGTACAGGAATATCAACCTGTTGT CCATCACCTACGCCTTTCGGCCTCGGCTTAGGTCCTGACTAACCCAGGGCAGAAGAACCT TCCCCTGGAAACCTTGGGTTGACGGCCCGTG

>gnl|SRA|SRR11085736.100.2 100 (Biological)

ATCCCACGGGCCGTAAACCCAAGGTTTCCAGGGGAAGGTTCGTCCGCCCTGGGTTAGTCA GGACCTAAGCCGACGCCGAAAGGCGTAGGTGATGGACAACAGGTTGATATTCCTGTAACC GCAATAAGCGTTTGAGAGATGGGATGACAGT

Figure 2: the first 100 reads in SRX7724693 did not show any Telomere-like repeats. In addition, SRX7724752 contained 6% all-N sequences that were exactly 35nt long, which is not found at levels any close in other datasets that had the same design section.

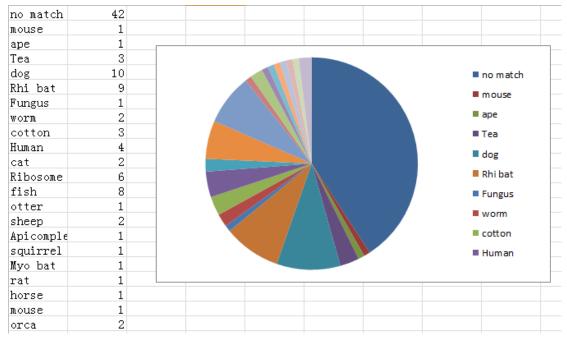
Reads (separated)

>gnl|SRA|SRR11085797.11.1 11 (Biological)

>gnl|SRA|SRR11085797.11.2 11 (Biological)

Figure 3: an example of All-N read in SRX7724752.

Anomalous enrichment of non-attributable and low-match



data within SRX7724752

Figure 4: BLAST result of 100 random reads obtained from RaTG13 using BLASTn.

In addition to the anomalous enrichment of repeats, The vast majority of the non-repeat sequences in SRX7724752 does not show any clear matches when examined using BLASTn. With matching results ranging from nearly all domains of life—all of which were partial and low-quality matches, including that of bats.

Only 2 out of 7 Non-repeat and non-PolyN sequences from the first 20 reads from SRX7724752 had any matches, and the match was only partial matches to certain hypothetical proteins

Description	gnl SRA SRR11085797.2.1 2 (Biological)	Percent Identity	E value	Qu	ery Coverage
Molecule type	dna	to	to		to
Query Length	151				
Other reports	Distance tree of results MSA viewer 🔞				Filter Reset
Descriptions	Graphic Summary Alignments Taxonomy				
Sequences	producing significant alignments	Download 🕥	Manage Co	l umns Y S	how 5000 🗸 🕄
🗹 select all	6 sequences selected		GenBank	<u>Graphics</u>	Distance tree of results
select all	6 sequences selected Description		Max Total	Query Cover E value	Distance tree of results Per. Ident Accession

		Organism only top 20 will appear						
Results for								
Program	BLASTN ② Citation ~							
Database	nt <u>See details</u> ~	+ Add organism						
Query ID	lcl Query_61914	Percent Identity E va	lue		Qu	ery Cov	erage	
Description	gnl SRA SRR11085797.7.17 (Biological)	to	to			t	0	
Molecule type	dna							
Query Length	150	Filter						
Other reports	Distance tree of results 🔞							
Descriptions	Graphic Summary Alignments Taxonomy							
Sequences p	producing significant alignments	Download 🗡	Manage C	olumns	s ~ s	show	5000 🗸 🕝	
🗹 select all	2 sequences selected		<u>GenBa</u>	nk <u>Gr</u>	aphics	Distan	ce tree of results	
	Description		Max Tota Score Score		E value	Per. Ident	Accession	
Eimeria mit	is hypothetical protein, conserved partial mRNA		73.1 73.	1 40%	3e-09	88.33%	XM_013494305.1	
Cyprinus ca	arpio genome assembly common carp genome, scaffold 000012284		73.1 73.	1 40%	3e-09	88.33%	LN595243.1	

Figure 5: the BLASTn result of the 2 non-repeat and non-PolyN sequences in the first 20 reads of SRX7724752. The rest can not be matched to any known organisms.

Depletion of bacterial-like reads in SRX7724272 which is inconsistent with fecal samples prepared using the methods as indicated by the "Design" section of the SRX7724752 metadata.

Fecal matter [3], is primarily bacteria by composition. All other fecal swabs prepared using the methods indicated by the metadata correctly showed the presence of bacteria as the majority of the reads. In contrast, SRX7724272 contained only 0.65% bacteria-like reads, all of which were 16S rRNA.

RNA-Seq of Rhinolophus affinis:Fecal swab (SRR11085797)

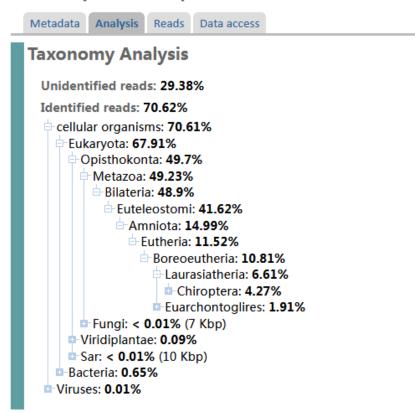


Figure 6: Phylogenetic analysis of SRX7724272.

RNA-Seq of Miniopterus schreibersii: Anal swab RNA-Seq of Hipposideros larvatus: Anal swab RNA-Seq of Tylonycteris pachypus: Anal swab

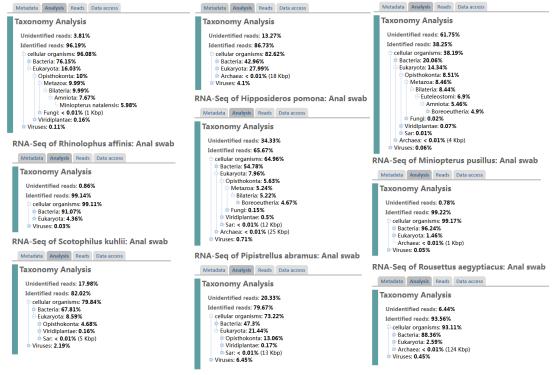


Figure 7: A set of 9 Swabs. The only ones that matches RaTG13 by metadata on Genbank. None of them had more Eukarya-like reads than Bacteria-like reads.

In addition to the anomalous depletion of bacterial-like reads, SRX7724272 also lacked

discernible reads from bacterial mRNA.

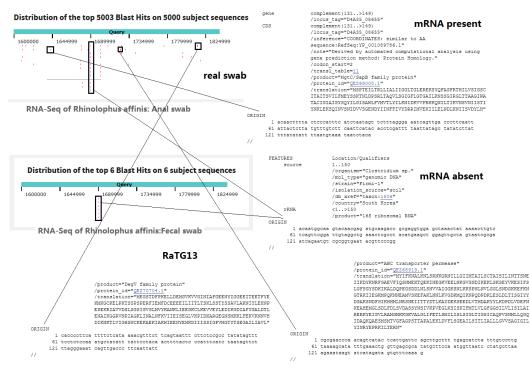


Figure 8: BLAST hits of bacterial non-ribosomal RNA genome on SRX7724272 and another swab from Rhinolophus Affinis under the same library preparation section.

Observation of anomalous and unexpected data within SRX7724752

>gnl|SRA|SRR11085797.11550005.1 11550005 (Biological)

GCCCGTATTTAGCCTTAGATGGAGTTTACCACCCGCTTTGGGCTGCATTCCCAAGCAACC CGACTCCGGGAAGACCCCGGGCCCGGCGCGCGCGGGGGCCGCTACCGGCCTCACACCGTCCA CGGGCTGGGCCTCGATCAGAAGGACTTGGGC

>gnl|SRA|SRR11085797.11550005.2 11550005 (Biological) CGGTGGGGCGCGGGGACATTTGGCGTACGGAAGACCCACTCCCCGGCGCCGCTCGTGGGGG CCCAAGTCCTTCTGATCGAGGCCCAGCCCGTGGACGGTGTGAGGCCGGTAGCGGCCCCG GCGCGCCGGGCCCGGGTCTTCCCCGGAGTCGG

ISRA SRR1108	35797.11550005.1 1155	50005 (B								
olecule type										
па										
uery Length										
51										
ther reports										
stance tree of r	results MSA viewer	3								
Descriptions	Graphic Summa	ary Alignmer	nts Taxor	iomy						_
Sequences	producing signific	ant alignments		ownload ~	Mana	age Colur	nns ~	Show	100 🗸	0
select all	100 sequences selected				Ge	<u>nBank</u>	Graphics	Distan	ice tree of re	sults
		Description				Total Quer Score Cove	y E r value	Per. Ident	Accessi	on
	ED: Phyllostomus discolor 3	28S ribosomal RNA (LO	C114512504), rRNA	<u>\</u>	279	279 1009	6 2e-71	100.00%	XR_0036858	309.1
	ED: Phyllostomus discolor	basic proline-rich protein	-like (LOC11451244	1 <u>2), mRNA</u>	279	279 1009	6 2e-71	100.00%	XM_0285314	404.1
Homo sap	iens IncAB370.3 IncRNA g	<u>ene, complete sequence</u>	1		274	274 1009	6 9e-70	99.34%	MK280359.1	1
Homo sap	iens IncAB366.1 IncRNA g	ene, complete sequence	1		274	274 1009	6 9e-70	99.34%	MK280356.1	1
Felis catus	<u>s Senzu DNA, chromosom</u> e	e: E1, American Shortha	iir breed		274 1	1372 1009	6 9e-70	99.34%	AP023165.1	L
	ED: Marmota flaviventris 28	<u>S ribosomal RNA (LOC1</u>	<u>17794687), rRNA</u>		274	274 1009	6 9e-70	99.34%	<u>XR_0046185</u>	<u>536.1</u>
_	ED: Arvicanthis niloticus 28				274	274 1009	6 9e-70	99.34%	XR_0046063	
PREDICTE	ED: Arvicanthis niloticus 28	S ribosomal RNA (LOC1	17704855), rRNA		274	274 1009	6 9e-70	99.34%	XR_0046063	368.1
escription	gnl SRA SRR11085797.1155	0005 1 11550005 (Riolog	ical)							
		0000.1 11000000 (BIOIOS	ical)	Percent Identity	,	Evalue		Oue	ry Coverage	
olecule type	dna	0005.1 11550005 (Blolog		Percent Identity	/	E value	to	Que	ry Coverage	
iery Length	151			Percent Identity	/	E value	to		to	
uery Length						E value	to		to	Reset
uery Length	151					E value	to		to	Reset
her reports	151 Distance tree of results MS	A viewer ? Alignments		to			to		to	
her reports	151 Distance tree of results MS Graphic Summary	A viewer ? Alignments		to			age Colur	nns × s	to Filter	v 0
her reports	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	ownload Max	Y Man	age Colur	nns × S Graphics Per.	Filter	v 0
ery Length in the reports in the secriptions sequences pro-	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	wnload Max Score S	Y Man	age Colur E I value k	nns × S Graphics Per.	to Filter	✓ O of results
ery Length 3 her reports 1 Descriptions Sequences provided and 10 select all 10 SRX1724752 SRX1724752	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	ownload Score S 279 279	Total Query Score Cover 279 100%	age Colur E I value I 2e-73 10 2e-73 10	mns × S Graphics Per. Jent 0.00% <u>SRA:</u>	to Filter I Filter I Filter F	✓ ♥ of results 1601488.1 1599592.1
erry Length 3 her reports 1 Descriptions 3 Sequences provide 10 Select all 10 SRX1724752 3 SRX1724752 3 SRX1724752 3	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	ownload Score 5 279 279 279	Total Query Score Cover 279 100% 279 100%	E value 2e-73 10 2e-73 10 2e-73 10	mns × S Graphics Per. 9.0.0% SRA: 0.00% SRA: 0.00% SRA:	to Filter 1 how 100 Construction for the second s	✓ ♥ of results 1601488.1 1599592.1 1599895.1
ery Length 3 her reports 1 Descriptions Sequences provided and 10 select all 10 SRX1724752 SRX1724752	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	ownload Score S 279 279 279 279 279	Total Query Score Cover 279 100%	E I 2e-73 100	nns × S Graphics Per. Jent 0.00% SRA: 0.00% SRA: 0.00% SRA:	to Filter I Filter I Filter F	 of results 1601488.1 1599592.1 1598985.1 1598757.1
ery Length 3 ber reports 1 Descriptions Sequences provided and 10 select all 10 SRX1724752 SRX1724752 SRX1724752	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	279 279 279 279 279 279	Man Total Query Score Cover 279 100% 279 100% 279 100% 279 100% 279 100%	E I 2e-73 10	mns × S Graphics Per. dent 0.00% SRA: 0.00% SRA: 0.00% SRA: 0.00% SRA:	to Filter 1 how 100 Construction for the second s	 fresults 601488.1 1599592.1 1598955.1 1558757.1 1558853.1
ery Length 2 her reports 2 Descriptions Sequences pro- Sector all 10 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	279 279 279 279 279 279 279 279 279 279	Total Query Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	E I value I 2e-73 10	nns S Graphics Per. Jent 0.00% SRA:	to Filter 1 Filter 1 Filter 1 Comparison of the second sec	of results 1601488.1 1599592.1 1598955.1 1598955.1 15885755.1 15885755.1 15885755.1
ery Length 2 bescriptions Sequences pro- select all 10 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	279 279 279 279 279 279 279 279 279 279	Total Query Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	E I 2e-73 10	nns S Graphics Per. Jent 0.00% SRA:	to Filter 1 Filter 1 Filter 1 Comparison of the second sec	 Image: Control of the second se
ery Length 3 her reports 1 Descriptions Sequences pr Select all 10 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	279 279 279 279 279 279 279 279 279 279	Total Query Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	E I 2e-73 10	nns S Graphics Per. Jent 0.00% SRA:	to Filter 1 Filter 1 Filter 1 Comparison of the second sec	of results for the second sec
ery Length 3 her reports 3 Descriptions Sequences pr Select all 10 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752 SRXT724752	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	279 279 279 279 279 279 279 279 279 279	Total Query Score Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	E I 2e-73 10	nns × S Graphics S Per. 1 Jent SRA: 0.00% SRA:	to Filter 1 Filter 1 Filter 1 Control	✔ ● ● of results ■ 1601488.1 ■ 1598952.1 ■ 1598757.1 ■ 15887555.1 ■ 15887555.1 ■ 15887555.1 ■ 15887555.1 ■ 15887555.1 ■ 15887555.1 ■ 15887555.1 ■ 15887555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 1588755555.1 ■ 1588755555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 158875555.1 ■ 1588755555.1 ■ 158875555555.1 ■ 1588755555.1 ■ 158875555555555555555555555555555555555
Lery Length 3 Constructions Sequences pro- Secret all 10 Select all 10 Secret all 10 Secre	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	279 279 279 279 279 279 279 279 279 279	Total Query Score Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	E I 2e-73 10	nns × S Graphics S Der. 1 Jent 0.00% 0.00% SRA.	to Filter I I Filter I	Control Contro
ther reports in the second sec	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	279 279 279 279 279 279 279 279 279 279	Total Query Score Cover 279 100%	E I 2e-73 10 2e-74 10	nns × S Graphics S Der. 1 Jent SRA 0.00% SRA	to Filter I I Filter I	C 011488.1 1599592.1 1599592.1 159985.1 1598757.1 15885419.1 1588542.1 1585542.1 1585542.1 1585542.1 1579566.1 1579566.1 1579566.1 1579566.1 157956.1
Lery Length 3 Constructions Sequences pro- Secret all 10 Select all 10 Secret all 10 Secre	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	279 279 279 279 279 279 279 279 279 279	Total Query Score Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	E I 2e-73 10 2e-73 10	nns × S Graphics S Jen 0.00% SRA: 0.00% SRA: 0.00%	to Filter I I Filter I	C C C C C C C C C C C C C C C C C C C
ery Length 3 cher reports 3 Sequences pro- select all 10 Select all 10 Select all 10 Secret all 10 Secre	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	Max Score S 279 279 279 279 279 279 279 279 279 279	Total Query Score Cover 279 100%	E I 2e-73 10 2e-73 10	nns × S Graphics S Der. 0.00% SRA- 0.00% SRA- 0.00%	to Filter I I I I I I I I I I I I I I I I I I I	C C C C C C C C C C C C C C C C C C C
uery Length 3 ther reports 1 Descriptions 1 Sequences provements 1 is select all 10	151 Distance tree of results MS Graphic Summary oducing significant ali	A viewer ? Alignments		to	Max Score S 279 279 279 279 279 279 279 279 279 279	Man Total Query Score Cover 279 100%	E I 2e-73 10 2e-73 10	mns S Graphics S Jent S Joord SRA: 0.00% SRA:	to Filter 1	✔ ♥ ● fresults ■ 1601488.1 ■ 155952.1 ■ 155952.1 ■ 155952.1 ■ 155952.1 ■ 155852.1



Pale spear-nosed bat

Figure 9: Phyllostomus Discolor, or Pale spear-nosed bat, a species of bat native to the Americas, is found in high abundance within SRX7724752.

>gnl|SRA|SRR11085797.11550023.2 11550023 (Biological) TTTGTTTTGTTTTTATAATTTATTTTTTAAAATTTATTGGGGTGACAATTGTTAGTAAAA TTACATAGATTTCAGGTGTACAATTCTGTATTACATGTGGACGGTCCAGCCGCCACGAGT TCAACGTTTTACATGAAAGGGGGGTGTGGGA

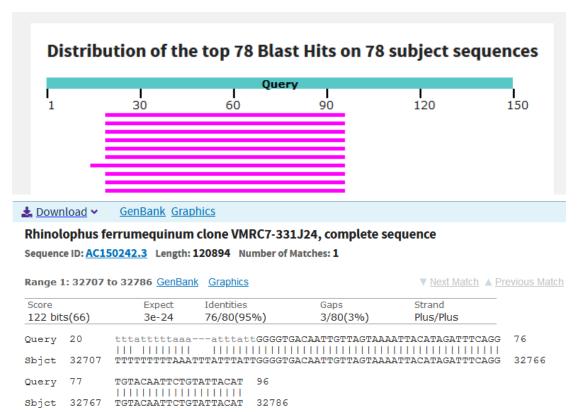


Figure 10: A sequence which was matched to a bat mRNA clone in the first 96 nucleotides, but then matching nothing on the later nucleotides. This match end with a T.

Reads (separated)

▲ Download ∨

>gnl|SRA|SRR11085797.8568962.1 8568962 (Biological)

>gnl|SRA|SRR11085797.8568962.2 8568962 (Biological)

GenBank Graphics

ACAAACCAACGAACTCTCGATCTCTTGTAGATCTGTAACTTCCTTTTGTCCCCTAGTTCA CCGATCTCATCGCCCCTGCCAGGGGTCCATGGACTGTGTGACCCCGTCCGCTGCTCCCC CAT

PREDICTED: Rhinolophus ferrumequinum zinc finger CCCH-type containing 12A (ZC3H12A), mRNA

Range 1:	2468	to 2550 GenBank	Graphics		Vext Match	Previous Match
Score 132 bits	(71)	Expect 4e-27	Identities 79/83(95%)	Gaps 0/83(0%)	Strand Plus/Minus	
Query Sbjct			11 11111111111		GGGCGATGAGATCGGTG	65 2491
Query	66	AACTAGGGGACA				
Sbjct	2490	AACTAGGGGACA		в		
L Downl	oad 🗸	<u>GenBank</u> Gra	aphics			
Bat core	onavir	us RaTG13, co	mplete genome			
		-	h: 29855 Number of M	latches: 1		
Range 1:	12 to	47 GenBank Gra	phics		▼ <u>Next Match</u> ▲	Previous Match
-		Expect	Identities	Gaps	Strand	
Score 67.6 bits			36/36(100%)	0/36(0%)	Plus/Minus	

Sbjct 47 ACAGATCTACAAGAGATCGAGAGATTCGTTGGTTTGT 12

Figure 11: a viral sequence fused to a mRNA-like sequence. Again overlapping on an A. Of the only 3 sequences within the viral reads within SRX7724752 that displays fusion of different sequences, only one sequence matches that of a canonical coronavirus subgenomic mRNA leader, another one was the read illustrated in Figure.11, while the third one was a non-canonical fusion of two non-TRS regions in the RaTG13 genome.

>gnl|SRA|SRR11085797.10676687.1 10676687 (*Biological*) GGTCCTTGATGTCACAGCGTCCTAGATGGTGTCCAGCAATACGAAGATGTCCACGAAGGA TGACAGCTCCGATTACAAGTTCACTCTCTAGAAGCGGTCTGGTCAAAATAGTGCCATGGA GTGGCACGTTGAGCAAAATGTTAGTTTCTGG

>gnl|SRA|SRR11085797.10676687.2 10676687 (Biological) ATGAAGGCAATTCACCATTCCATCCTCTAGCTGATAATAAATTTGCACTGACTTGCTTTA GCACTGATGTGGCTGAGCTACTTCATTGCTTCTTTCAGGCTATTTGCACGTACGCGTTCC ATGTGGTCATTCAATCCAGAAACTAACATTT

🛓 <u>Down</u>	nload 🗸	<u>GenBank</u> <u>Gr</u>	aphics Sort by:	E value	~	
Bat co	ronaviru	us RaTG13, co	mplete genom	e		
Sequenc	e ID: <u>MN9</u>	96532.1 Leng	th: 29855 Numbe	er of Matches: 2		
Range 1	L: 26736	to 26844 <u>GenB</u>	ank <u>Graphics</u>		▼ <u>Next M</u>	atch 🔺 Previous Match
Score		Expect	Identities	Gaps	Strand	
161 bit	s(87)	7e-36	102/109(94%)	2/109(1%	6) Plus/Plus	3
Query	45	GCACTGACTTO	-CTT-TAGCACTG	ATGTGGCTGAGCTAC	TTCATTGCTTCTTTC	AGGCTA 102
Sbjct	26736	GCAATGGCTTO	TCTTGTAGGCTTG	ATGTGGCTGAGCTAC	ITCATTGCTTCTTTC	AGGCTA 26795
Query Sbjct	103 26796	1111111111	1111111111111	TCATTCAATCCAGAA 		
2		to 27542 <u>GenB</u>	ank <u>Graphics</u>	▼ №	lext Match 🔺 Previou	is Match 🛦 First Match
Score		Expect	Identities	Gaps	Strand	
121 bit	s(65)	1e-23	65/65(100%)) 0/65(0%) Plus/Plus	
	-()		,(,	-,(/	
Query	1	ATGAAGGCAAT	TCACCATTCCATC	CTCTAGCTGATAATA	AATTTGCACTGACTT	GCTTTA 60
Sbjct	27478	ATGAAGGCAAT	TCACCATTCCATC	CTCTAGCTGATAATA	AATTTGCACTGACTT	GCTTTA 27537
Query	61	GCACT 65				
Sbjct	27538	GCACT 2754	2			

Figure 12: an anomalous fusion of two non-canonical regions of the RaTG13 genome. The fusion again happens on a T.

Bat coronavirus RaTG13, complete genome

Sequence ID: MN996532.1 Length: 29855 Number of Matches: 2

Range 1	L: 28217	to 28346 GenBa	nk Graphics		▼ <u>Next Match</u> ▲ Pr	evious Match
Score 233 bit	s(126)	Expect 2e-57	Identities 129/130(99%)	Gaps 1/130(0%)	Strand Plus/Plus	
Query	23	TTC-TCTAAACG	AACAAACTAAAATGTO	TGATAATGGACCCCAA	AACCAACGAAATGCAC	81
Bbjct	28217	TTCATCTAAACO	AACAAACTAAAATGTO	TGATAATGGACCCCAA	AACCAACGAAATGCAC	28276
Query	82	CCCGCATTACG	TTGGTGGACCCTCAG	ATTCAACTGGCAGTAAC	CAGAATGGAGAACGCA	141
Bbjct	28277	CCCGCATTACG	TTGGTGGACCCTCAGA	ATTCAACTGGCAGTAAC	CAGAATGGAGAACGCA	28336
Query	142	GTGGAGCACG	151			
Sbjct	28337	GTGGAGCACG	28346			
Range 2	2: 25 to 6	0 <u>GenBank</u> <u>Gra</u> j	<u>phics</u>	▼ <u>Next Ma</u>	tch ▲ Previous Match	First Match
Score	t=(2C)	Expect	Identities	Gaps	Strand	
67.6 bi	15(30)	2e-07	36/36(100%)	0/36(0%)	Plus/Plus	
Query	1 CT	CTCGATCTCTTGI	AGATCTGTTCTCTAAA	ACGAAC 36		
Sbjct	25 CT	CTCGATCTCTTG	AGATCTGTTCTCTAA	ACGAAC 60		

Figure 13: the only canonical sgRNA-like read* in SRX7724752.

Furthermore, SRX7724752 contained significant amount of reads that had higher query coverage on the DNA sequence than on the corresponding mRNA. This most likely indicate a clonal, rather than cDNA, library, was responsible for most of the bat-like reads observed in SRX7724752.

>gnl|SRA|SRR11085797.76.1 76 (Biological)

CATCAAACTGAGGTTTCAGCAAGGCAAAGATAGCCAGCAACAAAAACAAAAAGGCATCCTA CTGAATGGAAGCAGATAATTGCCAATAGTACATCAGTAAGGAGTTAATATTAAGAATTAG TTTTTAAAAAAGCTCTATATGATGTCAGAAAT

>gnl|SRA|SRR11085797.76.2 76 (Biological)

GTTTTCACTTGCATTTCTCTAATAATTAGTGATGTTGAGCATCTTTTCATATGTCTATTG GCCATCTGTATGTCGTCTTTGGAGAAATGTCTATTCAGATTTCTGCCCCAATTTTTAATTG GCTTGTTTGTTTTTGTTTTTGAATTGAGTT

	criptions	Graphic Summary	Alignments	Taxonomy							
Seq	juences pro	ducing significant a	alignments		Download 🗡	Mar	age C	olumn	s ~	Show	100 🗸 💡
•	select all 11 s	sequences selected				G	enBanl	<u>k Gra</u>	aphics	<u>Distan</u>	ce tree of results
			Des	cription		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	Rhinolophus fer	rumequinum clone VMRC7-7	1A7_complete_sequenc	2		267	267	99%	1e-67	98.68%	AC150307.3
<	Rhinolophus fer	rumequinum clone VMRC7-2	51C10, complete seque	nce		185	185	94%	4e-43	90.14%	AC149630.3
<	Myotis lucifugu	s clone CH235-427D16_comp	plete sequence			137	137	82%	1e-28	87.20%	AC174832.3
✓	Pteropus alecto	clone BAC P100M20, BAC I	P103A18, complete seq	uence		135	135	86%	4e-28	85.38%	KP862827.1
✓	Pteropus alecto	clone BAC P201M3, BAC P	216K21, complete sequ	ence		135	135	86%	4e-28	85.38%	KP862826.1
✓	Pteropus alecto	clone BAC P56N20, comple	te sequence			135	135	86%	4e-28	85.38%	KP862825.1
✓	Rhinolophus eu	ryale isolate REM0134 micro	satellite RM1198 seque	<u>1Ce</u>		128	128	47%	7e-26	98.61%	KC910215.1
≤	Pteropus alecto	clone BAC P21207-1, BAC	P229M21, complete se	quence		126	126	84%	3e-25	84.38%	KP862828.1
✓	PREDICTED: N	Ainiopterus natalensis zinc fin	ger protein 713 (ZNF713). transcript variant X	mRNA	106	106	90%	3e-19	81.02%	XM_016196283.1
✓	PREDICTED: N	Ainiopterus natalensis zinc fin	ger protein 713 (ZNF713). transcript variant X	mRNA	106	106	90%	3e-19	81.02%	XM_016196281.1
<	PREDICTED: N	<u>liniopterus natalensis zinc fin</u>	ger protein 713 (ZNF713) <u>, transcript variant X</u>	mRNA	106	106	90%	3e-19	81.02%	XM_016196280.1

Figure 14: a read from SRX7724752 which have higher coverage on the clone than on the corresponding mRNA. E.g. the read contained nucleotide sequences that were not supposed to be transcribed in actual cells/bats.

Inability of SRX8357956 to prove the promises claimed in [5]

Date	R	RaTG13 SARS-CoV-2		Sequence Blast Archive	Nama		
Date	Identity	Query Cover	Identity	Query Cover	sequence	Blast Archive	Name
4-Oct-18	95.53%	56%	93.15%	56%	24	archive.is/H107n	gnl SRA SRR11806578.24 RaTG13-9-5-5_9-5-f1_2018-10-14_B02
4-Oct-18	97.31%	76%	93.18%	75%	23	archive.is/8phs4	gnl SRA SRR11806578.23 RaTG13-9-5-4_9-5-r1_2018-10-14_C02
4-Oct-18	97.55%	53%	93.00%	52%	25	archive.is/ycQ89	gnl SRA SRR11806578.25 RaTG13-9-5-5_9-5-r1_2018-10-14_D02
4-Oct-18	99.43%	97%	96.39%	97%	22	archive.is/abSp6	gnl SRA SRR11806578.22 RaTG13-9-5-4_9-5-f1_2018-10-14_A02
1-Oct-18	97.37%	98%	92.23%	98%	20	archive.is/B20Et	gnl SRA SRR11806578.20 RaTG13-9-5-1_21230-F_2018-10-11_A12
1-Oct-18	98.54%	99%	88.26%	99%	21	archive.is/L2pTq	gnl SRA SRR11806578.21 RaTG13-9-5-1_23258-R_2018-10-11_B12
8-Oct-18	98.64%	98%	96.19%	99%	5	archive.is/W7Fxp	gnl SRA SRR11806578.5 RaTG13-11-2_18297-F_TSS20181008-027-0303_G10
8-Oct-18	99.19%	98%	91.83%	98%	7	archive.is/h8810	gnl SRA SRR11806578.7 RaTG13-12-2_24144-R_TSS20181008-027-0303_C11
8-Oct-18	99.89%	99%	87.86%	99%	6	archive.is/jVJY2	gnl SRA SRR11806578.6 RaTG13-12-2_22717-F_TSS20181008-027-0303_H10
0-Sep-18	99.50%	99%	92.71%	99%	9	archive.is/ON8UX	gnl SRA SRR11806578.9 RaTG13-2-3_RaTG13-2-R1_2018-09-30_B11
0-Sep-18	99.79%	99%	92.89%	99%	8	archive.is/udSil	gnl SRA SRR11806578.8 RaTG13-2-3_RaTG13-2-F_2018-09-30_A02
9-Sep-18	99.00%	99%	94.20%	99%	10	archive.is/jdzvN	gnl SRA SRR11806578.10 RaTG13-2-3_RaTG13-2-R2_2018-09-29_D05
9-Sep-18	99.09%	98%	94.97%	98%	3	archive.is/II96Y	gnl SRA SRR11806578.3 RaTG13-10-3_RaTG13-10-F_2018-09-29_G04
9-Sep-18	99.72%	98%	98.06%	98%	11	archive.is/7Kioa	gnl SRA SRR11806578.11 RaTG13-20-1_RaTG13-F_2018-09-29_H04
9-Sep-18	99.72%	98%	95.19%	98%	4	archive.is/tKTXg	gnl SRA SRR11806578.4 RaTG13-10-3_RaTG13-10-R_2018-09-29_E05
7-Sep-18	95.03%	98%	90.88%	98%	14	archive.is/NNfnm	gnl SRA SRR11806578.14 RaTG13-4-2_RaTG13-4-R_2018-09-27_G06
7-Sep-18	95.82%	98%	93.46%	93%	13	archive.is/Rmrhq	gnl SRA SRR11806578.13 RaTG13-4-2_RaTG13-4-F_2018-09-27_G05
7-Sep-18	98.08%	98%	94.50%	98%	1	archive.is/kcHAi	gnl SRA SRR11806578.1 RaTG13-1-2_RaTG13-1-F_2018-09-27_E05
7-Sep-18	98.81%	99%	96.90%	99%	17	archive.is/nhvd2	gnl SRA SRR11806578.17 RaTG13-6-2_RaTG13-6-R_2018-09-27_H06
7-Sep-18	98.91%	99%	94.54%	99%	2	archive.is/veLPW	gnl SRA SRR11806578.2 RaTG13-1-2_RaTG13-1-R_2018-09-27_F06
7-Sep-18	99.09%	99%	96.92%	99%	16	archive.is/0tqMp	gnl SRA SRR11806578.16 RaTG13-6-2_RaTG13-6-F_2018-09-27_A06
7-Sep-18	99.28%	98%	96.81%	98%	12	archive.is/ZHJmY	gnl SRA SRR11806578.12 RaTG13-3-2_RaTG13-3-F_2018-09-27_F05
7-Sep-18	99.46%	98%	96.74%	98%	15	archive.is/Epig7	gnl SRA SRR11806578.15 RaTG13-5-2_RaTG13-5-F_2018-09-27_H05
7-Sep-18	99.50%	98%	98.49%	98%	18	archive.is/NdyHK	gnl SRA SRR11806578.18 RaTG13-7-2_RaTG13-7-F_2018-09-27_B06
7-Sep-18	99.53%	99%	95.67%	97%	19	archive.is/2qg0a	gnl SRA SRR11806578.19 RaTG13-8-2_RaTG13-8-F_2018-09-27_C06
0-Jun-17	99.10%	99%	96.61%	99%	28	archive.is/ve7nN	gnl SRA SRR11806578.28 RaTG13-R-1-1_7896-1-F1_2017-06-20_E03
0-Jun-17	99.61%	99%	97.43%	99%	32	archive.is/ehzBr	gnl SRA SRR11806578.32 RaTG13-R-4-1_7896-4-F_2017-06-20_F03
0-Jun-17	99.87%	98%	97.42%	98%	33	archive.is/do9Rt	gnl SRA SRR11806578.33 RaTG13-R-4-1_7896-4-R_2017-06-20_H03
0-Jun-17	99.90%	98%	97.44%	98%	29	archive.is/HjQD8	gnl SRA SRR11806578.29 RaTG13-R-1-1_7896-1-R1_2017-06-20_G03
.7-Jun-17	98.56%	99%	95.85%	99%	26	archive.is/fqWWF	gnl SRA SRR11806578.26 RaTG13-ORF8-1-1_ORF8-F_2017-06-17_A05
.7-Jun-17	98.99%	98%	96.52%	98%	27	archive.is/N01Ah	gnl SRA SRR11806578.27 RaTG13-ORF8-1-1_ORF8-R1_2017-06-17_A06
3-Jun-17	99.07%	97%	97.49%	97%	30	archive.is/WwyWy	gnl SRA SRR11806578.30 RaTG13-R-2-1_7896-2-F1_2017-06-03_A07
)3-Jun-17	99.46%	99%	98.01%	99%	31	archive.is/tCLHu	gnl SRA SRR11806578.31 RaTG13-R-2-1 7896-2-R1 2017-06-03 A08

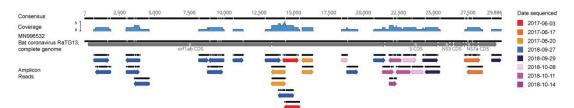


Figure 15: A complete analysis [4] of all Amplicon sequences in SRX8357956. Including the location of these amplicons and the similarity of such amplicon to the RaTG13 and SARS-CoV-2 genome.

Chuan Xiao et.al claimed that RaTG13 contained all the 3 S1 variable loops that were previously considered unique in SARS-CoV-2. [5] However, such claims can not be verified using the amplicons listed in SRX8357956.

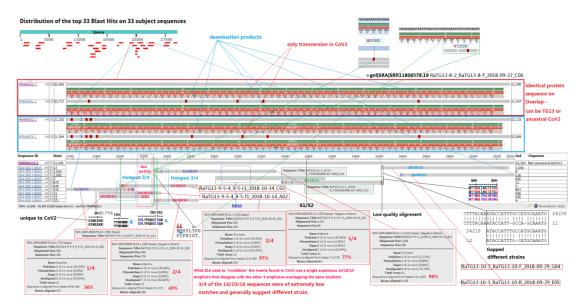


Figure 16: a thorough analysis of the amplicons located on the S locus of RaTG13 in SRX8357956. Notice that the last 4 amplicons sequenced in 14/10/2018 was of very low quality matches, and matched other organisms—including Mouse(mus musculus).

select all 99 sequences selected	Ge	enBank	<u>Gra</u>	aphics	Distance tree of results		
Description		Total Score		E value	Per. Ident	Accession	
Gadus morhua genome assembly, chromosome: 16	60.8	105	27%	4e-05	78.05%	LR633958.1	
Mus musculus BAC clone RP24-498P8 from chromosome 9, complete sequence	59.9	59.9	47%	1e-04	72.92%	AC168217.2	
Mus musculus chromosome 9_clone RP24-484G16_complete sequence	59.9	59.9	47%	1e-04	72.92%	AC137678.11	

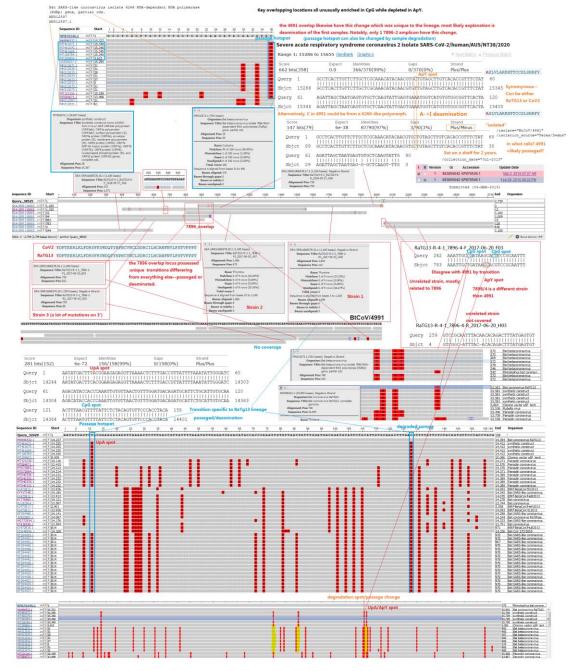
2 :	select all 100 sequences selected	GenBank Graphic				Distance tree of results		
	Description	Max Score	Total Score	,	E value	Per. Ident	Accession	
	Homo sapiens BAC clone RP11-792A8 from 7, complete sequence	54.5	54.5	44%	0.006	71.15%	AC027644.9	
✓	Coregonus sp. 'balchen' genome assembly, chromosome: 7	52.7	52.7	29%	0.022	73.68%	LR778259.1	
✓	Coregonus sp. 'balchen' genome assembly, chromosome: 15	50.9	50.9	16%	0.078	83.33%	LR778267.1	
✓	Salmo trutta genome assembly, chromosome: 21	50.9	50.9	37%	0.078	71.90%	LR584437.1	
✓	Xanthophyllomyces dendrorhous genome assembly Xden1, scaffold Scaffold_79	50.9	50.9	9%	0.078	96.88%	LN483167.1	
\checkmark	Coregonus sp. 'balchen' genome assembly, chromosome: 20	50.0	50.0	72%	0.078	67.49%	LR778272.1	
✓	Aquila chrysaetos chrysaetos genome assembly, chromosome: 14	50.0	141	42%	0.078	69.50%	LR606194.1	
✓	Bos mutus isolate yakQH1 chromosome 16	50.0	50.0	24%	0.078	74.68%	<u>CP027084.1</u>	
✓	Mus musculus BAC clone RP23-128D11 from 7, complete sequence	50.0	50.0	36%	0.078	71.90%	AC122222.6	
✓	Mus musculus BAC clone RP23-66E21 from 7, complete sequence	50.0	50.0	36%	0.078	71.90%	AC131741.4	

	select all 24 sequences selected	Ge	enBank	<u>c</u> <u>Gra</u>	<u>phics</u>	<u>Distan</u>	<u>ce tree of results</u>
	Description		Total Score	1 A A	E value	Per. Ident	Accession
✓	Mus musculus targeted KO-first, conditional ready, lacZ-tagged mutant allele Fabp4:tm1a(KOMP)Wts	50.9	50.9	23%	0.045	85.11%	JN963014.1
✓	Mus musculus targeted non-conditional, lacZ-tagged mutant allele Fabp4:tm1e(KOMP)Wtsi; transger	50.9	50.9	23%	0.045	85.11%	JN947213.1
✓	Mus musculus chromosome 3, clone RP23-436F15, complete sequence	50.9	50.9	23%	0.045	85.11%	AC123726.11
✓	Mus musculus chromosome 3, clone RP24-137C19, complete sequence	50.9	50.9	23%	0.045	85.11%	AC113990.10

Figure 16: BLAST result of the non-RaTG13 matched parts of Amplicons 25, 24 and 23 in SRX8357956

Using the remaining amplicons, the 3 variable loops, GTNGIKR, HKSNK and VIFSQ was obtained.

This is vastly different from the variable loops possessed by SARS-CoV-2, which were GTNGTKR, HKNNK and GDSSSG. Therefore, the promise of Chuan Xiao et. Al does not hold upon raw data analysis.



Probable discontinuities in RaTG13 sequencing in SRX8357956

Figure 17: Detailed analysis of the early amplicons located in the nsp12 RdRp region of RaTG13 in SRX8357956.

Within the amplicons labeled "7896", there were 2 sites of overlap—the first overlap, a region 158bp in length, contained only 2nt difference—all C-T transitions—to SARS-CoV-2. Such transitions easily arise in passage, and are probable sequencing errors from a degraded/passaged sample of DNA.

The second overlap, one with BtCoV/4991, contained only 1 C-T transition, which have a probable origin in the primers used to generate the amplicons in the first place.

DISCUSSIONS

Origins of the anomalies in SRX7724752

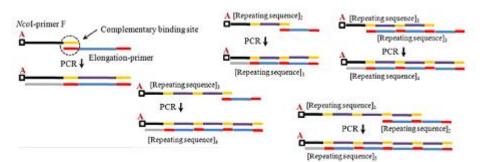


Figure 18A: Origin of repeating sequences in SRX7724752.

The only satiable explanation for the anomalous enrichment of the Telomere-like repeats in SRX7724752, involves the self-amplification of such sequences in a PCR reaction with little to no template.

Normally, with significant amount of template, the random primers normally used in RT-PCR amplifies most sequences evenly and outcompetes the repeat sequences in the reaction, and the result was an accurate reflection of such repeats within cellular samples—extremely poor. However, in samples that have little to no template, such that the random primers/random hexamers used in the reaction were not able to prime the amplification of most sequences—e.g. the amount of normal templates within the reaction falls below the timescale needed for the amplification of the repeating sequences, Repeating sequences, of which telomere-like repeats forms the vast majority of it in the environment and in most samples, can self-amplify in a primer-independent fashion, eventually reaching very high dominance, through repeated denaturing, sliding, reannealing and extension.

As this is a linear process, the self-amplification process is very slow, and is normally outcompeted by the normal amplicons as long as any usable amount of templates were present. Therefore, the presence of anomalously enriched telomere-like repeats within SRX7724752 indicate that the original sample couldn't have contained enough templates for the generation of the complete genome, through any means possible.

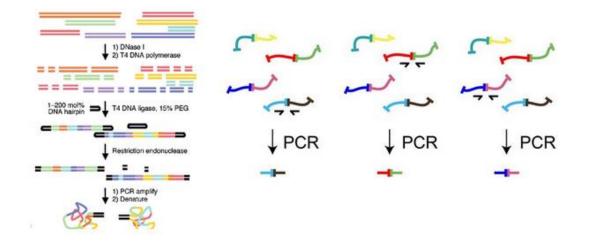


Figure 18B: Origin of the random matched sequences and partial sequences in SRX7724752 The majority of the sequences that were not repeats, when BLASTed, does not match any known organisms. There were also many sequences that matches—only partially, to many diverse organisms. What was striking, is, however, is that these matches often ends with an "A" or a "T". The most possible explanation of this anomaly is that the Library preparation process of ILLUMINA RNA-seq, which include strand synthesis and A-T ligation to adaptors, were fed dsDNA rather than ssRNA, as input. Such dsDNA input may be PCR products, or it may be a pre-made ILLUMINA sequencing library—Certain mRNA-like reads were inadvertently inverted, indicating double-stranded cDNA was likely used instead of single stranded mRNA.

RID	<u>J85DMESK016</u> Search expires on 08-01 20:59 pm <u>Download All</u> >
Results for	2:lcl Query_5937 gnl SRA SRR11085797.66.2 66 (Biological)(150bp)
Program	BLASTN 3 Citation V
Database	nt <u>See details</u> Y
Query ID	lcl Query_5937
Description	gnl SRA SRR11085797.66.2 66 (Biological)
Molecule type	dna
Query Length	150
Other reports	Distance tree of results 🔞

▲ <u>Download</u> **∨** <u>GenBank</u> <u>Graphics</u>

PREDICTED: Hipposideros armiger putative P2Y purinoceptor 10 (LOC109385656), mRNA

Sequence ID: XM_019648164.1 Length: 2682 Number of Matches: 1

Range 1: 2162 to 2293 GenBank Graphics

Next Match <u>Next Match</u> <u>Previous Match</u>

Score		Export	Identities	Gana	Strand	
	(00)	Expect		Gaps		
154 bit	S(83)	1e-33	116/132(88%)	1/132(0%)	Plus/Minus	
Query	12	TTT-TCATTATT.	AAGTATTATGTACTGT	ACATAATTGTATGTAG	TATACTTTTATACAAC	70
Sbjct	2293	TTTATCATTATC.	AAGTGTTATGTACTGT	ACAGTATTGTATGTG	TATACTTTTATATGAC	223
Query	71	TGGCAGCACAGC	AGGTTTGTTTATACCA	GCATCACCACAAAAA	GTGAGTAATGCATTAC	130
Sbjct	2233	TGACAGCATAGT	AGGCTTGTTTACACCA	GCATCACCACAAAAA	IGTGAGTAATGCATTAC	217
Query	131	ACTACAATGTTA	142			
Sbjct	2173	ACTATGATGTTA	2162			

Figure 19: An inverted mRNA-like read.

*: Analysis of the sole sgRNA-like read reveal the usage of a leader/F primer and the mispriming of Amplicon DNA

						tt	agat <mark>t</mark>	tcatctaaacgaacaaactaaaatgtctgataatggaccccaaaaccaacgaaatgcaccccgcattacgtttggtggaccct
			CT	CTCGATO	TCTT	TAGA	TCTG	TC TCTAAACGAAC
2	CAAACO	CAACG	AACT	CTCGATO	TCTT	TAGA	TCTG	
							1	AACCTCCTTTTGTCCCCTAGTTCACCGATCTCATCGCCCCTGCCAGGGGTCCATGGACTGTGTTGATCCTGTCCTCTGCTCCTCC

Figure 20a: the match analysis between different genomic fragments of RaTG13, of the sole sgRNA-like read* in SRX7724752. *:Figure 13

Despite being sgRNA-like in the first glance, analysis of the exact overlapping region of this particular read reveal that this region is identical to BtCoV/ZC45 and BtCoV/ZXC21—indicating it's identity as likely being a consensus primer.

select all 100 sequences selected	Gen	<u>Bank</u>	<u>Grap</u>	hics	Distance t	ree of results
Description	Max Score		Query Cover	E value	Per. Ident	Accession
Pangolin coronavirus isolate PCoV_GX-P3B_genomic sequence	67.6	67.6	100%	8e-09	100.00%	MT072865.1
Pangolin coronavirus isolate PCoV_GX-P2V, complete genome	67.6	67.6	100%	8e-09	100.00%	MT072864.1
Pangolin coronavirus isolate PCoV_GX-P5E, complete genome	67.6	67.6	100%	8e-09	100.00%	MT040336.1
Pangolin coronavirus isolate PCoV_GX-P5L_complete genome	67.6	67.6	100%	8e-09	100.00%	MT040335.1
Pangolin coronavirus isolate PCoV_GX-P1E, complete genome	67.6	67.6	100%	8e-09	100.00%	MT040334.1
Pangolin coronavirus isolate PCoV_GX-P4L, complete genome	67.6	67.6	100%	8e-09	100.00%	MT040333.1
Bat coronavirus RaTG13_complete_genome	67.6	67.6	100%	8e-09	100.00%	MN996532.1
Bat SARS-like coronavirus isolate bat-SL-CoVZXC21, complete genome	67.6	67.6	100%	8e-09	100.00%	MG772934.1
Bat SARS-like coronavirus isolate bat-SL-CoVZC45, complete genome	67.6	67.6	100%	8e-09	100.00%	MG772933.1
Bat coronavirus (BtCoV/279/2005), complete genome	65.8	65.8	97%	3e-08	100.00%	DQ648857.1
Mutant SARS coronavirus Urbani clone SARS-Urbani-MA_SHC014-spike, complete genome	63.9	63.9	94%	1e-07	100.00%	MT308984.1
Coronavirus BtRs-BetaCoV/YN2018D, complete genome	63.9	63.9	94%	1e-07	100.00%	MK211378.1

Figure 20b: BLAST result of CTCTCGATCTCTTGTAGATCTGTTCTCTAAACGAAC.

This particular sequence have extended overlap to the beginning of the N gene, which was coincidentally at the end of the last 2017/06/17 amplicon. This indicate it was most likely the product of mispriming, rather than a true sgRNA-like read.

SRX7724752 is a mixed library consists of a matrix of dried American bat guano, a bat WGS/RNA-seq ILLUMINA library, a synthetic 16S library and megaprimer PCR products from the SRX8357956 Amplicons and a degraded sample of SARS-CoV-2

cDNA

Phyllostomus Discolor, a species of bat native to Mexico and southern United states, leaves numerous Full-length 100% matched reads that don't match anything else. Coincidentally, Mexico is one of the major supplier of bat guano used for fertilizer and other commercial purposes[6]. The confirmed presence of this particular bat species, suggest the use of a commercial dried bat guano matrix as the bulk of the sample being sequenced. As in PRJNA494391[7] which synthetic metagenome samples were constructed using cDNA amplicons and a specific material matrix to simulate realistic metagenomic reads of a desired virus in a sample.

Traces of the original template used in the megaprimer PCR process can be seen as traces of low-matched virus-like reads within this dataset, which are found across the entire RaTG13 genome.



Figure 21: Read coverage of SRX7724752 on the RaTG13 genome. The red pixels represent significant mismatches on the reads in the dataset.

The Bacterial-like reads in SRX7724752 is also likely a synthetic 16S library—as the only other dataset with Telomere-like repeats(4%), still contained significant amount of bacterial mRNA.

>gnl|SRA|SRR11085733.2232944.1 2232944 (Biological)

GCCTTCGTTTGTATATAGTTTTAATGCAAATCCCCTAACATCTCTTTCAGCATCTGCTGC ACCTCTTTCACCAGCAACTGTAGAAAATCTTAAAAGGGCTTTTGTTTTTTACCAACTTT GTTAAAAATATCTGCTTTAGAATATTTTGT

>gnl|SRA|SRR11085733.2232944.2 2232944 (Biological)

AGAGGTCCTACTCTTTTACAAGATACTTGGCTTTTAGAAAAACTTGCACATTTCGATAGG GAAAGGATACCAGAAAGAGTTGTGCACGCTAAAGGAAGTGCTGCATACGGCGAATTAACA ATTACTAATGATATTACAAAATATTCTAAA

Helicobacter hepaticus ATCC 51449, complete genome

Sequence ID: AE017125.1 Length: 1799146 Number of Matches: 1

Range 1	L: 47720 t	o 47856 🤇	GenBank Grap	hics		▼ <u>Next</u>	Match 🔺 Pre	evious Match
Score 176 bit	s(95)	Exped 3e-40		(90%)	Gaps 0/137(0%)	Strand Plus/Min	us	
Query Sbjct	7 47856	11111				CACATTTCGATA 		66 47797
-	67					ACGGCGAATTAA		126
Query Sbjct	47796		1 11111 111			ACGGCGAATTAA ATGGTGAATTAA		47737
-							01111111011	1,,,,,,
Query	127		TTACAAAATA	143				
Sbjct	47736	AATGATA	TTACTCAATA	47720				
ç	gene		complement /gene="kati /locus_tage	A" ="HH_0043"	-			
(CDS		/old_locus_ complement	_tag="HH004: (<1 \\137)	3"			
Ì			/gene="kati					
			/locus_tag	_				
				_tag="HH004: ="1.11.1.6"	3"			
			/codon sta:					
			/transl_tak					
			/product="o					
				1=" <u>AAP76640</u>				
ORIGI	4		ERIPERVVHA AERDVRGFAL WDFWSLHPESI GIHNLTNKESI LTKVWSHKDYI SYGDTQRYRL(NALEPPLFVQI	KGSAAYGELTI KLYTNEGNWDI LHQVTILMSDR EAVIAKDRESH PLIEVGILELNI GINHTQLPVNA	INDITQYTKAE VGNNTPVFFIK GIPRSYREMHG QKDLFENIEKG KNPENYFAEVE PIVPVNNTHRD EYEDDYFVQAG	SITAGKKGPTLI LFNKVGKKTKAF DAIKFPDFIHTQ FGSHTYSFINAF NFPKWRFCIQVM QAAFNPANIVPG GFMQQGQFGDRF DLYRLMTAEQKF	TLRFSVVAGI OKRDPKTNMI (NERFWVKFI (SEKEAENYI SVGYSPDKVI (NYEPSYFNI	ERGAAD KSPTAM HFVCLQ RFNPFD LQGRLF DYVEDK
			_	-	_	ctgcacttcc		
				ctctctatca	aaatgtgcaa	gtttttctaa	aagccaagt	ta
11	121 000	tgtaaaa	guguagg					

Figure 22a: a bacterial mRNA read in SRX7724696, the only other dataset on NCBI that contained Telomere-like repeats in the first 100 reads of the dataset. Total amt. of repeats=4%

Job Title	AE017125:Helicobacter hepaticus ATCC 51449,
RID	<u>J88SF43U01R</u> Search expires on 08-01 21:56 pm <u>Download All</u> ✓
Program	Citation ✓
Database	SRA <u>See details</u> ✓
Query ID	AE017125.1
Description	Helicobacter hepaticus ATCC 51449, complete genome
Molecule type	nucleic acid
Query Length	934935
Other reports	0



No significant similarity found. For reasons why, click here

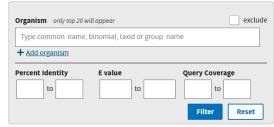
Figure 22b: the same species of bacteria in SRX7724752. No significant matches were found. This dataset is likely subjected to probe-capture sequencing similar to these other datasets—the use of a positive-sense CoV probe resulted in the selective presentation of the negative ssDNA strand of the ligation products to show up. This is supported by the observation that while most of the virus-like reads were on the negative strand, the Repeats does not show a bias in strand polarity, and the mRNA-like reads have a much higher chance of being on the wrong polarity for RNA-seq. This is likely due to the ligation process being used.

Probable signs of laboratory manipulation of SRX7724752

Job Title	gb AC097711.2	Filter Results						
ND	KJ2J1YCV01R Search expires on 08-17 18:28 pm Download All							
rogram	BLASTN ? Citation V	Organism only top 20 will appear exclu						
atabase	SRA <u>See details</u> ~	Type common name,	binomi	al, taxi	d or gro	oup nam	ie	
uery ID	AC097711.2	+ Add organism						
escription	Homo sapiens BAC clone RP11-162K6 from 4, complete sequence	Percent Identity	E va	lue			Query Coverage	
olecule type	nucleic acid	to			to		to	
uery Length	114657							
ther reports	Distance tree of results MSA viewer 😨						Filter Reset	
Descriptions	Graphic Summary Alignments							
Sequences p	producing significant alignments	Downloa	d ~	Ма	nage Co	olumns	∽ Show 100 ✔ 3	
🗹 select all	100 sequences selected					Graph	nics Distance tree of result	
	Description	Max Scor		Query Cover	E value	Per. Ident	Accession	
SRX772475	2	279	279	0%	1e-70	100.00%	SRA:SRR11085797.11044608.2	
	2	279	279	0%	1e-70	100.000	SRA:SRR11085797.11044608.1	

Job Title	gnl SRA SRR11085797.11044608.1 11044608
RID	KJ2FP0KE014 Search expires on 08-17 18:27 pm Download All ♥
Results for	2:lcl Query_54299 gnl SRA SRR11085797.11044608.2 11044608 (Biological 🗸
Program	BLASTN ? <u>Citation</u> ~
Database	nt <u>See details</u> •
Query ID	lcl Query_54299
Description	gnl SRA SRR11085797.11044608.2 11044608 (Biological)
Molecule type	dna
Query Length	151
Other reports	Distance tree of results 🔞

Filter Results



Descr	iptions	Graphic Summary	Alignments	Taxonomy							
Sequ	iences pro	oducing significant a	alignments		Download 🗡	Ма	nage (Colum	ns ~	Show	100 🗸
Select all 11 sequences selected GenBank Graphics Distance tree of results											
			Des	cription		Max Score			E value	Per. Ident	Accession
	Homo sapiens	BAC clone RP11-162K6 from	4. complete sequence			279	279	100%	2e-71	100.00%	AC097711.2
~ (Canis lupus fai	miliaris breed Labrador retrieve	er chromosome 32a			274	274	100%	9e-70	99.34%	CP050598.1
<u> </u>	Canis lupus fai	miliaris breed Labrador retrieve	er chromosome 32b			274	274	100%	9e-70	99.34%	CP050634.1
<u> </u>	Aquila chrysae	tos chrysaetos genome asser	mbly, chromosome: 1			257	257	100%	9e-65	97.35%	LR606181.1
2	Apteryx austra	<u>lis mantelli genome assembly</u>	AptMant0_scaffold_sca	ffold176		257	257	100%	9e-65	97.35%	LK064748.1
2	Erithacus rube	cula genome assembly, chron	nosome: 5			252	252	100%	4e-63	96.69%	LR812107.1
	Anas platyrhyr	tchos genome assembly, chro	mosome: 4			252	252	100%	4e-63	96.69%	LS423614.1
2 <u>8</u>	Streptopelia tu	rtur genome assembly, chrom	osome: 4			246	246	100%	2e-61	96.03%	LR594554.1
~ 1	Mus musculus	BAC clone RP24-204J10 from	n 5. complete sequence			243	243	98%	3e-60	95.97%	AC121929.2
2	Sciurus carolin	ensis genome assembly, chro	omosome: 15			204	204	78%	1e-48	97.48%	LR738605.1
•	PREDICTED: I	<u>Meleagris gallopavo uncharact</u>	erized LOC104910685 (L	_OC104910685), mRM	Δ	121	121	45%	1e-23	98.53%	<u>XM_01961511</u>

Filter Results Job Title gnl|SRA|SRR11085797.11044608.1 11044608 RID KJ2FP0KE014 Search expires on 08-17 18:27 pm Download All V **Results** for 1:lcl|Query_54298 gnl|SRA|SRR11085797.11044608.1 11044608 (Biological V Program BLASTN 😧 Citation 🗸 Database nt <u>See details</u> 🗸 Query ID lcl|Query_54298 gnl|SRA|SRR11085797.11044608.1 11044608 (Biological) Description dna Molecule type

Graphic Summary Alignments Taxonomy

Search Summary 🗙

< Edit Search

Query Length

Other reports

Descriptions

151

Distance tree of results (2)

Save Search

② How to read this report? ■ BLAST Help Videos ⑤Back to Traditional Results Page

Organism only top 20 v	vill appear	exclude
Type common name,	binomial, taxid or gr	oup name
+ Add organism		
Percent Identity	E value	Query Coverage
to	to	to
		Filter Reset

Seq	uences producing significant alignments	Download 🗡	Manag	e Colu	mns	∽ Sho	ow 10) 🗸 🔞	
Select all 4 sequences selected GenBank Graphics Distance tree of results									
	Description		Max Score		Query Cover	E value	Per. Ident	Accession	
✓	Homo sapiens BAC clone RP11-162K6 from 4, complete sequence		279	279	100%	2e-71	100.00%	AC097711.2	
✓	Canis lupus familiaris breed Labrador retriever chromosome 32a		274	274	100%	9e-70	99.34%	CP050598.1	
✓	Canis lupus familiaris breed Labrador retriever chromosome 32b		274	274	100%	9e-70	99.34%	CP050634.1	
<	Mus musculus BAC clone RP24-204J10 from 5, complete sequence		252	252	100%	4e-63	96.69%	AC121929.2	

Figure 23: Unique, fully-matched 100% read from Homo Sapiens is recovered from the dataset SRX7724752.

< Edit Search	Save Search Search Summary 💙	How to read this report? 💶 BLAST Help Videos Dack to Traditional Results Page								
Job Title	ref NW_015351248.1	Filter Results								
RID	KJ435D3801R Search expires on 08-17 18:54 pm Download All 💙									
Program	BLASTN 😧 Citation 🗸	Organism only top 20 will appear exclud								
Database	SRA See details ✓ Type common name, binomial, taxid or group name									
Query ID	NW_015351248.1 + Add organism									
Description	Marmota marmota marmota unplaced genomic scaffold, mai	Percent Identity E value Query Coverage								
Molecule type	dna	to to to								
Query Length	19578880									
Other reports	Distance tree of results MSA viewer 😨	Filter Reset								
Descriptions	Graphic Summary Alignments									
Sequences p	producing significant alignments	Download Y Manage Columns Y Show 100 V								
🔽 select all	100 sequences selected	Graphics Distance tree of results								
	Description	Max Total Query E Per. Score Score Cover value Ident								
SRX77247	<u>52</u>	279 279 0% 2e-68 100.00% <u>SRA:SRR11085797.7041148.2</u>								
SRX77247	<u>52</u>	278 278 0% 9e-68 100.00% <u>SRA:SRR11085797.4666666.2</u>								
SRX77247	52	276 276 0% 3e-67 100.00% <u>SRA:SRR11085797.8742622.2</u>								

< Edit Search Save Search Search Summary 🗙 How to read this report? BLAST Help Videos DBack to Traditional Results Page Filter Results Job Title 2 sequences (gnl|SRA|SRR11085797.10431565.1... RID KJ3PSXPK014 Search expires on 08-17 18:47 pm Download All ↔ exclude Organism only top 20 will appear 1:lcl|Query_51896 gnl|SRA|SRR11085797.10431565.110431565 (Biological 🗸 **Results** for Type common name, binomial, taxid or group name Program BLASTN ? Citation ~ + Add organism Database nt <u>See details</u> 🗸 Query ID lcl|Query_51896 Percent Identity E value Query Coverage gnl|SRA|SRR11085797.10431565.1 10431565 (Biological) Description to to to Molecule type dna Filter Reset 151 Query Length Other reports Distance tree of results (2) Alignments Descriptions Graphic Summary Taxonomy

Sec	uences producing significant alignments Down	load 🌱 M	anage	Colum	ns ~	Show	100 🗸 🔞
	select all 45 sequences selected		<u>GenBar</u>	<u>nk Gr</u>	aphics	Distan	ce tree of results
	Description	Ma Sco	c Total e Score			Per. Ident	Accession
	Canis lupus familiaris breed Labrador retriever chromosome 32a	27	279	100%	2e-71	100.00%	CP050598.1
	Canis lupus familiaris breed Labrador retriever chromosome 32b	27	279	100%	2e-71	100.00%	CP050634.1
	Homo sapiens VISTA enhancer hs712 (LOC110120752) on chromosome 4	27	279	100%	2e-71	100.00%	NG_053377.1
	Homo sapiens BAC clone RP11-476H13 from 4, complete sequence	27	279	100%	2e-71	100.00%	AC024192.6
	Aquila chrysaetos chrysaetos genome assembly, chromosome: 1	27	274	100%	9e-70	99.34%	LR606181.1
	Streptopelia turtur genome assembly, chromosome: 4	27	274	100%	9e-70	99.34%	LR594554.1
	PREDICTED: Cyanistes caeruleus uncharacterized LOC111928864 (LOC111928864), ncRNA	27	274	100%	9e-70	99.34%	XR_002864354.1
	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold564	27	274	100%	9e-70	99.34%	LK065221.1
	Anas platyrhynchos genome assembly, chromosome: 4	27	274	100%	9e-70	99.34%	LS423614.1

Job Title	2 sequences (gnl SRA SRR11085797.4967337.1
RID	KJ8SKF2E014 Search expires on 08-17 20:14 pm Download All V
Results for	2:lcl Query_37435 gnl SRA SRR11085797.4967337.2 4967337 (Biological)(: 🗸
Program	BLASTN 😧 <u>Citation</u> 🗸
Database	nt <u>See details</u> ♥
Query ID	Icl Query_37435
Description	gnl SRA SRR11085797.4967337.2 4967337 (Biological)
Molecule type	dna
Query Length	151
Other reports	Distance tree of results 🔞
Descriptions	Graphic Summary Alignments Taxonomy

Filter Results

Organism only top 20 v	vill appear	exclude
Type common name,	binomial, taxid or grou	up name
+ Add organism		
Percent Identity	E value	Query Coverage
to	to	to
		Filter Reset

	criptions	Graphic Summary	Alignments	Taxonomy								
Seq	juences pro	oducing significant a	lignments			Download 🗠	Mana	ge Col	umns	⊻ Sh	iow 1	00 🗸 🔞
~	select all 58	sequences selected					Gen	<u>Bank</u>	Grap	hics J	Distance t	ree of result
			De	scription			Max Score		Query Cover	E value	Per. Ident	Accession
~	Canis lupus familiaris breed Labrador retriever chromosome 32a									2e-71	100.00%	CP050598.1
~	Canis lupus far	miliaris breed Labrador retriever	chromosome 32b				279	279	100%	2e-71	100.00%	CP050634.1
~	Rhesus Macag	<u>ue BAC CH250-499F11 () com</u>	<u>plete sequence</u>				274	274	100%	9e-70	99.34%	AC204215.5
~	Homo sapiens	BAC clone RP11-678H22 from	4. complete sequence				274	274	100%	9e-70	99.34%	AC096766.3
~	Mus musculus	chromosome 5, clone RP24-31	15H14, complete seque	nce			263	263	100%	2e-66	98.01%	AC105976.1
~	Chrysemys pic	ta isolate 4965chr ultra consen	ved element locus chr4	<u>11164 genomic seq</u>	Jence		257	257	100%	9e-65	97.35%	<u>JQ873778.1</u>
~	Alligator missis	ssippiensis isolate 333all ultra o	conserved element locu	s chr4_11164_genom	ic sequence		257	257	100%	9e-65	97.35%	<u>JQ869146.1</u>
~	Apteryx austra	lis mantelli genome assembly /	AptMant0, scaffold scaf	fold564			241	241	100%	9e-60	95.36%	LK065221.1
~	Erithacus rubecula genome assembly, chromosome: 5								82%	6e-57	100.00%	LR812107.1
~	Streptopelia tur	rtur genome assembly, chromo	some: 4				231	231	82%	6e-57	100.00%	LR594554.1
~	Aquila chrysae	tos chrysaetos genome assem	bly, chromosome: 1				226	226	82%	3e-55	99.20%	LR606181.1

Figure. 24: Marmota Marmota genetic scaffold assemblies returned significant amount of 100% full-length matched reads that were sometimes also found in Homo Sapiens and Canis Lupus Famillaris.

Reads (separated)												
>gni SRA SRR11085797.6341838.1 6341838 <i>(Biological)</i> cgagaccatcctggctaacacggtgaaaccccgtctctactaaaaatacaaaaattagc cgggcgtgatggcgggcgcgctgtagtcccagctactcgggaggctgaggcaggagaatgg cgtgaacccgggaggcggagcntgcagtg												
>gnl SRA SRR11085797.6341838.2 6341838 (Biological) CTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAG CTGGGACTACAGGCGCCCGCCATCACGCCCGGCTAATTTTTTGTATTTTTAGTAGAGACG GGGTTTCACCGTGTTAGCCAGGATGGTCTCG												
Description gnl SRA SRR11085797.6341838.2.6341838 (Biological) to		to		to								
Molecule type dna			F	ilter Rese	et							
Other reports Distance tree of results 😨												
Descriptions Graphic Summary Alignments Taxonomy												
Sequences producing significant alignments	Download 🗡	Manage Colum	ns 🔨 Sh	iow 100 🗸	0							
Select all 100 sequences selected		<u>GenBank</u> <u>G</u>	raphics [Distance tree of re	<u>sults</u>							
Description	Max Score	Total Que Score Cove	ry E er value	Per. Accessi	ion							
Pongo abelii chromosome 10 clone CH276-32684, complete sequence	278	1689 100	% 7e-71 1	00.00% AC27064	<u>5.1</u>							
Pongo abelii chromosome 10 clone CH276-12G11, complete sequence	278			00.00% AC27053	3.1							
Pongo abelii chromosome 10 clone CH276-5H9, complete seguence	278		% 7e-71 1		_							
Homo sapiens beta-1,3-galactosyltransferase 1 (B3GALT1), mRNA Homo sapiens chromosome 1 clone VMRC53-455P10, complete sequence	274 274			99.34% <u>NM_0209</u> 99.34% AC27856								
Homo sapiens chromosome 1 clone VMRC53-455P10, complete sequence	274	3094 100	/0 98-70 5	55.3470 <u>MG2/830</u>								

Figure.25a: 100% full-length matched reads to Hominid(Pongo Albelii) genomic DNA.

Human endogenous retrovirus H HERV-H/env60 proviral copy, clone 734E12	252	25	2	100%	4e-63	96.69%	AJ289710.2
Synthetic human HSC3N1 Alu seguence	252	25	2	100%	4e-63	96.69%	<u>U02043.1</u>
Synthetic construct, complete sequence	250	363	35	100%	1e-62	96.69%	<u>JN255744.1</u>
Human artificial chromosome vector 21HAC4 DNA, isolated from the long arm, clone: YAC/BAC#26-2	250	196	64	100%	1e-62	96.69%	AB553834.1
Human ORFeome Gateway entry vector pENTR223-MGC2752, complete sequence	246	24	6	100%	2e-61	96.05%	LT735229.1
Expression vector pUMLIEP DNA, complete sequence	246	24	6	99%	2e-61	96.05%	LC175306.1
Synthetic construct Homo sapiens clone ccsbBroadEn_10246 MGC2752 gene, encodes complete protein	246	24	6	100%	2e-61	96.05%	KJ900852.1
HIV-1 isolate HK_JIDLNBL_S071 from Switzerland nonfunctional gag protein (gag) gene, complete sequence; and nonfunction	244	107	72	100%	7e-61	96.00%	MT154980.1
Cloning vector pSuper_7SL_AluAA 7SL enhancer and AluYa5 repeat element sequence	241	24	1	100%	9e-60	95.36%	EU092258.1
Cloning vector pSuper_7SL_AluA 7SL enhancer and AluYa5 repeat element sequence	241	24	1	100%	9e-60	95.36%	EU092257.1
Synthetic construct clone AluAU SRP promoter region and Alu repeat element sequence	241	24	1	100%	9e-60	95.36%	AF458115.1
Synthetic construct clone AluWD SRP promoter region and Alu repeat element sequence	241	24	1	100%	9e-60	95.36%	AF458112.1
Synthetic construct clone Alut253 SRP promoter region and Alu repeat element sequence	241	24	1	100%	9e-60	95.36%	AF458107.1
Synthetic construct clone Alu+A SRP promoter region and Alu repeat element sequence	241	24	1	100%	9e-60	95.36%	AF458106.1
Desmodus rotundus isolate DRU21DN04 contig68764, whole genome shotgun sequence	108	216	63%	2e-20	87.23	% PEH	R01068758.1
Myotis lucifugus cont2.6286, whole genome shotgun sequence	108	108	55%	2e-20	90.36	% AAP	E02006287.1
Artibeus jamaicensis isolate US092 ArtJam scaffold 27825, whole genome shotgun sequence	104	104	51%	2e-19	90.91	% PVK	R01013927.1
Macrotus californicus isolate US035 MacCal _ line_566643, whole genome shotgun sequence	102	102	51%	9e-19	90.79	% <u>VMD</u>	R010283404.1
Anoura caudifer isolate US021 AnoCau_scaffold_336054, whole genome shotgun sequence	102	102	61%	9e-19	86.96	% <u>PVK</u>	U01163203.1
Anoura caudifer isolate US021 AnoCau_scaffold_250162, whole genome shotgun sequence	102	102	61%	9e-19	86.96	% <u>PVK</u>	U01121529.1
Anoura caudifer isolate US021 AnoCau_scaffold_157416, whole genome shotgun sequence	102	102	61%	9e-19	86.96	% <u>PVK</u>	U01078866.1
Anoura caudifer isolate US021 AnoCau_scaffold_136788, whole genome shotgun sequence	102	102	61%	9e-19	86.96	% <u>PVK</u>	U01068554.1
Anoura caudifer isolate US021 AnoCau_scaffold_6229, whole genome shotgun sequence	102	102	51%	9e-19	90.79	% <u>PVK</u>	U01003121.1
Anoura caudifer isolate US021 AnoCau_scaffold_1146, whole genome shotgun sequence	102	102	71%	9e-19	84.26	% <u>PVK</u>	<u>U01000576.1</u>
Artibeus jamaicensis isolate US092 ArtJam_scaffold_590481, whole genome shotgun sequence	102	102	51%	9e-19	90.79	% <u>PVK</u>	R01295479.1
Artibeus jamaicensis isolate US092 ArtJam_scaffold_272373, whole genome shotgun sequence	102	102	51%	9e-19	90.79	% <u>PVK</u>	R01136397.1
Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000055F_070_arrow_arrow_whole genome shotgun sequence	101	101	51%	8e-19	88.46	% <u>RX</u>	PD01003063.1
Rhinolophus ferrumequinum isolate mRhiFer1 scaffold_m29_p_7, whole genome shotgun sequence	101	151	51%	8e-19	88.46	% <u>JA(</u>	CAGC010000007.1
Rhinolophus ferrumequinum RF_contig_107525, whole genome shotgun sequence	101	101	51%	8e-19	88.46	% <u>AW</u>	HA01101756.1
Rhinolophus ferrumequinum Isolate US033 RhiFer flattened line 8799, whole genome shotgun sequence	97.8	186	50%	1e-17	88.16	% <u>VM</u>	DN01004402.1
Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000061F_062_arrow_arrow_whole genome shotgun sequence	97.8	186	50%	1e-17	88.16	% <u>RX</u>	PD01001710.1
Rhinolophus ferrumeguinum isolate MPI-CBG mRhiFer1 chromosome 6, whole genome shotgun sequence	97.8	309	50%	1e-17	88.16	% <u>RX</u>	PC01000086.1
Rhinolophus ferrumequinum isolate mRhiFer1 scaffold_m29_p_8, whole genome shotgun sequence	97.8	309	50%	1e-17	88.16	5% <u>JA(</u>	CAGC01000008.1
Rhinolophus ferrumeguinum isolate MPI-CBG mRhiFer1 000003F_100_arrow_arrow, whole genome shotgun sequence	93.3	93.3	49%	1e-16	88.00	1% <u>RX</u>	PD01006157.1
Rhinolophus ferrumequinum isolate mRhiFer1 scaffold_m29_p_4, whole genome shotgun sequence	93.3	93.3	49%	1e-16	88.00	% JAC	CAGC010000004.1
Rhinolophus ferrumeguinum isolate US033 RhiFer flattened line 6166, whole genome shotgun sequence	90.6	90.6	44%	1e-15	89.55	5% <u>VM</u>	DN01003085.1
Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000061F_073_arrow_arrow, whole genome shotgun sequence	90.6	90.6	44%	1e-15	89.55	% <u>RX</u>	PD01006658.1

Fig.25b: BLAST search of this sequence revealed it to be a Homo Sapiens endogenous Retrovirus most similar to HIV-1, and is not found in any known bat genomic assemblies. This sequence is also found in several cloning vectors for mammalian DNA. Significance of these sequences are currently unknown.

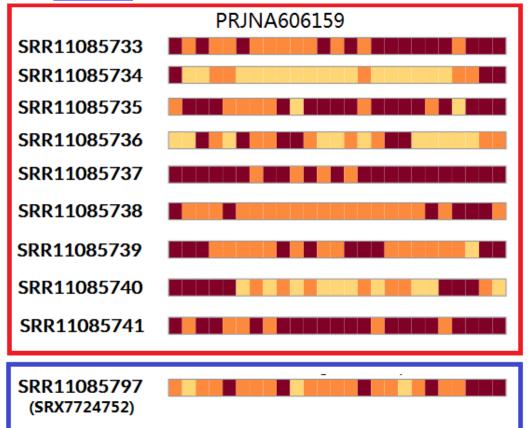
SRX7724752 contained Traces of confirmed contamination from other organisms, in particularly that of order Carnivora, Rodentia and Homo Sapiens. As such DNA contamination mostly happen during extensive manipulation of samples in the labs, This indicate that SRX7724752 Contained traces of laboratory manipulation, including Canis Lupus Famillaris DNA contamination which could not have been present in a fecal sample of a bat, even assuming normal lab manipulation for sequencing purposes.

This indicate the sample may have been subjected to in-vitro manipulation.

No evidence of methodological reasons for the generation of

anomalies in SRX7724752

In Order to test whether a specific sequencing technique was used for the sequencing of SRX7724752 which may have generated the anomalies observed above, we decided to use the sequencing depth of the Coronaviruses within SRX7724752 and compare it against another set of



mNGS sequencing data of identical sample, origin, institute and submitted at the same date, located in <u>PRJNA606159</u>.

Fig. 26a: the Coverage map of Coronaviridae within the datasets located in PRJNA606159, compared against SRR11085797.

We generated the sequencing depth Heatmap [8] of all datasets located within PRJNA606159, and the sequencing depth pattern of the Coronavirus reads within such dataset does not show any statistical differences from that of SRR11085797.

<u>SRX7724696</u>: RNA-Seq of Hipposideros larvatus: Anal swab 1 ILLUMINA (Illumina HiSeq 3000) run: 13.5M spots, 3.9G bases, 1.8Gb downloads

Design: Total RNA was extracted from bronchoalveolar lavage fluid using the QIAamp Viral RNA Mini Kit following the manufacturers instructions. An RNA library was then constructed using the TruSeq Stranded mRNA Library Preparation Kit (Illumina, USA). Paired-end (150 bp) sequencing of the RNA library was performed on the HiSeq 3000 platform (Illumina).

Submitted by: Wuhan Institute of Virology, Chinese Academy of Sciences

<u>SRX7724752</u>: RNA-Seq of Rhinolophus affinis:Fecal swab 1 ILLUMINA (Illumina HiSeq 3000) run: 11.6M spots, 3.3G bases, 1.7Gb downloads

Design: Total RNA was extracted from bronchoalveolar lavage fluid using the QIAamp Viral RNA Mini Kit following the manufacturers instructions. An RNA library was then constructed using the TruSeq Stranded mRNA Library Preparation Kit (Illumina, USA). Paired-end (150 bp) sequencing of the RNA library was performed on the HiSeq 3000 platform (Illumina).

Submitted by: Wuhan Institute of Virology, Chinese Academy of Sciences

Fig.26b: the experimental design section of the datasets within PRJNA606159 is identical to that of SRX7724752.

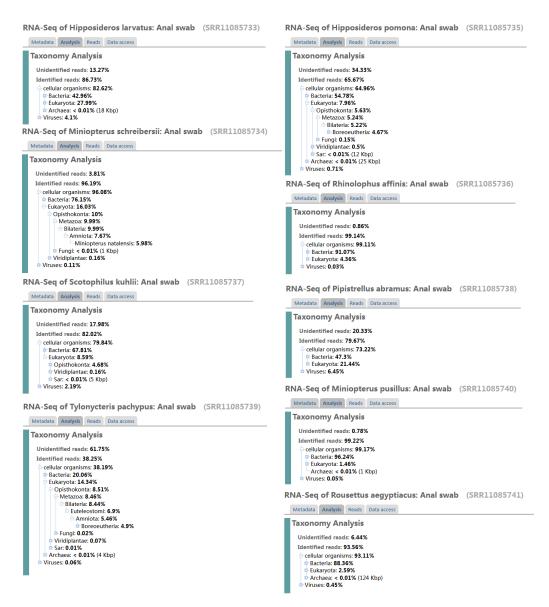


Fig.26c: No evidence of any anomalies were found within the datasets presented in PRJNA606159

We recently obtained a set of viral mNGS coverage data from a sequencing experiment that Uses PolyA enrichment for the selection of sequences [9].

Despite being isolated from the total RNA of freshly dissected and cleaned Bee Tissue samples, these PolyA enriched datasets displayed a heavy bias toward the 3'-end for all viral genomes that contained a polyA tail, and did not obtain any coverage past 8000nt to the 3'-end of such viral genomes. This is consistent with the fact that viral genomic RNA obtained from samples, even when freshly prepared, will always suffer from numerous RNA strand breaks, and therefore will be heavily biased toward the 3'-end as the enrichment process would have kept mostly the RNA that contained an intact polyA tail. As Coronaviruses have a PolyA tail, this is in sharp contrast to that found in SRX7724752, which does not show signs of such bias.

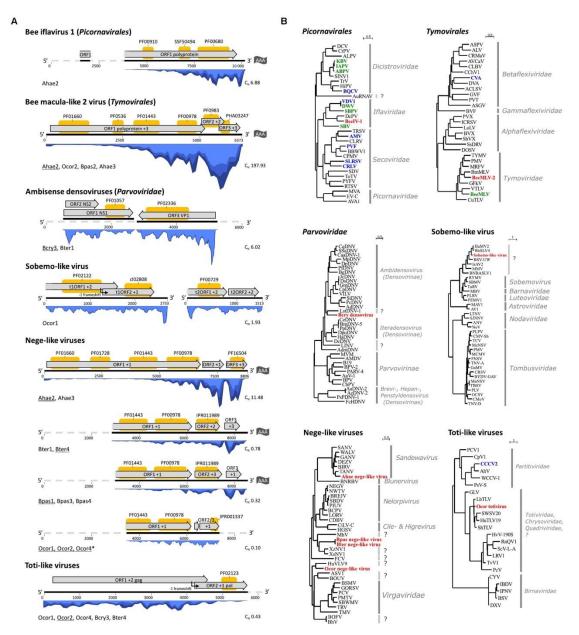


Figure 2 from [9]. A clear bias toward the 3'-end of RNA viral families that contained a polyA tail, was noticed.

The anomalies in SRX7724752 is associated with the absence of RNA viruses.

In order to further analyze the implications of the observed anomalies in SRX7724752, we performed a Keyword search on NCBI SRA using the Keyword "Bat feces" and "Bat fecal". We did not find any evidence of an RNA virus (Riboviria) within any of the returned datasets that contained less than 2.5% bacteria in total cellular organisms that can be confirmed by BLAST.

Description	Phaseolus vulgaris endornavirus 1 isolate PvEV-1_Brazil poly
Molecule type	nucleic acid
Query Length	14072
Other reports	0

A

No significant similarity found. For reasons why, click here

Figure 27: an example of a TRACE result that does not actually exist when BLASTed against the reference sequences of said virus.

Taxonomy Analysis
Unidentified reads: 46.05%
Identified reads: 53.95%
 cellular organisms: 53.95% Eukaryota: 52.61% Opisthokonta: 52.54% Vespertilionidae: 50.36% Myotis: 40.39% Viridiplantae: 0.03% Bacteria: 1.31%
Viruses: < 0.01% (0 Kbp)

Figure 28a: an example of a bacteria-depleted dataset. An absence of Riboviria reads was noted.

Taxonomy Analysis Unidentified reads: 39.21% Identified reads: 60.79% cellular organisms: 54.23% Bacteria: 53.42% Eukaryota: 0.15% • Viruses: 6.56% Caudovirales: 5.47% Myoviridae: 5.31% unclassified Myoviridae: 5.3% Tevenvirinae: 0.01% Seoulvirus: < 0.01% (21 Kbp)</p> Vequintavirinae: < 0.01% (6 Kbp)</p> Podoviridae: 0.14% Siphoviridae: 0.03% Satellites: 0.56% Riboviria: 0.41% Iridoviridae: 0.11% unclassified viruses: < 0.01% (101 Kbp)</p> Microviridae: < 0.01% (3 Kbp)</p>

Figure 28b: in contrast, Riboviria reads are found only in datasets that contained a significant amount of bacteria.

In addition, We found only 1 dataset that contained any significant levels of a Telomere-like repeat sequence. However, this dataset does not contain any evidence of an RNA virus(Riboviria).

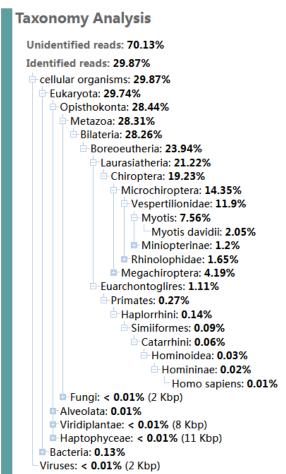


Figure 29: the only dataset with significant level of Telomere-like repeats (2%). There are no evidence of Riboviria(RNA viruses) within this dataset.

Signature of likely attenuation of the RaTG13 RBD.

The RaTG13 RBD have reduced binding affinity to ACE2 compared to

other viruses of the same clade.

Recently, a publication which tested the binding affinity and infection efficiency of RaTG13 S to human ACE2[10] have been published, which suggest that unlike the other RBDs within this clade (namely SARS-CoV-2 and pCoV_GX), RaTG13 can not bind to human ACE2 efficiently, and is incapable of entering cells through ACE2 that is expressed at physiological levels.

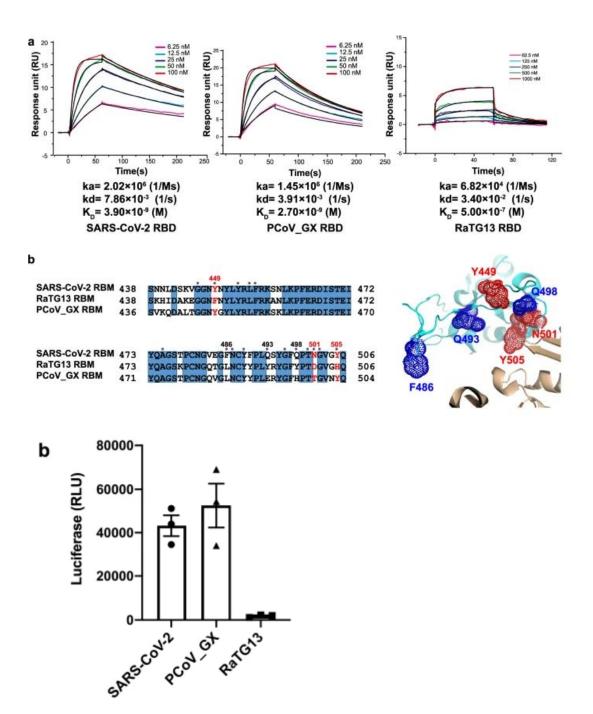


Fig.5a and 6b from [10]: RaTG13 RBD bound to human ACE2 very inefficiently, and did not show entry into hACE2-HEK293T cells at physiological level of ACE2 expression.

Initially considered as evidence of "bat specifity" for RaTG13, An recent test [11] did not find any higher binding affinity of the RaTG13 RBD to R.affinis ACE2 than to human ACE2—in fact, both the flow cytometry data and the pseudovirus entry data into HEK293T with Overexpressed ACE2 suggest a binding affinity of RaTG13 RBD to R.affinis ACE2 that is slightly lower than that of RaTG13 RBD to human ACE2.

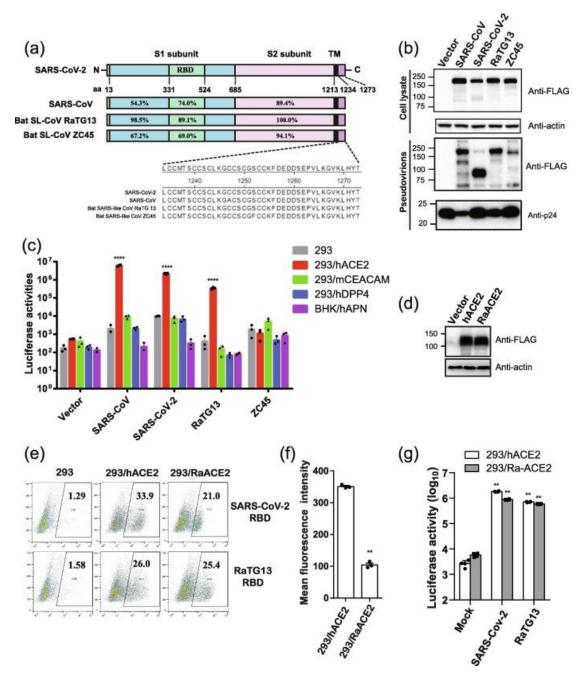


Fig.5 from [11]: d: The overexpression level for human and R.affinis ACE2 on the HEK293T cells are the same. e: The binding affinity of RaTG13 RBD-Fc on R.affinis ACE2(Lower Right) is slightly lower than the binding affinity of RaTG13 RBD-Fc on human ACE2(Lower Middle), and is significantly lower than the binding affinity of SARS-CoV-2 RBD to human ACE2(Upper middle). g: Pseudovirus entry assay using Lentivirus on HEK293T with Overexpressed ACE2 show that SARS-CoV-2-hACE2 > SARS-CoV-2-R.affinis ACE2 > RaTG13-hACE2 > RaTG13-R.affinis ACE2 by pseudovirus entry efficiency.

A multiple alignment of the RaTG13 Spike protein RBD to other Sarbecovirus Spike protein RBDs quickly revealed two specific residues—T403 and D501—that were never found in other Sarbecoviruses. In addition, H505 is found to be absent in all previous R.affinis infecting Sarbecoviruses.

		.0 20	30	40	50 ••	60	70	80	90	<u>100 y</u>	110	121
QHR63300,2:400-520 6ZGF_A:431-551 7CN4_A:400-520	FVITGDEVR FVITGDEVR	RQIAPGQTGKIADYN RQIAPGQTGKIADYN RDIAPGOTGKIADYN	IYKLPDDFTG	CYTANNSKHIDA		RKANLKPFER	DISTEIYQ	IGSKPCNGQ1	GLNCYYPL	YRYGFYPTDGY	'GHQPY <mark>RY\</mark>	VLSFELLNA
QJE37811.1:96-216 QRY71349.1:81-201 QNS17503.1:118-238	FVIRGDEVR FVIRGDEVR FVIRGDEVR	RQIAPGQTGKIADYN RQIAPGQTGKIADYN RQIAPGQTGKIADYN	IYKLPDDF T GI IYKLPDDF T GI IYKLPDDF T GI	CYTAHNSNNLDS	KYGGNYNYLYRLF Kyggnynylyrlf Kyggnynylyrlf	RKSNLKPFER	DISTEIYO	AGSTPCNGYE	GFNCYFPL	OSYGFOPTNGY	'G <mark>YO</mark> PY <mark>R</mark> V\	YLSFELLHA
QRV71341.1:104-224 QRV71340.1:81-201 7L7F_E:90-210	FYIRGDEVR	QIAPGQTGKIADYN QIAPGQTGKIADYN QIAPGQTGKIADYN	IYKLPDDFTG	CYTHANSNNLDS	KYGGNYNYL <mark>YR</mark> lf	RKSNLKPFER	UTZIFTAN	IGSTPUNGYE	:GFNCYFPL	USYGEUPINGY	GYUPYRYY	YLSFELLHH
QKY12177.1:404-524 QQN67583.1:410-530	FVIRGDEVR	RUTHPGUTGKIHDYN RUTHPGUTGKIADYN	IYKLPUUF IG IYKLPDDF IG	CYTHANSNNLDS	KYGGNYNYLYRLF Kyggnynylyrlf Kyggnynylyrlf	RKSNLKPFER	UTZIFTAN	IGSTPUNGYE	LIFNUTFPL	USYGEUPINGY	GYUPYRYY	YLSFELLHH
QIG55857.1:400-520 QQN67582.1:400-520 7E7B_A:400-520		XQIAPGQTGKIADYN XQIAPGQTGKIADYN XQIAPGQTGKIADYN	IYKLPDDF T GI IYKLPDDF T GI	CVIAHNSNNLDS CVIAHNSNNLDS	KYGGNYNYLYRLF Kyggnynyl yrlf	RKSNLKPFER	UISTELYQI DISTETYQI	IGSTPCNGYE	GENCYEPL	QSYGEQPTNGY OSYGEOPTNGY	GYQPYRY\ GYQPYRY\	VLSFELLHA
QJE37812.1:400-520 7E7D_R:400-520 7R4N_R:400-520	FVIRGDEVR FVIRGDEVR	QIAPGQTGKIADYN QIAPGQTGKIADYN QIAPGQTGKIADYN	IYKLPDDF T GI IYKLPDDF T GI	CVIANNSNNLDS Cviannsnnlds	KYGGNYNYL <mark>yr</mark> lf Kyggnynyl <mark>yr</mark> lf	RKSNLKPFER	DISTEIYQA DISTEIYQA	IGSTPCNGVE Igstpcngve	GFNCYFPL GFNCYFPL	.QSYGFQPTNGY .Osygfqptngy	'G <mark>YQ</mark> PY <mark>R</mark> V\ 'GYQPYRV\	VLSFELLHA VLSFELLHA
QLR06866.1:396-516 QLR06864.1:396-516	FYYRGDEVR FYYRGDEVR	RQIAPGQTGRIADYN ROIAPGOTGRIADYN	IYKLPDDF T GI IYKLPDDF T GI	CYTANNSNNLDS Cytannsnnlds	KYGGNYNYL <mark>yr</mark> lf Kyggnynyl <mark>yr</mark> lf	RKSNLKPFER RKSNLKPFER	DISTEIYQA DISTEIYQA	IGSTPCNGYE Igstpcngye	GFNCYFPL GFNCYFPL	.QSYGEHPTNGY .Osygehptngy	GYOPYRY\ GYOPYRY\	VLSFELLNA VLSFELLNA
QLR06867.1:396-516 QIG55945.1:396-516 7BBH_A:396-516	EVVRGNEVR	QIAPGQTGRIADYN QIAPGQTGRIADYN QIAPGQTGRIADYN QIAPGQTGRIADYN QIAPGQTGVIADYN	IYKI PODETGI	CVTAUNSNNI DS	KVGGNYNYI <mark>yr</mark> i f	RKSNI KPEER	NTSTETYOR	AGSTPCNGVE	GENCYEP	OSYGEHPTNGV	'G <mark>YO</mark> PY <mark>RV</mark> V	VI SEELI KA
6YH1_E:82-202 QIR48632.1:398-517 QIR48614.1:398-517	FVVKGDFVR	RATAPGOTGYTANYN	iyki podftgi	CVTAHNSVKODA	I TGGNYG <mark>yi yr</mark> i f	RKSKI KPFFR	NTSTFTYDE	AGSTPCNGQS	GI NCYYPI	FRYGEHPTTGY	NYOPERVV	VI SEELLN
7CN8_A:398-517 QIA48641.1:398-517 QIQ54048.1:400-519	FVVKGDFVR	QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	IYKI PODFTGI	CVTAUNSVKODA	i tgdnyg <mark>yi yr</mark> i f	RKSKI KPFFR	ITSTFTYDE	AGSTPCNGO	GI NCYYPI	FRYGEHPTTGV	'NYOPF <mark>R</mark> V\	VI SEELLN
QIA48623.1:398-515 AGZ48787.1:39-158		QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	YKLPDDFTG YKLPDDFTG	CVIAHNSYKODA CVLAHNTRNIDA	LTGGNYLYRLF Tqtgnynykyrsl	RKSKLKPFER	DISTEIYQ	GSTPCNGQ\ PDGKPCT-PF	GLNCYYPL	ERYGFHPTTGV NDYGFYITNGI	NYOPFRV GYOPYRV	VLSFELLN
AT098205.1:387-506 ALK02457.1:387-506 AGZ48828.1:388-507	FYYKGDDYR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFTG IYKLPDDFTG IYKLPDDFTG	CVLANNTRNIDA CVLANNTRNIDA CVLANNTRNIDA	TQTGNYNYKYRSL TQTGNYNYKYRSL TQTGNYNYKYRSL	RHGKLRPFER RHGKLRPFER	DISNYPFS	DGKPCT-PF DGKPCT-PF DGKPCT-PF	PAFNCYHPL PAFNCYHPL PAFNCYHPL	NDTGFTIINGI NDTGFTIINGI NDTGFTIINGI	GYQPYRVV GYQPYRVV	VLSFELLNA VLSFELLNA VLSFELLNA
AGZ48818.1:388-507 AHX37558.1:391-510 AHX37569.1:391-510		QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDF T GI IYKLPDDFHGI IYKLPDDFHGI	CYLAHNTRNIDA CYLAHNTRNIDA CYLAHNTRNIDA	TQTGNYNYKYRSL TSSGNFNYKYRSL TSSGNFHYKYRSL	.RHGKLRPFER .RHGKLRPFER .RHGKLRPFER	DISNYPFS DISNYPFS DISNYPFS	PDGKPCT-PF PDGKPCT-PF PDGKPCT-PF	Pafncywpl Pafncywpl Pafncywpl	NDYGFYITNGI Ndygfyttngi Ndygfyttngi	GYQPYRV\ GYQPYRV\ GYQPYRV\	VLSFELLNA VLSFELLNA VLSFELLNA
QDF43825.1:388-507 AT098218.1:388-507 AT098231.1:388-507	FVVKGDDVR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	iyki podfhgi	CVI AHNTRNTDA	TSTGNYNYKYRSI	RHGKI RPFFR	NTSNVPFS	PNGKPCT-PF	PAFNCYHPI	NDYGEFTTNGT	G <mark>YO</mark> PY <mark>R</mark> VV	VI SEELLNA
3BGF_A:70-189 2GHV_C:74-193	FVVKGNNVP	QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	IYKI PNNEHGI	CVI AUNTRNTNA	TSTGNYNYKYPYI	PHGKI PPEEP	TCNVPESI	POGKPCT-PF	PALNCYUP	NOYGEYTTTGT	FYNPYDUU	VISEFIINA
2DD8_S:71-190 QKY12178.1:404-523 ABE77216.1:387-506	FYYKGDDYR	QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	IYKLPDDFHG	CYLAANTRNIDA	TSTGNYNYKYRYL	.RHGKLRPFER	DISNYPFS	PDGKPCT-PF	PALNCYHPL	NDYGFYTTTGI	G <mark>YO</mark> PY <mark>R</mark> YV	YLSFELLNA
5X58_A:374-493 5HRG_A:387-506 6NB6_A:406-525	FYYKGDDYR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKL PODFHGI	CYLAHNTRNTDA	TSTGNYNYKYRYL	RHGKI RPFFR	DTSNYPFS	POGKPCT-PF	PALNCYHPL	NDYGFYTTTGT	GYOPYRYY	YLSEELLNA
6CRH_A:374-493 6ACC_A:387-506 ACJ60703.1:391-510	FVVKGUUVR	QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	YKI PHILENG	Y HHN RNI H	ISTGNYNYKYRYL	RHIKI RPFFR	SNYPESI	416KPI:1-PF	'HI NICYHPI	NUTHET	HYUPYRV	VESEELENH
ABD72985.1:387-506 ACZ72093.1:387-506	FVVKGDDVR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	YKI PROFING	CVI ANNTRNTAA	TSTGNYNYKYRYL	RHGKI RPFFR	ISNVPES	PIIGKPET-PE	PALNCYHP	NOYGEVEETGE	FYOPYRV	VI SEELI NA
ABD72982.1:387-506 AAR07630.1:387-506 AAR91586.1:387-506	FVVKGDDVR	RQIAPGQTGYIADYN RQIAPGQTGYIADYN RQIAPGQTGYIADYN	IYKLPDDFHG IYKLPDDFHG IYKLPDDFHG	CYLANNTRNIDA Cylanntrnida Cylanntrnida	TSTGNYNYKYRYL TSTGNYNYKYRYL	RHGKLRPFER RHGKLRPFER	DISNYPFS	PDGKPCT-PF PDGKPCT-PF PDGKPCT-PF	PALNCYHPL PALNCYHPL PALNCYHPL	NDTGFYTTTGI NDTGFYTTTGI NDTGFYTTTGI	GYQPYRV\ GYQPYRV\ GYQPYRV\	VLSFELLNA VLSFELLNA VLSFELLNA
ACB69883.1:387-506 AAX16192.1:387-506 ACZ72254.1:387-506	FYYKGDDYR FYYKGDDYR FYYKGDDYR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida Cylahntrnida	TSTGNYNYKYRYL Tstgnynykyryl Tstgnynykyryl	.RHGKLRPFER .RHGKLRPFER .RHGKLRPFER	DISNYPFS DISNYPFS DISNYPFS	PDGKPCT-PF PDGKPCT-PF PDGKPCT-PF	PALNCYMPL PALNCYMPL PALNCYMPL	NDYGFYTTTGI Ndygfytttgi Ndygfytttgi	GYQPYRV\ GYQPYRV\ GYQPYRV\	VLSFELLNA VLSFELLNA VLSFELLNA
AAU81608.1:387-506 6CRY_A:374-493 AAT76147.1:387-506		QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	iyklpddfhgi Iyklpddfhgi Iyklpddfhgi	CYLAHNTRNIDA CYLAHNTRNIDA CYLAHNTRNIDA	TSTGNYNYKYRYL Tstgnynykyryl Tstgnynykyryl	RHGKLRPFER	DISNVPFS DISNVPFS	PDGKPCT-PF PDGKPCT-PF PDGKPCT-PF	Palncyhpl Palncyhpl Palncyhpi	NDYGFYTTTGI NDYGFYTTTGI NDYGFYTTTGI		VLSFELLNA VLSFELLNA VLSFELLNA
ACQ82725.1:387-506 ACZ72020.1:387-506		QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	YKLPDDFHG YKLPDDFHG	CYLAHNTRNIDA CYLAHNTRNIDA	TSTGNYNYKYRYL Tstgnynykyryl		DISNYPFS	PDGKPCT-PF PDGKPCT-PF	PALNCYHPL	NDYGFYTTTGI NDYGFYTTTGI	GYQPYRV\ GYQPYRV\	VLSFELLNA
AAR07624.1:387-506 AAP13567.1:387-506 AAS75868.1:387-506	FVVKGDDVR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKI PROFING	CVI ANNTRNTAA	TSTGNYNYKYRYI	RHGKI RPFFR	TSNVPES	206KPCT-PF	PALNCYHP	NOYGEYTTTGT	GYOPYRV	VI SEELI NA
AFN43867.1:387-506 ABD73001.1:387-506 AAP33697.1:387-506	FYYKGDDYR	RQIAPGQTGYIADYN RQIAPGQTGYIADYN RQIAPGQTGYIADYN	IYKLPDDFHGI	CYLAHNTRNIDA	TSTGNYNYKYRYL	.RHGKLRPFER	DISNYPFS	PDGKPCT-PF	PALNCYHPL	NDYGFYTTTGI	G <mark>YO</mark> PY <mark>R</mark> YY	YLSFELLNA
ACB69860.1:387-506 ABD72984.1:387-506 ABD72977.1:387-506	FYYKGDDYR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHGI	CYLAHNTRNTDA	TSTGNYNYKYRYL	RHGKI RPFFR	DISNYPESI	PDGKPCT-PF	PALNCYHPL	NDYGFYTTTGI	GYOPYRVV	YLSEELLNA
YP_009825051.1:387-5 AFR58672.1:387-506 ADC35483.1:387-506	FYYKIIIIYK	QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	YKI PHILENG	Y HHN RNI H	ISTENYNYKYRYL	RHIKI RPFFR	SNYPESI	416KPI:1-PF	'HI NICYHPI	NUTHET	HYUPYRY	VESEELENH
ACZ71976.1:387-506 ABD72979.1:387-506	FVVKGNNVR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	YKI PNNENG	VIAUNTRNTNA	TSTGNYNYKYPYI	PHGKI PPEER	TSNVPES	PIIGKPET-PE	PALNCYUP	NOYGEYTTIGT	EVEPYDU	VISEELINA
ACZ72195.1:387-506 ABD72970.1:387-506 ACZ71826.1:387-506	FVVKGDDVR	QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	iyki ponfiigi	CVI AUNTRNTNA	TSTGNYNYKYRYI	RHGKI RPFFR	NTSNVPESI	РОСКРСТ-РЕ	PALNCYUP	NDYGEYTTTGT	G <mark>YO</mark> PY <mark>R</mark> VV	VI SEELLNA
ACB69894.1:387-506 ABD72969.1:387-506 ACZ71797.1:387-506	FYYKGDDYR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFMGI	CYLAHNTRNIDA	TSTGNYNYKYRYL	.RHGKLRPFER	DISNYPFS	PDGKPCT-PF	PALNCYHPL	NDYGFYTTTGI	G <mark>YO</mark> PY <mark>R</mark> YY	YLSFELLNA
BAF42873.1:387-506 ACZ72108.1:387-506 ABD72988.1:387-506	FYYKGDDYR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	YKLPDDFHG	CYLANNTRNIDA	TSTGNYNYKYRYL	RHGKLRPFER	DISNYPFS	PDGKPCT-PF	PALNCYMPL	NDYGFYTTTGI	GYQPYRY\	YLSFELLNA
ABD72995.1:387-506 ACZ71961.1:387-506 ACB69905.1:387-506		QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYK <mark>yr</mark> yl Tstgnynykyryl	.RHGKLRPFER	DISNYPFSI DISNYPFSI	PDGKPCT-PF PDGKPCT-PF	PAL <mark>NCYMPL</mark> PALNCYMPL	NDYGFYTTTGI NDYGFYTTTGI	GYOPYRYN GYOPYRYN	VLSFELLNA VLSFELLNA
AAS00003.1:387-506 AAR86775.1:387-506		RQIAPGQTGVIADYN RQIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYKYRYL Tstgnynykyryl	.RHGKLRPFER .RHGKLRPFER	DISNYPFS	PDGKPCT-PF PDGKPCT-PF	PAL <mark>NCYHPL</mark> PAL <mark>NCYHPL</mark>	NDYGFYTTTGI NDYGFYTTTGI	GYOPYRY\ GYOPYRY\	VLSFELLNA VLSFELLNA
AAT74874.1:387-506 AAP51227.1:387-506 7JN5_F:82-201		QIAPGQTGYIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYK <mark>yr</mark> yl Tstgnhnykyryl	.RHGKLRPFER	DISNYPFS	PDGKPCT-PF PDGKPCT-PF	PAL <mark>NCYMPL</mark> PAL <mark>NCYMPL</mark>	NDYGFYTTTGI NDYGFYTTTGI	GYOPYRYN GYOPYRYN	YLSYELLNA YLSFELLNA
QND76034.1:391-510 ACZ72122.1:387-506 AEA10473.1:387-506	FYYKGDDYR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG	CYLANNTRNIDA	TSTGNHNYK <mark>yr</mark> yl	RHGKLRPFER	DISNYPFS	PDGKPCT-PF	PALNCYMPL	NDYGFYTTTGI	G <mark>YQ</mark> PY <mark>R</mark> Y\	YLSFELLNA
ABF65836.1:387-506 ACZ71991.1:387-506 AAR07627.1:387-506	FYYKGDDYR	QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG	CYLANNTRNIDA	TSTGNHNYK <mark>yr</mark> yl	.RHGKLRPFER	DISNYPFS	PDGKPCT-PF	PALNCYHPL	NDYGFYTTTGI	G <mark>YQ</mark> PY <mark>R</mark> Y\	YLSFELLNA
AAR07626.1:387-506 AAR07628.1:387-506 AAR07625.1:387-506		RQIAPGQTGYIADYN RQIAPGQTGYIADYN RQIAPGQTGYIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYK <mark>yr</mark> yl Tstgnynykyryl	.RHGKLRPFER	DISNYPFSI DISNYPFSI	PDGKPCT-PF PDGKPCT-PF	PALNCYMP <mark>L</mark> PALNCYMPL	SDYGFYTTTGI SDYGFYTTTGI	GYOPYRYN GYOPYRYN	YLSFELLNA YLSFELLNA
AAR07631.1:387-506 AAR07629.1:387-506		QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA CYLAHNTRNIDA	TSTGNYNYKYRYL Tstgnynykyryl	RHGKLRPFER	DISNVPFS	PDGKPCT-PF PDGKPCT-PF	PAL <mark>NCYHPL</mark> PALNCYHPL	SDYGFYTTTGI SDYGFYTTTGI		VLSFELLNA
ABD72972.1:387-506 AFR58714.1:387-506 AAR23250.1:387-506		QIAPGQTGVIADYN QIAPGQTGVIADYN	iyklpddfhgi Iykl pddfhgi	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYK <mark>yr</mark> yl Tstgnynykyrsi	XHGKLRPFER		PDGKPCT-PF	PALNCYHPL PALNCYHPI	NDYGFYTTTGI NDYGFYTTTGI	GYOPYRYN GYOPYRYN	VLSFELLNA
AGT21078.1:387-506 3SCI_E:82-201 BAE93401.1:387-506		QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYKYRFL Tstgnynykyrcl	RHGKLRPFER	DISNYPFSI DISNYPFSI	PDGKPCT-PF PDGKPCT-PF	Pafncywpl Pafncywpl	NDYGFYTTTGI NDYGFYTTTGI	GYOPYRYN GYOPYRYN	VLSFELLNA VLSFELLNA
AEA10443.1:387-506 ABF68959.1:387-506 AAP82968.1:387-506		QIAPGQTGYIADYN QIAPGQTGYIADYN QIAPGQTGYIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNHNYK <mark>yr</mark> yl Tstgnynykyryl	XHGKLRPFER	DISNYPFSI DISNYPFSI	PDGKPCT-PF PDGKPCT-PF	PALNCYMPL PAPNCYMPL	NDYGFYTTTGI NDYGFYTTSGI	GYOPYRYN GYOPYRYN	YLSFELLNA YLSFELLNA
AAU04646.1:387-506 AAV91631.1:387-506		RQIAPGQTGVIADYN RQIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYKYRYL Tstgnynykyryl	.RHGKLRPFER .RHGKLRPFER	DISNVPFS	SDGKPCT-PF SDGKPCT-PF	Pap <mark>n</mark> cympl Papncympl	RGYGFYTTSGI RGYGFYTTSGI	GYOPYRY\ GYOPYRY\	VLSFELLNA VLSFELLNA
AAU04662.1:387-506 AAU04649.1:387-506 AAV49722.1:387-506		QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA CYLAHNTRNIDA	TSTGNYNYKYRYL Tstgnynykyryl	RHGKLRPFER	DISNVPFS	DGKPCT-PF	PAP <mark>NCYHPL</mark> PAPNCYHPL	RGYGFYTTSGI RGYGFYTTSGI		VLSFELLNA
AAU04664.1:387-506 AAU93319.1:387-506 AAV97985.1:387-506	FYYKGDDYR FYYKGDDYR FYYKGDDYR	XQIAPGQTGYIADYN XQIAPGQTGYIADYN XQIAPGQTGYIADYN	IYKLPDDFHG IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida Cylahntrnida	TSTGNYNYKYRYL Tstgnynykyryl Tstgnynykyryl	.RHGKLRPFER .RHGKLRPFER .RHGKLRPFER	DISNYPFS DISNYPFS DISNYPFS	PDGKPCT-PF PDGKPCT-PF PDGKPCT-PF	Papncyupl Papncyupl Papncyupl	RGYGFYTTSGI RGYGFYTTSGI RGYGFYTTSGI	GYQPYRV\ GYQPYRV\ GYQPYRV\	VLSFELLNA VLSFELLNA VLSFELLNA
AAV98000.1:387-506 AAV97998.1:387-506 AAV98002.1:387-506		QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYK <mark>yr</mark> yl Tstgnynykyryl	.RHGKLRPFER .RHGKLRPFER	DISNVPFS	PDGKPCT-PF PDGKPCT-PF	Pap <mark>ncympl</mark> Pap <mark>ncympl</mark>	RGYGFYTTSGI RGYGFYTTSGI	GYOPYRYN GYOPYRYN	VLSFELLNA VLSFELLNA
ABF68955.1:387-506 AAV98001.1:387-506		QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYK <mark>yr</mark> yl Tstgnynykyryl	.RHGKLRPFER .RHGKLRPFER	DISNYPFSI DISNYPFSI	PDGKPCT-PF PDGKPCT-PF	Pap <mark>n</mark> cymp <mark>l</mark> Papncympl	RGYGFYTTSGI RGYGFYTTSGI	GYOPYRYN GYOPYRYN	YLSFELLNA YLSFELLNA
ABF68956.1:387-506 AAV49720.1:387-506 AAU04661.1:387-506		QIAPGQTGVIADYN QIAPGQTGVIADYN QIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYKYRYL Tstgnynykyryl	.RHGKLRPFER .RHGKLRPFER	DISNYPFS	PDGKPCT-PF PDGKPCT-PF	Pap <mark>n</mark> cympl Papncympl	KGYGFYTTSGI KGYGFYTTSGI	GYOPYRY\ GYOPYRY\	VLSFELLNA VLSFELLNA
AAV97995.1:387-506 AAV49723.1:387-506	CHURCODIN											
ARV97989.1:387-506 ARV97989.1:387-506 ARV49719.1:387-506		RQIAPGQTGVIADYN RQIAPGQTGVIADYN RQIAPGQTGVIADYN RQIAPGQTGVIADYN	IYKLPDDFHG IYKLPDDFHG	CYLAHNTRNIDA Cylahntrnida	TSTGNYNYK <mark>yr</mark> yl Tstgnynykyryl	.RHGKLRPFER	DISNYPFSI DISNYPFSI	PDGKPCT-PF PDGKPCT-PF	Pap <mark>n</mark> cympl Papncympl	NGYGFYTTSGI	GYOPYRYN GYOPYRYN	YLSFELLNA YLSFELLNA

		20	30	40	50	6 0	70	80	90	100	110	121
AAS10463,1:387-506	FYYKGDDYRQIAPG		DDFHGCYL	RUNTRNIDATS	TGNYNYK	YRYLRHGKLR		IVPFSPDGKPCT	-PPAPNCYMPL	NGYGFYTTS	GIGYQPYRYYYLS	FELLNA
AAV97990.1:387-506	FYYKGDDYRQIAPG	TGYIADYNYKLF	PDDFHGCYL	ANNTRNIDATS	TGNYNYK	<mark>Yrylrhgkl</mark> r	PFERDIS	IVPFSPDGKPCT	-PPAPNCYHPL	NGYGFYTTS	GIGYQPYRYYYLS	SFELLNA
ABF68958,1:387-506	FYYKGDDYRQIAPG	TGYIADYNYKLF	PDDFHGCYLI	ANNTRNIDATS	TGNYNYK		PFERDIS	IVPFSPDGKPCT	-PPAPNCYUPL	NGYGFYTTS	GIGYQPYRYYYLS	SFELLNA
ABF68957.1:387-506 AAV97992.1:387-506	FYYKGDDYKQIHPG	ALGATHOTNIKU JIGATHOINNKU	YDDF NGC YLI	AMNIKNIUHIS	TGNYNYK	TRTLRHUKLK	PFERUISE	IYPESPOGKPU I	-PPHPNCTHPL	PGYGE YTTS	GTGYOPYPVVVI	SEELLNH
AAY97984,1:387-506	FYYKGDDYROIAPG	TGVIADYNYKLF	PDDFHGCYL	AUNTRNIDATS	TGNYNYK	XRYLRHGKLR	PFERDIS	IVPFSPXGKPCT	-PPAPNCYHPI		GIGYOPYRYYYL	SFELLNA
3D0G_E:64-179	FYYKGDDYRQIAPG	TGVIADYNYKLF	DDFHGCYL	ANNTRNIDATS	TGNYNYK	YRYLRHG <mark>kl</mark> r	PFERDIS	IVPFSPDGKPCT	-PPALNCYMPL	NDYGFYTTT	GIGYQPYRYYYLS	SFE
28JF_E:65-180	FVVKGDDVRQIAPG	TGVIADYNYKLF	PDDFHGCYL	ANNTRNIDATS	TGNYNYK		PFERDIS	IVPFSPDGKPCT	-PPALNCYHPL	NDYGFYTTT	GIGYQPYRYYYLS	SFE
3SCL_E:64-179 6HAQ_B:68-183	FVVKGDDVRQ1HPG	ALEATED ANALY	YUDE NGCYLI	HAN I KNIUH I S	TCNYNYK		PFERUIS	IVPESPICKPUT		NUTGETTI	GIGTOPYRYYLS	5FE 200
3D0H_E:64-179	FYYKGDDYRQIAPG	TGVIADYNYKLF	DDFHGCYL	ANTRNIDATS	TGNYNYK	YRYLRHGKLR	PFERDIS	IVPFSPDGKPCT	-PPALNCYHPL	KDYGFYTTS	GIGYQPYRYYYLS	FE
3D0H_E:64-179 3D0I_E:64-179	FYYKGDDYRQIAPG	QTGVIADYNYKLF	DDFHGCYL	ANNTRNIDATS	TGNYNYK	YRYLRHGKLR	PFERDIS	IVPFSPDGKPCT	-PPALNCYHPL	RGYGFYTTS	GIG <mark>YQ</mark> PY <mark>RYYYL</mark> S	SFE
3SCK_E:64-179 3SCJ_E:65-180	FYYKGDDYRQIAPG	UTGVIADYNYKLF	PODENGCYL	ANNTRNIDATS			PFERDIS	IVPESPDGKPCT	-PPAPNCYMPL	RGYG YTTT	GIGYQPYRYYYLS	SFE
AGZ48795,1:39-158	FYYKGDDYRQIAPG	JTGVIADYNYKLF	PDDFLGCYL	ANTNSKOSST	SGNYNYL	YRWYRRSKLN	PYERDLS	IDIYSPGGOSCS	-AYGPNCYNPL		GYGHOPYRYYYLS	SFELLNA
AGZ48785.1:39-158	FYYKGDDYRQIAPG	TGYIADYNYKLF	PDDFLGCYL	R <mark>hntn</mark> sk <mark>d</mark> sst	SGNYNYL	YRHYRRS <mark>kl</mark> n	PYERDLS	IDIYSPGGQSCS	-AYGPNCYNPL	RPYGFFTTA	IGYGHQPY <mark>RYYYL</mark> S	SFELLNA
AT098132,1:388-507	FYYKGDDYRQIAPG	TGVIADYNYKLF	PDDFLGCYL	ANNTNSKOSST	SGNYNYL	YRHYRRSKLN	PYERDLS	IDIYSPGGQSCS	-AYGPNCYNPL	RPYGFFTTA	GYGHQPYRYYYLS	SFELLNA
AGZ48806.1:388-507 AT098157.1:387-506	FVVKGDDVRQIAPG	ATGVTADYNYKI F	PODFLOCYLI	RUNTNSKOSST	SGNYNYI '	YRHVRRSKI N	PYERDI SI	INTYSPGGOSCS	-ATGPNCYNPI	RPYGEFTTA	GVGHOPYRVVVI 9	SEFLENA
QTJ30153,1:391-509	FYYKGDDYRQIAPA	TGVIADYNYKLF	DDFTGCYL	NUTINSYDSKO	GNNFY	YRLFRHGKIK	PYERDIS	IVLYNSAGGTCS	STSQLGCYEPL	KSYGFTPTY	GVGYQPYRVVVL	SFELLNA
AP040579.1:391-509	FYYKGDDYRQIAPA	QTGVIADYNYKLF	PDDFTGCYL	AMNTNSYDSKS	GNNFY	YRLFRHGKIK	PYERDIS	IVLYNSAGGTCS	SISQLGCYEPL	KSYGFTPTY	GYGYQPYRYYYLS	FELLNA
QTJ30135.1:391-508 YP_003858584.1:392-5	FYYKGDUYKUIHPH	ALEATED ANALY	YUDE TOCYLI	HAN I NSYUSKU BUNTNCI DCC.	UGN-FY -NEEE	TRLFRHGKIK	PTERUISM PTERUISM	IVLYNSHGGTCS		ACTOR TOCC	GYGTUPTRYYYLS	SFELLNH
ALJ94036.1:395-510	FIYKGDDYRQIAPS	TGVIADYNYKLF	PDDFTGCYI	ANTNALDSN-		YRLFRHGKIK	PYGRDLS	IPYSPSGTCST	-INNLNCFAPL	KSYGFTQSS	GISFOPYRVVVLS	SFELLNA
ALJ94036.1:395-510 BCG66627.1:374-485	FVVRGDEVRQIAPG	TGVIADYNYKLF	DDFTGCYL	Hunsrnodas t	SGNENYY	YRIHRSEKLR	PFERDIA	IYDYQYGTQFKS	Sl	KNYGFYSSA	GDSHQPYRVVVLS	SFELLNA
BBJ35999.1:82-186	FYYRGDEYRQIAPG	TGVIADYNYKLF	PODETECYL	HUNSRNODAST	SGNFNYY		PFERDIA	IYDYQYGTQFKS	SL	KNYGFYSSA		CELL NO
AGZ48800.1:39-140 AT098193.1:391-492	FLIRSSEVROVAPG	ETGYIADYNYKI F	PDFTGCVT	AMNTAKODOGO	T	RSSRKTKI K	PFERNI S	DE	NGVRTI	STYDE YPTV	PIEYOATRVVVI	SFELLNA
AT098169,1:391-492	FLIRSSEVRQVAPG	TGVIADYNYKLF	DDFTGCVI	ANNTAKQDQGC	і <mark>ү</mark> ү	YRSSRKTKLK	PFERDLSS	DE	NGYRTL	STYDFYPTY	PIEYQATRVVVLS	FELLNA
ANA96090.1:391-492	FLIRSSEVRQVAPG	TGVIADYNYKLF	DEFTGCVI	ANNTAKQDQGC	<mark>Y</mark> Y	TRSSRKTKLK	PFERDLSS	DE	NGVRTL	STYDFYPTV	PIEYQATRVVVLS	FELLNA
QDF43830.1:391-492 AT098108.1:391-492	FLIRSSEVRUVHPU FLIRSSEVRUVAPG	TGVIHUTNTKLF	PODETGCVT	HAN I HUUUKGU AUNTAOODKGO	 	TRSSRK I KLK Yrssrk tri k	PFERULSS	DE		STYNEYPTV	PIETUHIRVVVLS	SEFLENA
AGZ48794,1:39-140	FLIRSSEVROVAPG	TGVIADYNYKLF	DDFTGCVI	RHNTAKQDQGC	<u>י</u> יי	YRSSRKTKLK	PFERDLTS	DE	NGVRTL	STYDFYPNY	PIEYQATRVVVLS	FELLNA
ACU31032.1:391-492	FLIRSSEVRQVAPG	TGVIADYNYKLF	PDDFTGCVI	A <mark>hntakq</mark> dqgc	<mark>Y</mark> Y	YRSSRKTKLK	PFERDLTS	DE	NGVRTL	STYDFYPNY	PIEYQATRVVVLS	SFELLNA
QDF43835.1:391-492 AT098120.1:391-492	FLIRSSEVRUVHPG	TGVIHUYNYKLE	YUDE IGCVI	HANTAKUDUGU DUNTAKODOGO	T T		PEERUL IS	DE	NGVRIL	STYDEYPNY CTYNEYPNU		SFELLNH
AGC74176,1:391-492	FLIRFSEYRQIAPG	TGVIADYNYKLF	DEFTGCYL	ANNTANODRGO	¥Ÿ	YRSSRKTKLK	PFERDLSS	DE	NGYRTL	STYDFYPSY	PLEYQATRVVVLS	SFELLNA
AGC74176.1:391-492 ABD75332.1:391-492	FLIRSSEVRQVAPG	TGVIADYNYKLF	PDDFTGCVI	AMNTAQQDQGC	<mark>Y</mark> Y	YRSYRKEKLK	PFERDLSS	DE	NGYYTL	STYDFYPSI	PVEYQATRVVVL	FELLNA
AGC74165.1:390-491 AGZ48791.1:39-141	FLIRSSEVRUVHPG	TGVIHUYNYKLE	YUDE IGCYL	HAN I HNUUUGU DUNTOVODTCI	<u>Y</u> Y		PEERULSS	DE	NGYYTL	STYDENDAU	PLUYUHIRYYYLS	SFELLNH
AT098181,1:391-493	FLIRSSEVRQVAPG	TGVIADYNYKLF	DOFTGCYL	RUNTAKODTGI	YY	YRSHRKTKLK	PFERDLSS	DDG	NGYYTL	STYDENPNY	PVAYQATRVVVLS	SFELLNA
QDF43820.1:391-493	FLIRSSEVROVAPG	ETGVIADYNYKLF	PDDFTGCYI	R <mark>hntakq</mark> dtgh	<mark>Y</mark> Y	YRSHRKTKLK	PFERDLSS	DDG	NGVYTL	STYDENPNY.	PVAYQATRVVVLS	SFELLNA
AGZ48783.1:39-141 AIA62310.1:390-492	FLIRSSEVRQVAPG	TGVIADYNYKLF	PODETCCVI	RUNTAKODTG	Y Y		PFERDLSS	DDG	NGVYTL	STYDENPNY	PVAYQATRVVVLS	SFELLNA
AIA62320,1:391-493	FLIRSSEVROVAPG	TGVIADYNYKLE	PODETGCVI	RUNTAKODTGN	YY	YRSHRKTKLK	PFERDLSS	DDG	NGVYTL	STYDENPNY	PVAYOATRVVVL	SFELLNA
ADE34766.1:391-493	FLIRSSEVROVAPG	TGVIADYNYKLF	PDDFTGCVI	A <mark>hntakq</mark> dtgn	¥Ÿ	YRSHRKTKLK	PFERDLSS	DDG	NGYYTL	STYDENPNY	PVAYQATRVVVLS	SFELLNA
ADE34755.1:391-493 QDF43815.1:390-492	FLIRSSEVRQVAPG	TGVIADYNYKLF	PODETGCVI	ANNTAKODTGN	IYY	YRSHRKTKLK	PFERDLSS	DDG	NGVYTL	STYDENPNY	PVAYQATRVVVLS	SFELLNA
AAZ41329,1:391-493	FL TRSSEVROVAPG	TGYTADYNYKLF	PONETACYT	AHNTAKHDTGN	YY	YRSHRKTKI K	PFFRDL SS	006	NGVYTI	STYDENENY		SEFLENA
ADE34812,1:391-493	FLIRSSEVRQVAPG	ETGYIADYNYKLF	PDDFTGCVI	<mark>Runtakhdtg</mark> n	YY	YRSHRKTKLK	PFERDLSS	DDG	NGYYTL	STYDENPNY	PVAYQATRVVVLS	SFELLNA
ADE34722.1:391-493 ADE34823.1:391-493	FLIRSSEVRQVAPG	TGVIADYNYKLF	PODETCOVI	RUNTAKHDTGN	Y Y		PFERDLSS	DDG	NGVYTL	STYDENPNY CTYDENPNY	PVAYQATRVVVLS	SFELLNA
AID16716,1:391-493	FL TRSSEVROVAPG	TGVTADYNYKI F	PONETACYT	AHNTAKODTGN	YY	YRSHRKTKI K	PEERDLSS	DDG	NGVYTI	STYDENENY	PVAYOATRVVVI 9	SEFLENA
AGZ48788.1:39-141	FLIRSSEVRQVAPG	TGVIADYNYKLF	DDFTGCVI	AHNTAKQDTG1	¥Ÿ	YRSHRKTKLK	PFGRDLSS	DDG	NGVYTL	STYDENPNY	PVAYQATRVVVLS	FELLNA
ACU31051,1:391-492	FLIRSSEVRQVAPG	TGVIADYNYKLF	PDDFTGCVI	Runtakodtgi	<mark>Y</mark> Y	YRSHRKTKLK	PFERDLSS	DD	DGVYTL	STYDENPNY	PVAYQATRVVVLS	SFELLNA
AGZ48784.1:39-140 AR076382.1:386-487	FLIRFSEVRUVHPG	TGVINDINIKLI TGVINDINIKLI	PODETGCVI	ANNTAKODYGS	YF	TRSHRSSKLK	PFERDLSS	DE		STYDENPNY STYDENPNV	PLUTUHTRYYYL:	SEFLUNA
AT098145,1:384-485	FLIRFSEVRQVAPG	TGVIADYNYKLF	PDDFTGCVI	RHNTAKODVGS	ŸF'	YRSHRSSKLK	PFERDLSS	DE	NGVRTL	STYDENPNY	PLDYQATRVVVLS	FELLNA
QQM18864.1:379-480	FLIRFSEYRQIAPG	TGVIADYNYKLF	PDDFTGCYL	AMNTAKQDIGS	YF	YRSHRAYKLK	PFERDLSS	DE	NGYRTL	STYDENPNY	PLDYQATRVVVLS	FELLNA
AVP78031.1:396-497 QSQ01650.1:396-497	FLIRFSEVRUVHPG	ALEATED ANALY	YUDE IGCYL	HANTAKUUYGN	YF		PEERULSS	DE	NGVRIL	STYDENPNY CTYDENPNU		SFELLNH
AVP78042.1:395-496	FLIRFSEYROYAPG	TGYIADYNYKLF	PODETGCYI	ANNTAKODTG	YF	YRSHRSTKLK	PFERDLSS	DE	NGYRTL	STYDENPNY	PLEYOATRYYYLS	SFELLNA
AIA62330.1:383-484	FLIRFSEVRQVAPG	TGVIADYNYRLF	PDDFTGCYI	<mark>Rhntanq</mark> dy <mark>g</mark> s	YF	YRSHRSTKLK	PFERDLSS	DE	NGVRTL	STYDENPNY	PLDYQATRYYYLS	SFELLNA
ARI44799.1:383-484 AKZ19087.1:391-492	FLIRFSEVRQVAPG	TGVIADYNYRLF	PODETCCVI	AMNTANQDYG9	YF		PFERDLSS	DE	NGYRTL	STYDENPYY	PLOYQATRVVVLS	SFELLNA
AKZ19087.1:391-492 AKZ19076.1:391-492	FLIRFSEVROVAPG	ATGATHDINIKER Jigathdiniker	PODETGCVI	AANTAKTUYG: AANTAKTUYG:	YF	TRSHRSSKLK YRSHRSSKI K	PFERULS:	FF	NGARTI	STYDENONY	PLETUHTRYYYLS	SEFLENA
AIA62340,1:391-492	FLIRFSEVRQVAPG	TGVIADYNYKLF	DDFTGCVI	A <mark>hntakq</mark> dygs	¥F	YRSHRSSKLK	PFERDLSS	EE	NGVRTL	STYDFNQNY	PLEYQATRVVVLS	FELLNA
ABD75323.1:391-492	FLIRFSEVRQVAPG	QTGVIADYNYKLF	PDDFTGCVI	A <mark>hntakq</mark> dygs	YF	YRSHRSSKLK	PFERDLSS	EE	NGYRTL	STYDENONY	PLEYQATRVVVLS	FELLNA
AIA62290.1:391-492 AIA62300.1:389-490	FLIRESEVROVAPG	ALGATHOANAKI E	PODETGCVI	HAN I HKUUYGS AUNTAKONYGS	YF	TRSHRSSKLK VRSHRSSKL K	PFERULSS	EE		STYDENOTY STYDENOYV	PLETUHIRVVVLS	SEFLLNH
ARI44809,1:391-492	FLIRFSEVRQVAPG	TGVIADYNYKLF	PODFTGCVI	ANNTAKQDYGS	¥F'	YRSHRSSKLK	PFERDLSS	EE	NGVRTL	STYDENOYY	PLEYQATRVVVLS	SFELLNA
AIA62339.1:391-492	FLIRFSEVRQVAPG	TGVIADYNYKLF	PDDFTGCVI	A <mark>mntakq</mark> dygs	YF	YRSHRSSKLK	PFERDLSS	EE	NGVRTL	STYDENQYY	PLEYQATRVVVLS	FELLNA
AIA62277.1:386-487 ANA96027.1:386-487	FLIKESEVRQVAPG	ALGATHDANAKFL JICATHDANAKFL	UDFIGCVI DDFICCVI	HMNIHKQDYG9 AUNTAKODVCC	YF	TKSHKSSKLK	PEEPDLee	FF	NGYLTL	STYDENONY	PLETUHTRYYYLS	SEFLING
ASO66810,1:386-487	FLIRESEVROVAPG	OTGVIADYNYKI F	PDDFIGCVI	ANNTAKODVAS	YF	YRSHRSSKI K	PFERDL SS	EE	NGYI TI	STYDENONY	PLEYOATRYVVI	SFELLNA
ANH10613.1:386-487	FLIRFSEVRQVAPG	TGVIADYNYKLF	DDFIGCVI	A <mark>mntakq</mark> dygs	ŸF'	YRSHRSSKLK	PFERDLSS	EE	NGYLTL	STYDFNONY	PLEYQATRVVVL	SFELLNA
ABG47060,1:391-492	FLIRFSEYRQYAPG	TGVIADYNYKLF	PDDFTGCVI	A <mark>untako</mark> dygs	<u>YF</u>	YRSHRSSKLK	PFERDLSS	YE	ENGRTL	STYDFNQNY	PLEYQATRVVVLS	SFELLNA
QQM18908.1:379-478 AAS44787.1:387-444	LTKL2FAKATHLPP	TEATHOLNAKTE	PODENGCVI	ANNTRNTNATO	TGNYNYK	IAAAAAAAAAA YRYI R		~~~	XXXXXX	AATUENPNY	FLUTUHIKYYYLS	DECLL
AAS44779,1:386-443	FVVKGDDVRQIAPG FVVKGDDVRQIAPG FVVKGDDVRQIAPG	TGVIADYNYKLF	DDFHGCYL	ANTRNIDATS	TGNYNYK	YRYLR						
AAS44781.1:386-443	FVVKGDDVRQIAPG	TGVIADYNYKLF	DDFHGCYL	AMNTRNIDATS	TGNYNYK	YRYLR						
AAS44761.1:387-444 AAS01062.1:346-381	FYVKGDDVRQIAPG FYVKGDDVRQIAPG	ALOATHOTNAKEL JIOATHOINAKEL	PDDFMGCVL	HAN I KNTOH I 2 UMM I KNTOH I 2		TRTLK						
QQM18886.1:390-420	PG	TGYIADYNYKLF	PDDFTGCYL	ANNTAKOD								
Consensus	F1!r.s#YRQ!APG	#TGYIADYNYKLF	PDDFLGCVi	AWNtakqD.g.	••••• <mark>9</mark> •'	Yrs.rklk	pferdlss	••••••	ngv.tl	.stydf .p .v	pyqatrvvvls	sfellna

Fig.30a: multiple sequence alignment of all current known Sarbecovirus RBM sequences on NCBI. Orange arrows indicating critical residues for infection using ACE2[12][13]. Red square indicates nr and PDB sequences for the RaTG13 S.

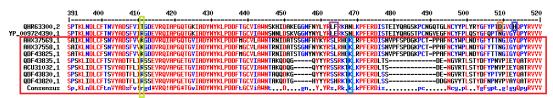


Fig.30b: multiple sequence alignment of R.affinis infecting Sarbecoviruses indicating that H505 is not found in other known Sarbecovirus strains infecting this species. Red square indicates previous Sarbecovirus with a host listed as "Rhinolophus Affinis".

As these sites were found to be unique in RaTG13, and since these 2 sites resulted in a significant change in the residue's general properties compared to the analogous position on all other

Sarbecovirus RBD known (Basic in all other RBDs->Neural polar for R403T, Neutral in all other RBDs -> Acidic for N501D), we set to deduce their effect in the binding of the RBD to ACE2. A publication using deep mutational scanning analysis[14] suggest that Y449F, N501D and Y505H resulting in the highest reduction of binding affinity to hACE2 when applied to SARS-CoV-2.

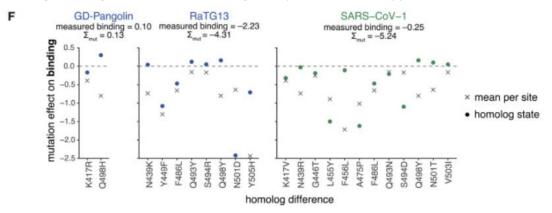


Figure 5F from [14] with different mutations from RaTG13 RBD applied to SARS-CoV-2 RBD. N501D resulted in the most severe reduction in binding affinity to ACE2, followed by Y449F and Y505H.

Using structural analysis, we discovered that the residues on ACE2 surrounding Y449, N501 and Y505 in SARS-CoV-2 are identical between Human and R.affinis ACE2, indicating that the reduction of binding affinity conferred by N501D, Y449F and Y505H would also cause the same reduction in binding affinity to R.affinis ACE2. Indeed, no sequence from R.affinis contained D501 or H505, implying that these 2 residues are also avoided in viruses that naturally circulating in this species, indicating that they cause a substantial reduction of viral fitness if introduced.

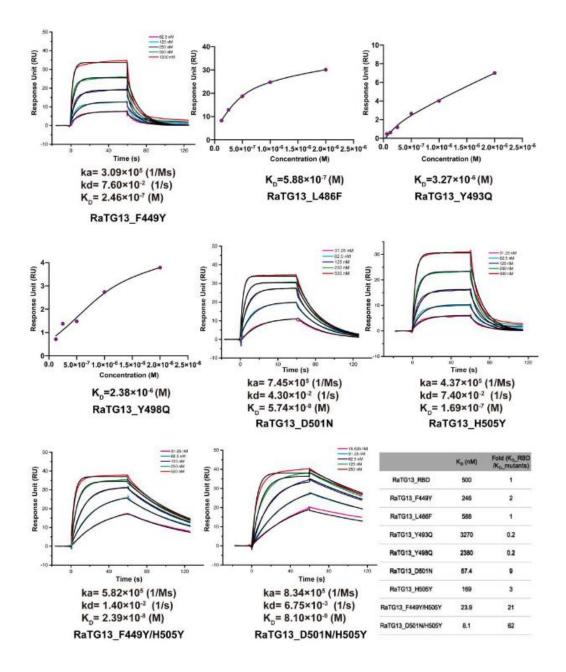


Fig. S10 from [10] denoting the effect of different amino acid substitutions on the RaTG13 RBD and their effects on binding to hACE2.

Indeed, upon mutating the position 501 and position 505 into N and Y, the consensus sequence within the R.affinis species, the binding affinity of RaTG13 RBD to ACE2 is fully restored, with an 82-fold increase compared to QHR63300.2 in term of ACE2 binding affinity. As these 2 residues interfaces with residues that were conserved between human and R.affinis ACE2, it is predicted that the same change will also improve the binding of the RaTg13 RBD to R.affinis ACE2, potentially to similar levels as SARS-CoV-2 to human ACE2.

Interestingly, the authors of [11] did not test the effect on ACE2 binding affinity of D501N or H505Y in the RaTG13 RBD, nor did they test any of the RBD mutants on R.affinis ACE2.

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
NP_001358344,1 QHQ39244,1 Consensus	MSGSSI	HELLSEVAY	raaqistted	RAKIFLONFN	HEAEDLSYQS	SLASHEYNTI	ITSDENVOKMD	Eagakusaft	_KEQSTLAQNYF EEQSKLAKNYF LeEQSklaqnyf	LEEIQTYPY	LQLQILQQSC	is <mark>p</mark> ylsedksk	(RLN <mark>s</mark> tlnams	TIYSTG
NP_001358344.1 0H039244.1									210 IRGDYEVNGVDG					
Consensus	KYCnP 261	#NPQEC1LLI 270	PGL##IMa 280	nS1DYNERLH 290	AMEgHRaEYGI 300	QLRPLYEEN 310	YaLKNEMARa 320	nHYEDYGDYI 330	<mark>IRrDYEt #esd6</mark> 340	isdYSRdQLie 350	DV#riFeEI) 360	PLYEHLHAY\ 370	/RaKLM#aYPf 380	hISPiG 390
NP_001358344,1 QHQ39244,1 Consensus	CLPAH	LLGDMAGRFI	ITNLYPLTY	PFGQKPNIDY	FDAMYNQGHD	INRIFKERE	KEEVSYGLPNM	TEGFANNSHL	_TDPGNYQKAYQ _TEPGDGRKYYQ _T#PG#grKaYQ	HPTANDLGK	OFRIKHCTK	THEDFLTAH	IENGHIQYDNA	YATOPY
	391	400	410	420	430	440	450	460	470	480	490	500	510	520
NP_001358344,1 QHQ39244,1 Consensus	LLRNG	ANEGEHEAV	JEYHSLSVA	TPKHLKTHGL	SPDFLEDNE	EINFLLKQ	ILNIVGTLPFT	YHLEKHRHH	/FKGEIPKDQHH /F <mark>R</mark> GEIPKEEHH /FrGEIPK##HH	IKKHHEMKRD	VGVVEPVPHE	DETYCOPASLE	HYANDYSFIE	YYTRTI
	521	530	540	550	560	570	580	590	600	610	620	630	640	650
NP_001358344,1 QHQ39244,1 Consensus	FEFQF	HEALCRIAD	IDGPLHKCD	ISNSTDAGKK	HQHLSYGKS	PHTYTLKD	evdsrnmdygp	LLRYFEPLY1	FHLKDQNKNSFY FHLQEQNRKSYY FHLg#QNrnS%Y	GHNTDHSPYS	DQSIKVRISL	KSALGEKAYE	HNDNEMYLFR	SSVAYA
	651	660	670	680	690	700	710	720	730	740	750	760	770	780
NP_001358344,1 QMQ39244,1 Consensus	MRQYF	LKYKNQHILI SK-KNQPILI	GEEDVRVA Gvenvhvs	NLKPRISFNF NLKPRISFNF	VTAPKNVSD VTSPGNVSD	LIPRTEVEK LIPRSEVEG	IRMSRSRIND IRMSRSRIND	AFRLNDNSLE	EFLGIQPTLGPF EFLGIQPTLGPF EFLGIQPTLGPF		CVFGVVHGVIV CVFGVVH <mark>N</mark> VVV	VGIVILIFTO	IRDRKK <mark>knk</mark> f Irdrrktdof	rsgenp Rs <mark>e</mark> enp
NP_001358344.1 QHQ39244.1 Consensus	YASID YSSVD	790 ISKGENNPGI LNKGENNPGI INKGENNPGI	QNTDDYQT QNGDDYQT	SF SF										

Fig.31: Alignment of Human and R.affinis ACE2.

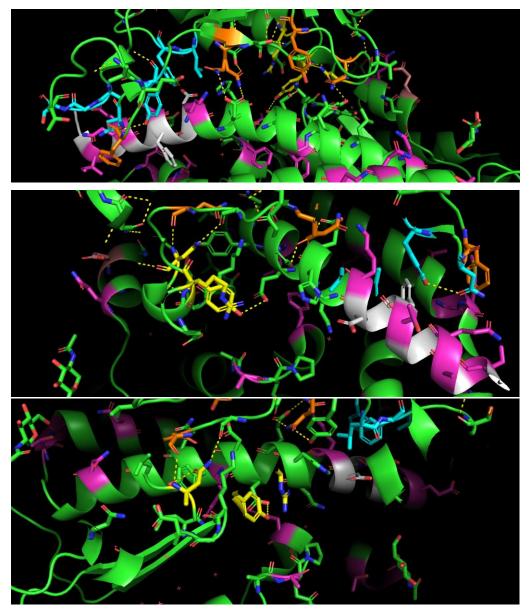


Fig.32: Structure of the pocket surrounding R403, Y449, N501 and Y505 on the SARS-CoV-2 RBD protein. Yellow is the R403, N501 and Y505 on SARS-CoV-2 RBD, green and white sticks denote residues that are identical between Human and R.affinis ACE2, Orange sticks denote residues that were different between SARS-CoV-2 and RaTG13 RBD and Magenta sticks denote residues that were different between human and R.affinis ACE2.

Consistent with the discovery of two residues (T403 and D501) with unique chemical properties that have never been recorded in any Sarbecoviruses, This reduction of binding affinity was found to be general for all animal ACE2 tested, with the highest recorded pseudovirus entry (RLU) being ~10^1 times lower than the entry efficiency of SARS-CoV-2 S on hACE2.

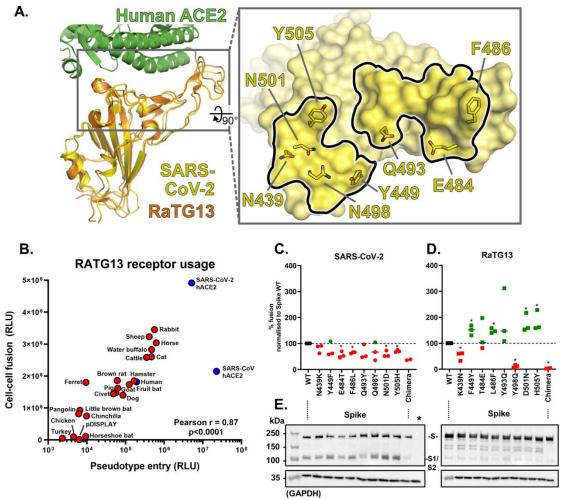
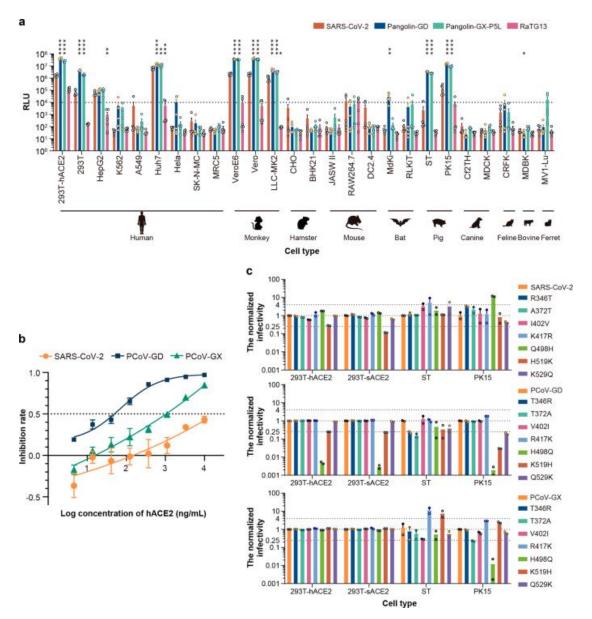


Fig.4 from [16]: Cell-to-cell fusion and pseudovirus entry assay of SARS-CoV-2, SARS-CoV and RaTG13 on different ACE2 orthologues overexpressed on HEK293T cells.



The RaTG13 S exhibits a restricted tropism and is specific to Immortalized Kidney cells.

Fig.1 from [15] comparing the tropism of SARS-CoV-2, pCoV-GD, pCoV-GX and RaTG13. While only HEK293T-ACE2 displayed an infectivity for RaTG13 pseudovirus with an RLU of over 10⁴, both HEK293T-ACE2 and PK15 displayed a roughly 10¹ difference between the infectivity for SARS-CoV-2 and RaTG13. All other cell line where there RaTG13 show above-background infectivity, except for mouse Macrophage cell line RAW264.7, had a difference of 10² to 10³ in term of pseudovirus entry for SARS-CoV-2 and RaTG13.

As ACE2 bind integrins through the KGD motif on position 353-355 which is conserved in human and R.affinis ACE2[17][18][19], It is modeled that ACE2 in physiological concentrations is bound to Integrin α 5 β 1 and is inaccessible to binding by the RBD of Sarbecovirus Spike proteins[18], unless it is displaced by an KGD/RGD motif that is found in all RBD sequences that lacked the two deletions in SL-CoVs RBDs that does not use ACE2 for entry.[20]

The primary feature of the HEK293T cells (ACE2) used in [11] is their substantially higher expression of ACE2 over their constitutional ITGB1 expression, which will result in large amount of free ACE2 that is physiologically unrealistic in real tissues like bat intestines or human lungs.

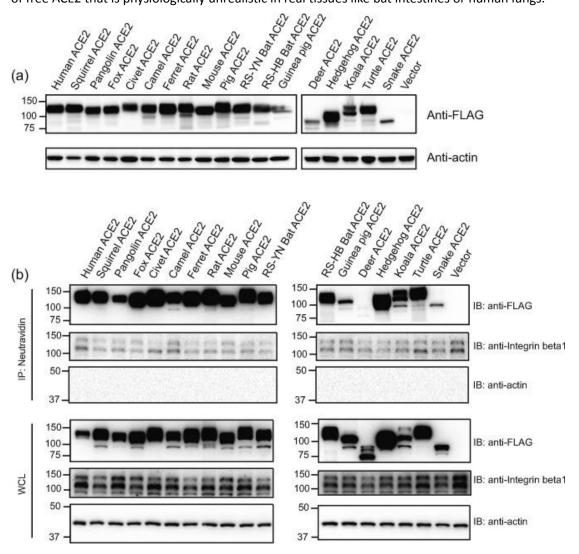


Fig.2 from [11]: Overexpression of ACE2 orthologues on HEK293T result in an overwhelmingly high amount of both whole-cell and surface ACE2 molecules comparing to the amount of ITGB1 molecules available for binding to ACE2.

Indeed, Inhibiting the interaction of the Spike to α 5 β 1integrins using an integrin-inhibiting peptide have been found to reduce the binding and entry efficiency of SARS-CoV-2 to hACE2 and VERO E6 cells[18], and Integrins have been speculated as a co-receptor for SARS-CoV-2[19][21].

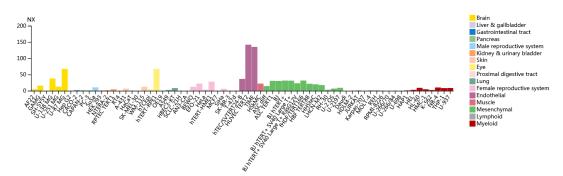


Fig.33a: expression level of ITGA5 on different cell lines.

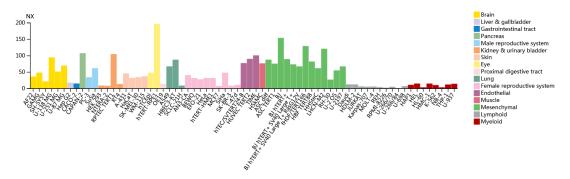


Fig.33b: expression level of ITGB1 on different cell lines.

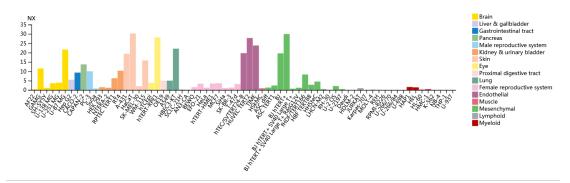


Fig.33c: expression level of ITGA3 on different cells lines.

The effect of integrin expression on the entry efficiency of RaTg13 and SARS-CoV-2 S pseudotyped lentivirus is demonstrated by the difference between the entry efficiency of the 3 Spike proteins to HeLa-ACE2 and HEK293T cells[22].

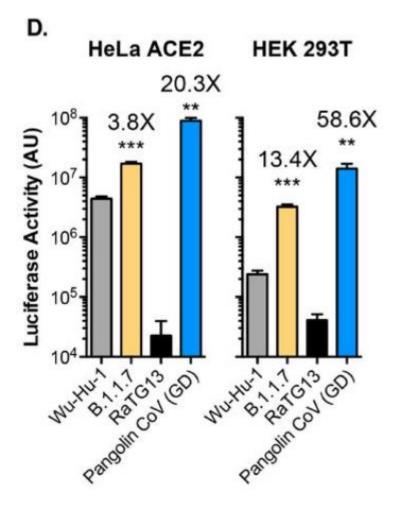


Fig.3d from[22]: HeLa ACE2 show substantially higher ratios between WuHu-1 or B1.1.7 entry and RaTG13 entry, compared to HEK293T.

As immortalized Kidney cells (HEK293T) are found to have much lower level of ITGB1, ITGA2 or ITGA5 expression, compounded with the fact that HEK293T-ACE2 cells is the only human cell line that support substantial entry by RaTG13 pseudotyped lentivirus with RLU above 10^4[15] despite all cell lines from humans expressing only human ACE2, This highly restricted tropism of RaTg13 toward Immortalized Kidney cells likely indicate serial passage within such cells.

Incidentally, the only cell line from R.affinis in possession of the Wuhan Institute Of Virology (WIV) was RaK4324 cells from the Kidneys, which would have been the laboratory passage host for RaTG13 if it have been cultured prior to sequencing.[23]

As both the neutral 501 and the basic (R/K)403 are found to be 100% conserved in all Sarbecovirus RBD proteins except for RaTG13, these 2 positions are likely indispensable for the in-vivo fitness of Sarbecoviruses in both reservoir hosts, in other animals and in humans.

Since experimental evidence have validated the broadly detrimental effect of both D501 and T403 on the RaTG13 S on viral fitness (RBD binding to ACE2, Spike entry into cells), these two positions can be considered as signature of attenuation in the RaTG13 RBD protein.

RaTG13 is an attenuated vaccine strain cultured in

immortalized bat kidney cells?

The only known cell line from R.affinis in possession of the Wuhan Institute of Virology was RaK4324 Primary Kidney cells[23], which were used in the isolation and culture of bat Coronaviruses.

Should the SRA dataset of RaTG13, SRX7724752, have been a cell culture of an attenuated virus within an Immortalized version of the RaK4324 cells, It would simultaneously explain nearly all the known anomalies associated with the raw read data and the nucleotide sequence of RaTG13, MN996532.

Bat telomeres are known to not shorten with age[24], which indicate that the mechanism of telomere erosion is likely absent in bat cells. Should a traditional telomerase based immortalization strategy being used on a culture of R.affinis Kidney cells, one of the likely outcome is that telomeric sequences will grow uncontrollably and accumulate to very high fractions after extensive passage of the cell line due to the TERT activity not being balanced by telomere erosion mechanisms that were found in other mammalian cells but not bat cells.

In addition, Cell cultures are normally kept under sterile conditions using a cocktail of antibiotics in combination with aseptic techniques to minimize microbial growth, which would have resulted in a sample that is mostly sterile with minimal to no bacterial sequences.

One of the defining feature of the SARS-CoV-2 S is the optimization of ACE2 binding and folding at 37° C, a feature that it shares with the RaTG13 S.[26][27] However, the body temperature of a Horseshoe bat can reach up to 41° C [28], where substantial unfolding of the RaTG13 Spike happens according to Differential Scanning Fluorimetry on the Spike trimer[27]. This is incompatible with the high body temperature of a horseshoe bat (as the virus will be inactivated by the heat generated by bat flight), but is compatible with a cell culture as most cells in laboratories are cultured at 37° C.

Traces of lentivirus- and HERV-like fragments found in SRX7724752 likely indicate the usage of retroviral- and lentiviral- vectors on the sample, which are frequently used for the delivery of a TERT gene for the immortalization of cell lines in-vitro, a pre-requisite for the subsequent culture and attenuation for a vaccine strain.

Indeed, the combined features of the RaTG13 genome resemble that of a Live Attenuated Vaccine (LAV) almost suspiciously, to the point that there are actual proposals for the usage of this sequence as a candidate vaccine against SARS-CoV-2[25].

In addition, evidence of mutagen usage during the divergent evolution of RaTG13 and SARS-CoV-2 from a common ancestor can be found in the Spike protein CDS of RaTG13, manifesting as a very large excess of C:T transitions compared to the substitutional pattern of SARS-CoV-2 WIV04:ZC45, RaTG13:ZC45 or SARS-CoV Tor2:WIV1 on the aligned sections of their Spike protein CDS sequences.[29] As Mutagens like 5-fluorouracil is not present in wild bats, The RaTG13 Spike protein CDS would have to be grown in a cell culture to be influenced by 5-fluorouracil and show such a peculiar substitutional pattern.

CONCLUSION

The raw data of BtCoV/RaTG13 Contained multiple anomalies that signifies that the original sample could not have contained enough RNA template for the extraction of a complete viral genome as in MN996532.1

Furthermore, many of these anomalies points toward the fraudulent use of a mixed DNA library, rather than genuine mRNA, for the sequencing of SRX7724752, evident by the presence of widespread A-T ligation of unrelated dsDNA fragments that can only happen if the same library preparation process have been ran on dsDNA instead of ssRNA. which would constitute Academic fraud.

The Spike glycoprotein of RaTG13 does not resemble that of a wild virus but instead possessed multiple signatures of artificial attenuation in a cell culture when compared to the SARS-CoV-2 Spike and the Spike sequences of other related viruses, indicating that the sequence did not derive from what the Wuhan Institute Of Virology claimed to be.

Therefore, the sequencing of BtCoV/RaTG13 cannot be considered to be valid or honest as is, and any publications, including [2], and other publications that cites or use RaTG13 as critical pieces of evidence or proof, must be immediately invalidated and retracted.

REFERENCES

[1] https://www.thetimes.co.uk/article/seven-year-covid-trail-revealed-I5vxt7jqp

[2] Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273. doi:10.1038/s41586-020-2012-7

[3] Rose C, Parker A, Jefferson B, Cartmell E. The Characterization of Feces and Urine: A Review of the Literature to Inform Advanced Treatment Technology. *Crit Rev Environ Sci Technol*. 2015;45(17):1827-1879. doi:10.1080/10643389.2014.1000761

[4] https://telegra.ph/RaTG13-07-06

[5] Xiao C, Li X, Liu S, Sang Y, Gao SJ, Gao F. HIV-1 did not contribute to the 2019-nCoV genome. *Emerg Microbes Infect*. 2020;9(1):378-381. Published 2020 Feb 14. doi:10.1080/22221751.2020.1727299

[6] https://www.planetnatural.com/product/mexican-bat-guano/

[7] Paskey, A.C., Frey, K.G., Schroth, G. *et al.* Enrichment post-library preparation enhances the sensitivity of high-throughput sequencing-based detection and characterization of viruses from complex samples. *BMC Genomics* **20**, 155 (2019). <u>https://doi.org/10.1186/s12864-019-5543-2</u>

[8] Edgar, R. C. *et al.* Petabase-scale sequence alignment catalyses viral discovery. *bioRxiv* 2020.08.07.241729 (2020) <u>doi:10.1101/2020.08.07.241729</u>

[9] Study of the Metatranscriptome of Eight Social and Solitary Wild Bee Species Reveals Novel Viruses and Bee Parasites

Karel Schoonvaere1,2*, Guy Smagghe3, Frédéric Francis2 and Dirk C. de Graaf1

1Laboratory of Molecular Entomology and Bee Pathology, Department of Biochemistry and Microbiology, Faculty of Sciences, Ghent University, Ghent, Belgium

2Functional and Evolutionary Entomology, Gembloux Agro-Bio Tech, University of Liege, Gembloux, Belgium

3Laboratory of Agrozoology, Department of Crop Protection, Faculty of Bioscience Engineering, Ghent University, Ghent, Belgium

[10] Zhang, S., Qiao, S., Yu, J. *et al.* Bat and pangolin coronavirus spike glycoprotein structures provide insights into SARS-CoV-2 evolution. *Nat Commun* **12**, 1607 (2021). <u>https://doi.org/10.1038/s41467-021-21767-3</u>

[11] Pei Li, Ruixuan Guo, Yan Liu, Yingtao Zhang, Jiaxin Hu, Xiuyuan Ou, Dan Mi, Ting Chen, Zhixia Mu, Yelin Han, Zihan Chen, Zhewei Cui, Leiliang Zhang, Xinquan Wang, Zhiqiang Wu, Jianwei Wang, Qi Jin, Zhaohui Qian,

The Rhinolophus affinis bat ACE2 and multiple animal orthologs are functional receptors for bat coronavirus RaTG13 and SARS-CoV-2,

Science Bulletin,

2021,

, ISSN 2095-9273, https://doi.org/10.1016/j.scib.2021.01.011.

[12] Andersen, K.G., Rambaut, A., Lipkin, W.I. *et al*. The proximal origin of SARS-CoV-2. *Nat Med* **26**, 450–452 (2020). https://doi.org/10.1038/s41591-020-0820-9

[13] Makowski, L.; Olson-Sidford, W.; W. Weisel, J. Biological and Clinical Consequences of Integrin Binding via a Rogue RGD Motif in the SARS CoV-2 Spike Protein. *Viruses* **2021**, *13*, 146. https://doi.org/10.3390/v13020146

[14] Tyler N. Starr, Allison J. Greaney, Sarah K. Hilton, Daniel Ellis, Katharine H.D. Crawford, AdamS. Dingens, Mary Jane Navarro, John E. Bowen, M. Alejandra Tortorici, Alexandra C. Walls, Neil P. King, David Veesler, Jesse D. Bloom,

Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding,

Cell, Volume 182, Issue 5, 2020, Pages 1295-1310.e20, ISSN 0092-8674,

https://doi.org/10.1016/j.cell.2020.08.012.

[15] Nie, J., Li, Q., Zhang, L. *et al.* Functional comparison of SARS-CoV-2 with closely related pangolin and bat coronaviruses. *Cell Discov* **7**, 21 (2021). https://doi.org/10.1038/s41421-021-00256-3

[16] Conceicao C, Thakur N, Human S, Kelly JT, Logan L, Bialy D, Bhat S, Stevenson-Leggett P, Zagrajek AK, Hollinghurst P, Varga M, Tsirigoti C, Tully M, Chiu C, Moffat K, Silesian AP, Hammond JA, Maier HJ, Bickerton E, Shelton H, Dietrich I, Graham SC, Bailey D. The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins. PLoS Biol. 2020 Dec 21;18(12):e3001016. doi: 10.1371/journal.pbio.3001016. PMID: 33347434; PMCID: PMC7751883.
[17] Clarke NE, Fisher MJ, Porter KE, Lambert DW, Turner AJ. Angiotensin converting enzyme (ACE) and ACE2 bind integrins and ACE2 regulates integrin signalling. PLoS One. 2012;7(4):e34747. doi: 10.1371/journal.pone.0034747. Epub 2012 Apr 16. PMID: 22523556; PMCID: PMC3327712.

[18] Brandon J. Beddingfield, Naoki Iwanaga, Prem P. Chapagain, Wenshu Zheng, Chad J. Roy, Tony Y. Hu, Jay K. Kolls, Gregory J. Bix,

The Integrin Binding Peptide, ATN-161, as a Novel Therapy for SARS-CoV-2 Infection,

JACC: Basic to Translational Science,

Volume 6, Issue 1,

2021,

Pages 1-8,

ISSN 2452-302X,

https://doi.org/10.1016/j.jacbts.2020.10.003.

[19] Factors Associated with Emerging and Re-emerging of SARS-CoV-2 Variants

Austin N. Spratt, Saathvik R. Kannan, Lucas T. Woods, Gary A. Weisman, Thomas P. Quinn, Christian L. Lorson, Anders Sönnerborg, Siddappa N. Byrareddy, Kamal Singh

bioRxiv 2021.03.24.436850; doi: <u>https://doi.org/10.1101/2021.03.24.436850</u>

[20]

Difference in Receptor Usage between Severe Acute Respiratory Syndrome (SARS) Coronavirus and SARS-Like Coronavirus of Bat Origin

Wuze Ren, Xiuxia Qu, Wendong Li, Zhenggang Han, Meng Yu, Peng Zhou, Shu-Yi Zhang, Lin-Fa Wang, Hongkui Deng, Zhengli Shi

Journal of Virology Jan 2008, 82 (4) 1899-1907; DOI: 10.1128/JVI.01085-07

[21] Tikam Chand Dakal,

SARS-CoV-2 attachment to host cells is possibly mediated via RGD-integrin interaction in a calcium-dependent manner and suggests pulmonary EDTA chelation therapy as a novel treatment for COVID 19,

Immunobiology, Volume 226, Issue 1, 2021, 152021, ISSN 0171-2985, https://doi.org/10.1016/j.imbio.2020.152021.

[22] Characterisation of B.1.1.7 and Pangolin coronavirus spike provides insights on the evolutionary trajectory of SARS-CoV-2

Samuel J. Dicken, Matthew J. Murray, Lucy G. Thorne, Ann-Kathrin Reuschl, Calum Forrest, Maaroothen Ganeshalingham, Luke Muir, Mphatso D. Kalemera, Machaela Palor, Laura E. McCoy, Clare Jolly, Greg J. Towers, Matthew B. Reeves, Joe Grove

bioRxiv 2021.03.22.436468; doi: https://doi.org/10.1101/2021.03.22.436468

[23] Wang N, Luo C, Liu H, Yang X, Hu B, Zhang W, Li B, Zhu Y, Zhu G, Shen X, Peng C, Shi Z. Characterization of a New Member of Alphacoronavirus with Unique Genomic Features in Rhinolophus Bats. Viruses. 2019 Apr 24;11(4):379. doi: 10.3390/v11040379. PMID: 31022925; PMCID: PMC6521148.

[24] Growing old, yet staying young: The role of telomeres in bats' exceptional longevity

By Nicole M. Foley, Graham M. Hughes, Zixia Huang, Michael Clarke, David Jebb, Conor V. Whelan, Eric J. Petit, Frédéric Touzalin, Olivier Farcy, Gareth Jones, Roger D. Ransome, Joanna Kacprzyk, Mary J. O'Connell, Gerald Kerth, Hugo Rebelo, Luísa Rodrigues, Sébastien J. Puechmaille, Emma C. Teeling Science Advances07 Feb 2018 : eaao0926

[25] Chougle, Afreen and Chougle, Humera and Sayyed, Rajat, Prospect of Using RaTG13 Sarbecovirus As a Candidate Vaccine for COVID-19 (September 2, 2020). Available at SSRN: https://ssrn.com/abstract=3685067 or http://dx.doi.org/10.2139/ssrn.3685067

[26] Zhonghua Zhou, Ziyi Yang, Junxian Ou, Hong Zhang, Qiwei Zhang, Ming Dong, Gong Zhang,

Temperature dependence of the SARS-CoV-2 affinity to human ACE2 determines COVID-19 progression and clinical outcome,

Computational and Structural Biotechnology Journal,

Volume 19,

2021,

Pages 161-167,

ISSN 2001-0370,

https://doi.org/10.1016/j.csbj.2020.12.005.

[27] Wrobel, A.G., Benton, D.J., Xu, P. *et al.* SARS-CoV-2 and bat RaTG13 spike glycoprotein structures inform on virus evolution and furin-cleavage effects. *Nat Struct Mol Biol* **27**, 763–767 (2020). <u>https://doi.org/10.1038/s41594-020-0468-7</u>

[28] O'Shea TJ, Cryan PM, Cunningham AA, Fooks AR, Hayman DT, Luis AD, Peel AJ, Plowright RK, Wood JL. Bat flight and zoonotic viruses. Emerg Infect Dis. 2014 May;20(5):741-5. doi: 10.3201/eid2005.130539. PMID: 24750692; PMCID: PMC4012789.

[29] Lv Longxian, Li Gaolei, Chen Jinhui, Liang Xinle, Li Yudong Comparative Genomic Analyses Reveal a Specific Mutation Pattern Between Human Coronavirus SARS-CoV-2 and Bat-CoV RaTG13 Frontiers in Microbiology vol.11 2020

https://www.frontiersin.org/article/10.3389/fmicb.2020.584717

https://doi.org/10.3389/fmicb.2020.584717

ISSN: 1664-302X