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)

GTTAGGGTTAGGGTTAGGGTTAGGGTTGGGGTTGGGATACAGGATATGGTTAGGGTTAGGG GTAGGGTCAGGGTTAGGATTGGAAACGAGATAGGTTACGTGATAGGGTTAGCGTTAGGGT 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)

GGTAAGGGTTAGGGTTAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG GTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGA 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			Query Co	/erage
Molecule type	dna	to	1	0			to
Query Length	151						
Other reports	Distance tree of results MSA viewer 😧					Filter	Reset
Descriptions	Graphic Summary Alignments Taxonomy						
Sequences p	producing significant alignments	Download 🗡	Manag	ge Colui	nns ~	Show	5000 🗸 😮
🗹 select all	6 sequences selected		Ger	n <u>Bank</u>	Graph	ics <u>Dista</u>	nce tree of results
	Description		Max Score	Total Q Score C	iery over va	E Per. Ilue Ident	Accession
-							

		Organism only top 20 will appea	ar				exclude			
Results for	1:lcl Query_61914 gnl SRA SRR11085797.7.1 7 (Biological)(150bp)	· · · · · · · · · · · · · · · · · · ·								
Program	BLASTN ② <u>Citation</u> ✓									
Database	nt <u>See details</u> •	nt See details Y								
Query ID	lcl Query_61914	Percent Identity E va	lue		Qu	ery Cov	erage			
Description	gnl SRA SRR11085797.7.1 7 (Biological)	to	to			t	0			
Molecule type	dna									
Query Length	150	Filter Rese								
Other reports	Distance tree of results 🔞									
Descriptions	Graphic Summary Alignments Taxonomy									
Sequences p	producing significant alignments	Download 🗡	Manage	Column	s ~ S	show	5000 🗸 🔞			
🗹 select all	2 sequences selected		GenBa	ink <u>Gr</u>	aphics	Distan	ce tree of results			
	Description		Max Tot Score Sco	al Query re Cover	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										
าล										
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			Max 1 Score S	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%	<u>XM_0285314</u>	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	S ribosomal RNA (LOC1	<u>17704856), rRNA</u>		274	274 1009	6 9e-70	99.34%	XR_0046063	<u>369.1</u>
	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	dna 151	0003.111330003 (Biolog		Percent Identity	/	E value	to	Que	to	
uery Length	dna 151 Distance tree of results MS	A viewer 😮		Percent Identity		E value	to	Que	to Filter	Reset
Descriptions	dna 151 Distance tree of results MS Graphic Summary	Alignments		to		E value	to	Que	to Filter	Reset
Decule type of the second seco	dna 151 Distance tree of results MS Graphic Summary oducing significant all	Alignments		Dercent Identity	ownload	E value	to	Que	ry Coverage to Filter	Reset
Descriptions Descriptions Sequences privations Select all 10	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments		Dercent Identity	ownload	E value	to	Que	ry Coverage to Filter	Reset
her reports Descriptions Sequences pro Select all 10	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Percent Identity	/ ownload	E value E value Man Total Query Score Cover	to	Que	ry Coverage to Filter I how 100 Distance tree of Accession	Reset
ery Length : her reports : Descriptions Sequences pr Select all 10 SRX1724762	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Percent Identity	y pwnload Score S 279	E value E value Man Total Query Score Cover 279 100%	to age Colur value k 2e-73 10	Que	ry Coverage to Filter 1 how 100 Distance tree of Accession SRR11085797.11	Reset
ecule type of ery Length in the reports in the reports in the section of the sect	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 00 sequences selected	Alignments Description		Percent Identity to to Do	y wmload Score S 279 279	E value E value Man Total Query Cover 279 100% 279 100%	to	Quei	ry Coverage to Filter 10 Now 100 C Accession SRR11085797.11 SRR11085797.11	 Reset of results 601488.1 659562.1
elecule type ery Length her reports	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Dercent Identity	y pwnload Score S 279 279 279 279 279	E value Warn Total Query Cover 279 100% 279 100% 270 1	to	Quei	to to Filter 100 Miles Services Service	Reset of results 1601488.1 1599592.1 1598965.1 160972.1
ery Length ery Length her reports	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279	E value Man Total Query Score Cover 279 100% 279 100% 279 100% 279 100%	to	Quei	ry Coverage to Filter 10 how 100 • Distance tree of Accession SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11	 Reset of results 601488.1 1599592.1 1598955.1 1598955.1 1598955.1 1598757.1 158857.1 158857.1
ery Length ery Length her reports	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 0 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279	E value Man Total Query Core Cove 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	to age Column value 2e-73 100 2e-73 100 2e-73 100 2e-73 100 2e-73 100 2e-73 100 2e-73 100	Quei Imms S Graphics Per. Jono% SRA: 0.00% SRA:	ry Coverage to Filter to Distance tree of Accession SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11	Reset of results 601488.1 1599521 15999521 1598955.1 1598757.1 15885757.1 15885765.1
etecule type ery Length erry Length err reports	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 00 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279 279	E value Man Total Query Total Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	to	Quei mns S Graphics S Per. 0 0.00% SRA:	ry Coverage to to Filter 10 how 100 • Distance tree : Accession SRR:11085797.11 SRR:11085797.11 SRR:11085797.11 SRR:11085797.11	Reset Contended Cont
iecule type i ery Length i her reports i Descriptions i Sequences pr Image: Select all 10 Imag	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 0 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279 279	E value Man Total Query Score Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	to age Column 2e-73 10 2e-73 10 2e-74 10 2e-	Quei nns × S Graphics - - Per. - - 0.00% SRA: -	ry Coverage to to Filter 10 how 100 Accession SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11	Reset of results 601488.1 159952.1 159952.1 159875.1 158875.1 158575.1 158575.1 158575.1 158575.1 1583582.1 1582582.1 1582582.1 1582582.1
 lecule type select all 10 select all 10 select all 20 sele	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279 279	E value Marrier Total Query Core Cover 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100% 279 100%	to	Quei Imms S Graphics S O.00% SRA. 0.00% SRA.	ry Coverage to Filter to Distance tree of Accession SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.13	Reset
 cecule type rery Length her reports Descriptions Sequences product select all 10 SRXT724752 	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279 279	E value Man Total Man Total Cover 279 100% 279 100% 270 100%	to	Quei Imms S Graphics S Par. S 0.00% SRA:	ry Coverage to Filter to Distance tree of Accession SRR11085797.11 SRR11085797.11 SRR11085797.13 SRR11085797.13 SRR11085797.13 SRR11085797.13 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11	Reset of results 1601488.1 159952.1 1598965.1 1598757.1 1588533.1 1588533.1 1588533.1 1588532.1 158262.1 158262.1 158262.1 158262.1 158262.1 158262.1 158262.1 158262.1 158262.1 158262.1 158262.1 158262.1 1582051.1 158262.1
Constant Section 2012 Constant Sect	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279 279	E value Man Total Queve Score Cover 279 100% 279 100% 270 10	to	Quei Imms S Statution S Graphics S Per. S O.00% SRA: 0.00% SRA:	ry Coverage to Filter to Distance tree of Accession SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11	Reset
Concule type Lery Length Lery Length Concurrent C	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 0 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279 279	E value Total Query Score Query Core Qu	to E Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value Value V	Quei Imms S Graphics S Jent S Jono% SRA: 0.00% SRA:	ry Coverage to to Filter to Distance tree of Accession SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11	Reset
Concurrent Length	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279 279	E value Man Total Core Cor	to	Quei Imms S Graphics S Graphics S O00% SRA 0.00% SRA	ry Coverage to to Filter to Distance tree of Accession SRR11085797.11	Reset
Colocitle type Lery Length Lery Length Colocitle type Col	dna 151 Distance tree of results MS Graphic Summary oducing significant ali 10 sequences selected	Alignments Description		Percent Identity	279 279 279 279 279 279 279 279 279 279	E value Man Total Correl C	to	Que: mns S Graphics S Graphics S 0.00% SRA: 0.00% SRA:	ry Coverage to to Filter 100 • Distance tree s Accession SRR11085797.11	Reset
otecule type Image: Constraint of the second seco	dna 151 Distance tree of results MS Graphic Summary oducing significant all 0 sequences selected	Alignments Description		Percent Identity	Amar Score S Score S 279 279 279 279 279 279 279 279 279 279	E value Image: Constant C	to	Quei Imms S Graphics S O.00% SRA: 0.00% SRA:	ry Coverage to to Filter 100 Filter Distance tree of Accession SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11 SRR11085797.11	Reset



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

Score 132 hits	(71)	Expect	Identities	Gaps 0/83(0%)	Strand Plus/Minus	
Diery	6		GGGGTCAACACAGTCC			65
Zucry	0					00
Sbjct	2550	GGAGCAGAGGAC	AGGATCAACACAGTCC	ATGGACCCCTGGCAG	GGGCGATGAGATCGGTG	2491
Query	66	AACTAGGGGACA	AAAGGAAGTTA 88			
Chiet	2490			0		
				-		
bown	load 🗸	GenBank Gra	aphics			
Pat cor	onavia		mploto gonomo			
Dat Cor	onavn	us kar615, co	inplete genome			
Sequence	ID: MN	996532.1 Lengt	h: 29855 Number of I	Matches: 1		
	: 12 to	47 GenBank Gra	<u>iphics</u>		V Next Match	revious Match
Range 1						
Range 1		Expect	Identities	Gaps	Strand	

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>Dowi</u>	nload 🗸	<u>GenBank</u>	Graphics Sort b	y: E value	~		
Bat co	ronaviru	us RaTG13,	complete gene	ome			
Sequen	ce ID: MN9	96532.1 Le	ength: 29855 Nur	nber of Matches: 2			
ocquein							
Range	1: 26736	to 26844 <u>G</u> e	enBank Graphics			▼ <u>Next Match</u> ▲ P	revious Match
Score		Expect	Identities	Gaps		Strand	
161 bit	ts(87)	7e-36	102/109(94	%) 2/109	(1%)	Plus/Plus	
Query	45	GCACTGAC	TTG-CTT-TAGCA	CTGATGTGGCTGAGCT	ACTTCATTGC	TTCTTTCAGGCTA	102
_							
Sbjct	26736	GCAATGGC	TTGTCTTGTAGGC	PTGATGTGGCTGAGCT.	ACTTCATTGC	TTCTTTCAGGCTA	26795
Query	103	TTTGCACG	TACGCGTTCCATG	IGGTCATTCAATCCAG	AAACTAACAT	TT 151	
		11111111				11	
Sbjet	26/96	TTTGCACG	TACGCGTTCCATG	IGGTCATTCAATCCAG	AAACTAACAT	"I'I' 26844	
_							
Range	2: 27478	to 27542 <u>Ge</u>	enBank Graphics		Next Match	Previous Match	First Match
Score		Exped	t Identities	Gaps		Strand	
121 bit	ts(65)	1e-2	3 65/65(100	0/65(0%)	Plus/Plus	
_						~~~~~~~~~~~	<u></u>
Query	T	ATGAAGGC.	AATTCACCATTCC	ATCCTCTAGCTGATAA	TAAATTTGCA		60
Sbjct	27478	ATGAAGGC.	AATTCACCATTCC	ATCCTCTAGCTGATAA	TAAATTTGCA	CTGACTTGCTTTA	27537
0	C1	CO1 00 0	-				
Query	10	GCACT 6	5				
Sbjct	27538	GCACT 2	7542				

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: 28217	to 28346 GenBar	k Graphics		▼ <u>Next Match</u> ▲ P	evious Match
Score 233 bit	ts(126)	Expect 2e-57	Identities 129/130(99%)	Gaps 1/130(0%)	Strand Plus/Plus	
Query	23	TTC-TCTAAACG	AACAAACTAAAATGTC	TGATAATGGACCCCAA	AACCAACGAAATGCAC	81
Sbjct	28217	TTCATCTAAACG	AACAAACTAAAATGTC	TGATAATGGACCCCAA	AACCAACGAAATGCAC	28276
Query	82	CCCGCATTACGT	TTGGTGGACCCTCAGA	TTCAACTGGCAGTAAC	CAGAATGGAGAACGCA	141
Sbjct	28277	CCCGCATTACGT	TTGGTGGACCCTCAGA	TTCAACTGGCAGTAAC	CAGAATGGAGAACGCA	28336
Query	142	GTGGAGCACG	151			
Sbjct	28337	GTGGAGCACG	28346			
Range	2: 25 to (50 <u>GenBank</u> <u>Grap</u>	<u>hics</u>	▼ <u>Next Ma</u>	tch A Previous Match	First Match
Score		Expect	Identities	Gaps	Strand	
67.6 b	its(36)	2e-07	36/36(100%)	0/36(0%)	Plus/Plus	
Query	1 C:	PCTCGATCTCTTGT	AGATCTGTTCTCTAAA	CGAAC 36		
Sbjct	25 C	PCTCGATCTCTTGT	AGATCTGTTCTCTAAA	CGAAC 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

Des	criptions	Graphic Summary	Alignments	Taxonomy							
Sec	luences p	roducing significant a	lignments		Download 🗡	Man	age C	olumn	s ~	Show	100 🗸 😮
	select all 1	1 sequences selected				Ge	enBanl	<u>Gra</u>	aphics	<u>Distan</u>	ce tree of results
			Des	cription		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	Rhinolophus	ferrumequinum clone VMRC7-7	1A7. complete sequence			267	267	99%	1e-67	98.68%	AC150307.3
	Rhinolophus	ferrumequinum clone VMRC7-2	51C10, complete sequen	ice		185	185	94%	4e-43	90.14%	AC149630.3
	Myotis lucifu	gus clone CH235-427D16. comp	lete sequence			137	137	82%	1e-28	87.20%	AC174832.3
	Pteropus ale	cto clone BAC P100M20, BAC I	P103A18, complete sequ	Jence		135	135	86%	4e-28	85.38%	KP862827.1
	Pteropus ale	cto clone BAC P201M3, BAC P	216K21, complete seque	ence		135	135	86%	4e-28	85.38%	KP862826.1
	Pteropus ale	cto clone BAC P56N20, comple	te sequence			135	135	86%	4e-28	85.38%	KP862825.1
	Rhinolophus	euryale isolate REM0134 micro	satellite RM1198 sequen	ce		128	128	47%	7e-26	98.61%	KC910215.1
	Pteropus ale	cto clone BAC P212O7-1, BAC	P229M21, complete seq	uence		126	126	84%	3e-25	84.38%	KP862828.1
	PREDICTED	Miniopterus natalensis zinc fin	ger protein 713 (ZNF713)	, transcript variant X	mRNA	106	106	90%	3e-19	81.02%	XM_016196283.1
	PREDICTED	Miniopterus natalensis zinc fin	ger protein 713 (ZNF713)	<u>, transcript variant X</u>	MRNA	106	106	90%	3e-19	81.02%	XM_016196281.1
	PREDICTED	Miniopterus natalensis zinc fin	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]

Data	R	aTG13	SAF	S-CoV-2	c		News
Date	Identity	Query Cover	Identity	Query Cover	Sequence	Blast Archive	Name
14-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
14-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
14-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
14-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
11-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
11-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
08-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
08-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
08-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
30-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
30-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
29-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
29-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
29-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
29-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
27-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
27-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
27-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
27-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
27-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
27-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
27-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
27-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
27-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
27-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
20-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
20-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
20-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
20-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
17-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
17-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
03-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
03-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>c Gra</u>	phics	Distance tree of resul		
	Description	Max Score	Total Score	Query Cover	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	

select all 100 sequences selected	Ge	enBank	<u>Gra</u>	aphics	Distan	<u>ce tree of results</u>
Description	Max Score	Total Score	Query Cover	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_Caffold_79	50.9	50.9	9%	0.078	96.88%	LN483167.1
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%	CP027084.1
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>Gra</u>	<u>iphics</u>	<u>Distan</u>	ce tree of results
	Description	Max Score	Total Score	Query Cover	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 😧 <u>Citation</u> 🗸
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

Score		Expect	Identities	Gaps	Strand	
154 bit	s(83)	1e-33	116/132(88%)	1/132(0%)	Plus/Minus	
Query	12	TTT-TCATTAT	FAAGTATTATGTACTGT	ACATAATTGTATGTAC	TATACTTTTATACAAC	70
Sbjct	2293	TTTATCATTAT(CAAGTGTTATGTACTGT	ACAGTATTGTATGTGT	TATACTTTTATATGAC	2234
Query	71	TGGCAGCACAG	CAGGTTTGTTTATACCA	GCATCACCACAAAAAT	GTGAGTAATGCATTAC	130
Sbjct	2233	TGACAGCATAG	IAGGCTTGTTTACACCA	GCATCACCACAAAAAT	GTGAGTAATGCATTAC	2174
Query	131	ACTACAATGTTA	A 142			
Sbict	2173	ACTATGATGTT	A 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

					t	tagat	ttcatctaaacgaacaaactaaaatgtctgataatggaccccaaaaccaacgaaatgcaccccgcattacgtttggtggaccct
		CT	CTCGAT	CTCTT	GTAG	ATCTG	TTC TCTAAACGAAC
ACAAACC	AACGI	AACT	CTCGAT	CTCTT	GTAG	ATCTG	I
							IAACCTCCTTTTGTCCCCTAGTTCACCGATCTCATCGCCCCTGCCAGGGGTCCATGGACTGTGTTGATCCTGTCCTCTGCTCCTCC

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	<u>Gen</u>	<u>Bank</u>	Graphics		Distance tree of results	
	Description	Max Score	Total 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
Z	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	1: 477	20 to 47856	GenBank Grap	hics		Vext Match 🔺 Pre	evious Match
Score 176 bit	s(95)	Exped 3e-40	t Identities 0 123/137	(90%)	Gaps 0/137(0%)	Strand Plus/Minus	
Query	7	CCTACTC	TTTTACAAGAT.	ACTTGGCTTTTA	GAAAAACTTGC	ACATTTCGATAGGGAAAGG	66
Sbjct	4785	6 CCTACAC	TTTTACAAGAT.	ACTTGGCTTTTA	 IGAAAAACTTGC	 ACATTTTGATAGAGAGAGA	47797
Query	67	ATACCAG	AAAGAGTTGTG	CACGCTAAAGGA	AGTGCTGCATA	CGGCGAATTAACAATTACT	126
Sbjct	4779	96 ATCCCAG	AGAGAGTGGTG	CACGCTAAAGGA	AGTGCAGCATA	II IIIIIIIIIIIIII TGGTGAATTAACAATTACA	47737
Query	127	AATGATA	ттасаааата	143			
Sbjct	4773	 36 AATGATA	 TTACTCAATA	47720			
(gene		complement /gene="kat; /locus_tag	(<1>137) A" ="HH_0043"	-		
(CDS		/old_locus complement /gene="kat]	_tag="HH0043 (<1>137) A"	3"		
			/locus_tag	="HH_0043"			
			/old_locus	_tag="HH0043	3"		
			/EC_number	="1.11.1.6"			
			/transl tal	hle=11			
			/product="	catalase"			
			/protein i	d="AAP76640.	1"		
			/translati	on="MSKKFTT	TGTPLGDNONS	SITAGKKGPTLLODTWLLEKI	LAHFDR
			ERIPERVVHA	KGSAAYGELTII	NDITQYTKAEI	FNKVGKKTKAFLRFSVVAG	ERGAAD
			AERDVRGFAL	KLYTNEGNWDIV	GNNTPVFFIKI	AIKFPDFIHTQKRDPKTNM	KSPTAM
			WDFWSLHPES	LHQVTILMSDRG	JIPRSYREMHGE	GSHTYSFINAKNERFWVKF	HFVCLQ
			GIHNLTNKES	EAVIAKDRESHÇ	QKDLFENIEKGN	IFPKWRFCIQVMSEKEAENYH	RFNPFD
			LTKVWSHKDY	PLIEVGILELNE	(NPENYFAEVE)	AAFNPANIVPGVGYSPDKVI	LQGRLF
			SYGDTQRYRL	GINHTQLPVNAB	PIVPVNNTHRDO	FMQQGQFGDRRNYEPSYFNI	DYVEDK
			NALEPPLFVQ	EGDVMYKYDHRE	YEDDYFVQAGI	DLYRLMTAEQKEALCQNIKE:	SMEGVP
			DEIKKRQLEH	FKKADKAYGKRV	AELLGL"		
ORIGI	N						
	1	tattgagtaa	tatcatttgt	aattgttaat	tcaccatatg	ctgcacttcc tttagcgt	gc
	61	accactctct	ctgggattct	ctctctatca	aaatgtgcaa	gtttttctaa aagccaagt	ta
	121	tcttgtaaaa	gtgtagg				
11							

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	<u>AE017125.1</u>
Description	Helicobacter hepaticus ATCC 51449, complete genome
Molecule type	nucleic acid
Query Length	934935
Other reports	8



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

< Edit Search	Save Search Search Summary 🕶	How to read this report? BLAST Help Videos Back to Traditional Results Page								
Job Title	gb AC097711.2	Filter Results								
RID	KJ2J1YCV01R Search expires on 08-17 18:28 pm Download All 🗸									
Program	BLASTN 😧 Citation 🗸	Organism only top 20 will appear exclude								
Database	SRA <u>See details</u> Y	Type common name, binomial, taxid or group name								
Query ID	AC097711.2 + Add organism									
Description	Homo sapiens BAC clone RP11-162K6 from 4, complete sequence	Percent Identity E value Query Coverage								
Molecule type	nucleic acid									
Query Length	114657									
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								
SRX772475	2	279 279 0% 1e-70 100.00% <u>SRA:SRR11085797.11044608.2</u>								
SRX772475	2	279 279 0% 1e-70 100.00% <u>SRA:SRR11085797.11044608.1</u>								

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 (Biologica) ✔
Program	BLASTN 🕄 Citation 🗸
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



Dese	riptions	Graphic Summary	Alignments	Taxonomy							
Seq	Sequences producing significant alignments Download × Manage Columns × Show 100 ×										100 🗸 🔞
Select all 11 sequences selected								<u>nk G</u>	raphics	Distan	ce tree of results
			Desc	cription		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	Homo sapier	IS BAC clone RP11-162K6 from	4. complete sequence			279	279	100%	2e-71	100.00%	AC097711.2
	Canis lupus	amiliaris breed Labrador retrieve	r chromosome 32a			274	274	100%	9e-70	99.34%	CP050598.1
	Canis lupus	amiliaris breed Labrador retrieve	r chromosome 32b			274	274	100%	9e-70	99.34%	CP050634.1
	Aquila chrys	aetos chrysaetos genome assen	nbly, chromosome: 1			257	257	100%	9e-65	97.35%	LR606181.1
	Apteryx aust	ralis mantelli genome assembly	AptMant0. scaffold scaf	fold176		257	257	100%	9e-65	97.35%	LK064748.1
	Erithacus rul	ecula genome assembly, chrom	osome: 5			252	252	100%	4e-63	96.69%	LR812107.1
	Anas platyrh	<u>ynchos genome assembly, chror</u>	mosome: 4			252	252	100%	4e-63	96.69%	LS423614.1
	<u>Streptopelia</u>	turtur genome assembly, chromo	osome: 4			246	246	100%	2e-61	96.03%	LR594554.1
	Mus musculi	us BAC clone RP24-204J10 from	5. complete sequence			243	243	98%	3e-60	95.97%	AC121929.2
	Sciurus caro	linensis genome assembly, chro	mosome: 15			204	204	78%	1e-48	97.48%	LR738605.1
	PREDICTED	Meleagris gallopavo uncharacte	rized LOC104910685 (L	<u>OC104910685), mR</u>	A	121	121	45%	1e-23	98.53%	XM_019615117.2

✓ Edit Search Save Search Search Summary ✓

How to read this report?
 Description
 BLAST Help Videos
 Description
 Descri

Job Title	gnl SRA SRR11085797.11044608.1 11044608	Filter Results							
RID	KJ2FP0KE014 Search expires on 08-17 18:27 pm Download All ↔								
Results for	1:lcl Query_54298 gnl SRA SRR11085797.11044608.1 11044608 (Biological 🗸	Organism only top 20 will appear							
Program	BLASTN ? <u>Citation</u> ~	Type common name, binomial, taxid or group name							
Database	nt See details 🗸	+ Add organism							
Query ID	lcl Query_54298	Percent Identity E value Query Covera	ge						
Description	gnl SRA SRR11085797.11044608.1 11044608 (Biological)	to to to							
Molecule type	dna								
Query Length	151	Filter	Reset						
Other reports	Distance tree of results 💡								
Descriptions	Graphic Summary Alignments Taxonomy								
Sequences p	producing significant alignments	Download Y Manage Columns Y Show 10	10 🗸 🔞						
🗹 select all	4 sequences selected	GenBank Graphics Distance to	ree of results						
	Description	Max Total Query E Per. Score Score Cover value Ident	Accession						
Homo sapie	ens 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 muscu	lus 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 thi	s repor	t? 🕨	BLAST	í Help V	lideos 🖁	Back to Traditional Res	ults Page	
Job Title	ref NW_015351248.1		Filter Results								
RID											
Program	BLASTN 😧 Citation 💙	organism only top 20 will appear exclud									
Database	SRA <u>See details</u> ✓	Type common name, binomial, taxid or group name									
Query ID	+ Add organis	<u>m</u>									
Description	Marmota marmota marmo	Percent Identit	y	E	value			Query Coverage			
Molecule type	Iolecule type dna				Γ		to		to		
Query Length	19578880										
Other reports	Distance tree of results MS	SA viewer 🔞	Filter Reset								
Descriptions	Graphic Summary	Alignments									
Sequences p	producing significant al	ignments	D	ownlo	ad ~	N	lanage	Column	s ~ Show 100 🗸	0	
🔽 select all	100 sequences selected							Gra	phics Distance tree of	results	
		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession				
SRX77247	<u>52</u>			279	279	0%	2e-68	100.00%	SRA:SRR11085797.704	1148.2	
SRX77247	52			278	278	0%	9e-68	100.00%	SRA:SRR11085797.4666	606.2	
SRX77247	<u>52</u>			276	276	0%	3e-67	100.00%	SRA:SRR11085797.8742	2622.2	

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

Seq	uences producing significant alignments D	ownload ~	Mai	nage C	olumr	ns ~	Show	100 🗸 😗
•	select all 45 sequences selected		G	ienBan	<u>k Gr</u>	aphics	Distanc	e tree of results
	Description		Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
\checkmark	Canis lupus familiaris breed Labrador retriever chromosome 32a		279	279	100%	2e-71	100.00%	CP050598.1
	Canis lupus familiaris breed Labrador retriever chromosome 32b		279	279	100%	2e-71	100.00%	CP050634.1
	Homo sapiens VISTA enhancer hs712 (LOC110120752) on chromosome 4		279	279	100%	2e-71	100.00%	<u>NG_053377.1</u>
≤	Homo sapiens BAC clone RP11-476H13 from 4_ complete sequence		279	279	100%	2e-71	100.00%	AC024192.6
	Aquila chrysaetos chrysaetos genome assembly, chromosome: 1		274	274	100%	9e-70	99.34%	LR606181.1
	Streptopelia turtur genome assembly, chromosome: 4		274	274	100%	9e-70	99.34%	LR594554.1
	PREDICTED: Cyanistes caeruleus uncharacterized LOC111928864 (LOC111928864), ncRNA		274	274	100%	9e-70	99.34%	XR_002864354.1
	Apteryx australis mantelli genome assembly AptMant0, scaffold scaffold564		274	274	100%	9e-70	99.34%	LK065221.1
≤	Anas platyrhynchos genome assembly, chromosome: 4		274	274	100%	9e-70	99.34%	LS423614.1

Job Title	bb Title 2 sequences (gnl SRA SRR11085797.4967337.1									
RID	KJ8SKF2E014 Search expires on 08-17 20:14 pm Download All									
Results for	2:lcl Query_37435 gnl SRA SRR11085797.4967337.2 4967337 (Biological)(: 🗸									
Program	BLASTN ? <u>Citation</u> ~									
Database	nt <u>See details</u> Y									
Query ID	lcl 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	vill appear	exclude
Type common name	, binomial, taxid or gro	up name
+ Add organism		
Percent Identity	E value	Query Coverage
to	to	to
		Filter Reset

escription	is Graphic Summary	Augminents	тахопотту									
Sequences producing significant alignments Download × Manage Columns × Show 100 • @												
Select all 58 sequences selected GenBank Graphics Distance tree of results												
		De	scription			Max Score	Total Score	Query Cover	E value	Per. Ident	Accession	
Canis I	upus familiaris breed Labrador retriev	er chromosome 32a				279	279	100%	2e-71	100.00%	CP050598.1	
Canis I	upus familiaris breed Labrador retriev	er chromosome 32b				279	279	100%	2e-71	100.00%	CP050634.1	
Rhesus	Macaque BAC CH250-499F11 () co	mplete sequence				274	274	100%	9e-70	99.34%	AC204215.5	
Homo s	apiens BAC clone RP11-678H22 from	m 4 <u>, complete sequence</u>				274	274	100%	9e-70	99.34%	AC096766.3	
Mus mi	usculus chromosome 5, clone RP24-	315H14, complete seque	nce			263	263	100%	2e-66	98.01%	AC105976.13	
Chryse	mys picta isolate 4965chr ultra cons	erved element locus chr4	<u>11164 genomic seq</u>	uence		257	257	100%	9e-65	97.35%	JQ873778.1	
Alligato	r mississippiensis isolate 333all ultra	a conserved element locu	s chr4_11164_genom	<u>iic sequence</u>		257	257	100%	9e-65	97.35%	JQ869146.1	
Apteryz	australis mantelli genome assembly	AptMant0. scaffold scaf	fold564			241	241	100%	9e-60	95.36%	LK065221.1	
Erithac	us rubecula genome assembly, chror	mosome: 5				231	231	82%	6e-57	100.00%	LR812107.1	
Strepto	pelia turtur genome assembly_chron	nosome: 4				231	231	82%	6e-57	100.00%	LR594554.1	
Aquila	chrysaetos chrysaetos genome asse	mbly, chromosome: 1				226	226	82%	3e-55	99.20%	LR606181.1	
	equence select a Canis In Canis In Canis In Canis In Rhesus Homo s Homo s Homo s Mus mi Chryse Alligato Apteryz Erithac Strepto	capite Summary coupies coupie	Capital Summary Augments Capital Summary Augments Select all 58 sequences selected Canis lupus familiaris breed Labrador retriever chromosome 32a Canis lupus familiaris breed Labrador retriever chromosome 32a Canis lupus familiaris breed Labrador retriever chromosome 32b Rhesus Macaque BAC Cht250499F11() complete sequence Homo sagiens BAC clone RP11-678H22 form 4, complete sequence Mus musculus chromosome 5, clone RP24-316H14, complete sequence Chrysenys picta isolate 495cch ultra conserved element locu Chrysenys picta isolate 495cch ultra conserved element locu Apteryx australis martelli genome assembly. chromosome.5 Streptopelia tutur genome assembly, chromosome.4 Aquila chrysaetos chrysaetos genome assembly, chromosome.1	Canis lupus familiaris breed Labrador retriever chromosome 32a Canis lupus familiaris breed Labrador retriever chromosome 32a Canis lupus familiaris breed Labrador retriever chromosome 32a Canis lupus familiaris breed Labrador retriever chromosome 32b Rhesus Macaque BAC Ch250-499F11 () complete sequence Homo sapiens BAC clone RP11-678H22 from 4, complete sequence Mus musculus chromosome 5, clone RP24-316H14, complete sequence Chrysenys picta isolate 4965chr ultra conserved element locus ch4_11164 genomic seq Alligator mississippiensis isolate 333all ultra conserved element locus ch4_11164 genom Apteryx australis mantelli genome assembly.chromosome: 5 Streptopelia tutur genome assembly.chromosome: 4 Aquila chrysaetos chrysaetos centome assembly.chromosome: 1	Canis lupus familiaris breed Labrador retriever chromosome 32a Canis lupus familiaris breed Labrador retriever chromosome 32b Rhesus Macaque BAC CH250-499F11 () complete sequence Homo sagiens BAC clone RP11-678H22 from 4, complete sequence Chrysenys picta isolate 395chr ultra conserved element locus chrl 11164 genomic sequence Alligator mississippiensis isolate 333al ultra conserved element locus chrl 11164 genomic sequence Alligator mississippiensis isolate 333al ultra conserved element locus chrl 11164 genomic sequence Atteryx australis mantelli genome assembly AptMant0. scaffold scaffold564 Erithacus rubecula genome assembly.chromosome.1	Sector producing significant alignments Download equences producing significant alignments Download select all 58 sequences selected association Canis lupus familiaris bread Labrador retriever chromosome 32a Canis lupus familiaris bread Labrador retriever chromosome 32a Canis lupus familiaris bread Labrador retriever chromosome 32a Canis lupus familiaris bread Labrador retriever chromosome 32b Rhesus Macaque BAC CH250-499F11 () complete sequence Homo sagiens BAC choe RP11-678H22 from 4, complete sequence Mus musculus chromosome 5. Chrysenyx picta isolate 333all ultra conserved element locus chrl 11164 genomic sequence Alligator mississippiensis isolate 333all ultra conserved element locus chrl 11164 genomic sequence Apting xustralis mantelli genome assembly. AptMant0. scaffold 564 Erithacus rubecula genome assembly. AptMant0. scaffold 564 Erithacus rubecula genome assembly. chromosome 1	Caribitions Chepine Summary Pagiments Pownload Manage equences producing significant alignments Download Manage select all 58 sequences selected Gen	Sector bound Complexe dummary Augments Taxing Sequences producing significant alignments Download Manage Col Select all 58 sequences selected GenBank Description Max Total Canis lupus familiaris breed Labrador retriever chromosome 32a 279 279 Canis lupus familiaris breed Labrador retriever chromosome 32b 279 279 Canis lupus familiaris breed Labrador retriever chromosome 32b 274 274 Canis lupus familiaris breed Labrador retriever chromosome 32b 274 274 Chromes BAC clone RP11-578H22 from 4, complete sequence 274 274 Mus musculus chromosome 5, clone RP24-315H14, complete sequence 263 263 Chrysenys picta isolate 305Chr ultra conserved element locus chd; 11164 genomic sequence 257 257 Alligator mississippiensis isolate 333all ultra conserved element locus chd; 11164 genomic sequence 257 257 Attention subscula genome assembly. AdMant0, scaffold 564 241 241 241 Erithacus rubecula genome assembly. AdMant0, scaffold 564 241 241 241 Erithacus rubecula genome assembly. chromosome: 5 231 231 231 231 </th <th>Canip Loganite Summary Regimients Download Manage Columns equences producing significant alignments Download Manage Columns select all 58 sequences selected GenBank Grap Description Max Total Ouery Canis lupus familiaris breed Labrador retriever chromosome 32a 279 279 100% Canis lupus familiaris breed Labrador retriever chromosome 32b 279 274 100% Rhesus Macaque BAC Ch250-499E11 () complete sequence 274 274 100% Homo sapiens BAC clone RP11-678H22 from 4, complete sequence 274 274 100% Mus musculus chromosome 5, clone RP24-315H14, complete sequence 277 277 100% Alligator mississippiensis isolate 333all ultra conserved element locus chd_11164 genomic sequence 257 257 100% Atteryx australis mantelli genome assembly. chromosome 5 231 231 82% Streptopelia tutur genome assembly. chromosome: 1 226 226 82%</th> <th>Sector of the summary angineering function in the summary indexing and the sector of the sector o</th> <th>Download Manage Columns V Marage Columns V Mov Identify equences producing significant alignments Download Marage Columns V Mov Identify Identify select all 58 sequences selected GenBank Graphic Sufficient Distance I Canis lupus familiaris breed Labrador retriever chromosome 32a 279 279 100% 2e-71 100.0% Canis lupus familiaris breed Labrador retriever chromosome 32b 279 279 100% 2e-71 100.0% Canis lupus familiaris breed Labrador retriever chromosome 32b 279 274 100 9e-70 99.34% Homo sagiens BAC clone RP11-678H22 from 4, complete sequence 274 274 100% 9e-70 99.34% Mus musculus chromosome 5, clone RP24-316H14, complete sequence 274 274 10% 9e-70 99.34% Alligator mississippiensis isolate 333all ultra conserved element locus chrd_ 11164 genomic sequence 277 277 10% 9e-50 9.35% Atteryx australis mantelli genome assembly. AdMant0, scaffold 664 241 241 10% 9e-50 9.35% Streptopelia tutur genome assembly.</th>	Canip Loganite Summary Regimients Download Manage Columns equences producing significant alignments Download Manage Columns select all 58 sequences selected GenBank Grap Description Max Total Ouery Canis lupus familiaris breed Labrador retriever chromosome 32a 279 279 100% Canis lupus familiaris breed Labrador retriever chromosome 32b 279 274 100% Rhesus Macaque BAC Ch250-499E11 () complete sequence 274 274 100% Homo sapiens BAC clone RP11-678H22 from 4, complete sequence 274 274 100% Mus musculus chromosome 5, clone RP24-315H14, complete sequence 277 277 100% Alligator mississippiensis isolate 333all ultra conserved element locus chd_11164 genomic sequence 257 257 100% Atteryx australis mantelli genome assembly. chromosome 5 231 231 82% Streptopelia tutur genome assembly. chromosome: 1 226 226 82%	Sector of the summary angineering function in the summary indexing and the sector of the sector o	Download Manage Columns V Marage Columns V Mov Identify equences producing significant alignments Download Marage Columns V Mov Identify Identify select all 58 sequences selected GenBank Graphic Sufficient Distance I Canis lupus familiaris breed Labrador retriever chromosome 32a 279 279 100% 2e-71 100.0% Canis lupus familiaris breed Labrador retriever chromosome 32b 279 279 100% 2e-71 100.0% Canis lupus familiaris breed Labrador retriever chromosome 32b 279 274 100 9e-70 99.34% Homo sagiens BAC clone RP11-678H22 from 4, complete sequence 274 274 100% 9e-70 99.34% Mus musculus chromosome 5, clone RP24-316H14, complete sequence 274 274 10% 9e-70 99.34% Alligator mississippiensis isolate 333all ultra conserved element locus chrd_ 11164 genomic sequence 277 277 10% 9e-50 9.35% Atteryx australis mantelli genome assembly. AdMant0, scaffold 664 241 241 10% 9e-50 9.35% Streptopelia tutur genome assembly.	

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)												
>gnl SRA SRR11085797.6341838.1 6341838 (Biological) CGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGC CGGGCGTGATGGCGGGCGCGCGTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGG CGTGAACCCGGGAGGCGGAGCNTGCAGTG												
>gnl SRA SRR11085797.6341838.2 6341838 (Biological) CTCACTGCAAGCTCCGGCCTCCCGGGTTCACGCCATTCTCCTGCCTCAGCCTCCCGAGTAG CTGGGACTACAGGCGCCCGCCATCACGCCCGGCTAATTTTTTGTATTTTTAGTAGAGAGGC GGGTTTCACCGTGTTAGCCAGGATGGTCTCG												
Description gnl SRA SRR11085797.6341838.2.6341838 (Biological) to to to												
Molecule type ona Filter Res												
Other reports Distance tree of results 😧												
Descriptions Graphic Summary Alignments Taxonomy												
Sequences producing significant alignments Download × Manage Columns × Show 100 • 9												
Select all 100 sequences selected		<u>GenBank</u>	<u>Gra</u>	<u>phics</u>	Distance	tree of results						
Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession						
Pongo abelii chromosome 10 clone CH276-326B4, complete sequence	278	1689	100%	7e-71	100.00%	AC270645.1						
Pongo abelii chromosome 10 clone CH276-12G11, complete sequence	278	1379	100%	7e-71	100.00%	AC270533.1						
Pongo abelii chromosome 10 clone CH276-5H9, complete sequence	278	1253	100%	7e-71	100.00%	AC270518.1						
Homo sapiens beta-1.3-galactosyltransferase 1 (B3GALT1), mRNA	274	274	100%	9e-70	99.34%	<u>NM_020981.4</u>						
Homo sapiens chromosome 1 clone VMRC53-455P10, complete sequence	274	3094	100%	9e-70	99.34%	AC278561.1						

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

2	Human endogenous retrovirus H HERV-H/env60 proviral copy, clone 734E12	252	25	2	100%	4e-63 9	16.69%	AJ289710.2
	Synthetic human HSC3N1 Alu sequence	252	25	2	100%	4e-63 9	16.69%	<u>U02043.1</u>
_	Synthetic construct, complete sequence	250	363	35	100%	1e-62 9	16.69%	JN255744.1
	Human artificial chromosome vector 21HAC4 DNA, isolated from the long arm, clone: YAC/BAC#26-2	250	196	64	100%	1e-62 9	16.69%	AB553834.1
_	Human ORFeome Gateway entry vector pENTR223-MGC2752, complete sequence	246	24	6	100%	2e-61 9	16.05%	LT735229.1
	Expression vector pUMLIEP DNA, complete sequence	246	24	6	99%	2e-61 9	16.05%	LC175306.1
	Synthetic construct Homo sapiens clone ccsbBroadEn_10246 MGC2752 gene, encodes complete protein	246	24	6	100%	2e-61 9	16.05%	KJ900852.1
	HIV-1 isolate HK_JIDLNBL_S071 from Switzerland nonfunctional gag.protein (gag) gene, complete sequence; and nonfunctional	244	107	72	100%	7e-61 9	16.00%	MT154980.1
	Cloning vector pSuper_7SL_AluAA 7SL enhancer and AluYa5 repeat element sequence	241	24	1	100%	9e-60 9	15.36%	EU092258.1
	Cloning vector pSuper_7SL_AluA 7SL enhancer and AluYa5 repeat element sequence	241	24	1	100%	9e-60 9	15.36%	EU092257.1
	Synthetic construct clone AluAU SRP promoter region and Alu repeat element sequence	241	24	1	100%	9e-60 9	15.36%	AF458115.1
	Synthetic construct clone AluWD SRP promoter region and Alu repeat element sequence	241	24	1	100%	9e-60 9	15.36%	AF458112.1
	Synthetic construct clone Alut253 SRP promoter region and Alu repeat element sequence	241	24	1	100%	9e-60 9	15.36%	AF458107.1
	Synthetic construct clone Alu+A SRP promoter region and Alu repeat element sequence	241	24	1	100%	9e-60 9	15.36%	<u>AF458106.1</u>
	Desmodus rotundus isolate DRU21DN04 contig68764, whole genome shotgun sequence	108	216	63%	2e-2	87.239	% <u>PE</u> ł	HR01068758.1
	Myotis lucifugus cont2.6286, whole genome shotgun sequence	108	108	55%	2e-2	90.369	% <u>AAP</u>	'E02006287.1
	Artibeus jamaicensis isolate US092 ArtJam_scaffold_27825, whole genome shotgun sequence	104	104	51%	2e-1	90.919	% <u>PV</u> K	R01013927.1
	Macrotus californicus isolate US035 MacCal line 566643, whole genome shotgun sequence	102	102	51%	9e-1	90.799	<u>% vm</u> г	DR010283404.1
	Anoura caudifer isolate US021 AnoCau_scaffold_336054, whole genome shotgun sequence	102	102	61%	9e-1!	86.969	% <u>PV</u> K	KU01163203.1
	Anoura caudifer isolate US021 AnoCau_scaffold_250162, whole genome shotgun sequence	102	102	61%	9e-1	86.969	% <u>PV</u> K	(U01121529.1
	Anoura caudifer isolate US021 AnoCau_scaffold_157416, whole genome shotgun sequence	102	102	61%	9e-1	86.96	% <u>PV</u> K	(U01078866.1
	Anoura caudifer isolate US021 AnoCau_scaffold_136788, whole genome shotgun sequence	102	102	61%	9e-1	86.969	% <u>PV</u> K	(U01068554.1
	Anoura caudifer isolate US021 AnoCau_scaffold_6229, whole genome shotgun sequence	102	102	51%	9e-1	90.799	% <u>PV</u> K	KU01003121.1
	Anoura caudifer isolate US021 AnoCau_scaffold_1146, whole genome shotgun sequence	102	102	71%	9e-1	84.269	% <u>PV</u> k	(U01000576.1
	Artibeus jamaicensis isolate US092 ArtJam_scaffold_590481, whole genome shotgun sequence	102	102	51%	9e-1	90.799	% <u>PV</u> K	R01295479.1
	Artibeus jamaicensis isolate US092 ArtJam_scaffold_272373, whole genome shotgun sequence	102	102	51%	9e-1	90.799	% <u>PV</u> K	(R01136397.1
								I
~	Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000055F_070_arrow_arrow, whole genome shotgun sequence	101	101	51%	8e-19	88.46	% RX	(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_107625, whole genome shotgun sequence	101	101	51%	8e-19	88.46	% <u>AV</u>	VHA01101756.1
	Rhinolophus ferrumequinum isolate US033 RhiFerflattened_line_8799, whole genome shotgun sequence	97.8	186	50%	1e-17	88.16	% <u>VN</u>	IDN01004402.1
	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 ferrumequinum Isolate MPI-CBG mRhiFer1 chromosome 6, whole genome shotgun sequence	97.8	309	50%	1e-17	88.16	% RX	(PC01000086.1
	Rhinolophus ferrumequinum isolate mRhiFer1 scaffold_m29_p_8, whole genome shotgun sequence	97.8	309	50%	1e-17	88.16	% <u>JA</u>	CAGC010000008.1
	Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000003F_100_arrow_arrow, whole genome shotgun sequence	93.3	93.3	49%	1e-16	5 88.00	% <u>R</u>	(PD01006157.1
	Rhinolophus ferrumequinum isolate mRhiFer1 scaffold_m29_p_4, whole genome shotgun sequence	93.3	93.3	49%	1e-16	88.00	% JA	CAGC010000004.1
	Rhinolophus ferrumequinum isolate US033 RhiFer flattened line 6166, whole genome shotgun sequence	90.6	90.6	44%	1e-15	89.55	% <u>v</u> M	IDN01003085.1
	Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000061F_073_arrow_arrow_whole genome shotgun sequence	90.6	90.6	44%	1e-18	89.55	% <u>R</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.

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.

Therefore, the sequencing of BtCoV/RaTG13 can not 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-l5vxt7jqp

[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). https://doi.org/10.1186/s12864-019-5543-2