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
ATCCTGTTTCCTCTCCCGAACATAACCCT

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

>**gnl|SRA|SRR11085797.8.1** 8 *(Biological)*
GTTAGGGTTAGGGTTAGGGTTAGGGTTGGGGTTGGATACAGGATATGGTTAGGGTTAGGG
GTAGGGTCAGGGTTAGGATTGGAAACGAGATAGGTTACGTGATAGGGTTAGCGTTAGGGT
TAGGTTTAGTAATCCGCAACGGCTTAGGGTT

>**gnl|SRA|SRR11085797.8.2** 8 *(Biological)*
CCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTATCCTGTTCCCAACCCTAACC
CTAACCCTAACCCTAACCCTAACACAAAACATAACCCTAACCCCAACCCAAACCCTAACC
CCATCTTTACTCACACCCTAACCCAAAACTC

>**gnl|SRA|SRR11085797.10.1** 10 *(Biological)*
GTTAGGGTTAGGGTTAGCGTTAGGGTTAGAAACAGGATAGGGTTAGGGTTAGGGTTAGGG
TTAGGGTTTTGGTTGGTCACAGTGTTGCGCTAGGCATAGGGATAGGGTTCGCGTTAGGGT
TAGGGTTAGGATTCGGAAGAGCTAGCTAAA

>**gnl|SRA|SRR11085797.10.2** 10 *(Biological)*
GTTCCCAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTTTCCTTTTTCCAACCCTA
ACACTAACCCTAACACTAAACCTAACCCCAACCCTACCACTATACTATATCCGACTCTCA
CGCTAACACTAAACATAAGTAATCACAAATT

>**gnl|SRA|SRR11085797.13.1** 13 *(Biological)*
TAACCCTAACCCTAAGCGTAAACCTAATCCAATCCTGTTCCCAACCCTAACCCTAACCCT
GACCCTAAGCTTTTCCCGACCCGAACGCCGACCCGATCCGCCACCCTAACCGTAACCCGT
TCCCAACCCCTCCTACTGCTCGATCCGCCT

>**gnl|SRA|SRR11085797.13.2** 13 *(Biological)*
GGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTGGGAACAGGATAGG
GTTAGGGTTAGGGTTGGGATCAGGATAGGGATAGGGATAGGGATAGGGATAGGGTTAGGG
TGGGGAACAGGAGAGCGTTAGGCAAGG

>**gnl|SRA|SRR11085797.14.1** 14 *(Biological)*
GGTTAGGGTTAGGGTGGAGAAGAGGATAGGTTTAGGTTTAGGGTTAGGGTTAGGGTGAGG
GTTAGGGAGAGGGTTAGCTACACGATAGGAGTAGGGTAACGATTAGGGTTAGGGTTAGGT
TTGGAAAAAGCATAGGCTATGAGGTACGGT

>**gnl|SRA|SRR11085797.14.2** 14 *(Biological)*
CTGCTTCCAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACC
CTATCCTGTTCCCAACCCTAACCCTAACCCTAACCCCAACCCTAACCCCACCCAAACCCT
AACCCTAACCCCAACCCATACCCCAACCAT

>**gnl|SRA|SRR11085797.15.1** 15 *(Biological)*
TGTTCCCAACCCTAAACCTAAACCTAAGCCGATCCTGTTCCCAACCCTAACCCTAACCCT
ATCCTGTAAACAACCCCCACCCTAAAAACATCCTCGTACAAACCCTAACCCAACCCCCAT
CCCAAACCACATACCCGTCACGAACCCACCC

>**gnl|SRA|SRR11085797.15.2** 15 *(Biological)*
GTTGGGGTTAGGGTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTGGGGTCAGGG
TTGGGACAAGGACGGGTAGGGGGTTAGGGTAGTGCACAGGGTAGGGAGTGGGGTAGGGTT
GGGAACAGAGAAGGGACAGTGGGGGGAGTGG

>**gnl|SRA|SRR11085797.16.1** 16 *(Biological)*
CTAACCCTAACCCTACCCCTATCCTGTTCCTAACCCGAACCCTAACCCTAACCCTAACCC
TAACCCTCACCTGTTCCAGACCGTAATGCTAACCTTAACACTATCCTGTGCGCTACCCCG
ACCCTAACCCTCAGCCGACGCGTCACGCCCG

>**gnl|SRA|SRR11085797.16.2** 16 *(Biological)*
GGTAAGGGTTAGGGTTAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
GTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTTGGAACAGGATAGGGTTAGGGA
TAGGGGTAGAGATAGGGTGAGGTGGTGGAA

>**gnl|SRA|SRR11085797.17.1** 17 *(Biological)*
AACCCTAACCCTAACCCTAACGCTATCATGATCCCATCCCTAACCCTAACCCTAACCCTA
ACCCTAATACTAACCCTACCCTTTTCATCTCCCTTACACTACCCCCAACACGCCACCCAT
CCCCAACCACTATGCATGCACTGTCCTAAAC

>**gnl|SRA|SRR11085797.17.2** 17 *(Biological)*
GGTTAGGGTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGAACAGGATAGGG
GTAGGGTTAGGGCTAGGGTGAGAAACAGGGTAGGGGTAGGGTGAGGATAAGGGATAGGGT
TGGGGTTGGGAACAGAGAAGGGGAAGGGCA

>**gnl|SRA|SRR11085797.18.1** 18 *(Biological)*
CTAACCTGTTCCCAAACTTAAATCCAATCCTAACCCTATCCTGTTCCCAACCCTAACCCT
AAACCTATACCTATCCTGCCCCACACACCGACCCTATACACCACCCTAAACGCAACCCTA
ACCCCATCCTGTTATCGAAGCATACCCCCAC

>**gnl|SRA|SRR11085797.18.2** 18 *(Biological)*
GTTAGGGTTAGGGTTAGGGTTGGTAACAGGATAGGGTTAGGGTTAGGGTTAGTGTTGGGA
ACAGGATAGGGGGAGGGAGAGGGTTAGGGTTGGGAAGAGGATAGGGATAGGGGTAGGGAG
AGGGTTGGGGATAGGGAAGAGAGAGGAG

>**gnl|SRA|SRR11085797.19.1** 19 *(Biological)*
CCTAACCCTAAACCTAACCCTCTCCTGTTTCCAACCATAACCCTAACCCTATCCCTAACC
CTAACCCACTCCTGTTCTTAACACTAACCTTAACTCTGAGCTCATCCCCAAACCTAACCA
TAACCCCACCAGTTCCGATACCATCACCCCC

>**gnl|SRA|SRR11085797.19.2** 19 *(Biological)*
AGGGTTAGGGTTGGAAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGTTTAGGGTTG
GGGACAGGATAGGGTTAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGTGTGGGGTTGG
GATTGGGTATGGGTAGTGGTCAGGGATAGTG

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.

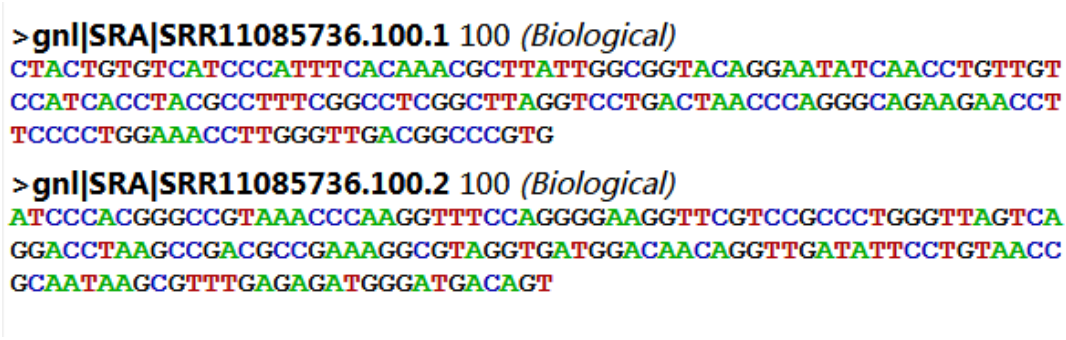


Figure 2: the first 100 reads in SRX7724693 did not show any Telomere-like repeats.

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.

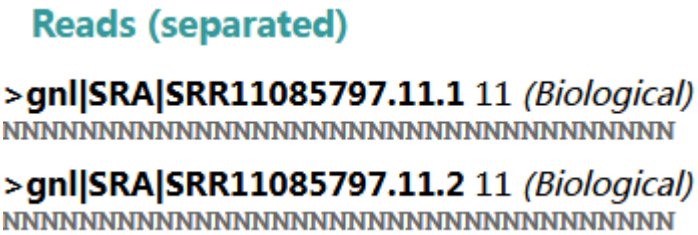


Figure 3: an example of All-N read in SRX7724752.

Anomalous enrichment of non-attributable and low-match data within SRX7724752

no match	42
mouse	1
ape	1
Tea	3
dog	10
Rhi bat	9
Fungus	1
worm	2
cotton	3
Human	4
cat	2
Ribosome	6
fish	8
otter	1
sheep	2
Apicomple	1
squirrel	1
Myo bat	1
rat	1
horse	1
mouse	1
orca	2

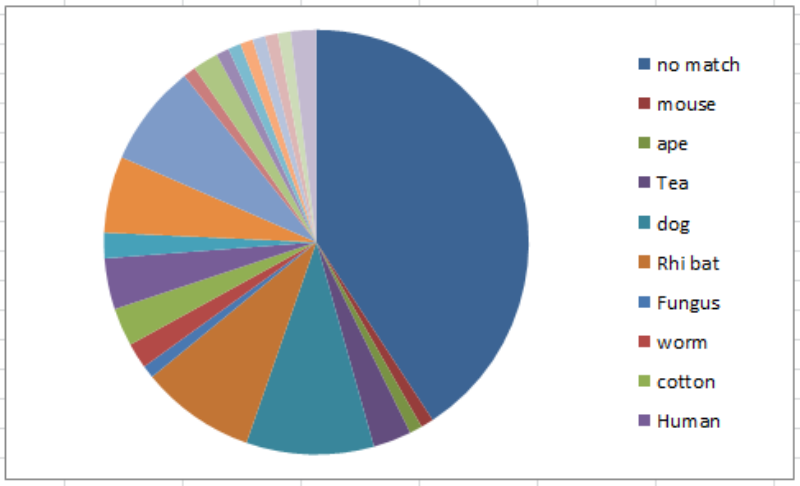
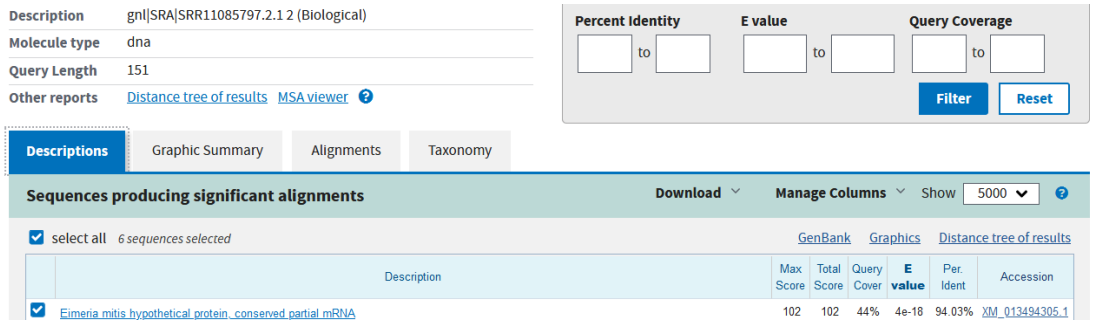


Figure 4: BLAST result of 100 random reads obtained from RaTG13 using BLASTn.

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



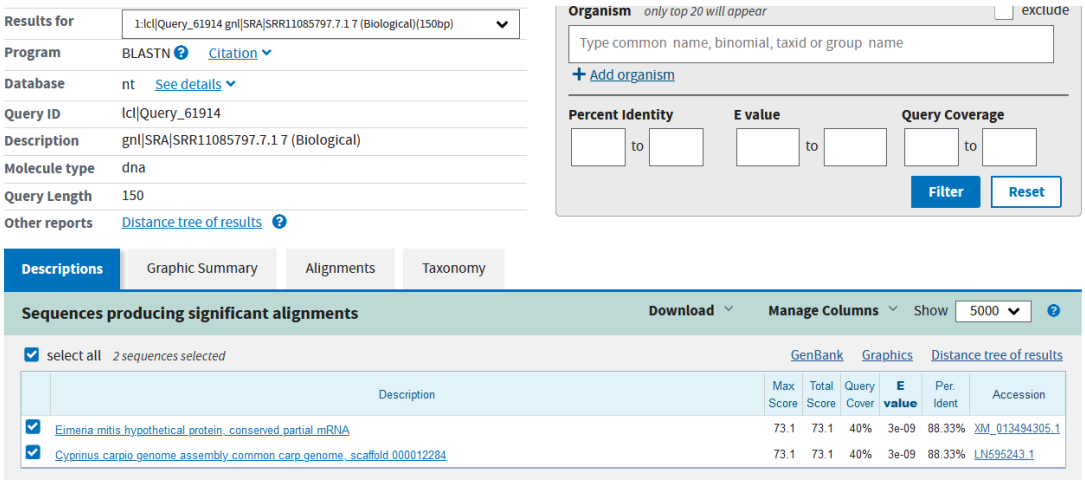


Figure 5: the BLASTn result of the 2 non-repeat and non-PolyN sequences in the first 20 reads of SRX7724752. The rest can not be matched to any known organisms.

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.

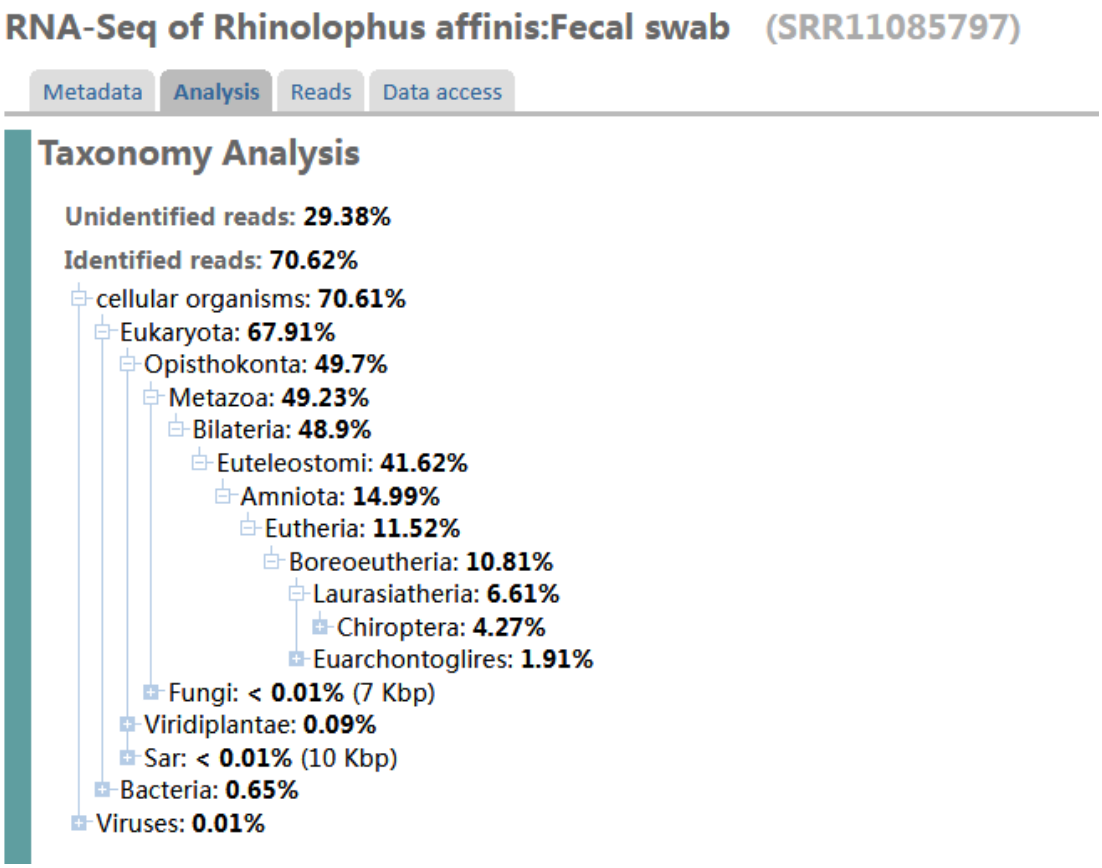


Figure 6: Phylogenetic analysis of SRX7724272.

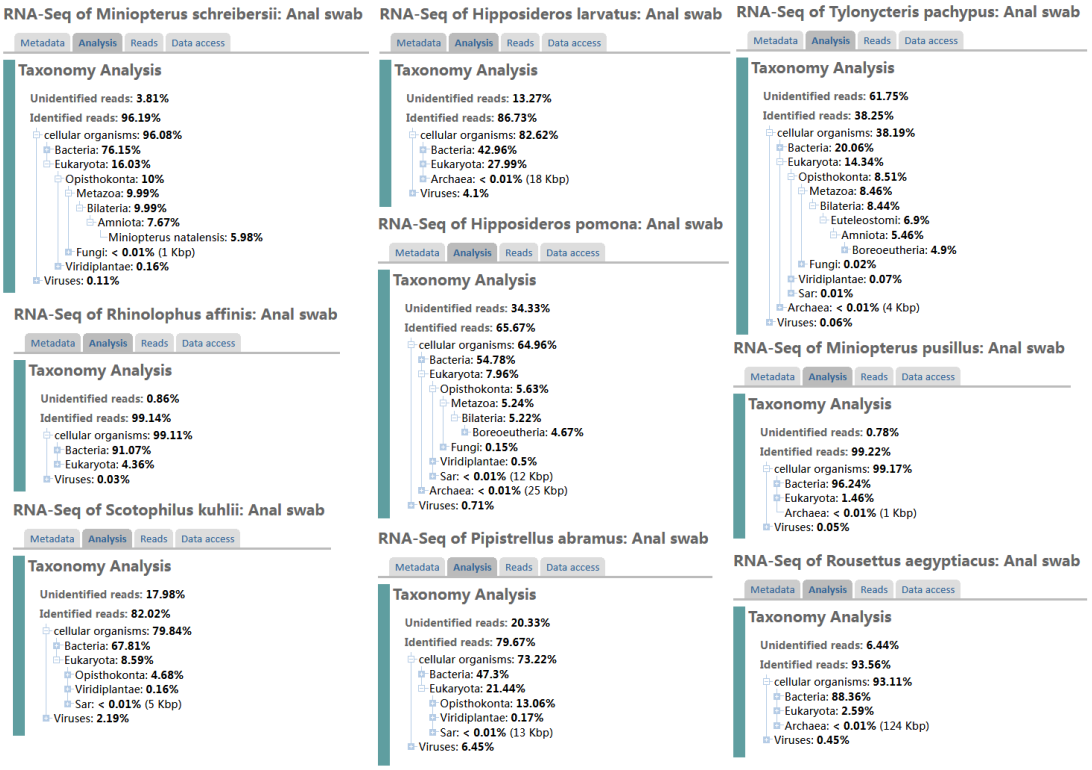


Figure 7: A set of 9 Swabs. The only ones that matches RaTG13 by metadata on Genbank. None of them had more Eukarya-like reads than Bacteria-like reads.

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

discernible reads from bacterial mRNA.

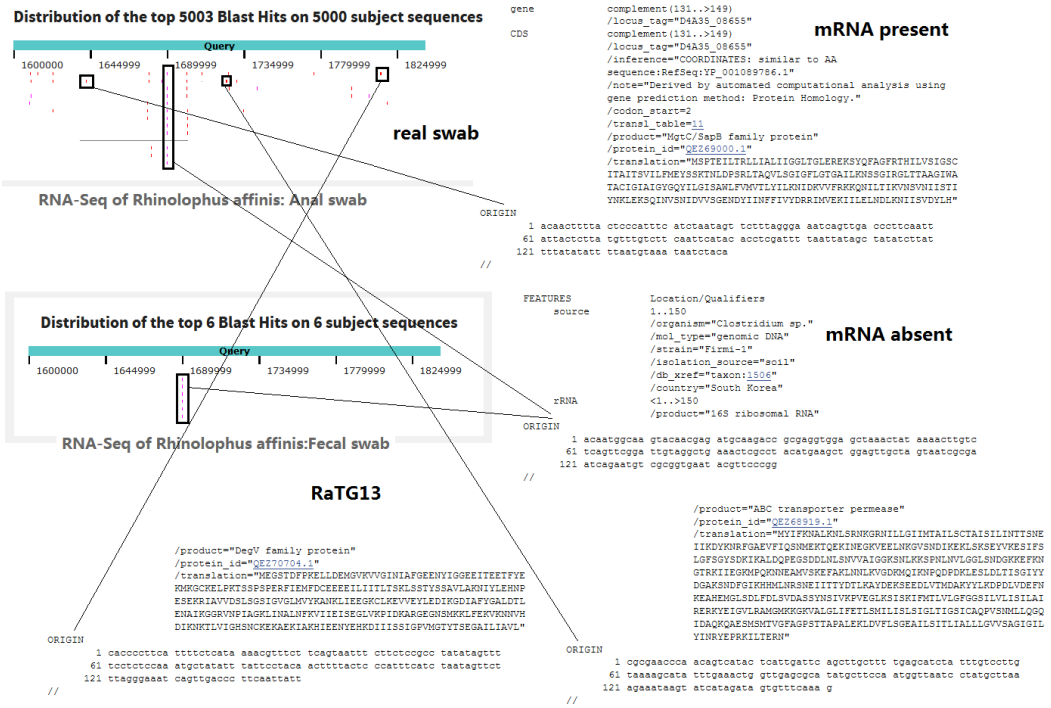


Figure 8: BLAST hits of bacterial non-ribosomal RNA genome on SRX7724272 and another swab from Rhinolophus Affinis under the same library preparation section.

Observation of anomalous and unexpected data within SRX7724752

>gnl|SRA|SRR11085797.11550005.1 11550005 *(Biological)*
GCCCGTATTTAGCCTTAGATGGAGTTTACCACCCGCTTTGGGCTGCATTCCCAAGCAACC
CGACTCCGGGAAGACCCGGGCCCGGCGCGCCGGGGGCCGCTACCGGCCTCACACCGTCCA
CGGGCTGGGCCTCGATCAGAAGGACTTGGGC

>gnl|SRA|SRR11085797.11550005.2 11550005 *(Biological)*
CGGTGGGGCGCGGGACATTTGGCGTACGGAAGACCCACTCCCCGGCGCCGCTCGTGGGGG
CCCAAGTCCTTCTGATCGAGGCCCAGCCCGTGGACGGTGTGAGGCCGGTAGCGGCCCCCG
GCGCGCCGGGCCCGGGTCTTCCCGGAGTCGG

Description

gnl|SRA|SRR11085797.11550005.1 11550005 (B ...

Molecule type

dna

Query Length

151

Other reports

Distance tree of results MSA viewer

Descriptions | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100

select all *100 sequences selected* GenBank Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
PREDICTED: Phyllostomus discolor 28S ribosomal RNA (LOC114512504), rRNA	279	279	100%	2e-71	100.00%	XR_003685809.1
PREDICTED: Phyllostomus discolor basic proline-rich protein-like (LOC114512442), mRNA	279	279	100%	2e-71	100.00%	XM_028531404.1
Homo sapiens lncAB370.3 lncRNA gene, complete sequence	274	274	100%	9e-70	99.34%	MK280359.1
Homo sapiens lncAB366.1 lncRNA gene, complete sequence	274	274	100%	9e-70	99.34%	MK280356.1
Felis catus Senzu DNA, chromosome: E1, American Shorthair breed	274	1372	100%	9e-70	99.34%	AP023165.1
PREDICTED: Marmota flaviventris 28S ribosomal RNA (LOC117794687), rRNA	274	274	100%	9e-70	99.34%	XR_004618536.1
PREDICTED: Arvicanthis niloticus 28S ribosomal RNA (LOC117704856), rRNA	274	274	100%	9e-70	99.34%	XR_004606369.1
PREDICTED: Arvicanthis niloticus 28S ribosomal RNA (LOC117704855), rRNA	274	274	100%	9e-70	99.34%	XR_004606368.1

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

Molecule type dna

Query Length 151

Other reports Distance tree of results MSA viewer

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

Descriptions | Graphic Summary | Alignments

Sequences producing significant alignments Download Manage Columns Show 100

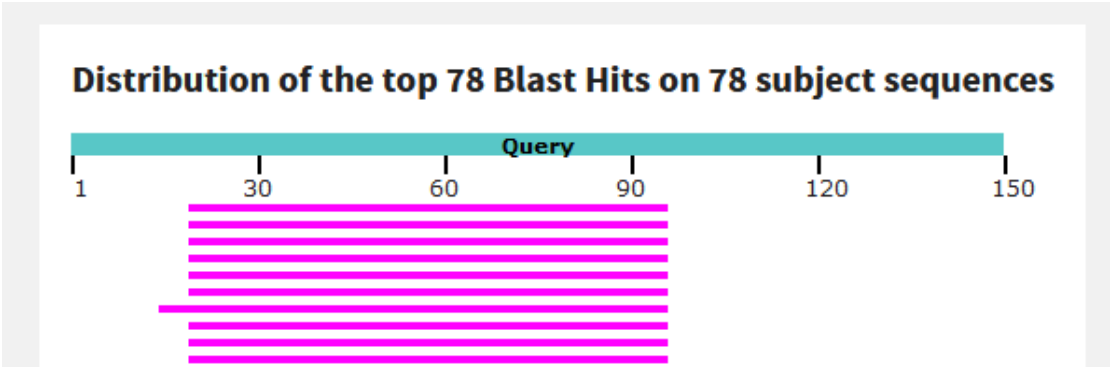
select all *100 sequences selected* Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11601488.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11599592.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11598985.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11598757.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11588533.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11585765.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11585419.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11583582.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11582051.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11580259.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11579956.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11579609.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11573984.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11570618.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11568464.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11567362.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11566985.1
SRX7724752	279	279	100%	2e-73	100.00%	SRA:SRR11085797.11558788.1

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)*

TTTGTTTTGTTTTTTATAATTTATTTTTAAAATTTATTGGGGTGACAATTGTTAGTAAAA
TTACATAGATTTCAGGTGTACAATTCTGTATTACATGTGGACGGTCCAGCCGCCACGAGT
TCAACGTTTTACATGAAAGGGGGTGTGGGA



Download GenBank Graphics

Rhinolophus ferrumequinum clone VMRC7-331J24, complete sequence

Sequence ID: AC150242.3 **Length:** 120894 **Number of Matches:** 1

Range 1: 32707 to 32786 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
122 bits(66)	3e-24	76/80(95%)	3/80(3%)	Plus/Plus

```
Query  20     tttatttttaaa---atttattGGGGTGACAATTGTTAGTAAAATTACATAGATTTCAGG  76
              ||| ||||||||    ||||||||||||||||||||||||||||||||||||||||||
Sbjct  32707  TTTTTTTTTAAATTTATTTATTGGGGTGACAATTGTTAGTAAAATTACATAGATTTCAGG  32766

Query  77     TGTACAATTCTGTATTACAT  96
              ||||||||||||||||||||
Sbjct  32767  TGTACAATTCTGTATTACAT  32786
```

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.

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

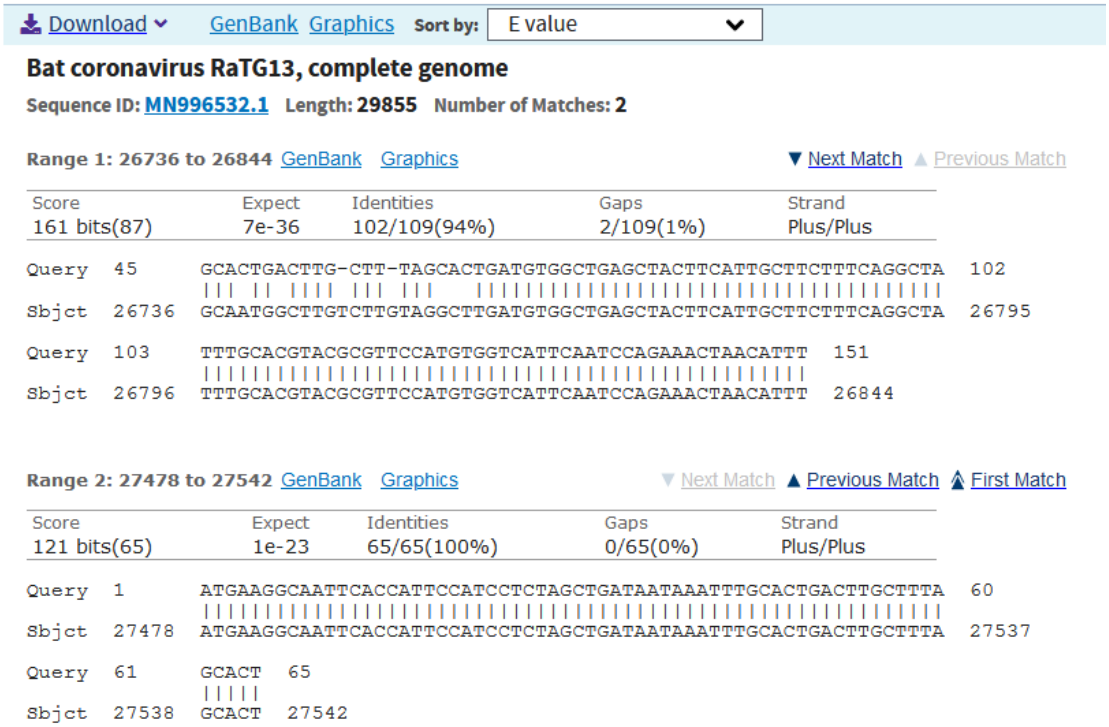


Figure 12: an anomalous fusion of two non-canonical regions of the RaTG13 genome. The fusion again happens on a T.

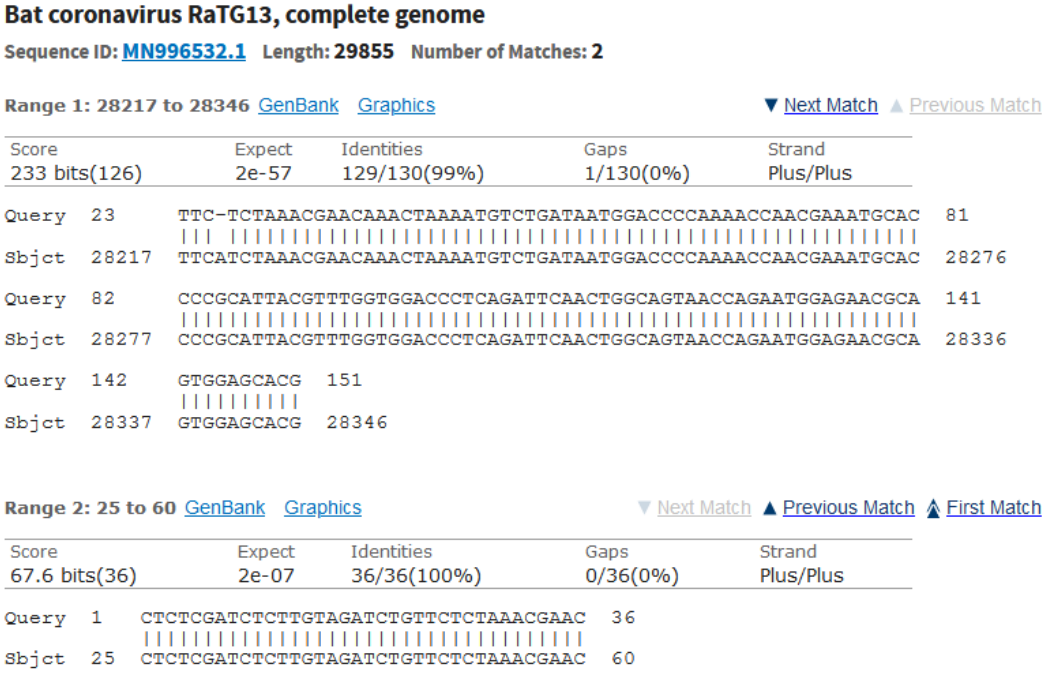


Figure 13: the only canonical sgRNA-like read* in SRX7724752.

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)*

CATCAAACTGAGGTTTCAGCAAGGCAAAGATAGCCAGCAACAAAACAAAAAGGCATCCTA
CTGAATGGAAGCAGATAATTGCCAATAGTACATCAGTAAGGAGTTAATATTAAGAATTAG
TTTTTAAAAAGCTCTATATGATGTCAGAAAT

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

GTTTTCACTTGCATTTCTCTAATAATTAGTGATGTTGAGCATCTTTTCATATGTCTATTG
GCCATCTGTATGTCGTCTTTGGAGAAATGTCTATTCAGATTTCTGCCCAATTTTTAATTG
GCTTGTTTGTTTTTTGTTTTTGAATTGAGTT

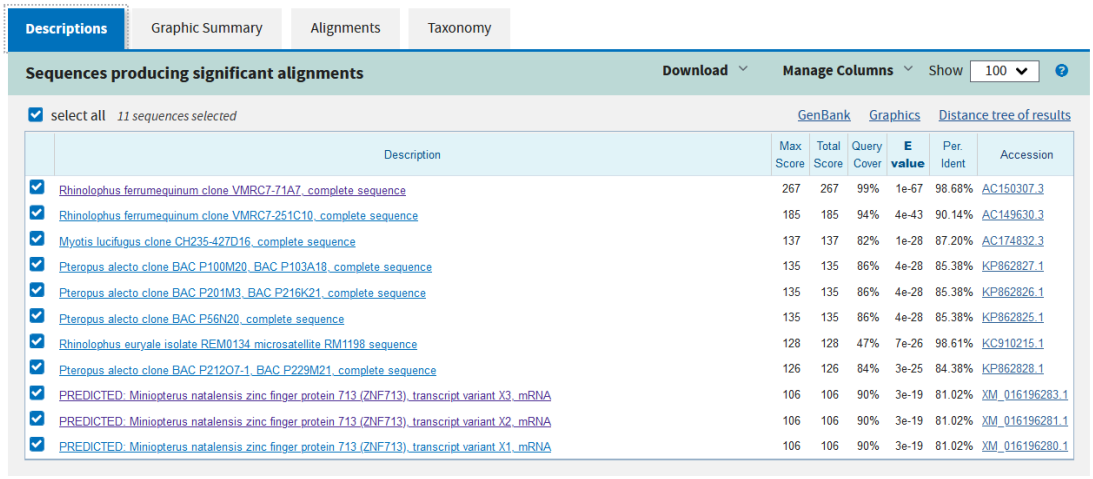


Figure 14: a read from SRX7724752 which have higher coverage on the clone than on the corresponding mRNA. E.g. the read contained nucleotide sequences that were not supposed to be transcribed in actual cells/bats.

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

Date	RaTG13		SARS-CoV-2		Sequence	Blast Archive	Name
	Identity	Query Cover	Identity	Query Cover			
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/Il96Y	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

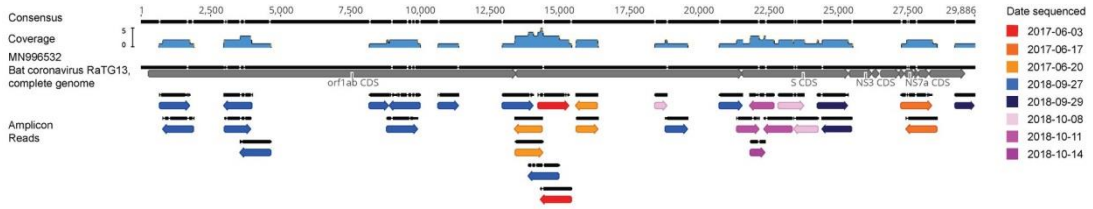


Figure 15: A complete analysis [4] of all Amplicon sequences in SRX8357956. Including the location of these amplicons and the similarity of such amplicon to the RaTG13 and SARS-CoV-2 genome.

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.

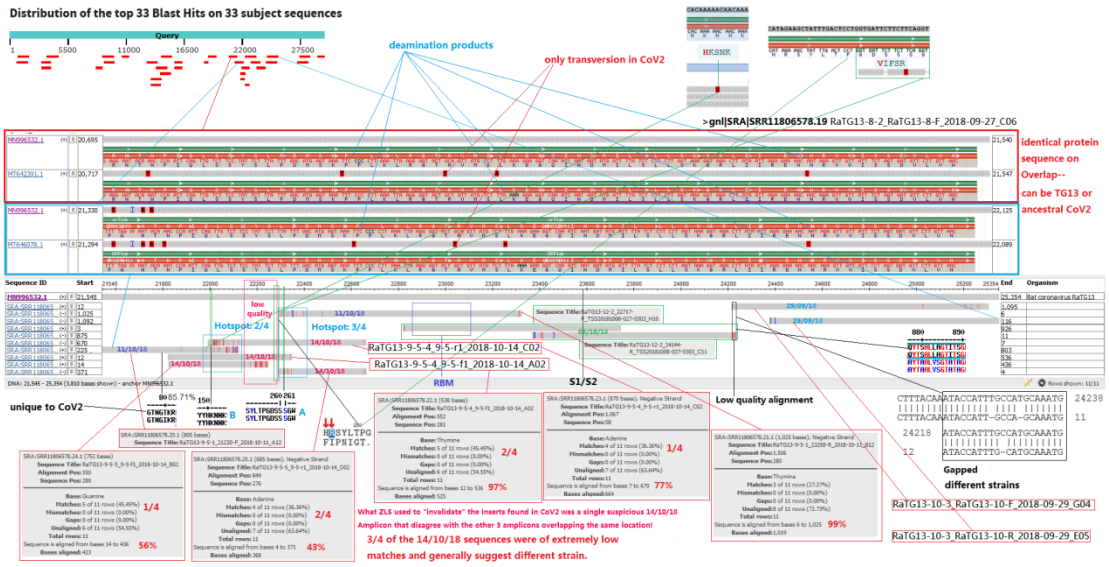


Figure 16: a thorough analysis of the amplicons located on the S locus of RaTG13 in SRX8357956. Notice that the last 4 amplicons sequenced in 14/10/2018 was of very low quality matches, and matched other organisms—including Mouse(mus musculus).

select all *99 sequences selected* GenBank Graphics Distance tree of results

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* GenBank Graphics Distance tree of results

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 Scaffold_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* GenBank Graphics Distance 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

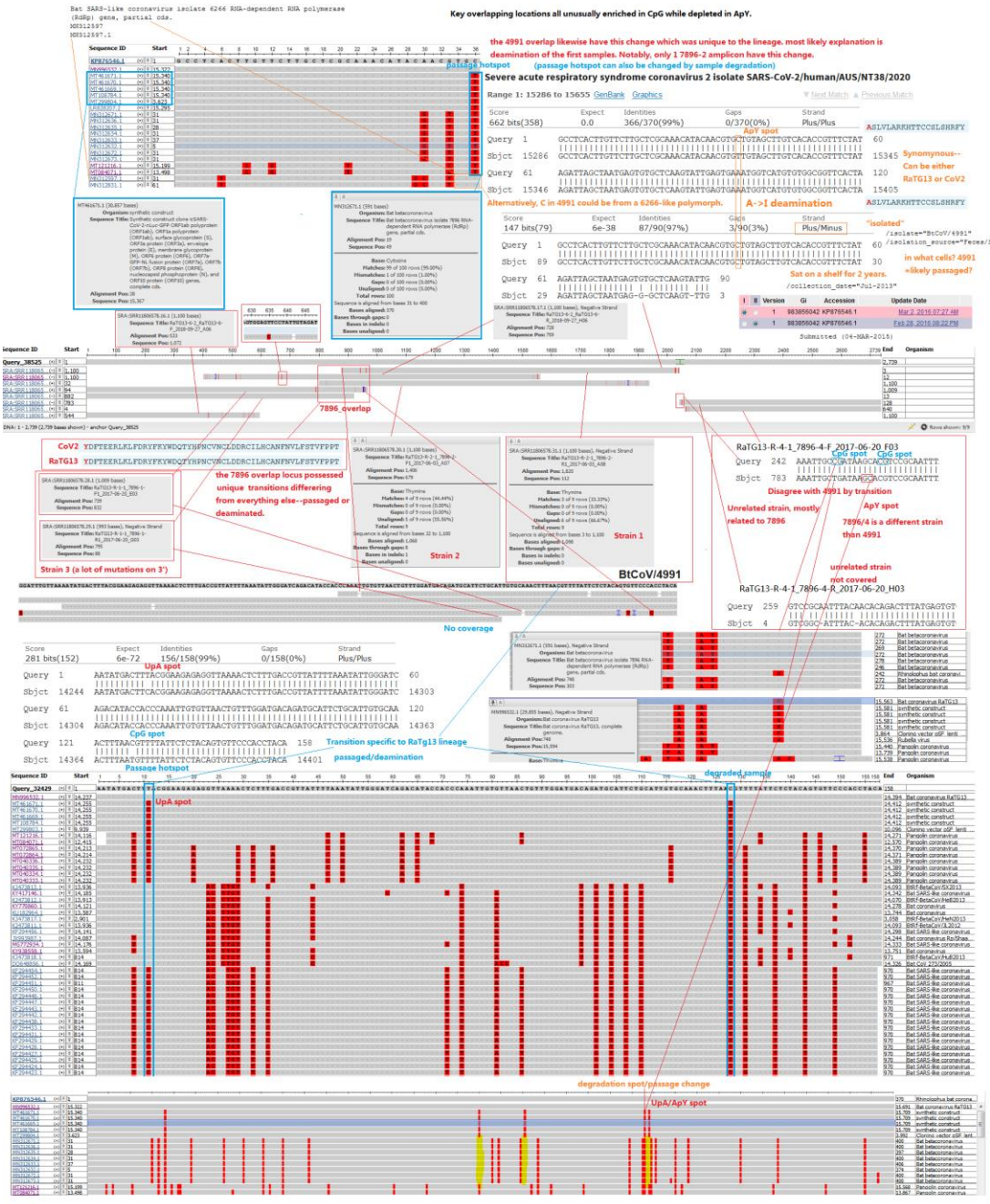


Figure 17: Detailed analysis of the early amplicons located in the nsp12 RdRp region of RaTG13 in SRX8357956.

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

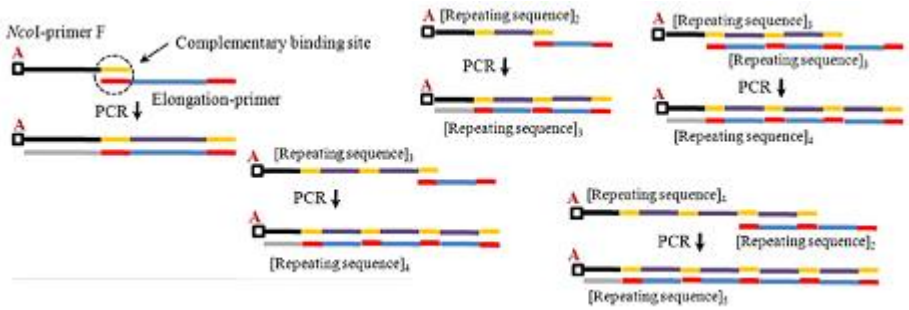


Figure 18A: Origin of repeating sequences in SRX7724752.

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.

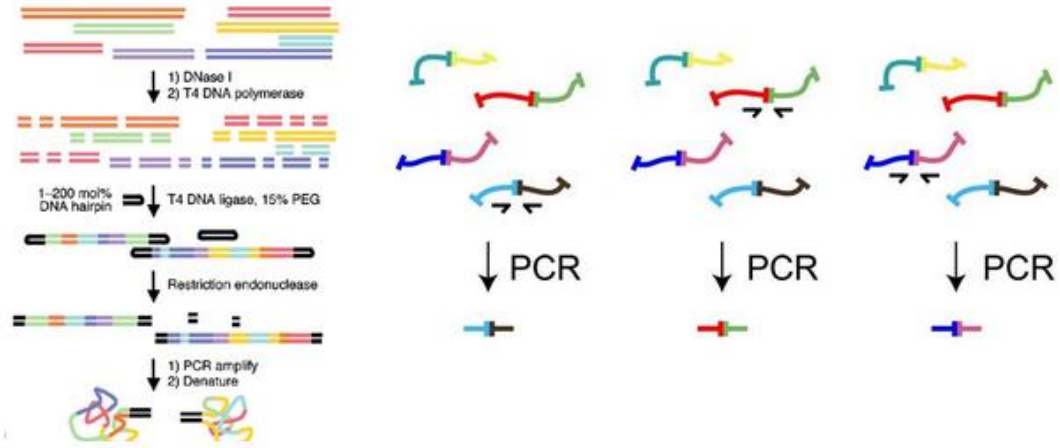


Figure 18B: Origin of the random matched sequences and partial sequences in SRX7724752

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	J85DMESK016 *Search expires on 08-01 20:59 pm* Download All
Results for	2:lcl|Query_5937 gnl|SRA|SRR11085797.66.2 66 (Biological)(150bp)
Program	BLASTN Citation
Database	nt See details
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

Download GenBank Graphics

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 ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
154 bits(83)	1e-33	116/132(88%)	1/132(0%)	Plus/Minus

```
Query  12    TTT-TCATTATTAAGTATTATGTACTGTACATAATTGTATGTACTATACTTTTATACAAC  70
             ||| ||||||| |||| ||||||||||||||   |||||||||  ||||||||||||   ||
Sbjct  2293  TTTATCATTATCAAGTGTTATGTACTGTACAGTATTGTATGTGTTATACTTTTATATGAC  2234

Query  71    TGGCAGCACAGCAGGTTTGTTTATACCAGCATCACCACAAAAATGTGAGTAATGCATTAC  130
             || ||||| || ||| ||||||| |||||||||||||||||||||||||||||||||||
Sbjct  2233  TGACAGCATAGTAGGCTTGTTTACACCAGCATCACCACAAAAATGTGAGTAATGCATTAC  2174

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

```
                    ttagatttcatctaaacgaacaaactaaaatgtctgataatggaccccaaaaccaacgaaatgcaccccgcattacgtttggtggaccct
          CTCTCGATCTCTTGTAGATCTGTTC TCTAAACGAAC
ACAAACCAACGAACTCTCGATCTCTTGTAGATCTGT
                                   TAACCTCCTTTTGTCCCCTAGTTCACCGATCTCATCGCCCCTGCCAGGGGTCCATGGACTGTGTTGATCCTGTCCTCTGCTCCTCC
```

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* GenBank 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
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
ACCTCTTTCACCAGCAACTGTAGAAAATCTTAAAAGGGCTTTTGTTTTTTTTACCAACTTT
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: 47720 to 47856 GenBank Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
176 bits(95)	3e-40	123/137(90%)	0/137(0%)	Plus/Minus

```
Query  7      CCTACTCTTTTACAAGATACTTGGCTTTTAGAAAAACTTGCACATTTCGATAGGGAAAGG  66
              ||||| ||||||||||||||||||||||||||||||||||||||| ||||| || ||
Sbjct  47856  CCTACACTTTTACAAGATACTTGGCTTTTAGAAAAACTTGCACATTTTGATAGAGAGAGA  47797

Query  67     ATACCAGAAAGAGTTGTGCACGCTAAAGGAAGTGCTGCATACGGCGAATTAACAATTACT  126
              || ||||| ||||| ||||||||||||||||||| ||||| || ||||||||||||||
Sbjct  47796  ATCCCAGAGAGAGTGGTGCACGCTAAAGGAAGTGCAGCATATGGTGAATTAACAATTACA  47737

Query  127    AATGATATTACAAAATA  143
              |||||||||||  ||||
Sbjct  47736  AATGATATTACTCAATA  47720
```

```
     gene            complement(<1..>137)
                     /gene="katA"
                     /locus_tag="HH_0043"
                     /old_locus_tag="HH0043"
     CDS             complement(<1..>137)
                     /gene="katA"
                     /locus_tag="HH_0043"
                     /old_locus_tag="HH0043"
                     /EC_number="1.11.1.6"
                     /codon_start=1
                     /transl_table=11
                     /product="catalase"
                     /protein_id="AAP76640.1"
                     /translation="MSKKFTTATGTPLGDNQNSITAGKKGPTLLQDTWLLEKLAHFDR
                     ERIPERVVHAKGSAAYGELTITNDITQYTKAELFNKVGKKTKAFLRFSVVAGERGAAD
                     AERDVRGFALKLYTNEGNWDIVGNNTPVFFIKDAIKFPDFIHTQKRDPKTNMKSPTAM
                     WDFWSLHPESLHQVTILMSDRGIPRSYREMHGFGSHTYSFINAKNERFWVKFHFVCLQ
                     GIHNLTNKESEAVIAKDRESHQKDLFENIEKGNFPKWRFCIQVMSEKEAENYRFNPFD
                     LTKVWSHKDYPLIEVGILELNKNPENYFAEVEQAAFNPANIVPGVGYSPDKVLQGRLF
                     SYGDTQRYRLGINHTQLPVNAPIVPVNNTHRDGFMQQGQFGDRRNYEPSYFNDYVEDK
                     NALEPPLFVQEGDVMYKYDHREYEDDYFVQAGDLYRLMTAEQKEALCQNIKESMEGVP
                     DEIKKRQLEHFKKADKAYGKRVAELLGL"
ORIGIN
        1 tattgagtaa tatcatttgt aattgttaat tcaccatatg ctgcacttcc tttagcgtgc
       61 accactctct ctgggattct ctctctatca aaatgtgcaa gtttttctaa aagccaagta
      121 tcttgtaaaa gtgtagg
//
```

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%

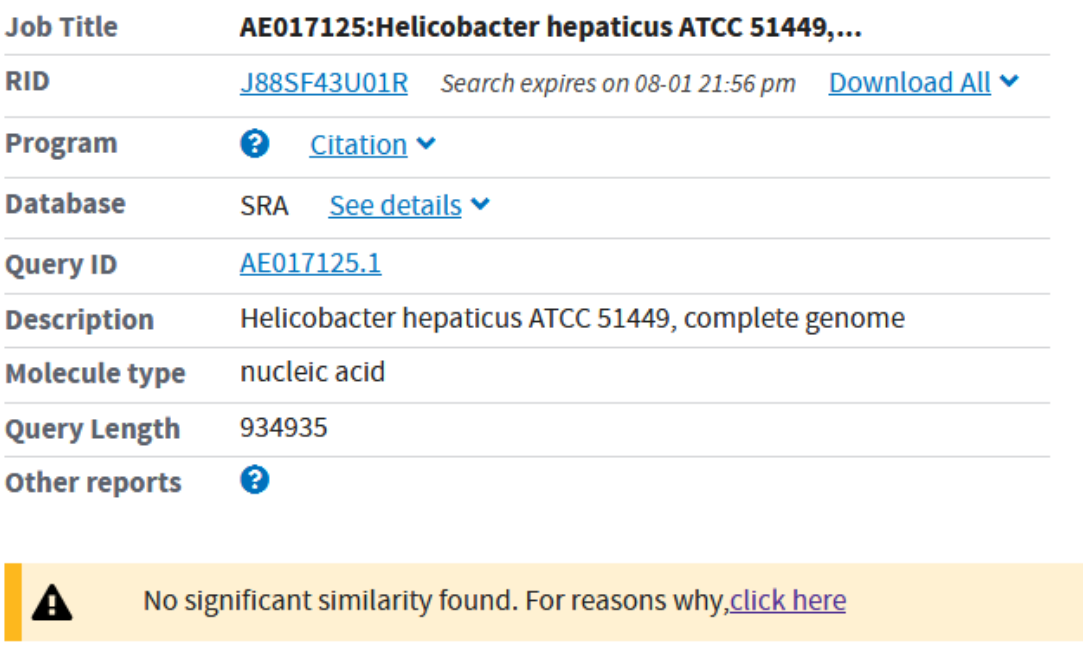
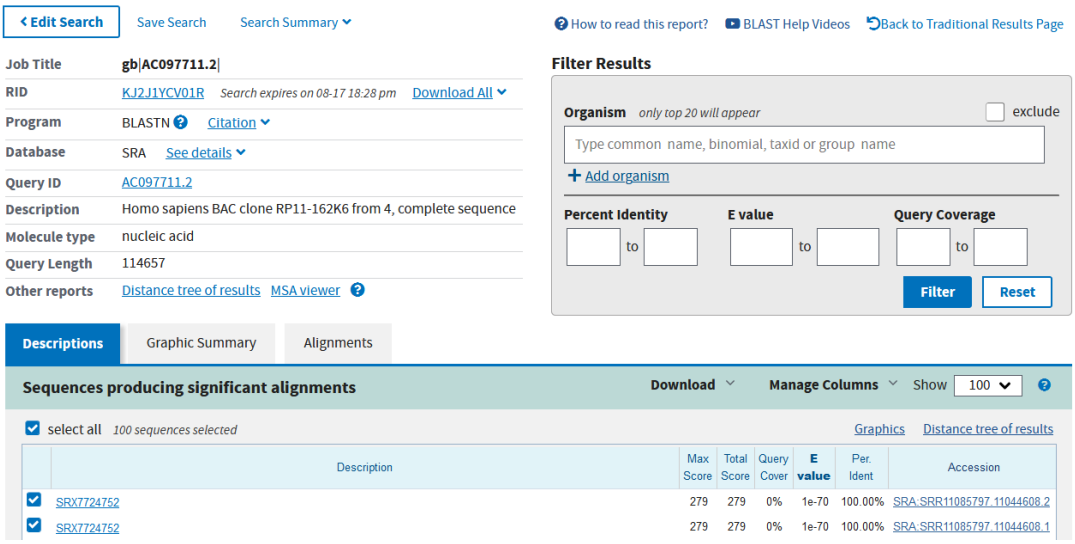


Figure 22b: the same species of bacteria in SRX7724752. No significant matches were found.

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



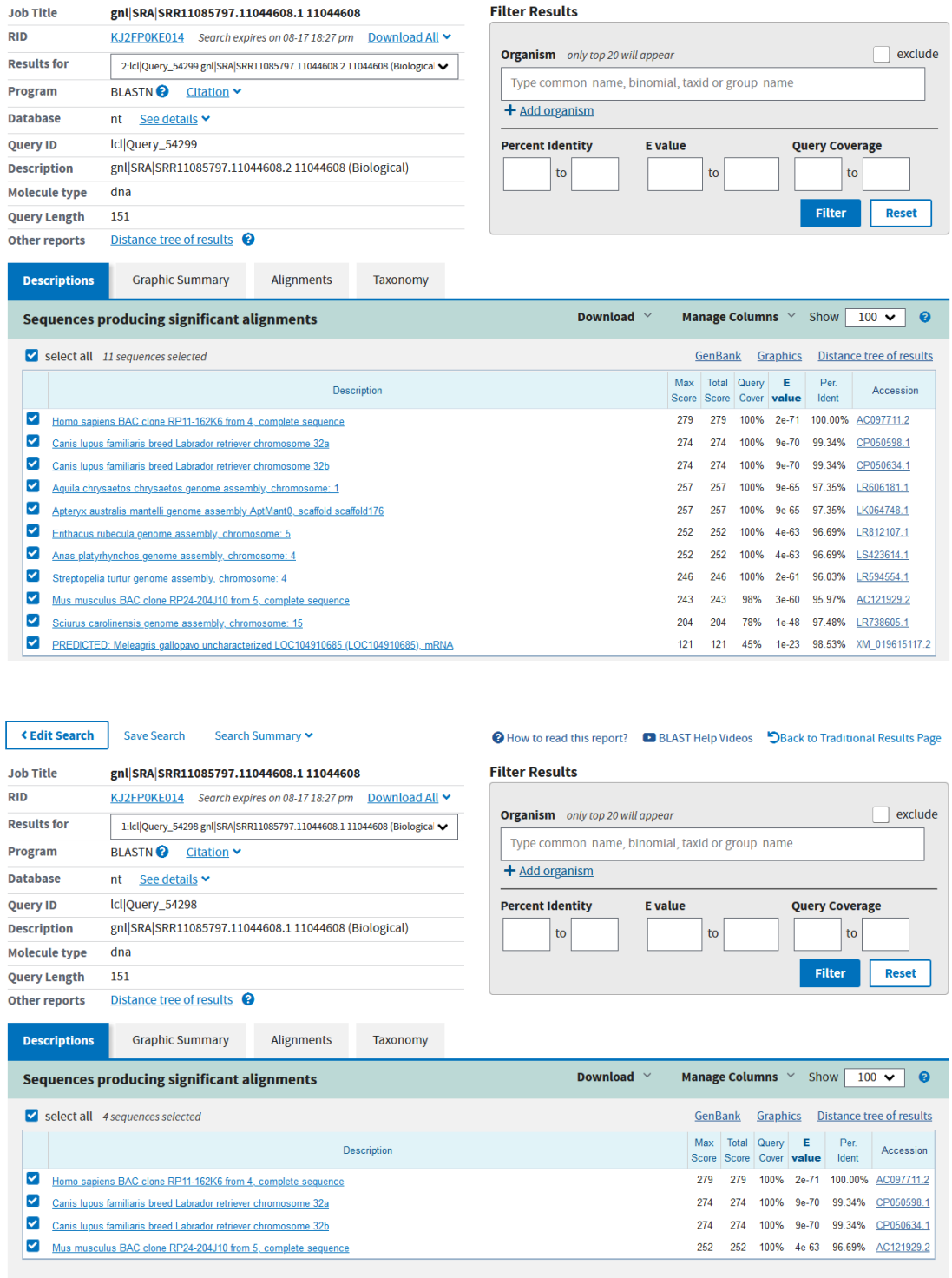
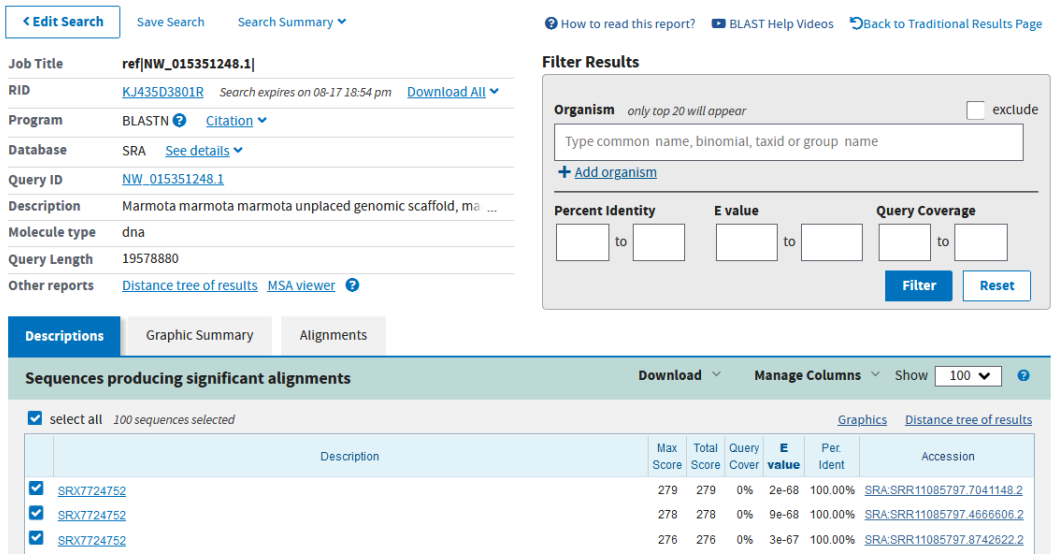
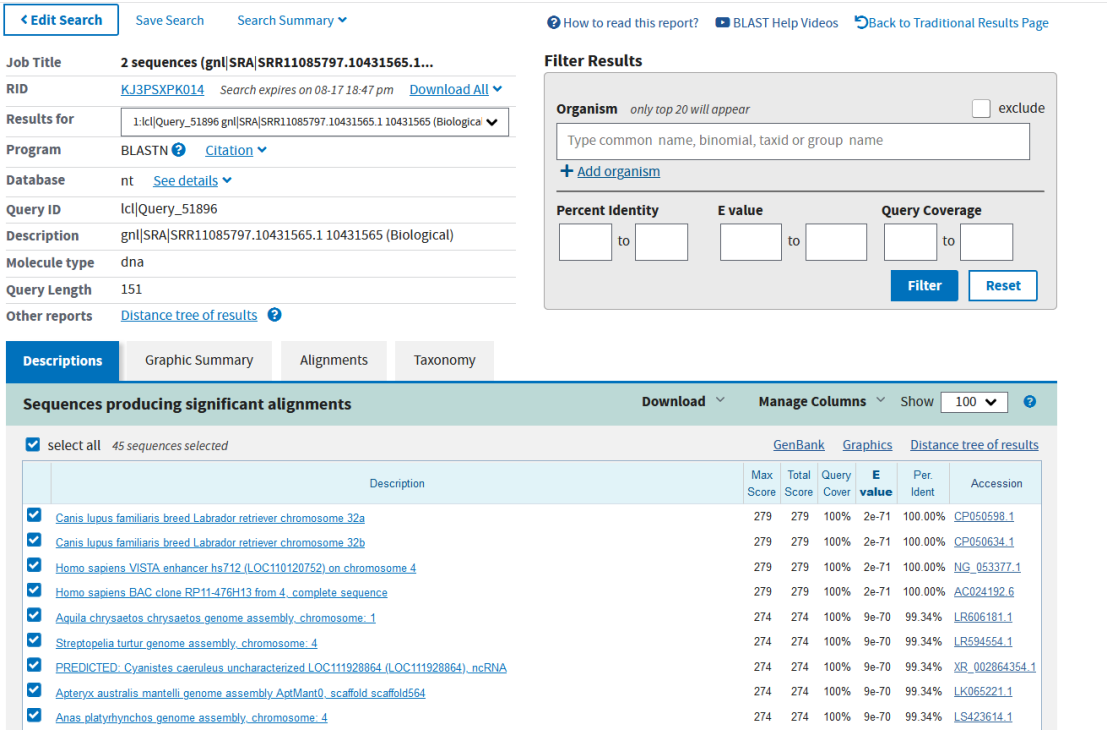


Figure 23: Unique, fully-matched 100% read from Homo Sapiens is recovered from the dataset SRX7724752.


Edit Search
Save Search
Search Summary
How to read this report?
BLAST Help Videos
Back to Traditional Results Page
Job Title ref|NW_015351248.1|
RID KJ435D3801R Search expires on 08-17 18:54 pm Download All
Program BLASTN Citation
Database SRA See details
Query ID NW_015351248.1
Description Marmota marmota marmota unplaced genomic scaffold, ma ...
Molecule type dna
Query Length 19578880
Other reports Distance tree of results MSA viewer
Filter Results
Organism only top 20 will appear
exclude
Type common name, binomial, taxid or group name
Add organism
Percent Identity
E value
Query Coverage
to
Filter
Reset
Descriptions
Graphic Summary
Alignments
Sequences producing significant alignments
Download
Manage Columns
Show 100
select all 100 sequences selected
Graphics
Distance tree of results
Description
Max Score
Total Score
Query Cover
E value
Per. Ident
Accession
SRX7724752 279 279 0% 2e-68 100.00% SRA:SRR11085797.7041148.2
SRX7724752 278 278 0% 9e-68 100.00% SRA:SRR11085797.4666606.2
SRX7724752 276 276 0% 3e-67 100.00% SRA:SRR11085797.8742622.2


Edit Search
Save Search
Search Summary
How to read this report?
BLAST Help Videos
Back to Traditional Results Page
Job Title 2 sequences (gnl|SRA|SRR11085797.10431565.1...
RID KJ3PSXPK014 Search expires on 08-17 18:47 pm Download All
Results for 1:lcl|Query_51896 gnl|SRA|SRR11085797.10431565.1 10431565 (Biological
Program BLASTN Citation
Database nt See details
Query ID lcl|Query_51896
Description gnl|SRA|SRR11085797.10431565.1 10431565 (Biological)
Molecule type dna
Query Length 151
Other reports Distance tree of results
Filter Results
Organism only top 20 will appear
exclude
Type common name, binomial, taxid or group name
Add organism
Percent Identity
E value
Query Coverage
to
Filter
Reset
Descriptions
Graphic Summary
Alignments
Taxonomy
Sequences producing significant alignments
Download
Manage Columns
Show 100
select all 45 sequences selected
GenBank
Graphics
Distance tree of results
Description
Max Score
Total Score
Query Cover
E value
Per. Ident
Accession
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% NG_053377.1
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

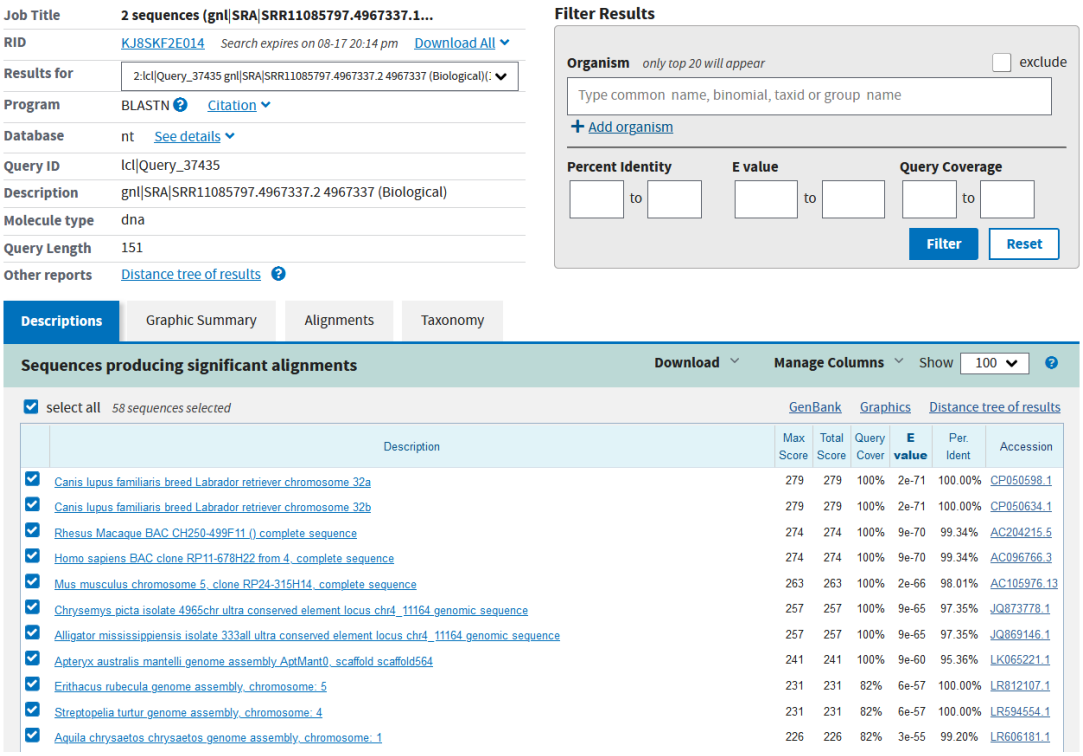


Figure. 24: Marmota Marmota genetic scaffold assemblies returned significant amount of 100% full-length matched reads that were sometimes also found in Homo Sapiens and Canis Lupus Famillaris.

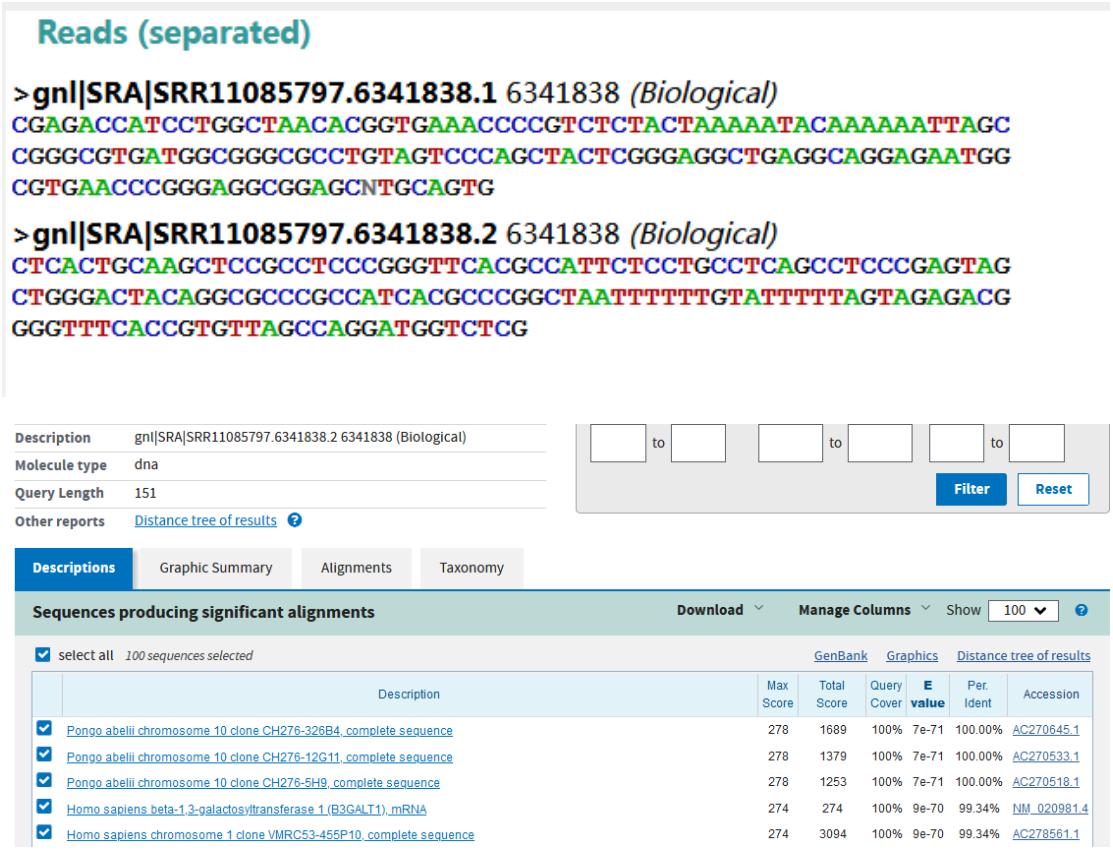
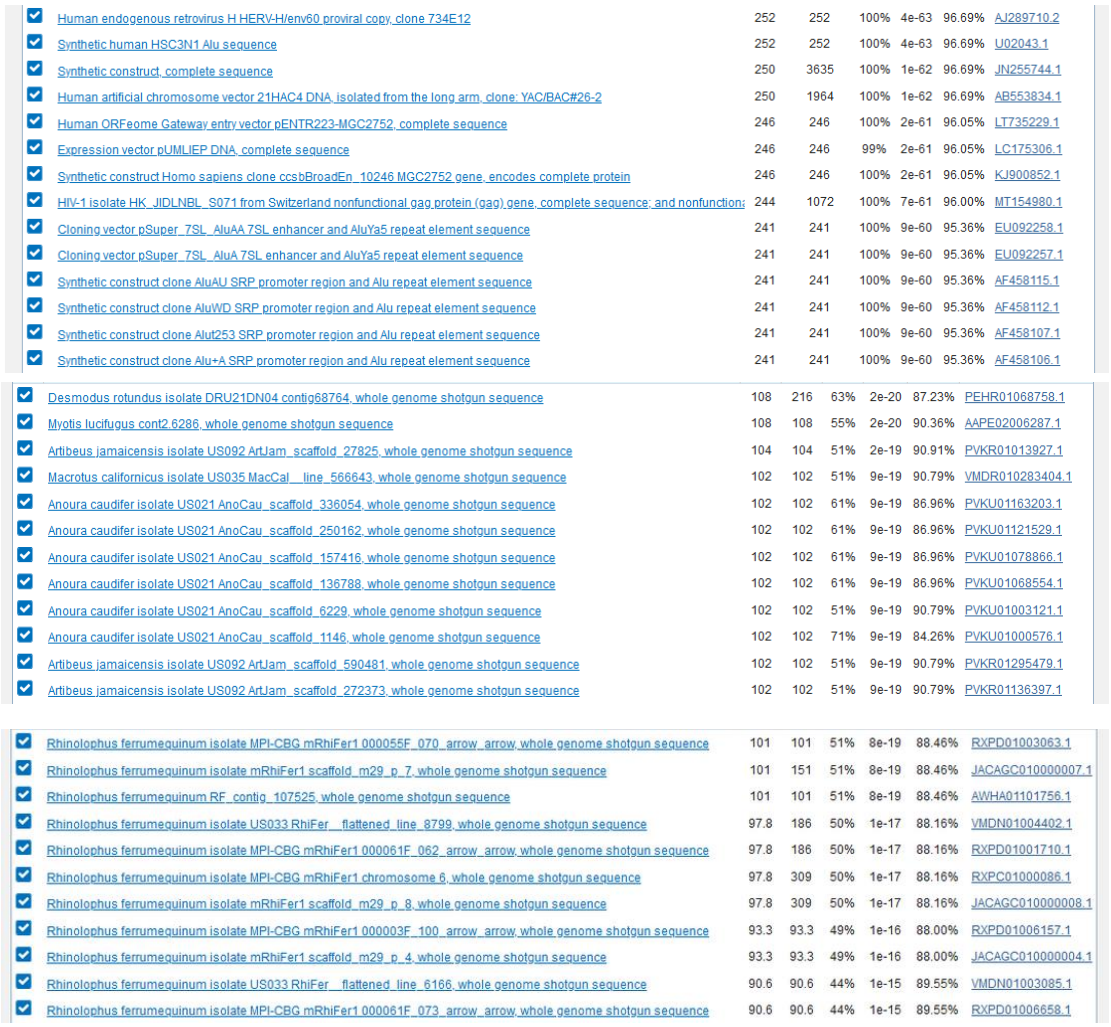


Figure.25a: 100% full-length matched reads to Hominid(Pongo Albelii) genomic DNA.



Human endogenous retrovirus H HERV-H/env60 proviral copy, clone 734E12	252	252	100%	4e-63	96.69%	AJ289710.2
Synthetic human HSC3N1 Alu sequence	252	252	100%	4e-63	96.69%	U02043.1
Synthetic construct, complete sequence	250	3635	100%	1e-62	96.69%	JN255744.1
Human artificial chromosome vector 21HAC4 DNA, isolated from the long arm, clone: YAC/BAC#26-2	250	1964	100%	1e-62	96.69%	AB553834.1
Human ORFeome Gateway entry vector pENTR223-MGC2752, complete sequence	246	246	100%	2e-61	96.05%	LT735229.1
Expression vector pUMLIEP DNA, complete sequence	246	246	99%	2e-61	96.05%	LC175306.1
Synthetic construct Homo sapiens clone ccsbBroadEn_10246 MGC2752 gene, encodes complete protein	246	246	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	1072	100%	7e-61	96.00%	MT154980.1
Cloning vector pSuper_7SL_AluAA 7SL enhancer and AluYa5 repeat element sequence	241	241	100%	9e-60	95.36%	EU092258.1
Cloning vector pSuper_7SL_AluA 7SL enhancer and AluYa5 repeat element sequence	241	241	100%	9e-60	95.36%	EU092257.1
Synthetic construct clone AluAU SRP promoter region and Alu repeat element sequence	241	241	100%	9e-60	95.36%	AF458115.1
Synthetic construct clone AluWD SRP promoter region and Alu repeat element sequence	241	241	100%	9e-60	95.36%	AF458112.1
Synthetic construct clone Alut253 SRP promoter region and Alu repeat element sequence	241	241	100%	9e-60	95.36%	AF458107.1
Synthetic construct clone Alu+A SRP promoter region and Alu repeat element sequence	241	241	100%	9e-60	95.36%	AF458106.1
Desmodus rotundus isolate DRU21DN04 contig68764, whole genome shotgun sequence	108	216	63%	2e-20	87.23%	PEHR01068758.1
Myotis lucifugus cont2.6286, whole genome shotgun sequence	108	108	55%	2e-20	90.36%	AAPE02006287.1
Artibeus jamaicensis isolate US092 ArtJam_scaffold_27825, whole genome shotgun sequence	104	104	51%	2e-19	90.91%	PVKR01013927.1
Macrotus californicus isolate US035 MacCal__line_566643, whole genome shotgun sequence	102	102	51%	9e-19	90.79%	VMDR010283404.1
Anoura caudifer isolate US021 AnoCau_scaffold_336054, whole genome shotgun sequence	102	102	61%	9e-19	86.96%	PVKU01163203.1
Anoura caudifer isolate US021 AnoCau_scaffold_250162, whole genome shotgun sequence	102	102	61%	9e-19	86.96%	PVKU01121529.1
Anoura caudifer isolate US021 AnoCau_scaffold_157416, whole genome shotgun sequence	102	102	61%	9e-19	86.96%	PVKU01078866.1
Anoura caudifer isolate US021 AnoCau_scaffold_136788, whole genome shotgun sequence	102	102	61%	9e-19	86.96%	PVKU01068554.1
Anoura caudifer isolate US021 AnoCau_scaffold_6229, whole genome shotgun sequence	102	102	51%	9e-19	90.79%	PVKU01003121.1
Anoura caudifer isolate US021 AnoCau_scaffold_1146, whole genome shotgun sequence	102	102	71%	9e-19	84.26%	PVKU01000576.1
Artibeus jamaicensis isolate US092 ArtJam_scaffold_590481, whole genome shotgun sequence	102	102	51%	9e-19	90.79%	PVKR01295479.1
Artibeus jamaicensis isolate US092 ArtJam_scaffold_272373, whole genome shotgun sequence	102	102	51%	9e-19	90.79%	PVKR01136397.1
Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000055F_070_arrow_arrow, whole genome shotgun sequence	101	101	51%	8e-19	88.46%	RXPD01003063.1
Rhinolophus ferrumequinum isolate mRhiFer1 scaffold_m29_p_7, whole genome shotgun sequence	101	151	51%	8e-19	88.46%	JACAGC010000007.1
Rhinolophus ferrumequinum RF_contig_107525, whole genome shotgun sequence	101	101	51%	8e-19	88.46%	AWHA01101756.1
Rhinolophus ferrumequinum isolate US033 RhiFer__flattened_line_8799, whole genome shotgun sequence	97.8	186	50%	1e-17	88.16%	VMDN01004402.1
Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000061F_062_arrow_arrow, whole genome shotgun sequence	97.8	186	50%	1e-17	88.16%	RXPD01001710.1
Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 chromosome 6, whole genome shotgun sequence	97.8	309	50%	1e-17	88.16%	RXPC01000086.1
Rhinolophus ferrumequinum isolate mRhiFer1 scaffold_m29_p_8, whole genome shotgun sequence	97.8	309	50%	1e-17	88.16%	JACAGC010000008.1
Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000003F_100_arrow_arrow, whole genome shotgun sequence	93.3	93.3	49%	1e-16	88.00%	RXPD01006157.1
Rhinolophus ferrumequinum isolate mRhiFer1 scaffold_m29_p_4, whole genome shotgun sequence	93.3	93.3	49%	1e-16	88.00%	JACAGC010000004.1
Rhinolophus ferrumequinum isolate US033 RhiFer__flattened_line_6166, whole genome shotgun sequence	90.6	90.6	44%	1e-15	89.55%	VMDN01003085.1
Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1 000061F_073_arrow_arrow, whole genome shotgun sequence	90.6	90.6	44%	1e-15	89.55%	RXPD01006658.1

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