

# Anomalies in BatCoV/RaTG13 sequencing and provenance

Daoyu Zhang

## ABSTRACT

To this date, the most critical piece of evidence on the purported “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 provenance of the promised claims of RaTG13 or its 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)
CTAACCCCTAACCCCTAGCACTATCCTGTTCCAACCCCCAACCTAACCCCTCACCCCTAACCC
TAACCCCCAGCCTGTTTCATAACCTTAACTCGCACCTCATCGCTAACCCCCAGCCCTCACCCG
ATCCTGTTCCCTCTCCGAACATAACCCCT

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

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

GTAGGGTTAGGGTTAGGGTTAGGGTTGGGATACAGGATATGGTTAGGGTAGGG  
TAGGGTCAGGGTTAGGATTGAAACGAGATAGGTTACGTGATAGGGTAGCGTTAGGGT  
TAGGTTAGTAATCCGCAACGGCTAGGGTT

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

CCTAACCTAACCTAACCTAACCTAACCTAACCTATCCTGTTCCCAACCCTAAC  
CTAACCTAACCTAACCTAACACAAAATAACCTAACCCAACCCAAACCCCTAAC  
CCATCTTACTCACACCTAACCCAAAATC

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

GTAGGGTTAGGGTTAGCGTTAGGGTTAGAAACAGGATAGGGTTAGGGTTAGGGTAGGG  
TTAGGGTTTGGTTGGTCACAGTGTTGCGCTAGGCATAGGGATAGGGTCGCGTTAGGGT  
TAGGGTTAGGATTCGGAAGAGCTAGCTAAA

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

GTTCCCAACCCTAACCTAACCTAACCTAACCTAACCTAACCTTTCTTTTCCAACCCTA  
ACACTAACCTAACACTAAACCTAACCCCAACCCCTACCACTATACTATATCCGACTCTCA  
CGCTAACACTAAACATAAGTAATCACAAATT

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

TAACCTAACCTAACGCTAAACCTAATCCAATCCTGTTCCCAACCCTAACCTAACCT  
GACCTAACGTTTCCCCACCCGAACGCCACCCGATCCGCCACCCCTAACCGTAACCCGT  
TCCCAACCCCTCCTACTGCTCGATCCGCCT

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

GGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTGGGAACAGGATAGG  
GTTAGGGTTAGGGTTGGGATCAGGATAGGGATAGGGATAGGGATAGGGTTAGGGTAGGG  
TGGGAAACAGGAGAGCGTTAGGCAAGG

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

GGTTAGGGTTAGGGTGGAGAAGAGGATAGGTTAGGTTAGGGTTAGGGTTAGGGTGAGG  
GTTAGGGAGAGGGTTAGCTACACCGATAGGAGTAGGGTAACGATTAGGGTTAGGGTTAGGT  
TTGGAAAAAGCATAGGCTATGAGGTACGGT

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

CTGCTTCAAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCTAACCTAAC  
CTATCCTGTTCCCAACCTAACCTAACCTAACCCCAACCCCTAACCCCACCCAAACCC  
AACCTAACCCCAACCCATAACCCCAACCAT

>gnl|SRA|SRR11085797.15.1 15 (*Biological*)  
TGTTCCCAACCTAAACCTAACCTAAGCCGATCCTGTTCCCACCCCTAACCTAACCCCT  
ATCCTGTAAACAACCCCCACCCCTAAAAACATCCTCGTACAAACCTAACCCAAACCCCAT  
CCCAAACACATACCCGTACGAAACCCACCC

>gnl|SRA|SRR11085797.15.2 15 (*Biological*)  
GTTGGGTTAGGGTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTCAGGG  
TTGGGACAAGGACGGGTAGGGGTAGGGTAGTGCACAGGGTAGGGAGTAGGGTAGGGTT  
GGGACAGAGAAGGGACAGTAGGGGGAGTAGG

>gnl|SRA|SRR11085797.16.1 16 (*Biological*)  
CTAACCTAACCTAACCCCTATCCTGTTCTAACCCGAACCCCTAACCTAACCCCTAACCC  
TAACCCCTCACCTGTTCCAGACCGTAATGCTAACCTAACACTATCCTGCGCTACCCCG  
ACCCTAACCTCAGCCGACCGTCACGCCCG

>gnl|SRA|SRR11085797.16.2 16 (*Biological*)  
GGTAAGGGTTAGGGTTAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG  
GTTGGGAAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGAA  
TAGGGTAGAGATAAGGGTGAGGTGGTGGAA

>gnl|SRA|SRR11085797.17.1 17 (*Biological*)  
AACCTAACCTAACCTAACGCTATCATGATCCCATCCCTAACCTAACCTAACCC  
ACCCTAAACTAACCTAACCTTTCATCTCCCTAACACTACCCCCAACACGCCACCCAT  
CCCCAACCAACTATGCATGCACGTGCTCTAAC

>gnl|SRA|SRR11085797.17.2 17 (*Biological*)  
GGTTAGGGTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGAACAGGATAGGG  
GTAGGGTTAGGGCTAGGGTGAGAAACAGGGTAGGGTAGGGTAGGGTAGGGATAAGGGTAGGGT  
TGGGGTTGGAACAGAGAAGGGAAAGGGCA

>gnl|SRA|SRR11085797.18.1 18 (*Biological*)  
CTAACCTGTTCCCAAACCTAACATCCAACCTATCCTGTTCCCACCCCTAACCTAACCC  
AACCTATAACCTATCCTCCCCACACACCGACCCCTATACACCACCCCTAACCGAACCTA  
ACCCCATCCTGTTATCGAACGATACCCCCAC

>gnl|SRA|SRR11085797.18.2 18 (*Biological*)  
GTTAGGGTTAGGGTTAGGGTTGGTAACAGGATAGGGTTAGGGTTAGGGTTAGTGTGGGA  
ACAGGATAGGGGGAGGGAGAGGGTTAGGGTTGGGAAAGAGGATAGGGATAGGGTAGGGAG  
AGGGTTGGGGATAGGGAGAGAGAGAGGAG

>gnl|SRA|SRR11085797.19.1 19 (*Biological*)  
CCTAACCTAACCTAACCTCTCCTGTTCCACCATAACCTAACCTATCCCTAACCC  
CTAACCCACTCCTGTTAACACTAACCTAACCTAACCTGAGCTCATCCCCAACCTAACCA  
TAACCCCCACCAAGTCCGATACCATCACCCCC

>gnl|SRA|SRR11085797.19.2 19 (*Biological*)  
AGGGTTAGGGTTGGAAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTG  
GGGACAGGATAGGGTTAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGGTAGGTGTGGGGTTGG  
GATTGGGTATGGTAGTGGTCAGGGATAGTG

```
>gnl|SRA|SRR11085797.20.1 20 (Biological)
CCTGTTCCAACCTCACCTGACACTGACCCTAACACTAACCTTAACCCCTAACCGATC
CTGTTCTGACCTAACGACAAGCCTGGCACTAAACTGATCGCGTTCCAATCGTTACCG
CTTCCCTAACACCGTCTGTGAAGATACTCCG

>gnl|SRA|SRR11085797.20.2 20 (Biological)
CTTTAGGTTAGTGTAGGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAG
GGTTGGGAACAGGATAGGGTTAGGGTTAGGGATAGGGTTGGGCTGGATAGGGTTGGGG
GTAGGGTTAGAGTGAGTGTGGGCAGCAGGCG
```

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)
CTACTGTGTCATCCCATTCACAAACGCTTATTGGCGGTACAGGAATATCAACCTGTTGT
CCATCACCTACGCCCTTCGGCCTCGGCTTAGGTCTGACTAACCCAGGGCAGAAGAACCT
TCCCCCTGGAAACCTTGGGTTGACGGCCCCGTG

>gnl|SRA|SRR11085736.100.2 100 (Biological)
ATCCCCACGGGCCGTAAACCCAAGGTTCCAGGGGAAGGTTCGTCCGCCCTGGTTAGTCA
GGACCTAAAGCCGACGCCGAAAGGCGTAGGTGATGGACAACAGGTTGATATTCCGTAAACC
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|SRR11085797.11.2 11 (Biological)

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

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

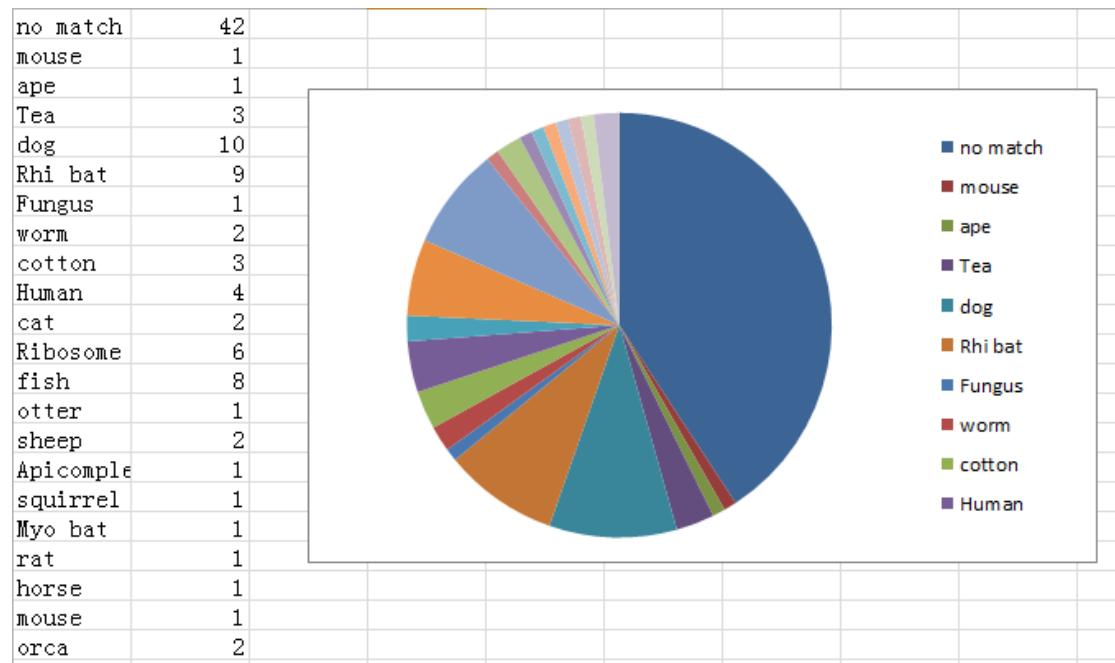


Figure 4: BLAST result of 100 random reads obtained from RaTG13 using BLASTn.

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

The screenshot shows the NCBI BLAST results page. At the top, there are search parameters: Description (gnl|SRA|SRR11085797.2.12 (Biological)), Molecule type (dna), Query Length (151), and Other reports (Distance tree of results, MSA viewer). To the right are filters for Percent Identity, E value, and Query Coverage, with 'Filter' and 'Reset' buttons. Below this is a navigation bar with tabs: Descriptions (selected), Graphic Summary, Alignments, and Taxonomy. The main content area displays a table titled 'Sequences producing significant alignments'. The table has columns: Description, Max Score, Total Score, Query Cover, E value, Per. Ident, and Accession. A checkbox 'select all' is checked, indicating 6 sequences selected. One row is highlighted in blue, showing the sequence 'Eimeria mitis hypothetical protein\_conserved\_partial mRNA' with a Max Score of 102, Total Score of 102, 44% Query Cover, an E value of 4e-18, 94.03% Per. Ident, and Accession XM\_013494305.1.

The screenshot shows a BLASTn search interface with the following details:

- Results for:** 1:lcl|Query\_61914.gnl|SRA|SRR11085797.7.1.7 (Biological)(150bp)
- Program:** BLASTN [?](#) [Citation](#)
- Database:** nt [See details](#)
- Query ID:** lcl|Query\_61914
- Description:** gnl|SRA|SRR11085797.7.1.7 (Biological)
- Molecule type:** dna
- Query Length:** 150
- Other reports:** [Distance tree of results](#) [?](#)

**Organism:** only top 20 will appear  exclude  
 Type common name, binomial, taxid or group name  
[+ Add organism](#)

**Percent identity**  to    **E value**  to    **Query Coverage**  to

**Buttons:** Filter [Reset](#)

**Navigation:** Descriptions [Graphic Summary](#) [Alignments](#) [Taxonomy](#)

**Download:** [GenBank](#) [Graphics](#) [Distance tree of results](#)

**Show:** 5000 [?](#)

**Sequences producing significant alignments**

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Eimeria mitis hypothetical protein_conserved partial mRNA</a>	73.1	73.1	40%	3e-09	88.33%	<a href="#">XM_013494305.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cyprinus carpio genome assembly_common carp genome_scaffold 000012284</a>	73.1	73.1	40%	3e-09	88.33%	<a href="#">LN595243.1</a>

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)

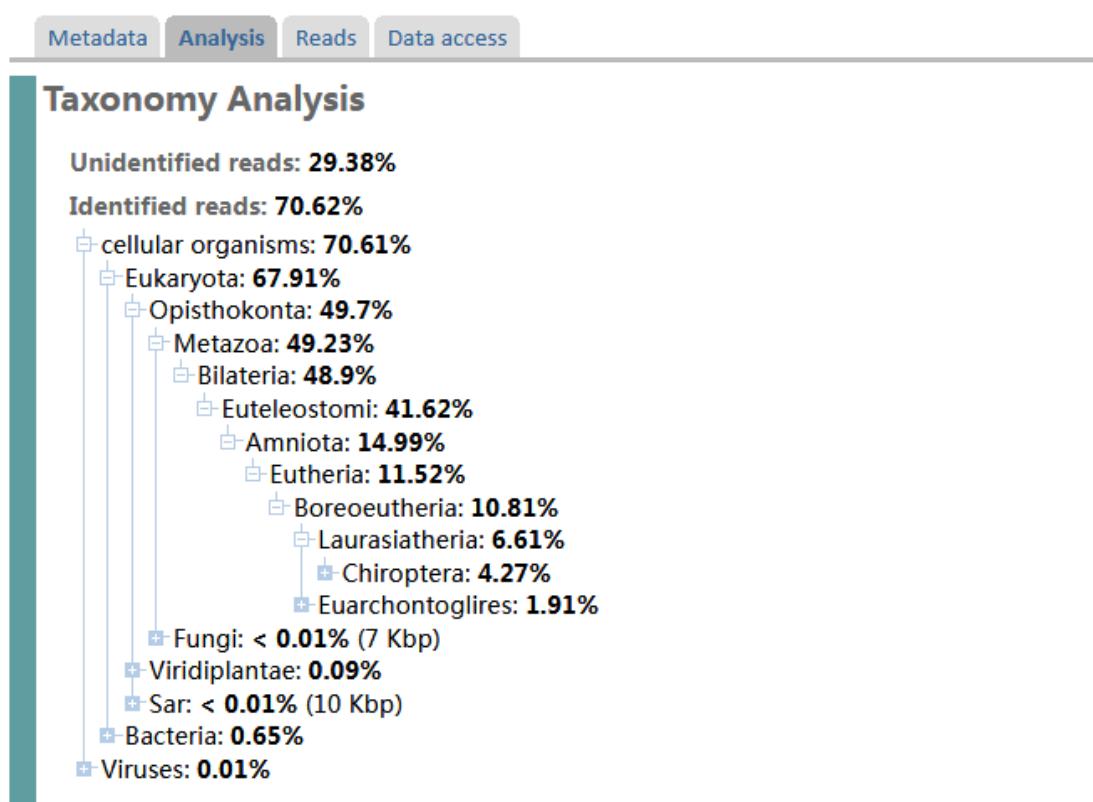


Figure 6: Phylogenetic analysis of SRX7724272.

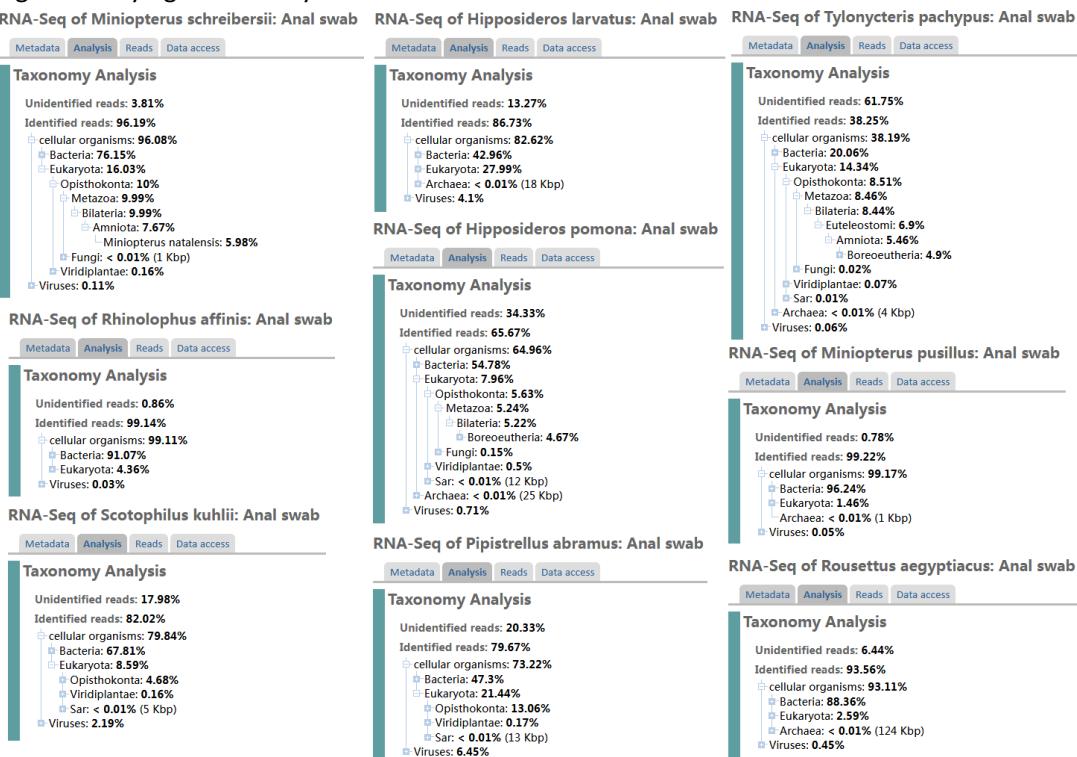


Figure 7: A set of 9 Swabs. The only ones that matches RaTG13 by metadata on Genbank. None of them had more Eukarya-like reads than Bacteria-like reads.

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

discernible reads from bacterial mRNA.

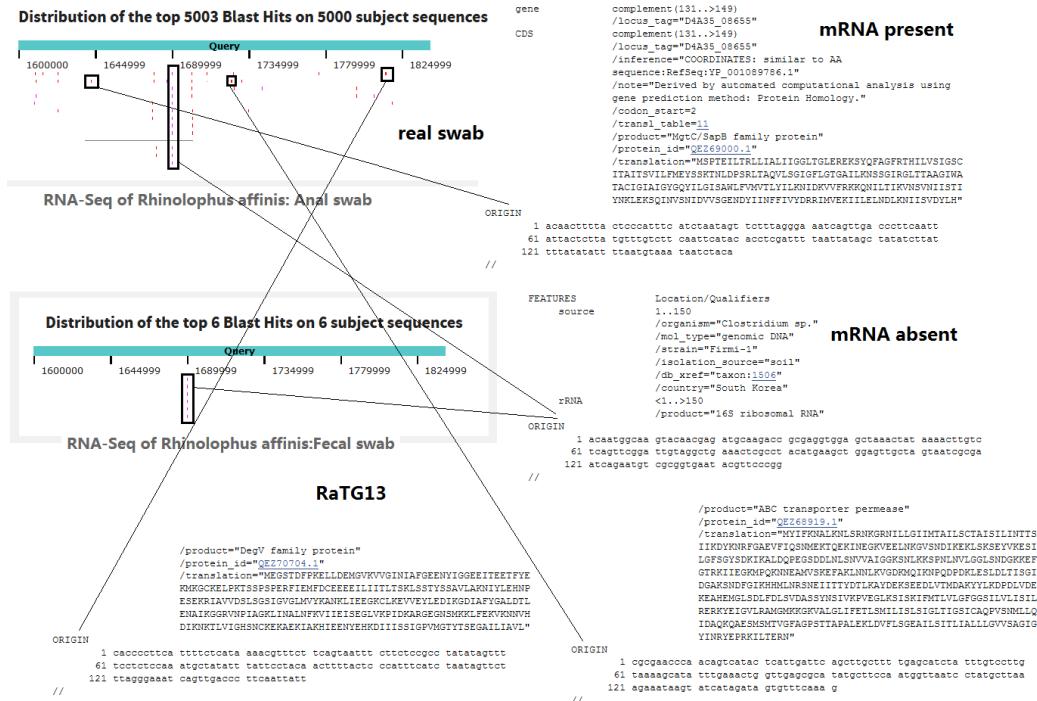


Figure 8: BLAST hits of bacterial non-ribosomal RNA genome on SRX7724272 and another swab from *Rhinolophus Affinis* under the same library preparation section.

## Observation of anomalous and unexpected data within SRX7724752

```
>gnl|SRA|SRR11085797.11550005.1 11550005 (Biological)
GCCCGTATTAGCCTTAGATGGAGTTACCAACCGCTTGGGCTGCATTCCAAGCAACC
CGACTCCGGGAAGACCCGGCCGCCGCGCCGGGGCGCTACCGGCCTCACACCGTCCA
CGGGCTGGGCCTCGATCAGAAGGACTTGGC

>gnl|SRA|SRR11085797.11550005.2 11550005 (Biological)
CGGTGGGGCGGGGACATTGGCGTACGGAAGACCCACTCCCCGGCGCGCTCGTGGGGG
CCCAAGTCCTCTGATCGAGGCCAGCCGTGGACGGTGTGAGGCCGGTAGCGGCCCG
GCGCGCCGGCCGGTCTTCCCAGAGTCGG
```

**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
<input checked="" type="checkbox"/>	PREDICTED: Phyllostomus discolor 28S ribosomal RNA (LOC114512504). rRNA	279	279	100%	2e-71	100.00%	<a href="#">XR_003685809.1</a>
<input checked="" type="checkbox"/>	PREDICTED: Phyllostomus discolor basic proline-rich protein-like (LOC114512442). mRNA	279	279	100%	2e-71	100.00%	<a href="#">XM_028531404.1</a>
<input checked="" type="checkbox"/>	Homo sapiens IncAB370.3 lncRNA gene, complete sequence	274	274	100%	9e-70	99.34%	<a href="#">MK280359.1</a>
<input checked="" type="checkbox"/>	Homo sapiens IncAB366.1 lncRNA gene, complete sequence	274	274	100%	9e-70	99.34%	<a href="#">MK280356.1</a>
<input checked="" type="checkbox"/>	Felis catus Senzu DNA_chromosome_E1_American Shorthair breed	274	1372	100%	9e-70	99.34%	<a href="#">AP023165.1</a>
<input checked="" type="checkbox"/>	PREDICTED: Marmota flaviventris 28S ribosomal RNA (LOC117794687). rRNA	274	274	100%	9e-70	99.34%	<a href="#">XR_004618536.1</a>
<input checked="" type="checkbox"/>	PREDICTED: Aricanthis niloticus 28S ribosomal RNA (LOC117704856). rRNA	274	274	100%	9e-70	99.34%	<a href="#">XR_004606369.1</a>
<input checked="" type="checkbox"/>	PREDICTED: Aricanthis niloticus 28S ribosomal RNA (LOC117704855). rRNA	274	274	100%	9e-70	99.34%	<a href="#">XR_004606368.1</a>

**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
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11601488.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11599592.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11598985.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11598757.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11588533.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11585765.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11585419.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11583582.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11582051.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11580259.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11579956.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11579609.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11573984.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11570618.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11568464.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11567362.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11566985.1</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	100%	2e-73	100.00%	<a href="#">SRA:SRR11085797.11558788.1</a>



## 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)  
 TTTGTTTGTAAAAATTATTTAAAATTATTGGGTGACAAATTGTTAGTAAAA  
 TTACATAGATTTCAGGTGTACAATTCTGTATTACATGTGGACGGTCCAGCCGACAG  
 TCAACGTTTACATGAAAGGGGGTGTGGGA

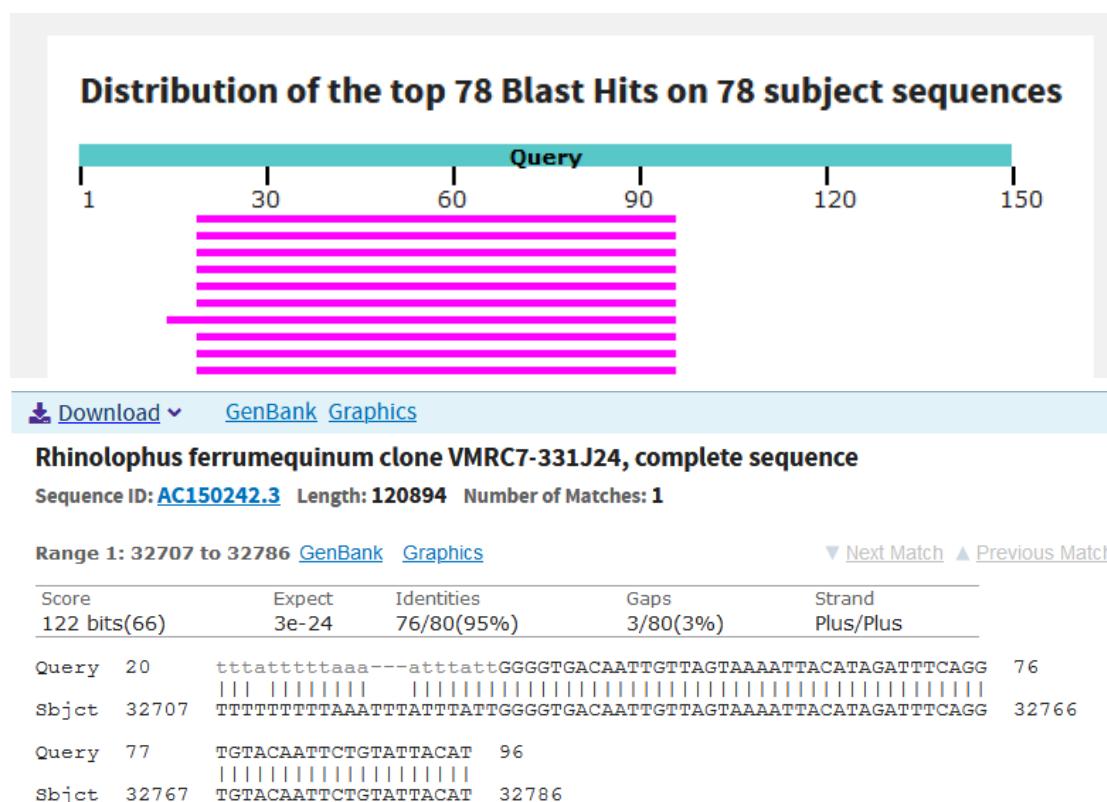


Figure 10: A sequence which was matched to a bat mRNA clone in the first 96 nucleotides, but then matching nothing on the later nucleotides. This match end with a T.

## Reads (separated)

>gnl|SRA|SRR11085797.8568962.1 8568962 (*Biological*)  
ATGGGGGAGCAGCGGACGGGGTCAACACAGTCCATGGACCCCTGGCAGGGCGATGAGAT  
CGGTGAACTAGGGGACAAAAGGAAGTTACAGATCTACAAGAGATCGAGAGTTCGTTGGTT  
TGT

>gnl|SRA|SRR11085797.8568962.2 8568962 (*Biological*)  
ACAAACCAACGAACCTCTCGATCTCTTGTAGATCTGTAACTTCCCTTTGTCCCCTAGTTCA  
CCGATCTCATGCCCTGCCAGGGGTCCATGGACTGTGTTGACCCCGTCCGCTGCTCCCC  
CAT

[Download](#) ▾ [GenBank](#) [Graphics](#)

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

Sequence ID: [XM\\_033115407.1](#) Length: 2670 Number of Matches: 1

Range 1: 2468 to 2550 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand	
132 bits(71)					Plus/Minus
Query 6	GGAGCAGCGGACGGGGTCAACACAGTCCATGGACCCCTGGCAGGGCGATGAGATCGGTG	65			
Sbjct 2550					
Query 66	GGAGCAGAGGACAGGATCAACACAGTCCATGGACCCCTGGCAGGGCGATGAGATCGGTG	2491			
Sbjct 2490	AACTAGGGACAAAAGGAAGTTA 88				
	AACTAGGGACAAAAGGAGGTTA 2468				

[Download](#) ▾ [GenBank](#) [Graphics](#)

Bat coronavirus RaTG13, complete genome

Sequence ID: [MN996532.1](#) Length: 29855 Number of Matches: 1

Range 1: 12 to 47 [GenBank](#) [Graphics](#)

▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand	
67.6 bits(36)					Plus/Minus
Query 88	ACAGATCTACAAGAGATCGAGAGTTCGTTGGTTGT	123			
Sbjct 47		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  
GTGGCACGTTGAGCAAAATGTTAGTTCTGG

>gnl|SRA|SRR11085797.10676687.2 10676687 (*Biological*)  
ATGAAGGCAATTCAACCATTCCATCCTCTAGCTGATAATAAAATTGCACTGACTTGCTTTA  
GCACTGATGTGGCTGAGCTACTTCATTGCTCTTCAGGCTATTGCACTGACGTACGCGTTCC  
ATGTGGTCATTCAATCCAGAAACTAACATT

<a href="#">Download</a>	<a href="#">GenBank</a>	<a href="#">Graphics</a>	Sort by:	E value	▼
<b>Bat coronavirus RaTG13, complete genome</b>					
Sequence ID: <a href="#">MN996532.1</a> Length: 29855 Number of Matches: 2					
Range 1: 26736 to 26844 <a href="#">GenBank</a> <a href="#">Graphics</a>					
Score	Expect	Identities	Gaps	Strand	
161 bits(87)	7e-36	102/109(94%)	2/109(1%)	Plus/Plus	
Query 45	GCAC TGACTT G- CTT -TAGCA CTGAT GTGG CTGAGC TACTTC ATTG CTTT CAGG CTA				102
Sbjct 26736	GCAATGGCTTGTCTTGTAGGCTTGTGGCTGAGCTACTTCATTGCTTTCAGGCTA				26795
Query 103	TTTG CACGTAC CGC TTCC ATG GGT CATT CAAT CC CAG AAA ACT AAC ATT				151
Sbjct 26796	TTTG CACGTAC CGC TTCC ATG GGT CATT CAAT CC CAG AAA ACT AAC ATT				26844
Range 2: 27478 to 27542 <a href="#">GenBank</a> <a href="#">Graphics</a>					
Score	Expect	Identities	Gaps	Strand	
121 bits(65)	1e-23	65/65(100%)	0/65(0%)	Plus/Plus	
Query 1	ATGA AGGC AATT CACC ATT CCAT CCT TAG CTGATA ATA AA ATT TGCA CTGACT TGCT TTA				60
Sbjct 27478	ATGA AGGC AATT CACC ATT CCAT CCT TAG CTGATA ATA AA ATT TGCA CTGACT TGCT TTA				27537
Query 61	GC ACT 65				
Sbjct 27538	GC ACT 27542				

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			<a href="#">GenBank</a>	<a href="#">Graphics</a>	<a href="#">▼ Next Match</a>	<a href="#">▲ Previous Match</a>
Score 233 bits(126)	Expect 2e-57	Identities 129/130(99%)	Gaps 1/130(0%)	Strand Plus/Plus		
Query 23	TTC-TCTAACGAACAAACTAAAAATGTCTGATAATGGACCCCAAAACCAACGAAATGCAC				81	
Sbjct 28217	TTTACGTTGGTGACCCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCA				28276	
Query 82	CCCCCATTACGTTGGTGACCCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCA				141	
Sbjct 28277	CCCCCATTACGTTGGTGACCCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCA				28336	
Query 142	GTGGAGCACG	151				
Sbjct 28337	GTGGAGCACG	28346				

Range 2: 25 to 60			<a href="#">GenBank</a>	<a href="#">Graphics</a>	<a href="#">▼ Next Match</a>	<a href="#">▲ Previous Match</a>	<a href="#">▲ First Match</a>
Score	Expect	Identities	Gaps		Strand		
67.6 bits(36)	2e-07	36/36(100%)	0/36(0%)		Plus/Plus		
Query 1	CTCTCGATCTTGTAGATCTGTTCTCTAAACGAAC	36					
Subject 25	CTCTCGATCTTGTAGATCTGTTCTCTAAACGAAC	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)  
 CATCAAACGTAGGGTTTCAGCAAGGCAAAAGATAGCCAGCAACAAAACAAAAAGGCATCCTA  
 CTGAATGGAAGCAGATAATTGCCAATAGTACATCAGTAAGGAGTTAATATTAAGAATTAG  
 TTTTTAAAAAGCTCTATATGATGTCAGAAAT

>gnl|SRA|SRR11085797.76.2 76 (Biological)  
 GTTTTCACTTGCATTTCTCTAATAATTAGTGATGTTGAGCATCTTTCATATGTCTATTG  
 GCCATCTGTATGTCGTCTTGGAGAAATGCTCTATTAGATTTCTGCCAATTTTAATTG  
 GCTTGTGTTGTTTTGTTTTGAATTGAGTT

Descriptions	Graphic Summary	Alignments	Taxonomy	Download	Manage Columns	Show 100	?	
Sequences producing significant alignments				GenBank	Graphics	Distance tree of results		
<input checked="" type="checkbox"/> select all 11 sequences selected								
	Description			Max Score	Total Score	Query Cover	E value	
				Per. Ident			Accession	
<input checked="" type="checkbox"/>	Rhinolophus ferrumequinum clone VMRC7-71A7, complete sequence			267	267	99%	1e-67	98.69% AC150307_3
<input checked="" type="checkbox"/>	Rhinolophus ferrumequinum clone VMRC7-251C10, complete sequence			185	185	94%	4e-43	90.14% AC149630_3
<input checked="" type="checkbox"/>	Myotis lucifugus clone CH235-427D16, complete sequence			137	137	82%	1e-28	87.20% AC174832_3
<input checked="" type="checkbox"/>	Pteropus alecto clone BAC P100M20, BAC P103A18, complete sequence			135	135	86%	4e-28	85.38% KP862827_1
<input checked="" type="checkbox"/>	Pteropus alecto clone BAC P201M3, BAC P216K21, complete sequence			135	135	86%	4e-28	85.38% KP862826_1
<input checked="" type="checkbox"/>	Pteropus alecto clone BAC P56N20, complete sequence			135	135	86%	4e-28	85.38% KP862825_1
<input checked="" type="checkbox"/>	Rhinolophus euryale isolate REM0134 microsatellite RM1198 sequence			128	128	47%	7e-26	98.61% KC910215_1
<input checked="" type="checkbox"/>	Pteropus alecto clone BAC P21207-1, BAC P229M21, complete sequence			126	126	84%	3e-25	84.38% KP862828_1
<input checked="" type="checkbox"/>	PREDICTED: Miniopterus natalensis zinc finger protein 713 (ZNF713) transcript variant X3, mRNA			106	106	90%	3e-19	81.02% XM_016196283_1
<input checked="" type="checkbox"/>	PREDICTED: Miniopterus natalensis zinc finger protein 713 (ZNF713) transcript variant X2, mRNA			106	106	90%	3e-19	81.02% XM_016196281_1
<input checked="" type="checkbox"/>	PREDICTED: Miniopterus natalensis zinc finger protein 713 (ZNF713) transcript variant X1, mRNA			106	106	90%	3e-19	81.02% XM_016196280_1

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

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

Date	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_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/VJY2	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/l96Y	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/tKhoa	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/nhv2	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/velPWV	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/2ggoA	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 RaTG13-1-1_7896-1-F1_2017-06-20_E03
20-Jun-17	99.61%	99%	97.43%	99%	32	archive.is/ehzBr	gnl SRA SRR11806578.32 RaTG13-4-1_7896-4_F_2017-06-20_F03
20-Jun-17	99.87%	98%	97.42%	98%	33	archive.is/do9Rt	gnl SRA SRR11806578.33 RaTG13-4-1_7896-4_R_2017-06-20_H03
20-Jun-17	99.90%	98%	97.44%	98%	29	archive.is/HjQD8	gnl SRA SRR11806578.29 RaTG13-1-1_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-R_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

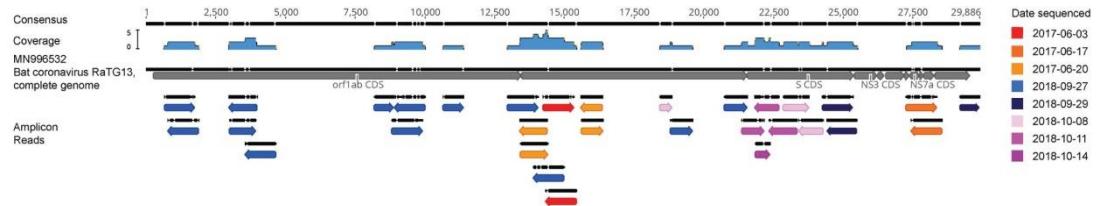


Figure 15: A complete analysis [4] of all Amplicon sequences in SRX8357956. Including the location of these amplicons and the similarity of such amplicon to the RaTG13 and SARS-CoV-2 genome.

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.

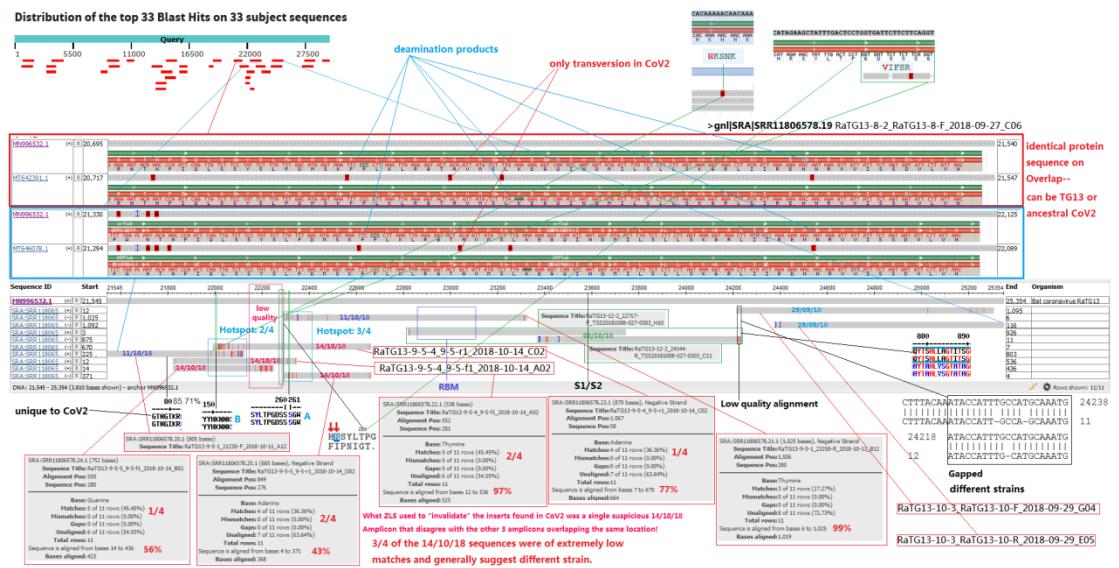


Figure 16: a thorough analysis of the amplicons located on the S locus of RaTG13 in SRX8357956. Notice that the last 4 amplicons sequenced in 14/10/2018 was of very low quality matches, and matched other organisms—including Mouse(*mus musculus*).

<input checked="" type="checkbox"/> select all	99 sequences selected	GenBank		Graphics		Distance tree of results	
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	<a href="#">Gadus morhua genome assembly_chromosome_16</a>	60.8	105	27%	4e-05	78.05%	<a href="#">LR633958.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mus musculus BAC clone RP24-49B8P8 from chromosome 9_complete sequence</a>	59.9	59.9	47%	1e-04	72.92%	<a href="#">AC168217.2</a>
<input checked="" type="checkbox"/>	<a href="#">Mus musculus chromosome 9_clone RP24-484G16_complete sequence</a>	59.9	59.9	47%	1e-04	72.92%	<a href="#">AC137678.11</a>

<input checked="" type="checkbox"/> select all	100 sequences selected			GenBank		Graphics		Distance tree of results	
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident		Accession	
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens BAC clone RP11-792A8 from 7. complete sequence</a>	54.5	54.5	44%	0.006	71.15%	<a href="#">AC027644.9</a>		
<input checked="" type="checkbox"/>	<a href="#">Coregonus sp. 'balchen' genome assembly_chromosome: 7</a>	52.7	52.7	29%	0.022	73.68%	<a href="#">LR778259.1</a>		
<input checked="" type="checkbox"/>	<a href="#">Coregonus sp. 'balchen' genome assembly_chromosome: 15</a>	50.9	50.9	16%	0.078	83.33%	<a href="#">LR778267.1</a>		
<input checked="" type="checkbox"/>	<a href="#">Salmo trutta genome assembly_chromosome: 21</a>	50.9	50.9	37%	0.078	71.90%	<a href="#">LR584437.1</a>		
<input checked="" type="checkbox"/>	<a href="#">Xanthophylomyces dendrophorus genome assembly_Xden1_scaffold_Scaffold_79</a>	50.9	50.9	9%	0.078	96.88%	<a href="#">LN483167.1</a>		
<input checked="" type="checkbox"/>	<a href="#">Coregonus sp. 'balchen' genome assembly_chromosome: 20</a>	50.0	50.0	72%	0.078	67.49%	<a href="#">LR778272.1</a>		
<input checked="" type="checkbox"/>	<a href="#">Aquila chrysaetos chrysaetos genome assembly_chromosome: 14</a>	50.0	141	42%	0.078	69.50%	<a href="#">LR606194.1</a>		
<input checked="" type="checkbox"/>	<a href="#">Bos mutus isolate yakQH1 chromosome 16</a>	50.0	50.0	24%	0.078	74.68%	<a href="#">CP027084.1</a>		
<input checked="" type="checkbox"/>	<a href="#">Mus musculus BAC clone RP23-128D11 from 7. complete sequence</a>	50.0	50.0	36%	0.078	71.90%	<a href="#">AC122222.6</a>		
<input checked="" type="checkbox"/>	<a href="#">Mus musculus BAC clone RP23-66E21 from 7. complete sequence</a>	50.0	50.0	36%	0.078	71.90%	<a href="#">AC131741.4</a>		

<input checked="" type="checkbox"/> select all	24 sequences selected	GenBank		Graphics		Distance tree of results		
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession	
<input checked="" type="checkbox"/>	<a href="#">Mus musculus targeted KO-first, conditional ready, lacZ-tagged mutant allele Fabp4:tm1a(KOMP)Wtsi</a>	50.9	50.9	23%	0.045	85.11%	<a href="#">JN963014.1</a>	
<input checked="" type="checkbox"/>	<a href="#">Mus musculus targeted non-conditional, lacZ-tagged mutant allele Fabp4:tm1e(KOMP)Wtsi_transger</a>	50.9	50.9	23%	0.045	85.11%	<a href="#">JN947213.1</a>	
<input checked="" type="checkbox"/>	<a href="#">Mus musculus chromosome 3 clone RP23-436F15_complete sequence</a>	50.9	50.9	23%	0.045	85.11%	<a href="#">AC123726.11</a>	
<input checked="" type="checkbox"/>	<a href="#">Mus musculus chromosome 3 clone RP24-137C19_complete sequence</a>	50.9	50.9	23%	0.045	85.11%	<a href="#">AC113990.10</a>	

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

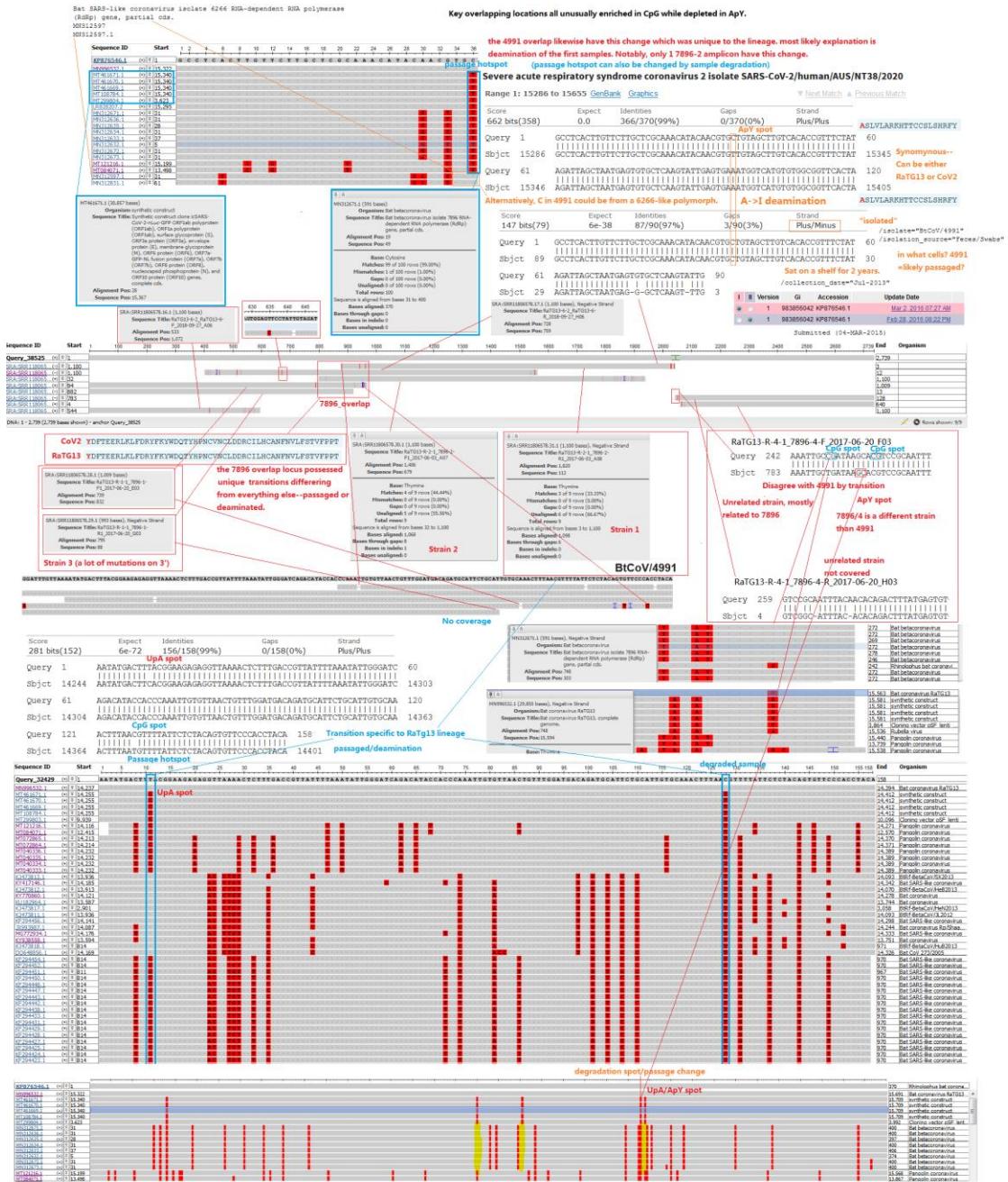


Figure 17: Detailed analysis of the early amplicons located in the nsp12 RdRp region of RaTG13 in SRX8357956.

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

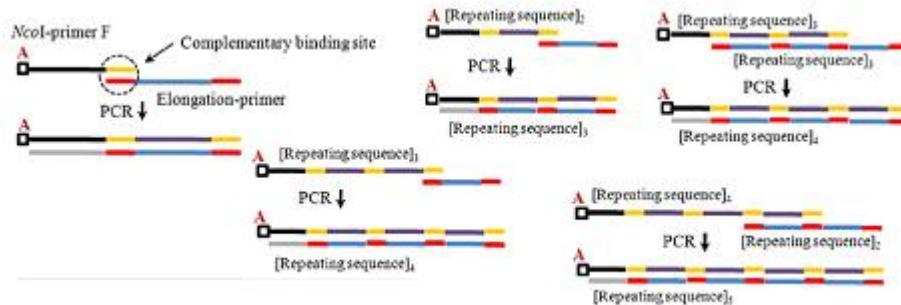


Figure 18A: Origin of repeating sequences in SRX7724752.

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.

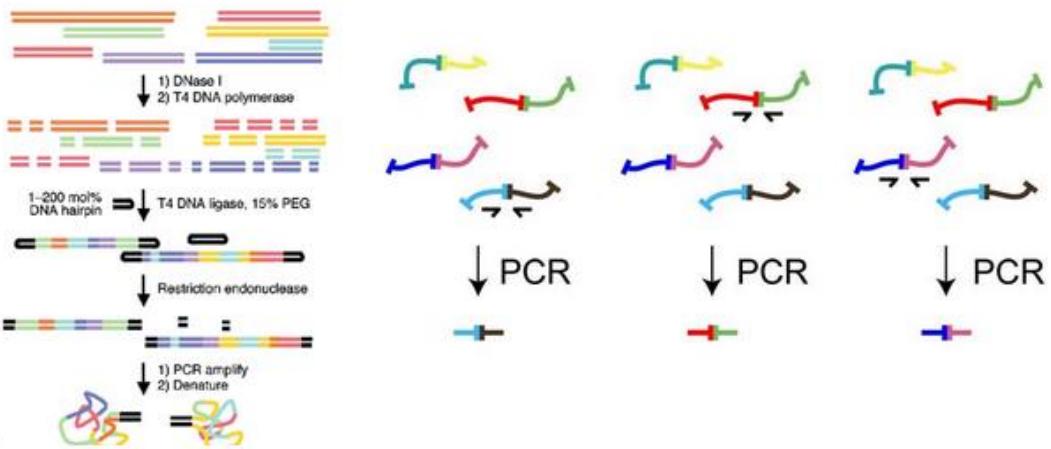


Figure 18B: Origin of the random matched sequences and partial sequences in SRX7724752  
The majority of the sequences that were not repeats, when BLASTed, does not match any known organisms. There were also many sequences that matches—only partially, to many diverse organisms. What was striking, is, 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	<a href="#">J85DMESK016</a>	Search expires on 08-01 20:59 pm	<a href="#">Download All</a> ▾
Results for	2:lcl Query_5937 gnl SRA SRR11085797.66.2 66 (Biological)(150bp) ▾		
Program	BLASTN	?	<a href="#">Citation</a> ▾
Database	nt	<a href="#">See details</a> ▾	
Query ID	lcl Query_5937		
Description	gnl SRA SRR11085797.66.2 66 (Biological)		
Molecule type	dna		
Query Length	150		
Other reports	<a href="#">Distance tree of results</a>	?	

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

ttagattttctactaaacgaaaaactaaaatgtctgtataatggaccggaaaaccacaaacgaaatgcaccccgattacgttttgtggaccct  
CTCTCGATCTCTGTAGATCTGTC TCTAAACGAC  
ACAAACCAACGAACTCTCGATCTCTGTAGATCTGT  
TAACCTCTTCTGTCCCTAGTCACCGATCTCATGCCCTGCGCAGGGTCCATGGACTGTGTGATCTGTCTCTGCTCTGCTCTGCTCTGCT

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.

<input checked="" type="checkbox"/> select all	100 sequences selected	GenBank		Graphics		Distance tree of results	
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Pangolin coronavirus isolate PCoV_GX-P3B_genomic_sequence	67.6	67.6	100%	8e-09	100.00%	MT072865_1
<input checked="" type="checkbox"/>	Pangolin coronavirus isolate PCoV_GX-P2V_complete_genome	67.6	67.6	100%	8e-09	100.00%	MT072864_1
<input checked="" type="checkbox"/>	Pangolin coronavirus isolate PCoV_GX-P5E_complete_genome	67.6	67.6	100%	8e-09	100.00%	MT040336_1
<input checked="" type="checkbox"/>	Pangolin coronavirus isolate PCoV_GX-P5L_complete_genome	67.6	67.6	100%	8e-09	100.00%	MT040335_1
<input checked="" type="checkbox"/>	Pangolin coronavirus isolate PCoV_GX-P1E_complete_genome	67.6	67.6	100%	8e-09	100.00%	MT040334_1
<input checked="" type="checkbox"/>	Pangolin coronavirus isolate PCoV_GX-P4L_complete_genome	67.6	67.6	100%	8e-09	100.00%	MT040333_1
<input checked="" type="checkbox"/>	Bat coronavirus RaTG13_complete_genome	67.6	67.6	100%	8e-09	100.00%	MN996532_1
<input checked="" type="checkbox"/>	Bat SARS-like coronavirus isolate bat-SL-CoVZXC21_complete_genome	67.6	67.6	100%	8e-09	100.00%	MG772934_1
<input checked="" type="checkbox"/>	Bat SARS-like coronavirus isolate bat-SL-CoVZC45_complete_genome	67.6	67.6	100%	8e-09	100.00%	MG772933_1
<input checked="" type="checkbox"/>	Bat coronavirus (RtCoV/B/279/2005)_complete_genome	65.8	65.8	97%	3e-08	100.00%	DQ648857_1
<input checked="" type="checkbox"/>	Mutant SARS coronavirus Urbani clone SARS-Urbani-MA_SHC014-spike_complete_genome	63.9	63.9	94%	1e-07	100.00%	MT308984_1
<input checked="" type="checkbox"/>	Coronavirus_BiRs-BetaCoV/YN2018D_complete_genome	63.9	63.9	94%	1e-07	100.00%	MK211378_1

Figure 20b: BLAST result of CTCTCGATCTCTTGTAGATCTGTTCTAAACGAAC.

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)
GCCTTCGTTTGTATATAGTTTAATGCAAATCCCCTAACATCTCTTTCAGCATCTGCTGC
ACCTCTTCAACCAGCAACTGTAGAAAATCTTAAAGGGCTTTGTTTTTACCAACTT
GTTAAAAATATCTGCTTGAATATTTGT

>gnl|SRA|SRR11085733.2232944.2 2232944 (Biological)
AGAGGTCCTACTCTTTACAAGATACTTGGCTTTAGAAAAACTTGCACATTTCGATAGG
GAAAGGATACCAGAAAGAGTTGTGCACGCTAAAGGAAGTGCTGCATACGGCGAATTAAACA
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 176 bits(95)	Expect 3e-40	Identities 123/137(90%)	Gaps 0/137(0%)	Strand Plus/Minus	
Query 7	CCTACTCTTTACAAGATACTGGCTTTAGAAAAACTTGCACATTCGATAGGGAAAGG				66
Sbjct 47856					47797
Query 67	ATACCAGAAAAGAGTTGTGCACGCTAAAGGAAGTGCTGCATACGGCGAATTAAACAATTACT				126
Sbjct 47796					47737
Query 127	AATGATATTACAAAATA 143				
Sbjct 47736	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="MSKKFTTATGTPLDNQNSITAGKKGPTLLQDTWLLEKLAHFDR
ERIPERVVHAKGSAYGELITITNDITQYTKAELFNVKGKTKAFLRFSVVAGERGAAD
AERDVVRGFALKLYTNEGNWDIVGNNTPVFFIKDAIKPPDFIHTQKRDPKTNMKSPTAM
WDFWSLHPESILHQVTILMSDRGIPRSYREMHGFGSHTYSFINAKNERFWVKFHVCLQ
GIHNLTNKESEAVIAKDRESHQKDLFENIEKGKGNFPKWRFCIQVMSEKEAENYRFNPFD
LTKVWSHKDYPLIEVGILELNKNPENYFAEVEQAAFNPANIVPGVGYSYPDVKVLQGRLF
SYGDTQRYRLGINHTQLPVNAPIVPVNNTHRDGFMQQQGFGDRRNYESPYFNDYVEDK
NALEPPLFVQEVDVMYKYDHREYEDDYFVQAGDLYRLMTAEQKEALCQNIKESMEGVP
DEIKKRQLEHFKKADKAYGKRVAELLGL"
ORIGIN
1 tatttagtaa tatcattttt aattgttaat tcaccatatg ctgcacttcc tttagcgtgc
61 accactctct ctgggattct ctctctatca aaatgtcaa gttttctaa 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%

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



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](#) ▾

Job Title	gb AC097711.2
RID	<a href="#">KJ2J1YCV01R</a> Search expires on 08-17 18:28 pm <a href="#">Download All</a> ▾
Program	BLASTN ? <a href="#">Citation</a> ▾
Database	SRA <a href="#">See details</a> ▾
Query ID	<a href="#">AC097711.2</a>
Description	Homo sapiens BAC clone RP11-162K6 from 4, complete sequence
Molecule type	nucleic acid
Query Length	114657
Other reports	<a href="#">Distance tree of results</a> <a href="#">MSA viewer</a> ?

[How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

**Filter Results**

Organism *only top 20 will appear*  exclude

+ Add organism

Percent identity	E value	Query Coverage
<input type="text"/> to <input type="text"/>	<input type="text"/> to <input type="text"/>	<input type="text"/> to <input type="text"/>
<input type="button" value="Filter"/> <input type="button" value="Reset"/>		

Descriptions
Graphic Summary
Alignments

**Sequences producing significant alignments**

select all 100 sequences selected
[Download](#) ▾
[Manage Columns](#) ▾
Show 100 ▾
?

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	SRX7724752	279	279	0%	1e-70	100.00%	<a href="#">SRA:SRR11085797.11044608.2</a>
<input checked="" type="checkbox"/>	SRX7724752	279	279	0%	1e-70	100.00%	<a href="#">SRA:SRR11085797.11044608.1</a>

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 See details

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**

**Organism** only top 20 will appear  exclude  
Type common name, binomial, taxid or group name  
+ Add organism

Percent Identity E value Query Coverage

[ ] to [ ] [ ] to [ ] [ ] to [ ]

Filter Reset

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100 ?

select all 11 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-162K6 from 4 .complete sequence	279	279	100%	2e-71	100.00%	AC097711.2
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 32a	274	274	100%	9e-70	99.34%	CP050598.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 32b	274	274	100%	9e-70	99.34%	CP050634.1
<input checked="" type="checkbox"/>	Aquila chrysaetos chrysaetos genome assembly_chromosome_1	257	257	100%	9e-65	97.35%	LR606181.1
<input checked="" type="checkbox"/>	Apteryx australis mantelli genome assembly_AptMant0_scaffold scaffold176	257	257	100%	9e-65	97.35%	LK064748.1
<input checked="" type="checkbox"/>	Erithacus rubecula genome assembly_chromosome_5	252	252	100%	4e-63	96.69%	LR812107.1
<input checked="" type="checkbox"/>	Anas platyrhynchos genome assembly_chromosome_4	252	252	100%	4e-63	96.69%	LS423614.1
<input checked="" type="checkbox"/>	Streptopelia turtur genome assembly_chromosome_4	246	246	100%	2e-61	96.03%	LR594554.1
<input checked="" type="checkbox"/>	Mus musculus BAC clone RP24-204J10 from 5 .complete sequence	243	243	98%	3e-60	95.97%	AC121929.2
<input checked="" type="checkbox"/>	Sciurus carolinensis genome assembly_chromosome_15	204	204	78%	1e-48	97.48%	LR738605.1
<input checked="" type="checkbox"/>	PREDICTED: Meleagris gallopavo uncharacterized LOC104910685 (LOC104910685). mRNA	121	121	45%	1e-23	98.53%	XM_019615117.2

← Edit Search Save Search Search Summary ?

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

Job Title gnl|SRA|SRR11085797.11044608.1 11044608

RID KJ2FP0KE014 Search expires on 08-17 18:27 pm Download All

Results for 1:lcl|Query\_54298 gnl|SRA|SRR11085797.11044608.1 11044608 (Biologica)

Program BLASTN ? Citation

Database nt See details

Query ID lcl|Query\_54298

Description gnl|SRA|SRR11085797.11044608.1 11044608 (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 [ ] [ ] to [ ] [ ] to [ ]

Filter Reset

**Descriptions** Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Manage Columns Show 100 ?

select all 4 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-162K6 from 4 .complete sequence	279	279	100%	2e-71	100.00%	AC097711.2
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 32a	274	274	100%	9e-70	99.34%	CP050598.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 32b	274	274	100%	9e-70	99.34%	CP050634.1
<input checked="" type="checkbox"/>	Mus musculus BAC clone RP24-204J10 from 5 .complete sequence	252	252	100%	4e-63	96.69%	AC121929.2

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

[Edit Search](#) Save Search Search Summary

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, m...  
Molecule type dna  
Query Length 19578880  
Other reports [Distance tree of results](#) [MSA viewer](#)

**Descriptions** Graphic Summary Alignments

**Sequences producing significant alignments** Download Manage Columns Show 100 [?](#)

select all 100 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	SRX7724752	279	279	0%	2e-68	100.00%	SRA:SRR11085797.7041148.2
<input checked="" type="checkbox"/>	SRX7724752	278	278	0%	9e-68	100.00%	SRA:SRR11085797.4686606.2
<input checked="" type="checkbox"/>	SRX7724752	276	276	0%	3e-67	100.00%	SRA:SRR11085797.8742622.2

[Edit Search](#) Save Search Search Summary

Job Title 2 sequences (gnl|SRA|SRR11085797.10431565.1...)

RID KJ3PSXP014 Search expires on 08-17 18:47 pm [Download All](#)

Results for 1:cl|Query\_51896 gnl|SRA|SRR11085797.10431565 (Biological)

Program BLASTN [Citation](#)

Database nt [See details](#)

Query ID cl|Query\_51896

Description gnl|SRA|SRR11085797.10431565.1 10431565 (Biological)

Molecule type dna

Query Length 151

Other reports [Distance tree of results](#)

**Descriptions** Graphic Summary Alignments Taxonomy

**Sequences producing significant alignments** Download Manage Columns Show 100 [?](#)

select all 45 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome_32a	279	279	100%	2e-71	100.00%	CP050598.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome_32b	279	279	100%	2e-71	100.00%	CP050634.1
<input checked="" type="checkbox"/>	Homo sapiens Vista enhancer hs712 (LOC110120752) on chromosome 4	279	279	100%	2e-71	100.00%	NG_053377.1
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-476H13 from 4..complete sequence	279	279	100%	2e-71	100.00%	AC024192.6
<input checked="" type="checkbox"/>	Aquila chrysaetos chrysaetos genome assembly_chromosome_1	274	274	100%	9e-70	99.34%	LR606181.1
<input checked="" type="checkbox"/>	Streptopelia tutur genome assembly_chromosome_4	274	274	100%	9e-70	99.34%	LR594564.1
<input checked="" type="checkbox"/>	PREDICTED: Cyanistes caeruleus uncharacterized LOC111928864 (LOC111928864).ncRNA	274	274	100%	9e-70	99.34%	XR_002864354.1
<input checked="" type="checkbox"/>	Apteryx australis mantelli genome assembly AptMant0_scaffold scaffold564	274	274	100%	9e-70	99.34%	LK065221.1
<input checked="" type="checkbox"/>	Anas platyrhynchos genome assembly_chromosome_4	274	274	100%	9e-70	99.34%	LS423614.1

Job Title 2 sequences (gnl|SRA|SRR11085797.4967337.1...)

RID KJ8SKF2E014 Search expires on 08-17 20:14 pm Download All

Results for 2:lcl|Query\_37435 gnl|SRA|SRR11085797.4967337.2.4967337 (Biological) [?]

Program BLASTN ? Citation

Database nt See details

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 ?

**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 [ ] [ ] to [ ] [ ] to [ ]

Filter Reset

Descriptions		Graphic Summary	Alignments	Taxonomy																																																																																																
<b>Sequences producing significant alignments</b>																																																																																																				
Download Manage Columns Show 100 ?																																																																																																				
<input checked="" type="checkbox"/> select all 58 sequences selected																																																																																																				
<table border="1"> <thead> <tr> <th></th> <th>Description</th> <th>Max Score</th> <th>Total Score</th> <th>Query Cover</th> <th>E value</th> <th>Per. Ident</th> <th>Accession</th> </tr> </thead> <tbody> <tr><td><input checked="" type="checkbox"/></td><td>Canis lupus familiaris breed Labrador retriever chromosome 32a</td><td>279</td><td>279</td><td>100%</td><td>2e-71</td><td>100.00%</td><td>CP050598.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Canis lupus familiaris breed Labrador retriever chromosome 32b</td><td>279</td><td>279</td><td>100%</td><td>2e-71</td><td>100.00%</td><td>CP050634.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Rhesus Macaque BAC CH250-499F11() complete sequence</td><td>274</td><td>274</td><td>100%</td><td>9e-70</td><td>99.34%</td><td>AC204215.5</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Homo sapiens BAC clone RP11-678H22 from 4_ complete sequence</td><td>274</td><td>274</td><td>100%</td><td>9e-70</td><td>99.34%</td><td>AC096766.3</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Mus musculus chromosome 5_ clone RP24-31H14_ complete sequence</td><td>263</td><td>263</td><td>100%</td><td>2e-66</td><td>98.01%</td><td>AC105976.13</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Chrysemys picta isolate 4965chr ultra conserved element locus chr4_11164 genomic sequence</td><td>257</td><td>257</td><td>100%</td><td>9e-65</td><td>97.35%</td><td>JQ873778.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Alligator mississippiensis isolate 333all ultra conserved element locus chr4_11164 genomic sequence</td><td>257</td><td>257</td><td>100%</td><td>9e-65</td><td>97.35%</td><td>JQ869146.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Apteryx australis mantelli genome assembly AptMant0_scaffold scaffold564</td><td>241</td><td>241</td><td>100%</td><td>9e-60</td><td>95.36%</td><td>LK065221.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Erithacus rubecula genome assembly_chromosome_5</td><td>231</td><td>231</td><td>82%</td><td>6e-57</td><td>100.00%</td><td>LR812107.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Streptopelia tutur genome assembly_chromosome_4</td><td>231</td><td>231</td><td>82%</td><td>6e-57</td><td>100.00%</td><td>LR594554.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Aquila chrysaetos chrysaetos genome assembly_chromosome_1</td><td>226</td><td>226</td><td>82%</td><td>3e-55</td><td>99.20%</td><td>LR606181.1</td></tr> </tbody> </table>						Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession	<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 32a	279	279	100%	2e-71	100.00%	CP050598.1	<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 32b	279	279	100%	2e-71	100.00%	CP050634.1	<input checked="" type="checkbox"/>	Rhesus Macaque BAC CH250-499F11() complete sequence	274	274	100%	9e-70	99.34%	AC204215.5	<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-678H22 from 4_ complete sequence	274	274	100%	9e-70	99.34%	AC096766.3	<input checked="" type="checkbox"/>	Mus musculus chromosome 5_ clone RP24-31H14_ complete sequence	263	263	100%	2e-66	98.01%	AC105976.13	<input checked="" type="checkbox"/>	Chrysemys picta isolate 4965chr ultra conserved element locus chr4_11164 genomic sequence	257	257	100%	9e-65	97.35%	JQ873778.1	<input checked="" type="checkbox"/>	Alligator mississippiensis isolate 333all ultra conserved element locus chr4_11164 genomic sequence	257	257	100%	9e-65	97.35%	JQ869146.1	<input checked="" type="checkbox"/>	Apteryx australis mantelli genome assembly AptMant0_scaffold scaffold564	241	241	100%	9e-60	95.36%	LK065221.1	<input checked="" type="checkbox"/>	Erithacus rubecula genome assembly_chromosome_5	231	231	82%	6e-57	100.00%	LR812107.1	<input checked="" type="checkbox"/>	Streptopelia tutur genome assembly_chromosome_4	231	231	82%	6e-57	100.00%	LR594554.1	<input checked="" type="checkbox"/>	Aquila chrysaetos chrysaetos genome assembly_chromosome_1	226	226	82%	3e-55	99.20%	LR606181.1
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession																																																																																													
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 32a	279	279	100%	2e-71	100.00%	CP050598.1																																																																																													
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 32b	279	279	100%	2e-71	100.00%	CP050634.1																																																																																													
<input checked="" type="checkbox"/>	Rhesus Macaque BAC CH250-499F11() complete sequence	274	274	100%	9e-70	99.34%	AC204215.5																																																																																													
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-678H22 from 4_ complete sequence	274	274	100%	9e-70	99.34%	AC096766.3																																																																																													
<input checked="" type="checkbox"/>	Mus musculus chromosome 5_ clone RP24-31H14_ complete sequence	263	263	100%	2e-66	98.01%	AC105976.13																																																																																													
<input checked="" type="checkbox"/>	Chrysemys picta isolate 4965chr ultra conserved element locus chr4_11164 genomic sequence	257	257	100%	9e-65	97.35%	JQ873778.1																																																																																													
<input checked="" type="checkbox"/>	Alligator mississippiensis isolate 333all ultra conserved element locus chr4_11164 genomic sequence	257	257	100%	9e-65	97.35%	JQ869146.1																																																																																													
<input checked="" type="checkbox"/>	Apteryx australis mantelli genome assembly AptMant0_scaffold scaffold564	241	241	100%	9e-60	95.36%	LK065221.1																																																																																													
<input checked="" type="checkbox"/>	Erithacus rubecula genome assembly_chromosome_5	231	231	82%	6e-57	100.00%	LR812107.1																																																																																													
<input checked="" type="checkbox"/>	Streptopelia tutur genome assembly_chromosome_4	231	231	82%	6e-57	100.00%	LR594554.1																																																																																													
<input checked="" type="checkbox"/>	Aquila chrysaetos chrysaetos genome assembly_chromosome_1	226	226	82%	3e-55	99.20%	LR606181.1																																																																																													

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

### Reads (separated)

>gnl|SRA|SRR11085797.6341838.1 6341838 (Biological)  
 CGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAAATACAAAAAAATTAGC  
 CGGGCGTGTGGCGGGCGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATGG  
 CGTGAACCCGGGAGGCCGGAGCNTGCACTG

>gnl|SRA|SRR11085797.6341838.2 6341838 (Biological)  
 CTCACTGCAAGCTCCGCCTCCGGGTTCACGCCATTCTCCTGCCTAGCCTCCCGAGTAG  
 CTGGGACTACAGGCGCCGCCATCACGCCCGCTAATTTTTGTATTTTAGAGAGACG  
 GGGTTCACCGTGTAGCCAGGATGGCTCG

Description gnl|SRA|SRR11085797.6341838.2 6341838 (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 [ ] [ ] to [ ] [ ] to [ ]

Filter Reset

Descriptions		Graphic Summary	Alignments	Taxonomy																																																
<b>Sequences producing significant alignments</b>																																																				
Download Manage Columns Show 100 ?																																																				
<input checked="" type="checkbox"/> select all 100 sequences selected																																																				
<table border="1"> <thead> <tr> <th></th> <th>Description</th> <th>Max Score</th> <th>Total Score</th> <th>Query Cover</th> <th>E value</th> <th>Per. Ident</th> <th>Accession</th> </tr> </thead> <tbody> <tr><td><input checked="" type="checkbox"/></td><td>Pongo abelii chromosome 10 clone CH276-326B4_ complete sequence</td><td>278</td><td>1689</td><td>100%</td><td>7e-71</td><td>100.00%</td><td>AC270645.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Pongo abelii chromosome 10 clone CH276-12G11_ complete sequence</td><td>278</td><td>1379</td><td>100%</td><td>7e-71</td><td>100.00%</td><td>AC270533.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Pongo abelii chromosome 10 clone CH276-5H9_ complete sequence</td><td>278</td><td>1253</td><td>100%</td><td>7e-71</td><td>100.00%</td><td>AC270518.1</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Homo sapiens beta-1,3-galactosyltransferase 1 (B3GALT1).mRNA</td><td>274</td><td>274</td><td>100%</td><td>9e-70</td><td>99.34%</td><td>NM_020981.4</td></tr> <tr><td><input checked="" type="checkbox"/></td><td>Homo sapiens chromosome 1 clone VMRC53-455P10_ complete sequence</td><td>274</td><td>3094</td><td>100%</td><td>9e-70</td><td>99.34%</td><td>AC278561.1</td></tr> </tbody> </table>						Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession	<input checked="" type="checkbox"/>	Pongo abelii chromosome 10 clone CH276-326B4_ complete sequence	278	1689	100%	7e-71	100.00%	AC270645.1	<input checked="" type="checkbox"/>	Pongo abelii chromosome 10 clone CH276-12G11_ complete sequence	278	1379	100%	7e-71	100.00%	AC270533.1	<input checked="" type="checkbox"/>	Pongo abelii chromosome 10 clone CH276-5H9_ complete sequence	278	1253	100%	7e-71	100.00%	AC270518.1	<input checked="" type="checkbox"/>	Homo sapiens beta-1,3-galactosyltransferase 1 (B3GALT1).mRNA	274	274	100%	9e-70	99.34%	NM_020981.4	<input checked="" type="checkbox"/>	Homo sapiens chromosome 1 clone VMRC53-455P10_ complete sequence	274	3094	100%	9e-70	99.34%	AC278561.1
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession																																													
<input checked="" type="checkbox"/>	Pongo abelii chromosome 10 clone CH276-326B4_ complete sequence	278	1689	100%	7e-71	100.00%	AC270645.1																																													
<input checked="" type="checkbox"/>	Pongo abelii chromosome 10 clone CH276-12G11_ complete sequence	278	1379	100%	7e-71	100.00%	AC270533.1																																													
<input checked="" type="checkbox"/>	Pongo abelii chromosome 10 clone CH276-5H9_ complete sequence	278	1253	100%	7e-71	100.00%	AC270518.1																																													
<input checked="" type="checkbox"/>	Homo sapiens beta-1,3-galactosyltransferase 1 (B3GALT1).mRNA	274	274	100%	9e-70	99.34%	NM_020981.4																																													
<input checked="" type="checkbox"/>	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.

<input checked="" type="checkbox"/> <a href="#">Human endogenous retrovirus H HERV-H/env60 proviral copy clone 734E12</a>	252	252	100%	4e-63	96.69%	<a href="#">AJ289710.2</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic human HSC3N1 Alu sequence</a>	252	252	100%	4e-63	96.69%	<a href="#">U02043.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct complete sequence</a>	250	3635	100%	1e-62	96.69%	<a href="#">JN255744.1</a>
<input checked="" type="checkbox"/> <a href="#">Human artificial chromosome vector 21HAC4 DNA isolated from the long arm clone YAC/BAC#26-2</a>	250	1964	100%	1e-62	96.69%	<a href="#">AB553824.1</a>
<input checked="" type="checkbox"/> <a href="#">Human ORFeome Gateway entry vector pENTR223-MGC2752 complete sequence</a>	246	246	100%	2e-61	96.05%	<a href="#">LT735229.1</a>
<input checked="" type="checkbox"/> <a href="#">Expression vector pJMLIEP DNA complete sequence</a>	246	246	99%	2e-61	96.05%	<a href="#">LC175306.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct Homo sapiens clone ccsBroadEn_10246 MGC2752 gene encodes complete protein</a>	246	246	100%	2e-61	96.05%	<a href="#">KJ900852.1</a>
<input checked="" type="checkbox"/> <a href="#">HIV-1 isolate HK_JIDLNL_S071 from Switzerland nonfunctional gag_protein (gag) gene complete sequence and nonfunctional</a>	244	1072	100%	7e-61	96.00%	<a href="#">MT154980.1</a>
<input checked="" type="checkbox"/> <a href="#">Cloning vector pSuper_7SL_AluAA_7SL enhancer and AluYa5 repeat element sequence</a>	241	241	100%	9e-60	95.36%	<a href="#">EU092258.1</a>
<input checked="" type="checkbox"/> <a href="#">Cloning vector pSuper_7SL_AluA_7SL enhancer and AluYa5 repeat element sequence</a>	241	241	100%	9e-60	95.36%	<a href="#">EU092257.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct clone AluAU SRP promoter region and Alu repeat element sequence</a>	241	241	100%	9e-60	95.36%	<a href="#">AF458115.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct clone AluWD SRP promoter region and Alu repeat element sequence</a>	241	241	100%	9e-60	95.36%	<a href="#">AF458112.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct clone Alu253 SRP promoter region and Alu repeat element sequence</a>	241	241	100%	9e-60	95.36%	<a href="#">AF458107.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct clone Alu+A SRP promoter region and Alu repeat element sequence</a>	241	241	100%	9e-60	95.36%	<a href="#">AF458106.1</a>
<input checked="" type="checkbox"/> <a href="#">Desmodus rotundus isolate DRU21DN04 contig68764 whole genome shotgun sequence</a>	108	216	63%	2e-20	87.23%	<a href="#">PEHR01068758.1</a>
<input checked="" type="checkbox"/> <a href="#">Myotis lucifugus cont2_6286 whole genome shotgun sequence</a>	108	108	55%	2e-20	90.36%	<a href="#">AAPE02006287.1</a>
<input checked="" type="checkbox"/> <a href="#">Artibeus jamaicensis isolate US092 ArtJam_scaffold_27825 whole genome shotgun sequence</a>	104	104	51%	2e-19	90.91%	<a href="#">PVKR01013927.1</a>
<input checked="" type="checkbox"/> <a href="#">Macrotus californicus isolate US035 MacCal_line_566643 whole genome shotgun sequence</a>	102	102	51%	9e-19	90.79%	<a href="#">VMDR010283404.1</a>
<input checked="" type="checkbox"/> <a href="#">Anoura caudifer isolate US021 AnoCau_scaffold_336054 whole genome shotgun sequence</a>	102	102	61%	9e-19	86.96%	<a href="#">PVKU01163203.1</a>
<input checked="" type="checkbox"/> <a href="#">Anoura caudifer isolate US021 AnoCau_scaffold_250182 whole genome shotgun sequence</a>	102	102	61%	9e-19	86.96%	<a href="#">PVKU01121529.1</a>
<input checked="" type="checkbox"/> <a href="#">Anoura caudifer isolate US021 AnoCau_scaffold_157416 whole genome shotgun sequence</a>	102	102	61%	9e-19	86.96%	<a href="#">PVKU01078866.1</a>
<input checked="" type="checkbox"/> <a href="#">Anoura caudifer isolate US021 AnoCau_scaffold_136788 whole genome shotgun sequence</a>	102	102	61%	9e-19	86.96%	<a href="#">PVKU01068554.1</a>
<input checked="" type="checkbox"/> <a href="#">Anoura caudifer isolate US021 AnoCau_scaffold_6229 whole genome shotgun sequence</a>	102	102	51%	9e-19	90.79%	<a href="#">PVKU01003121.1</a>
<input checked="" type="checkbox"/> <a href="#">Anoura caudifer isolate US021 AnoCau_scaffold_1146 whole genome shotgun sequence</a>	102	102	71%	9e-19	84.26%	<a href="#">PVKU01000576.1</a>
<input checked="" type="checkbox"/> <a href="#">Artibeus jamaicensis isolate US092 ArtJam_scaffold_590481 whole genome shotgun sequence</a>	102	102	51%	9e-19	90.79%	<a href="#">PVKR01295479.1</a>
<input checked="" type="checkbox"/> <a href="#">Artibeus jamaicensis isolate US092 ArtJam_scaffold_272373 whole genome shotgun sequence</a>	102	102	51%	9e-19	90.79%	<a href="#">PVKR01136397.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1_000055F_070_arrow_arrow_whole_genome_shotgun_sequence</a>	101	101	51%	8e-19	88.46%	<a href="#">RXPD01003063.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate mRhiFer1_scaffold_m29_p_7 whole genome shotgun sequence</a>	101	151	51%	8e-19	88.46%	<a href="#">JACAGC010000007.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum RF contig_107525 whole genome shotgun sequence</a>	101	101	51%	8e-19	88.46%	<a href="#">AVWA01101756.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate US033 RhiFer1 flattened_line_8799 whole genome shotgun sequence</a>	97.8	180	50%	1e-17	88.16%	<a href="#">VMDN01004402.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1_000061F_062_arrow_arrow_whole_genome_shotgun_sequence</a>	97.8	186	50%	1e-17	88.16%	<a href="#">RXPD01001710.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1_chromosome_6 whole genome shotgun sequence</a>	97.8	309	50%	1e-17	88.16%	<a href="#">RXPC01000086.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate mRhiFer1_scaffold_m29_p_8 whole genome shotgun sequence</a>	97.8	309	50%	1e-17	88.16%	<a href="#">JACAGC010000008.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1_000003F_100_arrow_arrow_whole_genome_shotgun_sequence</a>	93.3	93.3	49%	1e-16	88.00%	<a href="#">RXPD01006157.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate mRhiFer1_scaffold_m29_p_4 whole genome shotgun sequence</a>	93.3	93.3	49%	1e-16	88.00%	<a href="#">JACAGC010000004.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate US033 RhiFer1 flattened_line_6166 whole genome shotgun sequence</a>	90.6	90.6	44%	1e-15	89.55%	<a href="#">VMDN01003085.1</a>
<input checked="" type="checkbox"/> <a href="#">Rhinolophus ferrumequinum isolate MPI-CBG mRhiFer1_000061F_073_arrow_arrow_whole_genome_shotgun_sequence</a>	90.6	90.6	44%	1e-15	89.55%	<a href="#">RXPD01006658.1</a>

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

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

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

## No evidence of methodological reasons for the generation of anomalies in SRX7724752

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

mNGS sequencing data of identical sample, origin, institute and submitted at the same date, located in [PRJNA606159](#).

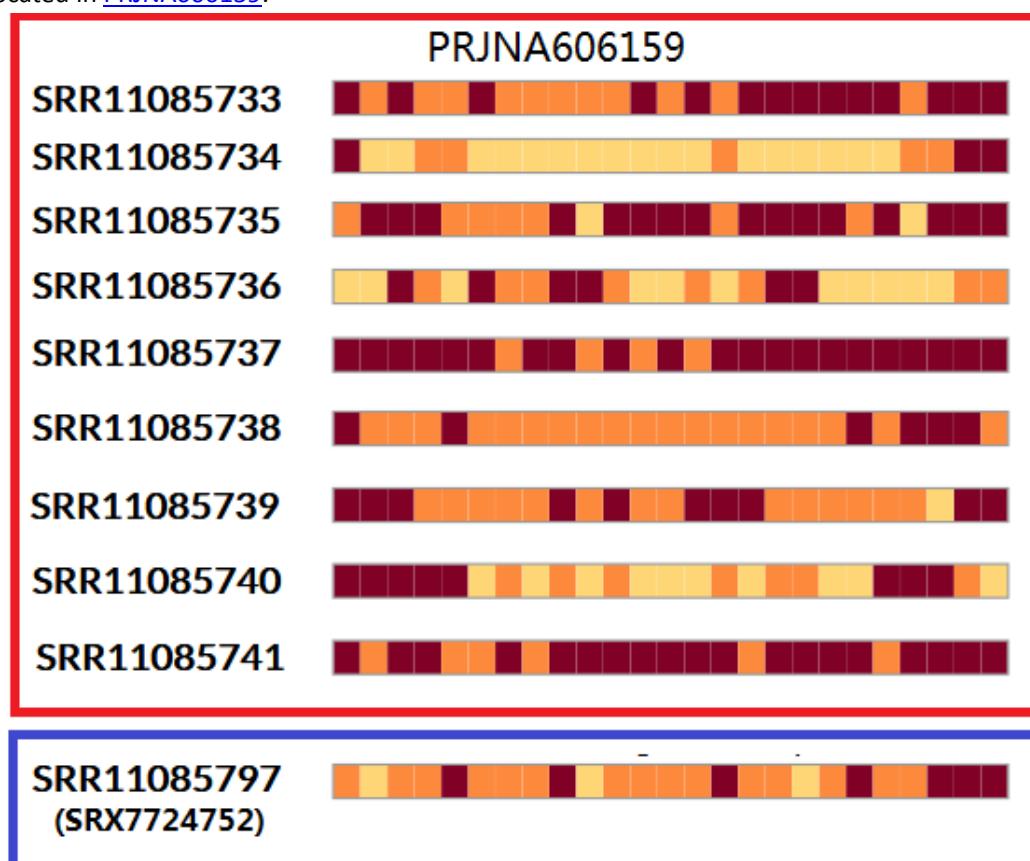


Fig. 26a: the Coverage map of Coronaviridae within the datasets located in PRJNA606159, compared against SRR11085797.

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

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

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

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

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

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

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

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

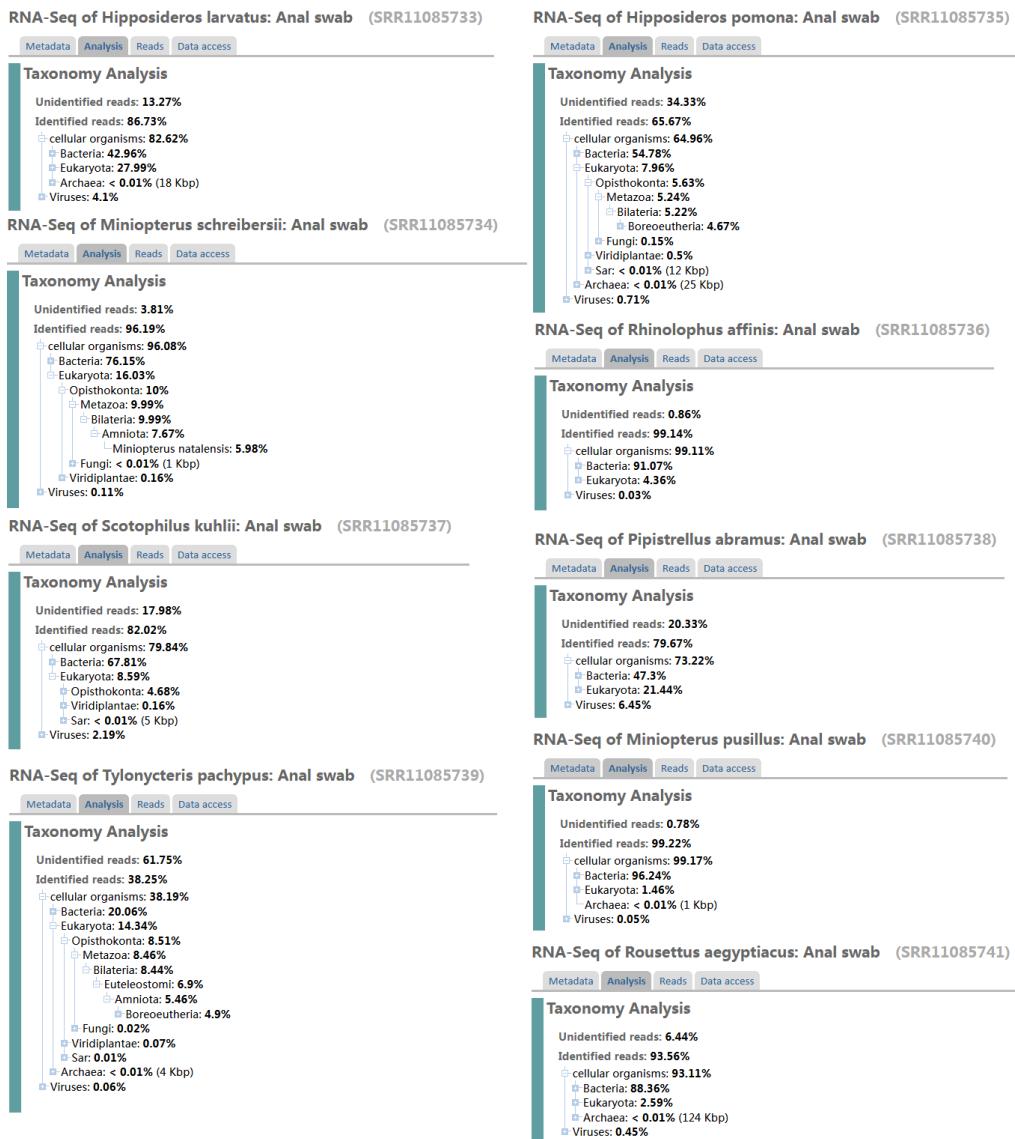


Fig.26c: No evidence of any anomalies were found within the datasets presented in PRJNA606159

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

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

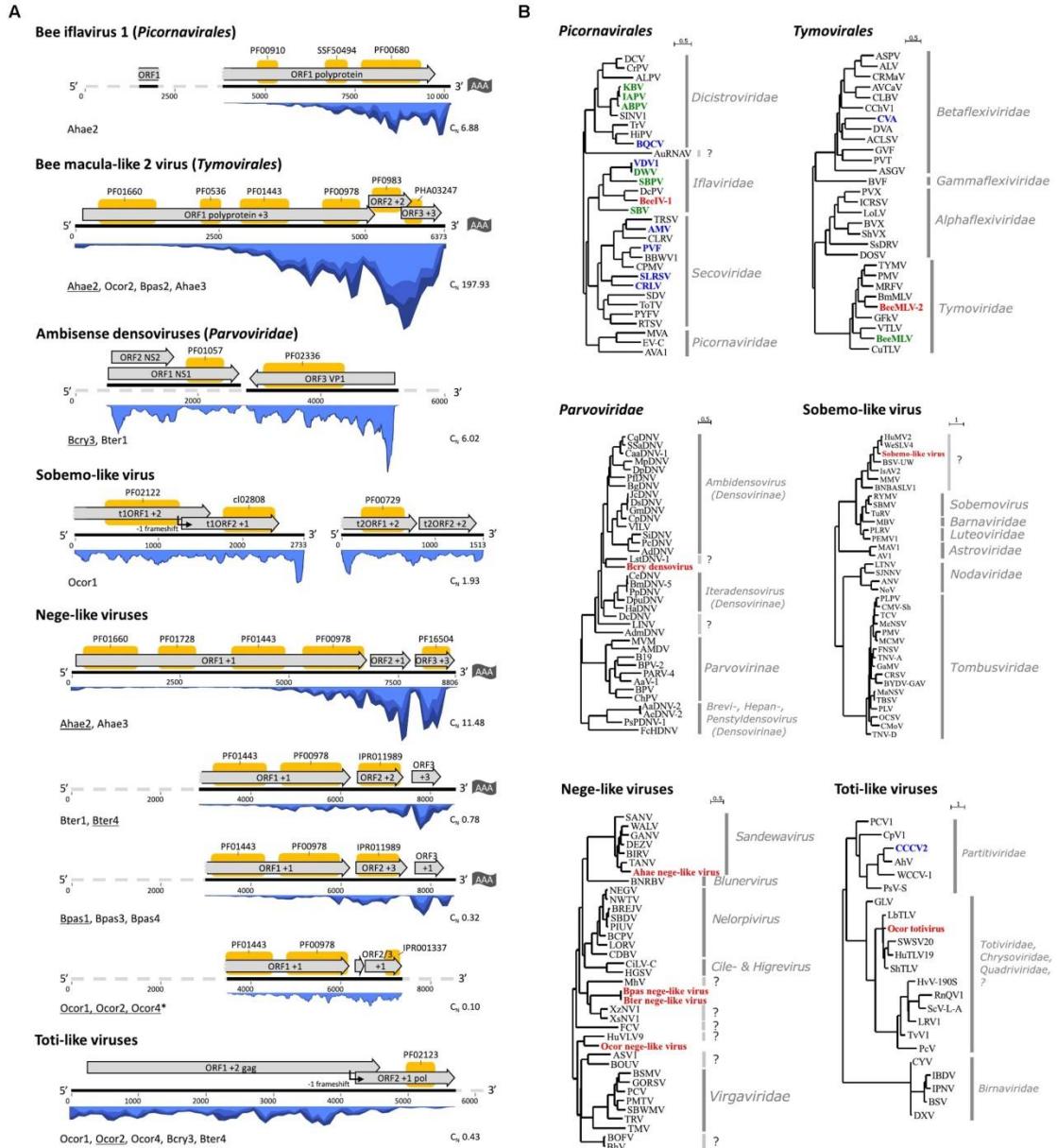


Figure 2 from [9]. A clear bias toward the 3'-end of RNA viral families that contained a polyA tail, was noticed.

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

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

Description	Phaseolus vulgaris endornavirus 1 isolate PvEV-1_Brazil poly ...
Molecule type	nucleic acid
Query Length	14072
Other reports	<a href="#">?</a>



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

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

## Taxonomy Analysis

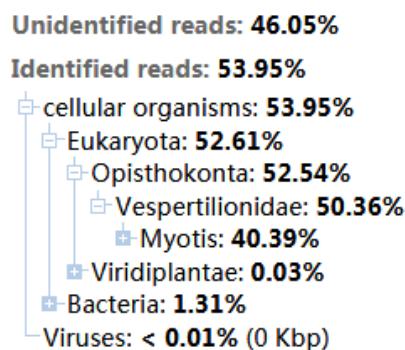


Figure 28a: an example of a bacteria-depleted dataset. An absence of Riboviria reads was noted.

## Taxonomy Analysis

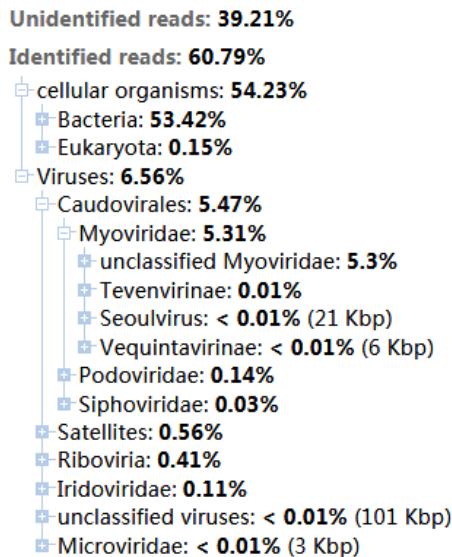


Figure 28b: in contrast, Riboviria reads are found only in datasets that contained a significant amount of bacteria.

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

## Taxonomy Analysis

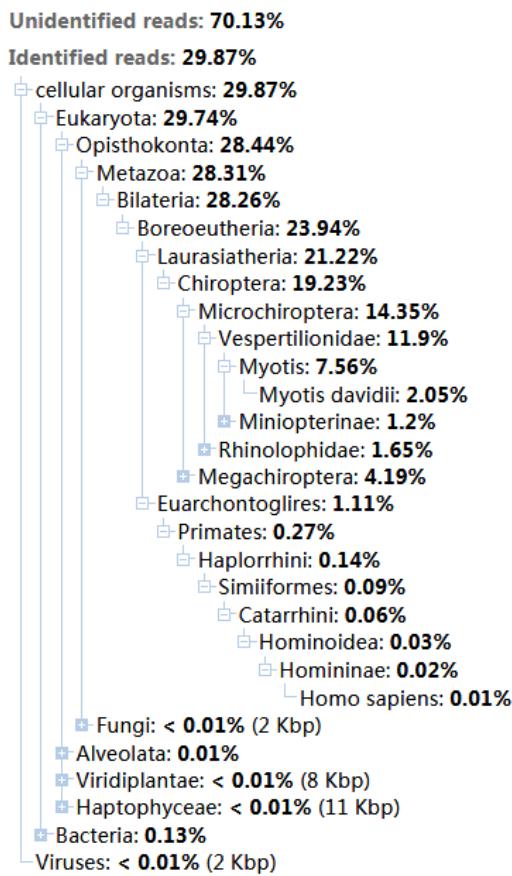


Figure 29: the only dataset with significant level of Telomere-like repeats (2%). There are no evidence of Riboviria(RNA viruses) within this dataset.

## 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>
- [8] Edgar, R. C. et al. Petabase-scale sequence alignment catalyses viral discovery. *bioRxiv* 2020.08.07.241729 (2020) [doi:10.1101/2020.08.07.241729](https://doi.org/10.1101/2020.08.07.241729)
- [9] Study of the Metatranscriptome of Eight Social and Solitary Wild Bee Species Reveals Novel Viruses and Bee Parasites  
Karel Schoonvaere<sup>1,2\*</sup>, Guy Smagghe<sup>3</sup>, Frédéric Francis<sup>2</sup> and Dirk C. de Graaf<sup>1</sup>

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

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

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