

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 provenance 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)
```

```
CTAACCCCTAACCCCTAGCACTATCCTGTTTCCAACCCCAACCCCTAACCCCTACCCCTAACCC  
TAACCCAGCCGTGTTTCATACCTTAACCTCGCACCTCATCGCTAACCCAGCCCTCACCCG  
ATCCTGTTTCTCTCCCGAACATAACCCCT
```

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

```
GGTTAGGGTTAGGGTTAGGGTTGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGG  
GTTAGAGTTAGGGTGGGAACAGGATAGGGTTAGGGTTAGGGCGAGGGATAGGGATAGGG  
AGGGAACAGGATAGTGGGAGGGCTAGGGGT
```

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

GTTAGGGTTAGGGTTAGGGTTAGGGTTGGGATACAGGATATGGTTAGGGTTAGGG
GTAGGGTCAGGGTTAGGATTGGAAACGAGATAGGTTACGTGATAGGGTTAGCGTTAGGGT
TAGGTTTAGTAATCCGCAACGGCTTAGGGTT

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

CCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCATCCTGTTCCCAACCCTAACC
CTAACCCTAACCCTAACCCTAACACAAAACATAACCCTAACCCCAACCCAAACCCTAACC
CCATCTTTACTCACACCCTAACCCAAAATC

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

GTTAGGGTTAGGGTTAGCGTTAGGGTTAGAAACAGGATAGGGTTAGGGTTAGGGTTAGGG
TTAGGGTTTTGGTTGGTCACAGTGTTCGCTAGGCATAGGGATAGGGTTCCCGTTAGGGT
TAGGGTTAGGATTTCGGAAGAGCTAGCTAAA

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

GTTCCCAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTTTCCTTTTTCCAACCCTA
ACACTAACCCTAACACTAAACCTAACCCCAACCCTACCCTATACTATATCCGACTCTCA
CGCTAACACTAAACATAAGTAATCACAAATT

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

TAACCCTAACCCTAAGCGTAAACCTAATCCAATCCTGTTCCCAACCCTAACCCTAACCCCT
GACCCTAAGCTTTTTCCCGACCCGAACGCCGACCCGATCCGCCACCCTAACCGTAACCCGT
TCCCAACCCCTCCTACTGCTCGATCCGCCT

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

GGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTGGGAACAGGATAGG
GTTAGGGTTAGGGTTGGGATCAGGATAGGGATAGGGATAGGGATAGGGATAGGGTTAGGG
TGGGGAACAGGAGAGCGTTAGGCAAGG

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

GGTTAGGGTTAGGGTTGGAGAAGAGGATAGGTTTAGGGTTTAGGGTTAGGGTTAGGGTGAGG
GTTAGGGAGAGGGTTAGCTACACGATAGGAGTAGGGTAACGATTAGGGTTAGGGTTAGGT
TTGGAAAAAGCATAGGCTATGAGGTACGGT

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

CTGCTTCCAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCCTAACCC
CTATCCTGTTCCCAACCCTAACCCTAACCCTAACCCCAACCCTAACCCCAACCCAAACCC
AACCCTAACCCCAACCCATACCCCAACCAT

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

TGTTCCCAACCCATAACCTAAACCTAAGCCGATCCTGTTCCCAACCCATAACCTAACCCT
ATCCTGTAAACAACCCCCACCCATAAAACATCCTCGTACAAACCCATAACCAACCCCCAT
CCCAAACCAACATACCCGTACGAAACCCACCC

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

GTTGGGGTTAGGGTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTGGGGTCAGGG
TTGGGACAAGGACGGGTAGGGGGTTAGGGTAGTGACACAGGGTAGGGAGTGGGGTAGGGTT
GGGAACAGAGAAGGGACAGTGGGGGGAGTGG

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

CTAACCCATAACCTACCCCTATCCTGTTCCCTAACCCGAACCCATAACCCATAACCC
TAACCCTCACCTGTTCCAGACCGTAATGCTAACCTTAACACTATCCTGTGCGCTACCCCG
ACCCATAACCCTCAGCCGACGCGTCACGCCCG

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

GGTAAGGGTTAGGGTTAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG
GTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTTGGAAACAGGATAGGGTTAGGGA
TAGGGGTAGAGATAGGGTGAGGTGGTGGAA

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

AACCCATAACCCTAACCCTAACGCTATCATGATCCCATCCCTAACCCATAACCCTAACCC
ACCCATAACTATAACCCTACCCTTTTTCATCTCCCTTACACTACCCCCAACACGCCACCCAT
CCCCAACCACTATGCATGCACTGTCCTAAAC

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

GGTTAGGGTTGGGAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGAACAGGATAGGG
GTAGGGTTAGGGCTAGGGTGAGAAACAGGGTAGGGGTAGGGTGAGGATAAGGGATAGGGT
TGGGGTTGGGAACAGAGAAGGGGAAGGGCA

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

CTAACCTGTTCCCAAACTTAAATCCAATCCTAACCCCTATCCTGTTCCCAACCCATAACCC
AAACCTATACCTATCCTGCCCCACACCCGACCCATACACCACCCATAAACGCAACCCTA
ACCCCATCCTGTTATCGAAGCATACCCCCAC

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

GTTAGGGTTAGGGTTAGGGTTGGTAACAGGATAGGGTTAGGGTTAGGGTTAGTGTGGGA
ACAGGATAGGGGGAGGGAGAGGGTTAGGGTTGGGAAGAGGATAGGGATAGGGGTAGGGAG
AGGGTTGGGGATAGGGAAGAGAGAGGAG

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

CCTAACCCATAACCTAACCCCTCTCCTGTTTCCAACCATAACCCATAACCCATATCCCTAACCC
CTAACCCACTCCTGTTCTTAACACTAACCTTAACCTCTGAGCTCATCCCCAAACCTAACCA
TAACCCCAACAGTTCCGATACCATCACCCCC

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

AGGGTTAGGGTTGGAAACAGGATAGGGTTAGGGTTAGGGTTAGGGTTAGGTTTAGGGTTG
GGGACAGGATAGGGTTAGGGTTAGGGTTAGGGTTGGGGTTAGGGTTAGGTGTGGGGTTGG
GATTGGGTATGGGTAGTGGTCAGGGATAGTG

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

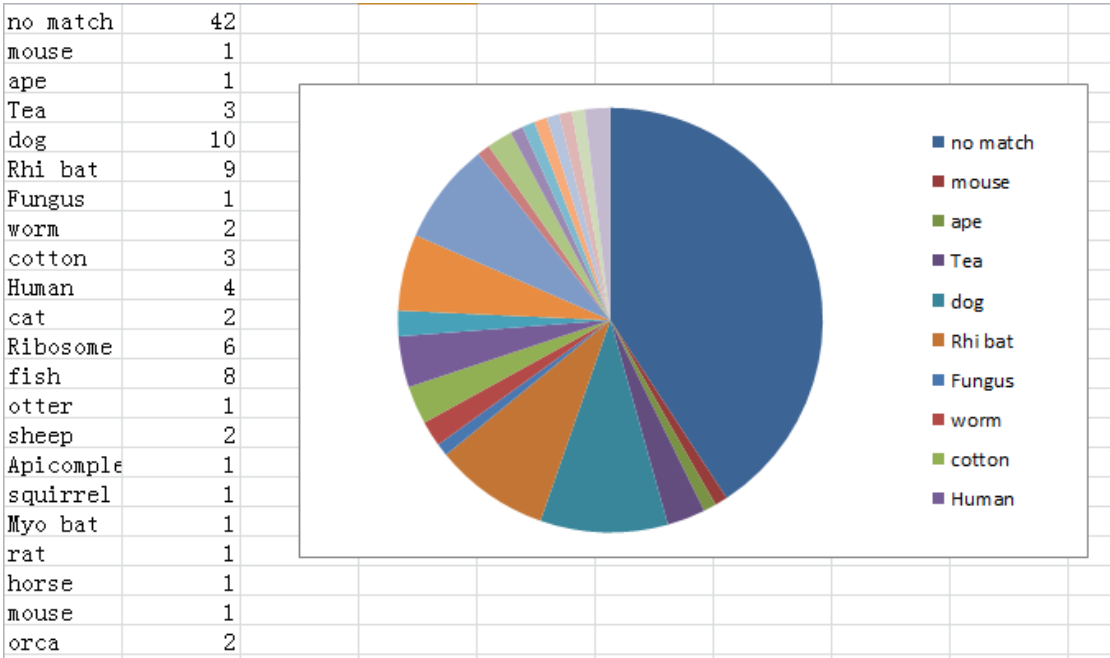
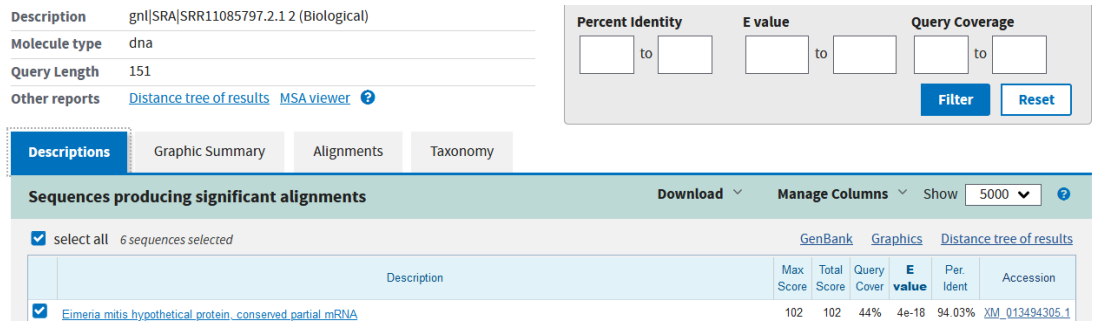


Figure 4: BLAST result of 100 random reads obtained from RaTG13 using BLASTn.

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



Results for

1:lc|Query_61914 gnl[SRA|SRR11085797.7.1.7 (Biological)](150bp) ▼

Program

BLASTN [Citation](#) ▼

Database

nt [See details](#) ▼

Query ID

lc|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

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download ▼

Manage Columns ▼

Show 5000 [?](#)

☒ select all 2 sequences selected

[GenBank](#)
[Graphics](#)
[Distance tree of results](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Eimeria mitis hypothetical protein conserved partial mRNA	73.1	73.1	40%	3e-09	88.33%	XM_013494305.1
<input checked="" type="checkbox"/>	Cyprinus carpio genome assembly common carp genome scaffold 000012284	73.1	73.1	40%	3e-09	88.33%	LN595243.1

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

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

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

Taxonomy Analysis

Unidentified reads: **29.38%**

Identified reads: **70.62%**

- cellular organisms: **70.61%**
 - Eukaryota: **67.91%**
 - Opisthokonta: **49.7%**
 - Metazoa: **49.23%**
 - Bilateria: **48.9%**
 - Euteleostomi: **41.62%**
 - Amniota: **14.99%**
 - Eutheria: **11.52%**
 - Boreoeutheria: **10.81%**
 - Laurasiatheria: **6.61%**
 - Chiroptera: **4.27%**
 - Euarchontoglires: **1.91%**
 - Fungi: **< 0.01%** (7 Kbp)
 - Viridiplantae: **0.09%**
 - Sar: **< 0.01%** (10 Kbp)
 - Bacteria: **0.65%**
 - Viruses: **0.01%**

RNA-Seq of *Miniopterus schreibersii*: Anal swab

Metadata Analysis Reads Data access

Taxonomy Analysis

Unidentified reads: 3.81%
 Identified reads: 96.19%
 cellular organisms: 96.08%
 Bacteria: 76.15%
 Eukaryota: 16.03%
 Opisthokonta: 10%
 Metazoa: 9.99%
 Bilateria: 9.99%
 Amniota: 7.67%
 Miniopterus natalensis: 5.98%
 Fungi: < 0.01% (1 Kbp)
 Viridiplantae: 0.16%
 Viruses: 0.11%

RNA-Seq of *Rhinolophus affinis*: Anal swab

Metadata Analysis Reads Data access

Taxonomy Analysis

Unidentified reads: 0.86%
 Identified reads: 99.14%
 cellular organisms: 99.11%
 Bacteria: 91.07%
 Eukaryota: 4.36%
 Viruses: 0.03%

RNA-Seq of *Scotophilus kuhlii*: Anal swab

Metadata Analysis Reads Data access

Taxonomy Analysis

Unidentified reads: 17.98%
 Identified reads: 82.02%
 cellular organisms: 79.84%
 Bacteria: 67.81%
 Eukaryota: 8.59%
 Opisthokonta: 4.68%
 Viridiplantae: 0.16%
 Sar: < 0.01% (5 Kbp)
 Viruses: 2.19%

RNA-Seq of *Hipposideros pumona*: Anal swab

Metadata Analysis Reads Data access

Taxonomy Analysis

Unidentified reads: 13.27%
 Identified reads: 86.73%
 cellular organisms: 82.62%
 Bacteria: 42.96%
 Eukaryota: 27.99%
 Archaea: < 0.01% (18 Kbp)
 Viruses: 4.1%

RNA-Seq of *Pipistrellus abramus*: Anal swab

Metadata Analysis Reads Data access

Taxonomy Analysis

Unidentified reads: 20.33%
 Identified reads: 79.67%
 cellular organisms: 73.22%
 Bacteria: 47.3%
 Eukaryota: 21.44%
 Opisthokonta: 13.06%
 Viridiplantae: 0.17%
 Sar: < 0.01% (13 Kbp)
 Viruses: 6.45%

RNA-Seq of *Tylonycteris pachypus*: Anal swab

Metadata Analysis Reads Data access

Taxonomy Analysis

Unidentified reads: 61.75%
 Identified reads: 38.25%
 cellular organisms: 38.19%
 Bacteria: 20.06%
 Eukaryota: 14.34%
 Archaea: 8.51%
 Opisthokonta: 8.46%
 Bilateria: 8.44%
 Euteleostomi: 6.9%
 Amniota: 5.46%
 Boreoeutheria: 4.9%
 Fungi: 0.02%
 Viridiplantae: 0.07%
 Sar: 0.01%
 Archaea: < 0.01% (4 Kbp)
 Viruses: 0.06%

RNA-Seq of *Rousettus aegyptiacus*: Anal swab

Metadata Analysis Reads Data access

Taxonomy Analysis

Unidentified reads: 0.78%
 Identified reads: 99.22%
 cellular organisms: 99.17%
 Bacteria: 96.24%
 Eukaryota: 1.46%
 Archaea: < 0.01% (1 Kbp)
 Viruses: 0.05%

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

discernible reads from bacterial mRNA.

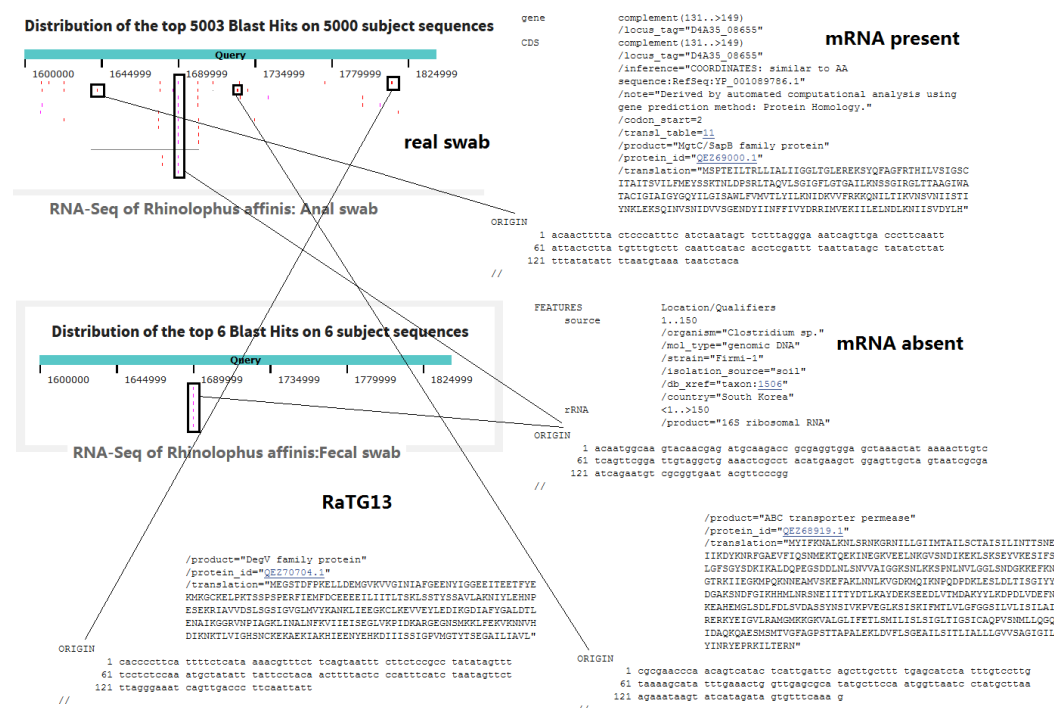


Figure 8: BLAST hits of bacterial non-ribosomal RNA genome on SRX7724272 and another swab from Rhinolophus Affinis under the same library preparation section.

Observation of anomalous and unexpected data within SRX7724752

```
>gnl|SRA|SRR11085797.11550005.1 11550005 (Biological)
GCCCGTATTTAGCCTTAGATGGAGTTTACCACCCGCTTTGGGCTGCATTCCCAAGCAACC
CGACTCCGGGGAAGACCCGGCGCCGGGGCCGCTACCGGCCCTCACACCGTCCA
CGGGCTGGGCCTCGATCAGAAGGACTTTGGGC

>gnl|SRA|SRR11085797.11550005.2 11550005 (Biological)
CGGTGGGGCGGGGACATTTGGCGTACGGAAGACCCACTCCCCGGCGCCGCTCGTGGGGG
CCCAAGTCCCTTCGATCGAGGCCAGCCCGTGGACGGTGTGAGGCCGGTAGCGGCCCGG
CGCGCGCGGGCCCGGGTCTTCCCGGAGTCCG
```


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 | | | | |
| <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"/> | PREDICTED: Phyllostomus discolor 28S ribosomal RNA (LOC114512504) .rRNA | 279 | 279 | 100% | 2e-71 | 100.00% | XR_003685809.1 |
| <input checked="" type="checkbox"/> | PREDICTED: Phyllostomus discolor basic proline-rich protein-like (LOC114512442) .mRNA | 279 | 279 | 100% | 2e-71 | 100.00% | XM_028531404.1 |
| <input checked="" type="checkbox"/> | Homo sapiens lncAB370.3 lncRNA gene .complete sequence | 274 | 274 | 100% | 9e-70 | 99.34% | MK280359.1 |
| <input checked="" type="checkbox"/> | Homo sapiens lncAB366.1 lncRNA gene .complete sequence | 274 | 274 | 100% | 9e-70 | 99.34% | MK280356.1 |
| <input checked="" type="checkbox"/> | Felis catus Senzu DNA .chromosome: E1 . American Shorthair breed | 274 | 1372 | 100% | 9e-70 | 99.34% | AP023165.1 |
| <input checked="" type="checkbox"/> | PREDICTED: Marmota flaviventris 28S ribosomal RNA (LOC117794687) .rRNA | 274 | 274 | 100% | 9e-70 | 99.34% | XR_004618536.1 |
| <input checked="" type="checkbox"/> | PREDICTED: Avicanthis niloticus 28S ribosomal RNA (LOC117704856) .rRNA | 274 | 274 | 100% | 9e-70 | 99.34% | XR_004606369.1 |
| <input checked="" type="checkbox"/> | PREDICTED: Avicanthis 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 | 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

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

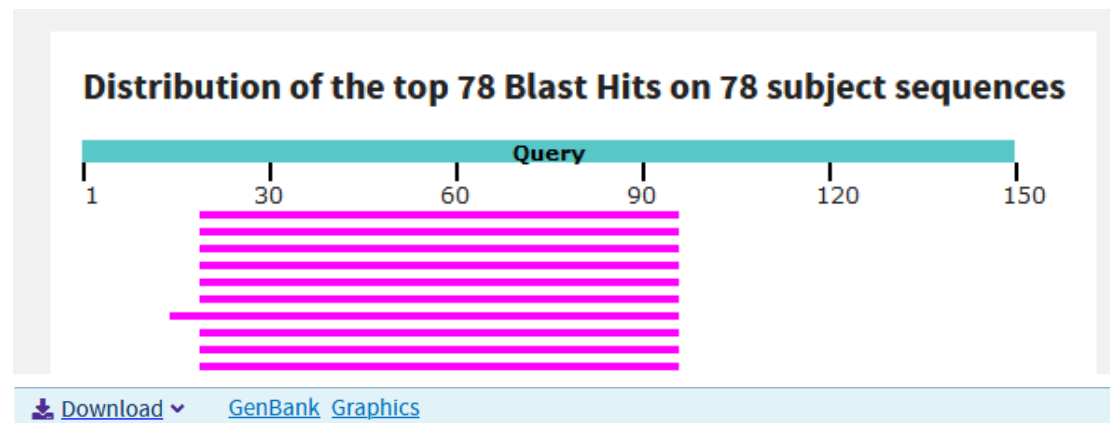


Pale spear-nosed bat

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

>gnl|SRA|SRR11085797.11550023.2 11550023 (Biological)

TTTGTTTTGTTTTTATAATTTATTTTAAAAATTTATTGGGGTGACAATTGTTAGTAAAA
TTACATAGATTTTCAGGTGTACAATTCGTATTACATGTGGACGGTCCAGCCGCCACGAGT
TCAACGTTTTACATGAAAGGGGGTGTGGGA



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 | tttatTTTTTAA--atTTATTGGGGTGACAATTGTTAGTAAAATTACATAGATTCAGG | 76 | | |
| | | | | |
| Sbjct 32707 | TTTTTTTTTAAATTTATTTATTGGGGTGACAATTGTTAGTAAAATTACATAGATTCAGG | 32766 | | |
| Query 77 | TGTACAATTCGTATTACAT | 96 | | |
| | | | | |
| Sbjct 32767 | TGTACAATTCGTATTACAT | 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.

Reads (separated)

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

ATGGGGGAGCAGCGGACGGGGTCAACACAGTCCATGGACCCCTGGCAGGGGCGATGAGAT
CGGTGAACTAGGGGACAAAAGGAAGTTACAGATCTACAAGAGATCGAGAGTTCTGTTGGTT
TGT

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

ACAAACCAACGAACCTCTCGATCTCTTGTAGATCTGTAACCTCCTTTTGTCCCTAGTTCA
CCGATCTCATCGCCCTGCCAGGGTCCATGGACTGTGTTGACCCCGTCCGCTGCTCCCC
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) | 4e-27 | 79/83(95%) | 0/83(0%) | Plus/Minus |
| Query 6 | GGAGCAGCGGACGGGGTCAACACAGTCCATGGACCCCTGGCAGGGGCGATGAGATCGGTG | 65 | | |
| | | | | |
| Sbjct 2550 | GGAGCAGAGGACAGGATCAACACAGTCCATGGACCCCTGGCAGGGGCGATGAGATCGGTG | 2491 | | |
| Query 66 | AACTAGGGGACAAAAGGAAGTTA | 88 | | |
| | | | | |
| Sbjct 2490 | AACTAGGGGACAAAAGGAGGTTA | 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) | 1e-07 | 36/36(100%) | 0/36(0%) | Plus/Minus |
| Query 88 | ACAGATCTACAAGAGATCGAGAGTTCGTTGGTTTGT | 123 | | |
| | | | | |
| Sbjct 47 | ACAGATCTACAAGAGATCGAGAGTTCGTTGGTTTGT | 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)

GGTCCTTGATGTCACAGCGTCCTAGATGGTGTCCAGCAATACGAAGATGTCACGAAGGA
TGACAGCTCCGATTACAAGTTCACTCTCTAGAAGCGGTCTGGTCAAATAGTGCCATGGA
GTGGCACGTTGAGCAAAATGTTAGTTTCTGG

>gnl|SRA|SRR11085797.10676687.2 10676687 (Biological)

ATGAAGGCAATTCACCATTCCATCCTCTAGCTGATAATAAATTTGCACTGACTTGCTTTA
GCACTGATGTGGCTGAGCTACTTCATTGCTTCTTCAGGCTATTTGCACGTACGCGTTCC
ATGTGGTCATTCAATCCAGAACTAACATTT

| | | | | |
|---|--|--|-----------|-----------|
| Download GenBank Graphics Sort by: E value | | | | |
| Bat coronavirus RaTG13, complete genome
Sequence ID: MN996532.1 Length: 29855 Number of Matches: 2 | | | | |
| Range 1: 26736 to 26844 GenBank Graphics ▼ Next Match ▲ Previous Match | | | | |
| Score | Expect | Identities | Gaps | Strand |
| 161 bits(87) | 7e-36 | 102/109(94%) | 2/109(1%) | Plus/Plus |
| Query 45 | GCAC | TGACTTG-CTT-TAGCACTGATGTGGCTGAGCTACTTCATTGCTTCTTTCAGGCTA | 102 | |
| Sbjct 26736 | GCAATGGCTTGTCTTGTAGGCTTGATGTGGCTGAGCTACTTCATTGCTTCTTTCAGGCTA | 26795 | | |
| Query 103 | TTTGCACGTACGCGTTCCATGTGGTCATTCAATCCAGAACTAACATTT | 151 | | |
| Sbjct 26796 | TTTGCACGTACGCGTTCCATGTGGTCATTCAATCCAGAACTAACATTT | 26844 | | |
| Range 2: 27478 to 27542 GenBank Graphics ▼ Next Match ▲ Previous Match ▲ First Match | | | | |
| Score | Expect | Identities | Gaps | Strand |
| 121 bits(65) | 1e-23 | 65/65(100%) | 0/65(0%) | Plus/Plus |
| Query 1 | ATGAAGGCAATTCACCATTCATCCTCTAGCTGATAATAAATTGCACTGACTTGCTTTA | 60 | | |
| Sbjct 27478 | ATGAAGGCAATTCACCATTCATCCTCTAGCTGATAATAAATTGCACTGACTTGCTTTA | 27537 | | |
| Query 61 | GCACT 65 | | | |
| Sbjct 27538 | GCACT 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 GenBank Graphics ▼ Next Match ▲ Previous Match | | | | |
| Score | Expect | Identities | Gaps | Strand |
| 233 bits(126) | 2e-57 | 129/130(99%) | 1/130(0%) | Plus/Plus |
| Query 23 | TTC-TCTAAACGAACAACTAAATGTCTGATAATGGACCCCAAAACCAACGAAATGCAC | 81 | | |
| Sbjct 28217 | TTTCATCTAAACGAACAACTAAATGTCTGATAATGGACCCCAAAACCAACGAAATGCAC | 28276 | | |
| Query 82 | CCCGCATTACGTTTGGTGGACCCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCA | 141 | | |
| Sbjct 28277 | CCCGCATTACGTTTGGTGGACCCCTCAGATTCAACTGGCAGTAACCAGAATGGAGAACGCA | 28336 | | |
| Query 142 | GTGGAGCACG 151 | | | |
| Sbjct 28337 | GTGGAGCACG 28346 | | | |
| Range 2: 25 to 60 GenBank Graphics ▼ Next Match ▲ Previous Match ▲ First Match | | | | |
| Score | Expect | Identities | Gaps | Strand |
| 67.6 bits(36) | 2e-07 | 36/36(100%) | 0/36(0%) | Plus/Plus |
| Query 1 | CTCTCGATCTCTTGTAGATCTGTTCTCTAAACGAAC | 36 | | |
| Sbjct 25 | CTCTCGATCTCTTGTAGATCTGTTCTCTAAACGAAC | 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)
CATCAAACTGAGGTTTCAGCAAGGCAAAAGATAGCCAGCAACAAAACAAAAAGGCATCCTA
CTGAATGGAAGCAGATAATTGCCAATAGTACATCAGTAAGGAGTTAATATTAAGAATTAG
TTTTTAAAAAGCTCTATATGATGTCAGAAAT

>gnl|SRA|SRR11085797.76.2 76 (Biological)
GTTTTTCACTTGCATTTCTCTAATAATTAGTGATGTTGAGCATCTTTTCATATGTCATTG
GCCATCTGTATGTCGTCTTTGGAGAAATGTCATTTCAGATTTCTGCCCAATTTTAATTG
GCTTGTTTGTTTTTGTGTTTTGAATTGAGTT
```

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download Manage Columns Show 100

select all

11 sequences selected

GenBank

Graphics

Distance tree of results

| | 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 | 98.68% | 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 P212O7-1 .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_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/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 |
| 29-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/ll96Y | 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/dtqMp | 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 |

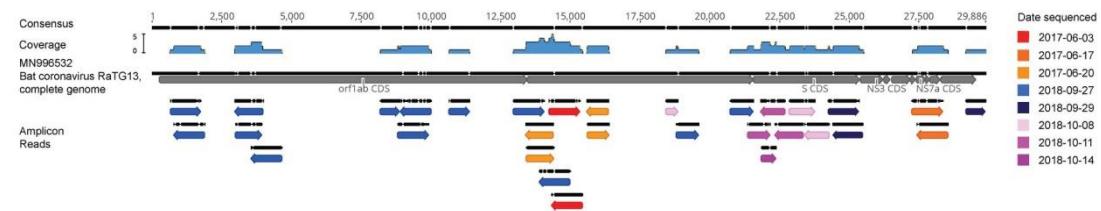


Figure 15: A complete analysis [4] of all Amplicon sequences in SRX8357956. Including the location of these amplicons and the similarity of such amplicon to the RaTG13 and SARS-CoV-2 genome.

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.

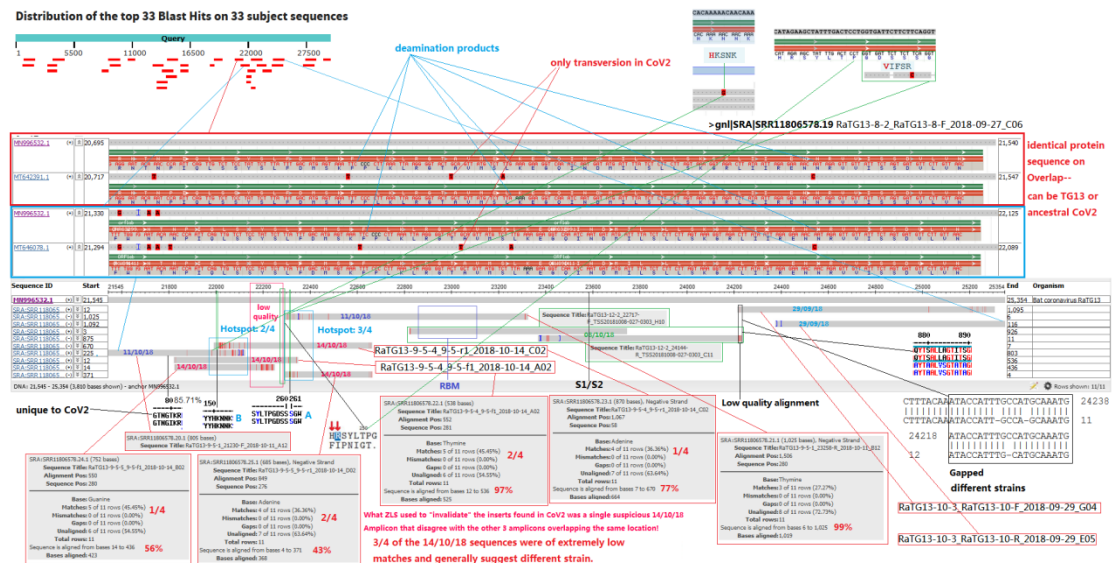


Figure 16: a thorough analysis of the amplicons located on the S locus of RaTG13 in SRX8357956. Notice that the last 4 amplicons sequenced in 14/10/2018 was of very low quality matches, and matched other organisms—including Mouse(mus musculus).

| | | | | | | | |
|--|--|-------------------------|-------------|--------------------------|---------|--|-----------------------------|
| <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"/> | Gadus morhua genome assembly_chromosome: 16 | 60.8 | 105 | 27% | 4e-05 | 78.05% | LR633958.1 |
| <input checked="" type="checkbox"/> | Mus musculus BAC clone RP24-498P8 from chromosome 9, complete sequence | 59.9 | 59.9 | 47% | 1e-04 | 72.92% | AC168217.2 |
| <input checked="" type="checkbox"/> | 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 |
| <input checked="" type="checkbox"/> | Homo sapiens BAC clone RP11-792A8 from 7, complete sequence | 54.5 | 54.5 | 44% | 0.006 | 71.15% | AC027644.9 |
| <input checked="" type="checkbox"/> | Coregonus sp. 'balchen' genome assembly_chromosome: 7 | 52.7 | 52.7 | 29% | 0.022 | 73.68% | LR778259.1 |
| <input checked="" type="checkbox"/> | Coregonus sp. 'balchen' genome assembly_chromosome: 15 | 50.9 | 50.9 | 16% | 0.078 | 83.33% | LR778267.1 |
| <input checked="" type="checkbox"/> | Salmo trutta genome assembly_chromosome: 21 | 50.9 | 50.9 | 37% | 0.078 | 71.90% | LR584437.1 |
| <input checked="" type="checkbox"/> | Xanthophyllomyces dendrorhous genome assembly_Xden1_scaffold Scaffold_79 | 50.9 | 50.9 | 9% | 0.078 | 96.88% | LN483167.1 |
| <input checked="" type="checkbox"/> | Coregonus sp. 'balchen' genome assembly_chromosome: 20 | 50.0 | 50.0 | 72% | 0.078 | 67.49% | LR778272.1 |
| <input checked="" type="checkbox"/> | Aquila chrysaetos chrysaetos genome assembly_chromosome: 14 | 50.0 | 141 | 42% | 0.078 | 69.50% | LR606194.1 |
| <input checked="" type="checkbox"/> | Bos mutus isolate yakQH1 chromosome 16 | 50.0 | 50.0 | 24% | 0.078 | 74.68% | CP027084.1 |
| <input checked="" type="checkbox"/> | Mus musculus BAC clone RP23-128D11 from 7, complete sequence | 50.0 | 50.0 | 36% | 0.078 | 71.90% | AC122222.6 |
| <input checked="" type="checkbox"/> | 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

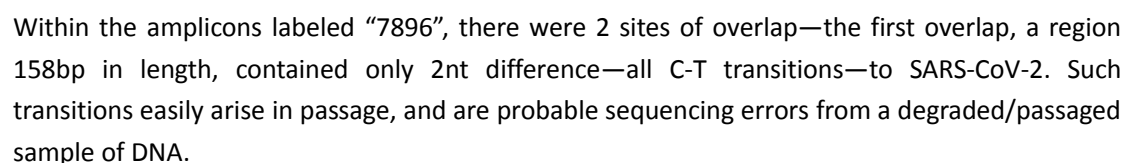
[GenBank](#)
[Graphics](#)
[Distance tree of results](#)

| | Description | Max Score | Total Score | Query Cover | E value | Per. Ident | Accession |
|-------------------------------------|--|-----------|-------------|-------------|---------|------------|-----------------------------|
| <input checked="" type="checkbox"/> | 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 |
| <input checked="" type="checkbox"/> | Mus musculus targeted non-conditional_lacZ-tagged mutant allele Fabp4:tm1e(KOMP)Wts; transger | 50.9 | 50.9 | 23% | 0.045 | 85.11% | JN947213.1 |
| <input checked="" type="checkbox"/> | Mus musculus chromosome 3_clone RP23-436F15, complete sequence | 50.9 | 50.9 | 23% | 0.045 | 85.11% | AC123726.11 |
| <input checked="" type="checkbox"/> | 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, HKSNN and VIFSQ was obtained.

Probable discontinuities in RaTG13 sequencing in SRX8357956



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

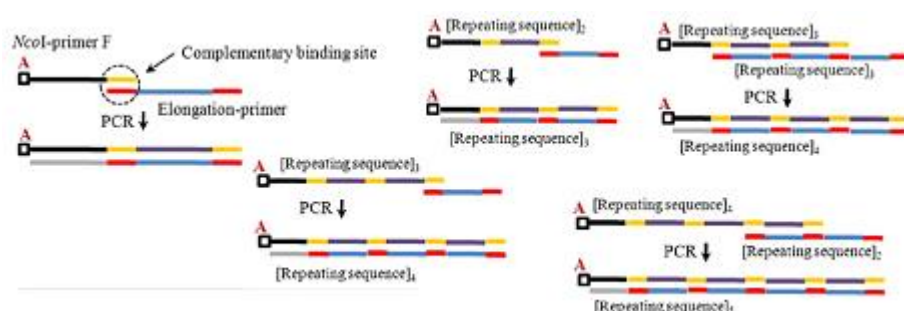


Figure 18A: Origin of repeating sequences in SRX7724752.

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.

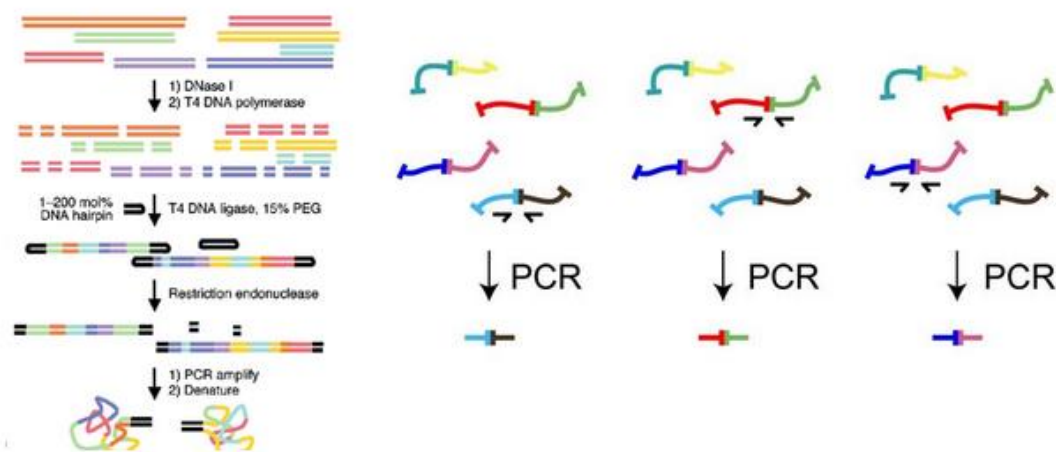


Figure 18B: Origin of the random matched sequences and partial sequences in SRX7724752

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-TCATTATTAAGTATTATGTACTGTACATAAATGTATGTACTATACTTTTATACAAAC | 70 | | |
| Sbjct 2293 | TTTATCATATATCAAGTGTATGTACTGTACAGTATGTATGTGTATATACTTTTATATGAC | 2234 | | |
| Query 71 | TGGCAGCACAGCAGGTTTGTATTATACCAGCATCACCACAAAAATGTGAGTAATGCATTAC | 130 | | |
| Sbjct 2233 | TGACAGCATAGTAGGCTTGTATTACACCAGCATCACCACAAAAATGTGAGTAATGCATTAC | 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

ttagatttcacatctaaacgaacaaactaaaatgtctgataatggaccccaaaacacgaaatgcaccccgattacgttttggtggaccct
 CTCTCGATCTCTGTAGATCTGTTCTCTAAACGAAC
 ACAAACCAACGAACCTCTCGATCTCTGTAGATCTGTTCTCTAAACGAAC
 TAACCTCCTTTTIGTCCCTAGTTTACCGATCTCATCGCCCTGCCAGGGTCCATGGACTGTGTTGATCTGCTCTGCTCCTC

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

| | 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/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 BtR-BetaCoV/YN2018D_complete_genome | 63.9 | 63.9 | 94% | 1e-07 | 100.00% | MK211378.1 |

Figure 20b: BLAST result of CTCTCGATCTCTGTAGATCTGTTCTCTAAACGAAC.

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)
GCCTTCGTTTGATATAGTTTTAATGCAAATCCCCTAACATCTCTTTCAGCATCTGCTGC
ACCTCTTTCACCAGCAACTGTAGAAAATCTTAAAAGGGCTTTTGTTTTTTTACCAACTTT
GTTAAAAATATCTGCTTTAGAATATTTTGT

>gnl|SRA|SRR11085733.2232944.2 2232944 (Biological)
AGAGGTCCACTCTTTTACAAGATACCTTGGCTTTTAGAAAACTTGACATTTTCGATAGG
GAAAGGATACCAGAAAGAGTTGTGCACGCTAAAGGAAGTGCTGCATACGGCGAATTAAACA
ATTACTAATGATATTACAAAATATTCTAAA
```


Sequence ID: [AE017125.1](#) Length: 1799146 Number of Matches: 1

▼ [Next Match](#) ▲ [Previous Match](#)

| Score | Expect | Identities | Gaps | Strand | |
|--------------|--------|---|-----------|------------|-------|
| 176 bits(95) | 3e-40 | 123/137(90%) | 0/137(0%) | Plus/Minus | |
| Query 7 | | CCTACTCTTTTACAAGATACTTGGCTTTTAGAAAAACTTGCACATTTTCGATAGGGAAAGG | | | 66 |
| | | | | | |
| Sbjct 47856 | | CCTACACTTTTACAAGATACTTGGCTTTTAGAAAAACTTGCACATTTTGATAGAGAGAGA | | | 47797 |
| Query 67 | | ATACCAGAAAGAGTTGTGCACGCTAAAGGAAGTGTGCATACGGCGAATTAACAATTACT | | | 126 |
| | | | | | |
| Sbjct 47796 | | ATCCCAGAGAGAGTGGTGCACGCTAAAGGAAGTGCAGCATATGGTGAATTAACAATTACA | | | 47737 |
| Query 127 | | AATGATATTACAAATA | 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
ERIPERVVHAKGSAAYGELTIITNDITQYTKAELFNKVGGKTKAFLRFSVVAGERGAAD
AERDVRGFGALKLYTNEGNWDIVGNNTIPVFFIKDAIKFPDFIHTQKRDPKTNMKSPTAM
WDFWSLHPESLHQVTILMSDRGIPRSYREMHGFGSHTYSFINAKNERFWVKFHFVCLQ
GIHNLTNKESEAVIAKDRESHQKDLFENIEKGNFPKWRFCIQVMSEKEAENYRFPNPF
LTKVWSHKDYPLIEVGILELNKNPENYFAEVEQAAFNPANIVPGVGYSPDKVLQGRLF
SYGDTQRYRLGINHTQLFPVNAPIVPVNNTHRDGFMQQGQFGDRRNYEPSYFNDYVEDK
NALEPPLFVQEGDVMYKYDHREYEDDYFVQAGDLYRLMTAEQKEALCQNIKESMEGVP
DEIKKRQLEHFKKADKAYGKRVAELLGL"

ORIGIN
      1  tattgagtaa  tatcatttgt  aattgttaat  tcaccatattg  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%

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



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

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

Probable signs of laboratory manipulation of SRX7724752

[← Edit Search](#)
[Save Search](#)
[Search Summary](#) ▼

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

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

Filter Results

Organism only top 20 will appear ☐ exclude

[+ Add organism](#)

Percent Identity to
E value to
Query Coverage to

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

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

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

☒ 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 | 0% | 1e-70 | 100.00% | SRA:SRR11085797.11044608.2 |
| <input checked="" type="checkbox"/> | SRX7724752 | 279 | 279 | 0% | 1e-70 | 100.00% | SRA:SRR11085797.11044608.1 |

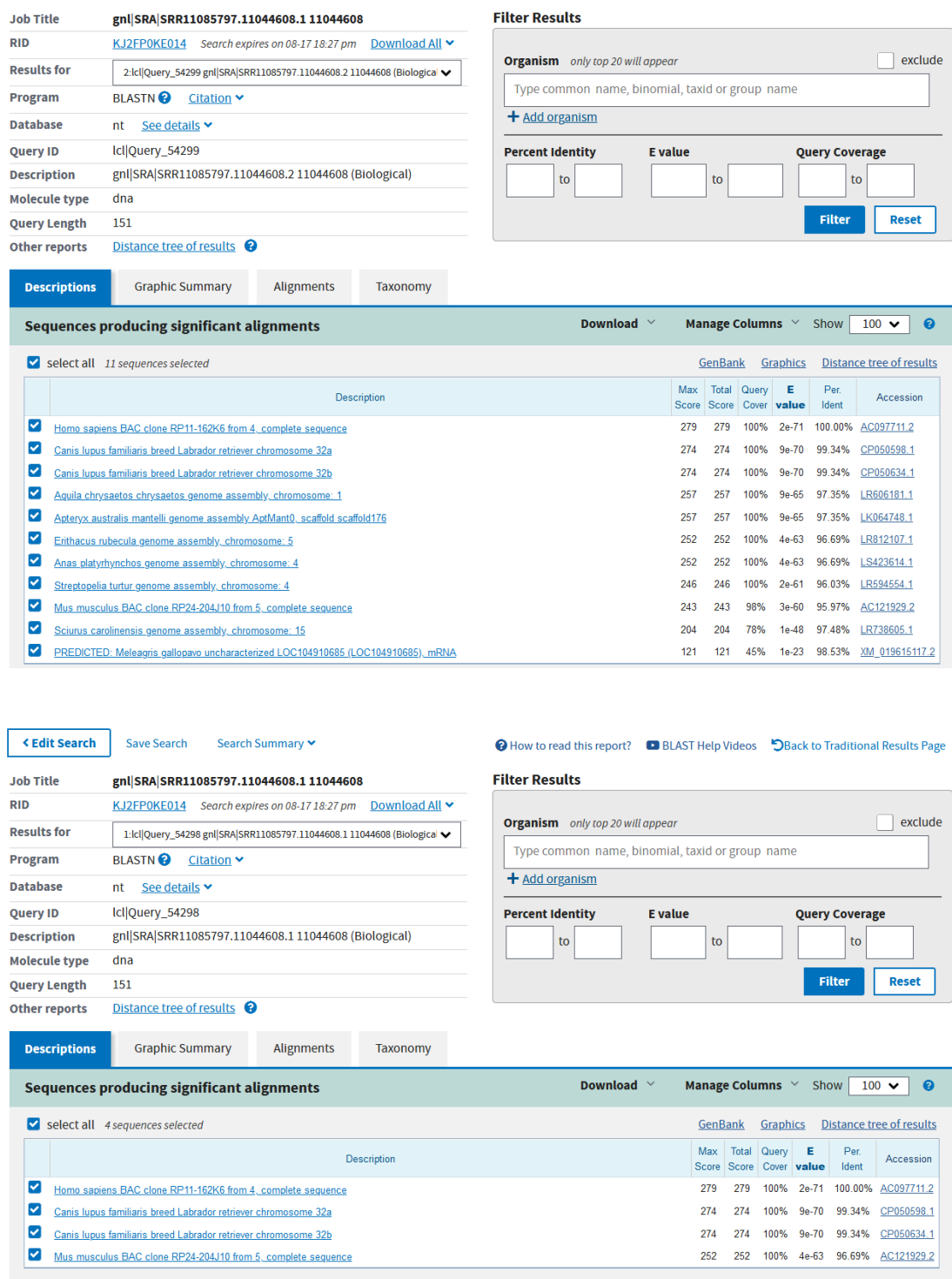


Figure 23: Unique, fully-matched 100% read from Homo Sapiens is recovered from the dataset SRX7724752.

← 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, mai ...

Molecule type

dna

Query Length

19578880

Other reports

[Distance tree of results](#)

[MSA viewer](#)

?

How to read this report?

BLAST Help Videos

Back to Traditional Results Page

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

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 | 0% | 2e-68 | 100.00% | SRA:SRR11085797.70411148.2 |
| <input checked="" type="checkbox"/> | SRX7724752 | 278 | 278 | 0% | 9e-68 | 100.00% | SRA:SRR11085797.4666606.2 |
| <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

KJ3PSXPK014

Search expires on 08-17 18:47 pm

Download All ▾

Results for

1:|c|Query_51896 gnl[SRA|SRR11085797.10431565.1 10431565 (Biologica ▾

Program

BLASTN

Citation ▾

Database

nt

See details ▾

Query ID

|c|Query_51896

Description

gnl[SRA|SRR11085797.10431565.1 10431565 (Biological)

Molecule type

dna

Query Length

151

Other reports

[Distance tree of results](#)

?

How to read this report?

BLAST Help Videos

Back to Traditional Results Page

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

45 sequences selected

[GenBank](#)

[Graphics](#)

[Distance tree of results](#)

| | 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% | HG_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 turtur genome assembly chromosome: 4 | 274 | 274 | 100% | 9e-70 | 99.34% | LR594554.1 |
| <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 |

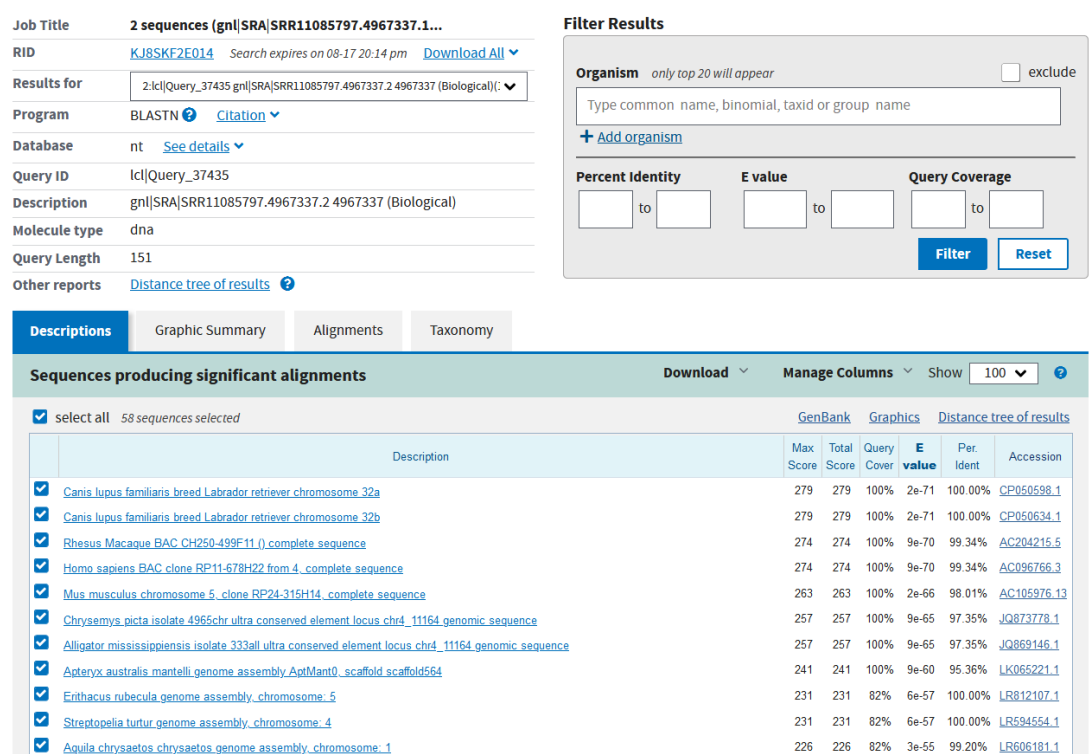


Figure. 24: Marmota Marmota genetic scaffold assemblies returned significant amount of 100% full-length matched reads that were sometimes also found in Homo Sapiens and Canis Lupus Familiaris.

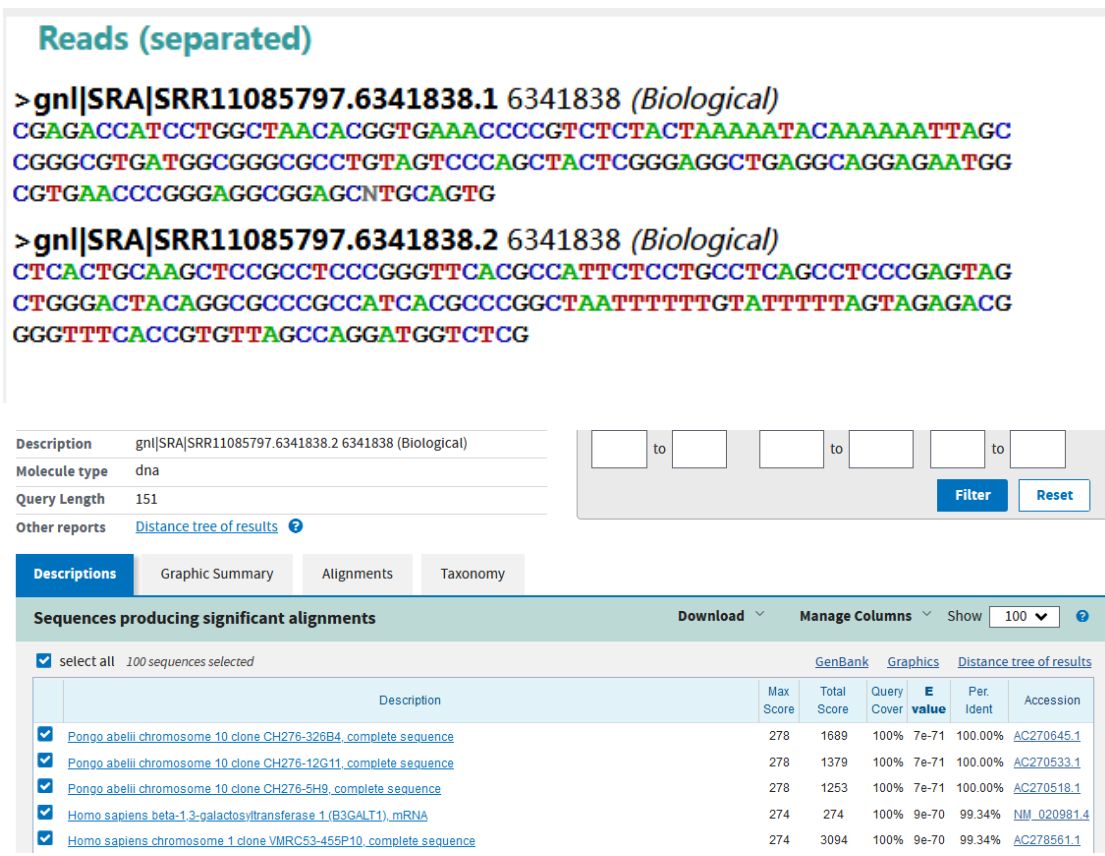


Figure.25a: 100% full-length matched reads to Hominid(Pongo Albelii) genomic DNA.

| | | | | | | |
|---|------|------|------|-------|--------|-----------------------------------|
| ✓ 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 ORFome 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 Alu253 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% | AVHA01101756.1 |
| ✓ 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% | RXPC010000086.1 |
| ✓ 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.

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](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA606159).

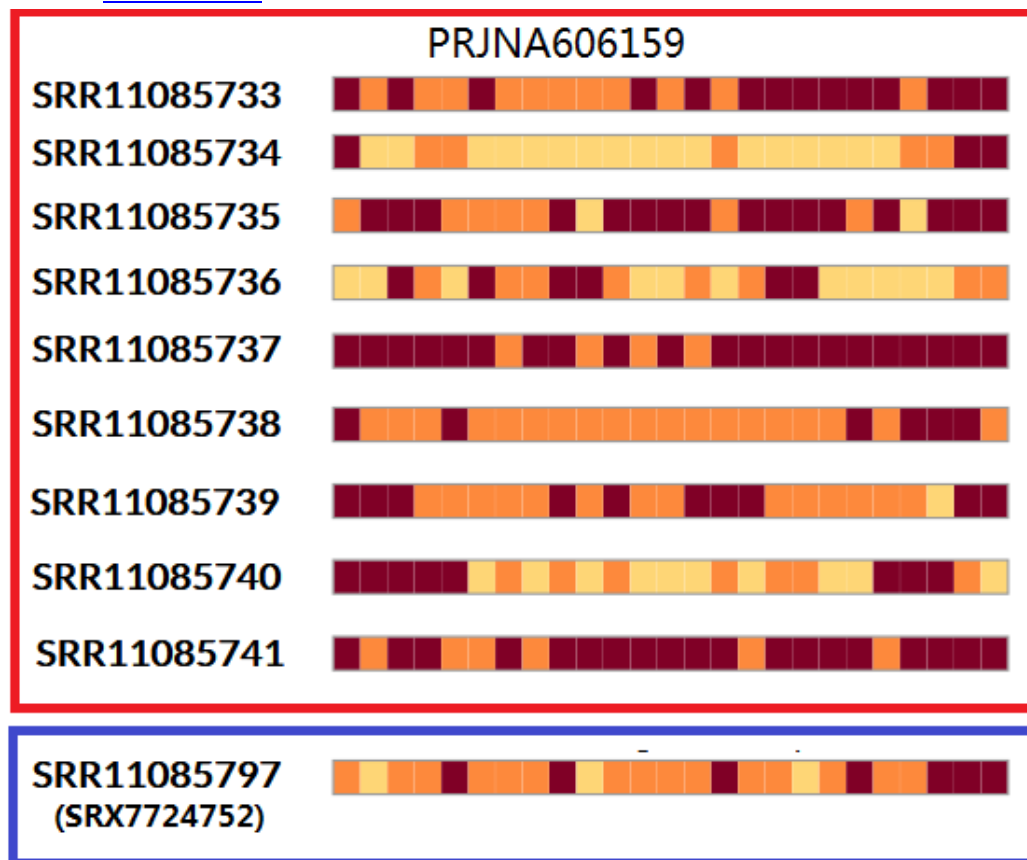


Fig. 26a: the Coverage map of Coronaviridae within the datasets located in PRJNA606159, compared against SRR11085797.

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](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA606159): 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

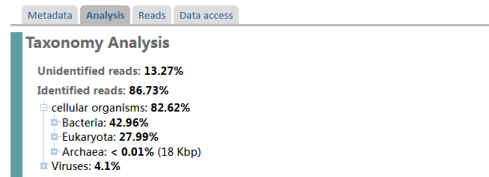
[SRX7724752](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA606159): RNA-Seq of *Rhinolophus affinis*: Fecal swab
1 ILLUMINA (Illumina HiSeq 3000) run: 11.6M spots, 3.3G bases, 1.7Gb downloads

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).

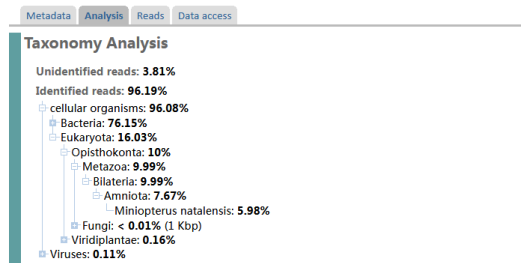
Submitted by: Wuhan Institute of Virology, Chinese Academy of Sciences

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

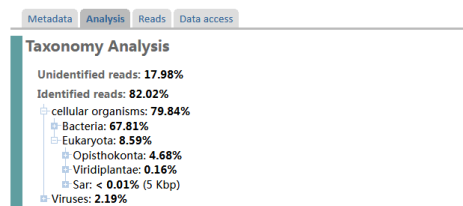
RNA-Seq of *Hipposideros larvatus*: Anal swab (SRR11085733)



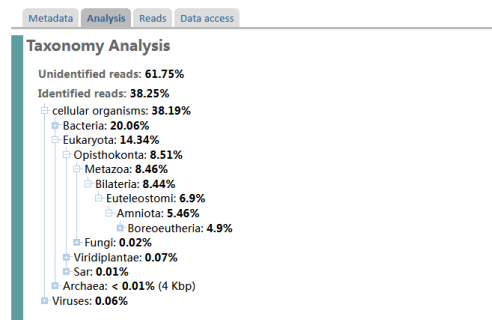
RNA-Seq of *Miniopterus schreibersii*: Anal swab (SRR11085734)



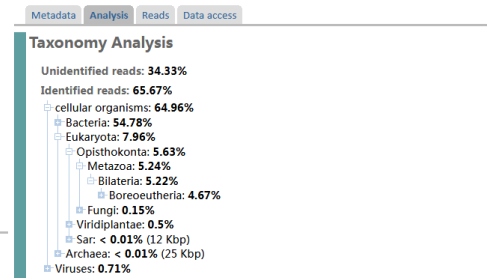
RNA-Seq of *Scotophilus kuhlii*: Anal swab (SRR11085737)



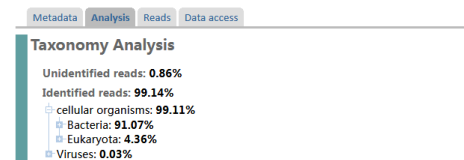
RNA-Seq of *Tylonycteris pachypus*: Anal swab (SRR11085739)



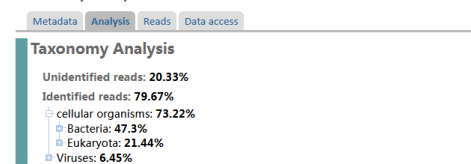
RNA-Seq of *Hipposideros pomona*: Anal swab (SRR11085735)



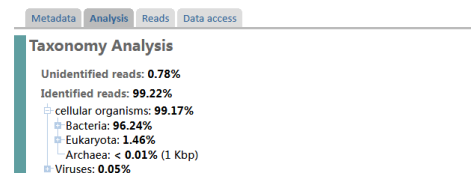
RNA-Seq of *Rhinolophus affinis*: Anal swab (SRR11085736)



RNA-Seq of *Pipistrellus abramus*: Anal swab (SRR11085738)



RNA-Seq of *Miniopterus pusillus*: Anal swab (SRR11085740)



RNA-Seq of *Rousettus aegyptiacus*: Anal swab (SRR11085741)

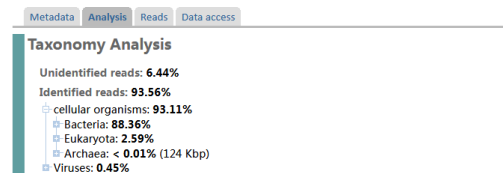


Fig.26c: No evidence of any anomalies were found within the datasets presented in PRJNA606159

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.

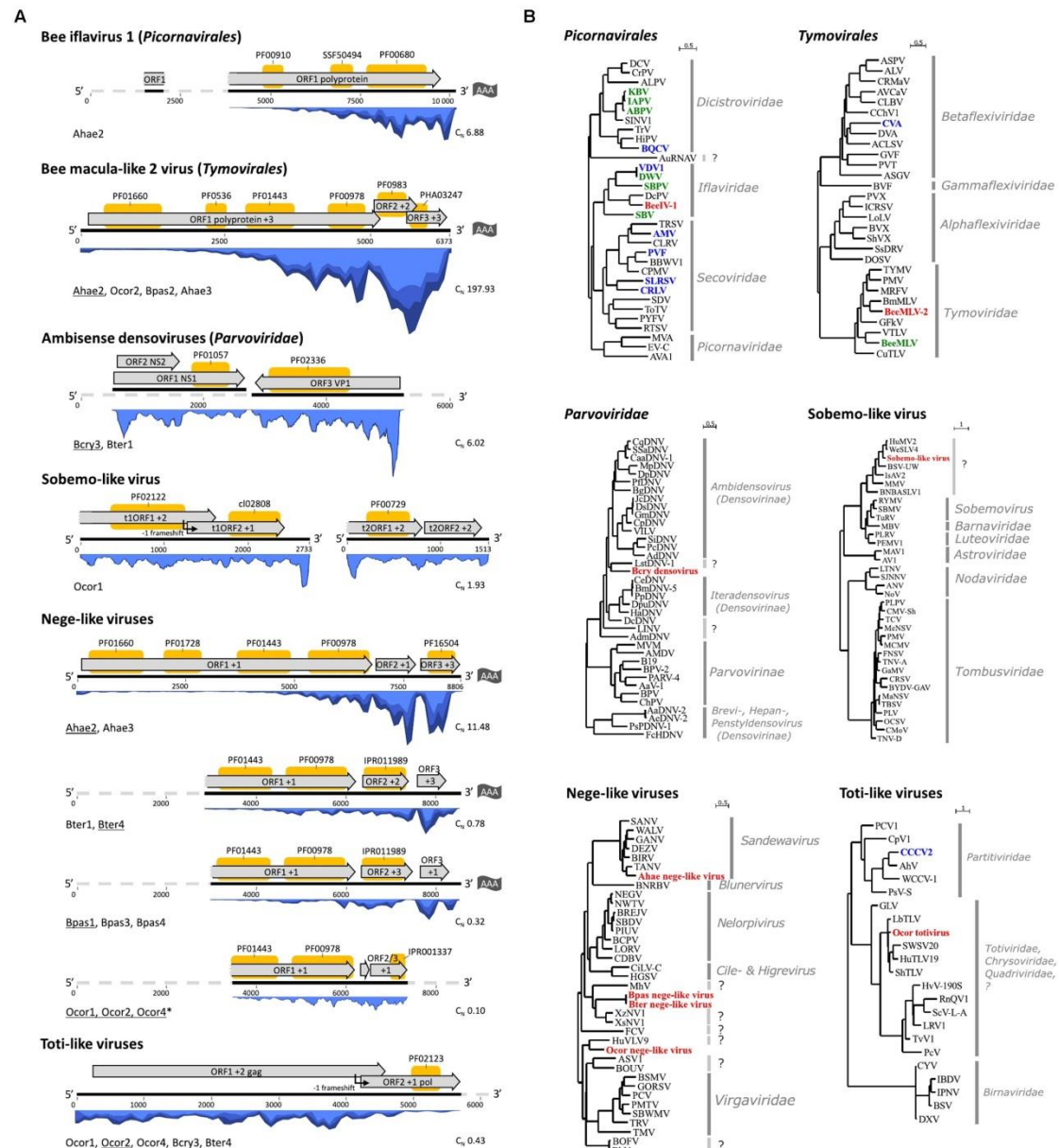


Figure 2 from [9]. A clear bias toward the 3'-end of RNA viral families that contained a polyA tail, was noticed.

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 | ? |



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.

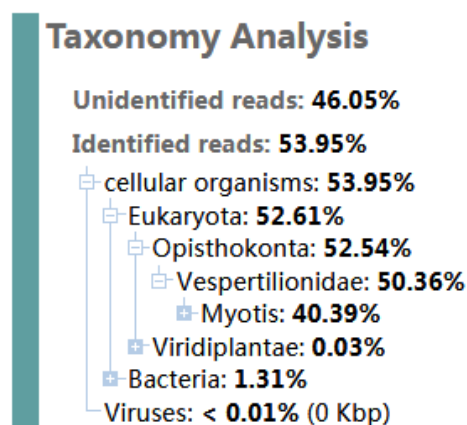


Figure 28a: an example of a bacteria-depleted dataset. An absence of Riboviria reads was noted.

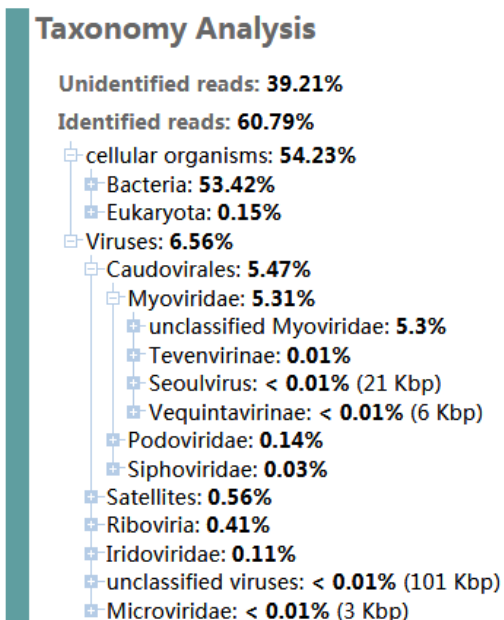


Figure 28b: in contrast, Riboviria reads are found only in datasets that contained a significant amount of bacteria.

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

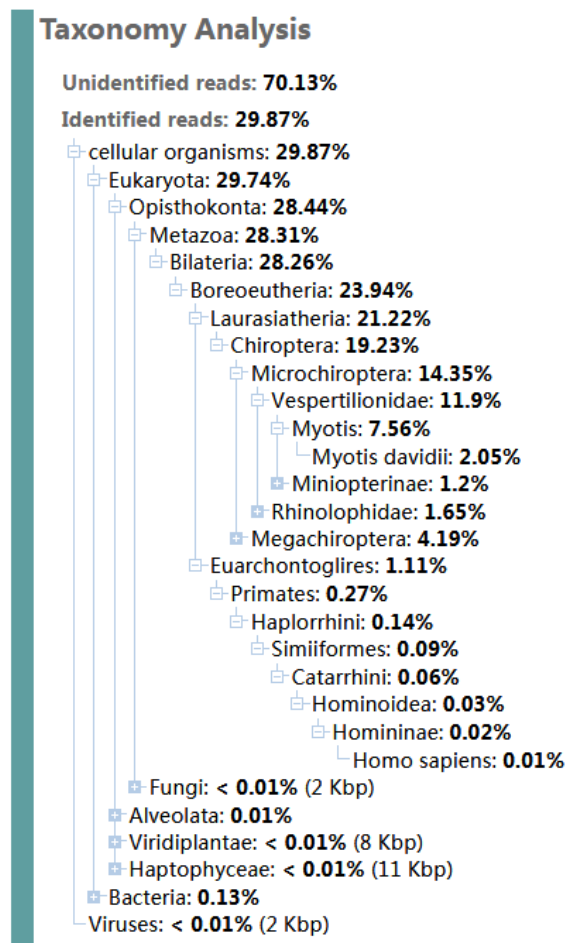


Figure 29: the only dataset with significant level of Telomere-like repeats (2%). There are no evidence of Riboviria(RNA viruses) within this dataset.

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^{1,2*}, Guy Smagghe³, Frédéric Francis² and Dirk C. de Graaf¹

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

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

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