

The Pan-SL-CoV/GD sequences may be from contamination.

Daoyu Zhang.

ABSTRACT

Recently, There were much hype about an alleged SARS-like coronavirus being found in samples of Malayan pangolins (*Manis Javanica*) possessing nearly identical RBD to the SARS-CoV-2 coronavirus. Prominent journals cite the alleged discovery to claim that pangolins may be one of a possible intermediate host for the zoonotic transmission of SARS-CoV-2 to humans.

Here, we report that all databases used to support such a claim, upon which metagenomic analysis was possible, contained unexpected reads and was in serious risk of contamination. Here we also report that the presence of unexpected reads are directly related to the presence of coronavirus reads. Finally, we deduced the actual causative agent of the death of the pangolins sampled in GuangDong 2019 where the claim of coronavirus detections was made.

METHODS

The NCBI Trace tool

The NCBI SRA archive come with it's own tool called Trace, which identifies the origin or reads within the SRA dataset through the recognition of unique K-mers within the nucleotide sequence. Multiple reads of 32 nucleotides is taken from each read to identify the reads toward an origin by comparison with a large database of reference sequences, which produces a classification signal. Then read of 64 nucleotides are taken from each of the read for definitive mapping toward species in the reference database. If any one of the 32nt or 64nt K-mers are found in more than one reference sequence, the reads are instead classified at the lowest phylogenetic classification node where reference sequences containing such a K-mer is found.

The 32nt TRACE generate a "strong signal" classification of sequence origin useful for the deduction of the content of the sample by organism of origin, accessed via the NCBI Krona charting tool,

While the 64nt TRACE generate a definitive classification signal used for the exact tracing of reads to the origin from a specific Species/Taxon, used for the exact classification of reads.

Both the 32nt and 64nt TRACE analysis classify their reads according to the lowest common taxonomical node where K-mers from said read are present in the reference sequence database, a strategy known as "lowest non-ambiguous mapping". Such a strategy avoids the problem with RNA degradation or sequencing errors by excluding potential errors in reads, without introducing potential ambiguous classification by clustering ambiguous reads under the lowest common classification node such ambiguity is found.

Therefore, if TRACE gives an identification to a specific taxonomical node for a sequence read, it could be from any of the taxonomical nodes and species classified under the node, but it could not be from a taxonomical node or species that is not under said node. E.g. if TRACE says hominoidea which was classified under Catarrhini; Simiiformes; Haplorrhini; Primates; Euarchontoglires, Then it can't be from a pangolin since pangolins (Manis Spp.) are classified under Pholidota; Laurasiatheria. The lowest common classification node between Primates and Pangolins is Boreoeutheria—reads from parts of the genomes shared between Primates and Pangolins will only be classified to Boreoeutheria, but not further classified down toward either Laurasiatheria or Euarchontoglires. And definitely will not be classified individually toward Pholidota or Primates, or any child nodes or phylogenetic nodes under them.

Specific BLAST analysis

Whenever a genus or species is provided by analysis, a specific BLAST analysis is performed to confirm the presence of reads toward the exact species by a search of the database in question with representative reference sequences of the specific species in question in look for matches that is either: 100% match, or: contained no 100% matches on BLAST when queried against the Pangolin reference sequences available on GanBank.

RESULTS

The Accession numbers and contents of all Pan-SL-CoV/GD related sequencing experiments are listed under the following table.

Table 1: List of available GD Pangolin sample datasets as provided under NCBI GenBank. By Accession number, size and citation by thesis (if claimed to have SARS-CoV-2 related reads by paper).

Accession number	Size	SARS-CoV-2-like Coronavirus Identified and Cited?
SRX6893158	16,491,648	
SRX6893157	9,275,501	Lung12 [3] SRR10168374
SRX6893156	22,220,187	Lung11 [1]
SRX6893155	18,067,615	Lung09 [1] [3] SRR10168376
SRX6893154	16,414,925	Lung08 [1] [3] [4] SRR10168377
SRX6893153	19,045,923	Lung07 [1] [3] [4] SRR10168378
SRX6893152	13,527,964	
SRX6893151	16,068,654	
SRX6893150	12,967,281	
SRX6893149	12,590,769	
SRX6893148	15,273,939	

SRX6893147	15,975,904	
SRX6893146	19,038,817	
SRX6893145	19,055,973	
SRX6893144	15,350,468	
SRX6893143	11,527,782	
SRX6893142	20,045,443	
SRX6893141	18,903,834	
SRX6893140	19,986,780	
SRX6893139	39,738,679	Lung02 [3] SRR10168392
SRX6893138	22,900,426	
SRX7756769	107,267,359 PRJNA607174**	M1[2]***
SRX7756766	273,651,431 PRJNA607174**	
SRX7756765	196,761,202 PRJNA607174**	
SRX7756764	222,286,763 PRJNA607174**	
SRX7756763	212,161,250 PRJNA607174**	
SRX7756762	232,433,120 PRJNA607174**	M6[2]***
SRX7756761	113,900,941 PRJNA607174**	
SRX7732094	2,633*	"P2S"[3]

*: "Design: This dataset contains coronavirus-like sequence reads, based on BLAST search."

**: All available SRA datasets from PRJNA607174

***: Actual SRA datasets identified from the "Extended Data Table 3" of [2]

Article

Extended Data Table 3 | Identification of SARSr-CoV sequence reads in metagenomes from the lung of pangolins using the SARS-CoV-2 sequence (GenBank accession No. MN908947) as the reference

Sample ID	Animal species	Total reads*	No. mapped	
M1	Malayan pangolin	107,267,359	496	←SRX7756769 "pangolin 9"
M2	Malayan pangolin	38,091,846	302	Not available
M3	Malayan pangolin	79,477,358	14	
M4	Malayan pangolin	32,829,850	1,100	
M5	Malayan pangolin	547,302,862	56	
M6	Malayan pangolin	232,433,120	10	←SRX7756762 "pangolin 2"
M8	Malayan pangolin	44,440,374	12	Not available
M10	Malayan pangolin	227,801,882	0	
Z1	Chinese pangolin	444,573,526	0	

Fig.1 the "Extended Data Table 3" of [2]. SRA datasets identified in the available database is pointed out by an arrow, while SRA "runs" that failed to be identified in known datasets are outlined in a red square.

Analysis of reads from The Available datasets using NCBI Trace.

Table 2. The Trace result of Known GD Pangolin datasets when examined using NCBI Trace SRA.

Accession number and registration date	Primary Mammalian Trace results and percentage	Primate-related results in Krona and read size by Kbp	Identification of "Coronaviridae" as by Trace and total read size
SRX6893158 20-Sep-2019	Manis javanica: 14.66%	N/D	N/D
SRX6893157 20-Sep-2019	Boreoeutheria: 1.24%	Catarrhini 644546	N/D***
SRX6893156 20-Sep-2019	Manis javanica: 7.51% Homo sapiens: 0.03%	Homo sapiens 81948	Pangolin coronavirus 2Kbp
SRX6893155 20-Sep-2019	Homo sapiens: 0.37%	Homininae 3534150	Pangolin coronavirus 5Kbp
SRX6893154 20-Sep-2019	Homo sapiens: 0.02%	Hominoidea 356003	Pangolin coronavirus 154Kbp
SRX6893153 20-Sep-2019	Homo sapiens: 0.01%	Homo sapiens 162180	Pangolin coronavirus 41Kbp
SRX6893152 20-Sep-2019	Manis javanica: 2.87% Euarchontoglires: 1.37%	N/D	N/D
SRX6893151 20-Sep-2019	Manis javanica: 7.47%	N/D	N/D
SRX6893150 20-Sep-2019	Boreoeutheria: 1.91%	N/D	N/D
SRX6893149 20-Sep-2019	Manis javanica: 1%	Simiiformes 313069	N/D
SRX6893148 20-Sep-2019	Manis javanica: 0.4%	Catarrhini 194320	N/D
SRX6893147 20-Sep-2019	Manis javanica: 2.71%	Catarrhini 69937	N/D
SRX6893146 20-Sep-2019	Boreoeutheria: 1.72%	Hominoidea 231755	N/D
SRX6893145 20-Sep-2019	Homininae: 0.27% Manis javanica: 1.01%	Homininae 2536765	N/D
SRX6893144 20-Sep-2019	Manis javanica: 0.62%	Hominoidea 166628	N/D
SRX6893143 20-Sep-2019	Manis javanica: 1.63%	N/D	N/D
SRX6893142	Manis javanica: 1.28%	Simiiformes 57084	N/D

20-Sep-2019			
SRX6893141 20-Sep-2019	Boreoeutheria: 1.41%	N/D	N/D
SRX6893140 20-Sep-2019	Boreoeutheria: 1.56%	N/D	N/D
SRX6893139 20-Sep-2019	Homo sapiens: 0.01%	Homo sapiens 491120	Pangolin coronavirus 2Kbp
SRX6893138 20-Sep-2019	Boreoeutheria: 1.67%	Homininae 2761176	N/D
SRX7756769 18-Feb-2020	Homo sapiens: 0.03%	Homo sapiens 5457929	Bat SARS-like coronavirus 2Kbp Wuhan seafood market pneumonia virus 2Kbp
SRX7756766 18-Feb-2020	Manis javanica: 78.6%	Cercopithecidae 3116	Betacoronavirus 2Kbp**
SRX7756765 18-Feb-2020	Manis javanica: 87.17%	Cercopithecinae 11339	N/D****
SRX7756764 18-Feb-2020	Manis javanica: 48.39%	Cercopithecidae 22600	N/D
SRX7756763 18-Feb-2020	Manis javanica: 94.95%	Cercopithecidae 5076	N/D
SRX7756762 18-Feb-2020	Manis javanica: 95.37%	Catarrhini* 2831	Nidovirales 0Kbp
SRX7756761 18-Feb-2020	Manis javanica: 13.63%	Chlorocebus sabaeus 498506	N/D
SRX7732094 15-Feb-2020	N/A***	N/A	Pangolin coronavirus***

*: Chlorocebus Sabaeus

** :Not claimed as being SARS-CoV-2 related in the original publication. Likely unrelated.

***Not analyzable. All Non-Coronavirus data filtered out. Leaving only 2,633 reads, all of which can be mapped to the SARS-CoV-2 reference genome.

****8 reads as claimed by [10]

Specific BLAST analysis

In order to determine the authenticity of the Primate-related reads in the datasets, Specific BLAST analysis is carried out for all datasets that possessed claimed or analyzed reads of coronaviridae-related viruses. An 100% full-length match that does not map to non-primates confirms Authenticity of read.

☒ 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"/>	SRX7756762	279	1047	0%	4e-68	100.00%	SRA-SRR11119766.160125840.2
<input checked="" type="checkbox"/>	SRX7756762	279	1366	0%	4e-68	100.00%	SRA-SRR11119766.138036805.1
<input checked="" type="checkbox"/>	SRX7756762	279	967	0%	4e-68	100.00%	SRA-SRR11119766.101239747.1
<input checked="" type="checkbox"/>	SRX7756762	279	1624	0%	4e-68	100.00%	SRA-SRR11119766.46413326.2

Chlorocebus sabaeus isolate 1994-021 unplaced genomic sca ...
dna
1339488
[Distance tree of results](#) [MSA viewer](#) [?](#)

Fig.2a Specific BLAST analysis on the PRJNA607174 dataset, [SRX7756762](#), that contained claimed SARS-CoV-2 related coronavirus reads. The 100% full-length matches clearly indicate presence of Primate-derived material.

☒ 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"/>	Macaca mulatta isolate Rh22777_5890-1b major histocompatibility complex genomic sequence	279	279	100%	2e-71	100.00%	KT332833.1
<input checked="" type="checkbox"/>	Macaca mulatta isolate Rh22335_5775-3 major histocompatibility complex genomic sequence	279	279	100%	2e-71	100.00%	KT332608.1
<input checked="" type="checkbox"/>	Macaca mulatta isolate Rh22335_5725-2 major histocompatibility complex genomic sequence	279	279	100%	2e-71	100.00%	KT332521.1
<input checked="" type="checkbox"/>	Macaca mulatta isolate Rh22335_5702-1a major histocompatibility complex genomic sequence	279	279	100%	2e-71	100.00%	KT332463.1

>gnl|SRA|SRR11119766.160125840.2 160125840 (Biological)
TAATCCTTTGGGTATATACCCAGTAATGGGATGGCTGGGTCATATGGTACATCTAGTTCCT
AGATCCTTGAGGAATCGCCATACTGTTTTCCTAATGGTTGAAGTCTTACATCCAC
CAACAGTGTAAGTGTCTCTATTCTCCAC

<input checked="" type="checkbox"/>	Uncultured organism clone VC1C968TR genomic sequence	230	230	100%	2e-56	94.04%	GQ879596.1
<input checked="" type="checkbox"/>	Synthetic construct Pan troglodytes LINE-1L1P1 retrotransposon tagged with EGFP sequence	219	219	100%	4e-53	92.72%	KF661301.1
<input checked="" type="checkbox"/>	Human artificial chromosome vector 21HAC4 DNA, isolated from the long arm, clone: YACBAC#26-2	219	219	100%	4e-53	92.72%	AB553834.1
<input checked="" type="checkbox"/>	Mus musculus NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length enriched library, clone F630221F08 product un	204	204	86%	1e-48	94.66%	AK171052.1
<input checked="" type="checkbox"/>	Mus musculus bone marrow macrophage cDNA, RIKEN full-length enriched library, clone G530008A19 product hypothetical pr	204	204	86%	1e-48	94.66%	AK149653.1
<input checked="" type="checkbox"/>	Ralstonia solanacearum genome assembly 9 genomes_chromosome - V	202	202	100%	4e-48	90.73%	LN899823.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 06a	154	3044	98%	1e-33	86.43%	CP050586.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 04a	154	4569	100%	1e-33	85.23%	CP050572.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 06b	154	3042	98%	1e-33	86.43%	CP050622.1

Fig.2b BLAST result on the returned sequence revealed it as a Primate-derived MHC complex gene that is not found in non-primates, confirming Primate origin.

☒ 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"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.269072261.2
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.255768440.2
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.255768440.1
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.255318754.2
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.254520929.1
<input checked="" type="checkbox"/>	SRX7756766	279	6344	0%	5e-67	100.00%	SRA-SRR11119762.251645135.1
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.234036838.2
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.211208832.2
<input checked="" type="checkbox"/>	SRX7756766	279	9108	0%	5e-67	100.00%	SRA-SRR11119762.199583624.1
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.198110623.2
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.196936636.2
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.196936636.1
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.133631622.2
<input checked="" type="checkbox"/>	SRX7756766	279	279	0%	5e-67	100.00%	SRA-SRR11119762.108819247.2

Description Macaca mulatta isolate AG07107 chromosome 3 genomic sca ...
Molecule type dna
Query Length 17855752
Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

Fig.3a Specific BLAST analysis of [SRX7756766](#) revealed large amount of 100% full-length matches with Macaca Mulatta.

✓	Macaca mulatta Major Histocompatibility Complex BAC MMU370002, complete sequence	279	279	100%	2e-71	100.00%	AC148706.1
✓	Macaca mulatta Major Histocompatibility Complex BAC MMU122H23, complete sequence	279	521	100%	2e-71	100.00%	AC148677.1
✓	Macaca mulatta Major Histocompatibility Complex BAC MMU065H09, complete sequence	279	279	100%	2e-71	100.00%	AC148671.1
✓	Macaca mulatta Major Histocompatibility Complex BAC MMU038L02, complete sequence	279	279	100%	2e-71	100.00%	AC148668.1
✓	Papio anubis clone rp41-22m16, complete sequence	279	554	100%	2e-71	100.00%	AC113268.8
✓	Papio anubis clone rp41-280n2, complete sequence	279	465	100%	2e-71	100.00%	AC091778.13
✓	Papio anubis clone rp41-5m22, complete sequence	279	279	100%	2e-71	100.00%	AC136143.4
✓	Papio anubis clone rp41-192i11, complete sequence	279	526	100%	2e-71	100.00%	AC091671.28
✓	Macaca mulatta chromosome 9 CH250-18D2, complete sequence	279	548	100%	2e-71	100.00%	CT573219.3
✓	Rhesus Macaque CHR4 BAC CH250-23P16 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Libr	279	279	100%	2e-71	100.00%	AC169807.2
✓	Rhesus Macaque CHR4 BAC CH250-476F18 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Libr	279	279	100%	2e-71	100.00%	AC171646.5
✓	Chlorocebus aethiops BAC clone CH252-163P9 from chromosome 5, complete sequence	278	556	99%	7e-71	100.00%	AC239684.4
✓	MACACA MULATTA BAC clone CH250-192J17 from chromosome unknown, complete sequence	278	552	100%	7e-71	100.00%	AC215693.3
✓	Macaca mulatta isolate Rh22335_5702-1a major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	KT332463.1
✓	Macaca mulatta isolate Rh9_6570-3 major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	KT331777.1
✓	Macaca mulatta isolate Rh9_6550-1b major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	KT331733.1
✓	Macaca mulatta isolate Rh9_6526-2 major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	KT331675.1
✓	Macaca mulatta isolate Rh18665_5547-1b major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	KT329509.1
✓	Macaca mulatta isolate Rh23717 clone 4777 major histocompatibility complex-B genomic sequence	274	274	100%	9e-70	99.34%	KJ913523.1
✓	Macaca mulatta isolate Rh23108 clone 4769-2 major histocompatibility complex-B genomic sequence	274	274	100%	9e-70	99.34%	KJ913420.1
✓	Eukaryotic synthetic construct chromosome 18	202	4103	96%	4e-48	91.78%	CP034496.1
✓	Eukaryotic synthetic construct chromosome 19	196	2020	96%	2e-46	92.09%	CP034522.1
✓	Eukaryotic synthetic construct chromosome 19	196	2130	96%	2e-46	92.09%	CP034497.1
✓	Eukaryotic synthetic construct chromosome 16	196	3654	96%	2e-46	91.61%	CP034494.1
✓	Eukaryotic synthetic construct chromosome 15	196	4994	97%	2e-46	92.09%	CP034493.1
✓	Eukaryotic synthetic construct chromosome 14	196	7731	96%	2e-46	92.09%	CP034492.1
✓	Eukaryotic synthetic construct chromosome 13	191	7988	97%	9e-45	90.91%	CP034516.1
✓	Eukaryotic synthetic construct chromosome Y	191	2781	96%	9e-45	91.37%	CP034510.1
✓	Eukaryotic synthetic construct chromosome 20	191	4517	96%	9e-45	91.37%	CP034499.1
✓	Eukaryotic synthetic construct chromosome 13	191	7988	97%	9e-45	90.91%	CP034491.1
✓	Eukaryotic synthetic construct chromosome 21	185	1053	96%	4e-43	90.65%	CP034500.1
✓	Eukaryotic synthetic construct chromosome 17	185	2450	96%	4e-43	90.65%	CP034495.1
✓	Eukaryotic synthetic construct chromosome 22	183	933	96%	2e-42	90.58%	CP034501.1

Fig.3b BLASTing such matches gives 1005 matches to only Primates, and with no matches outside of Primates. This indicate that [SRX7756766](#) also contained significant amount of material derived from primates.

Results for

2:lc|Query_13045 gnl|SRA|SRR11119762.182596220.2 182596220 (Biologi

Program

BLASTN [Citation](#)

Database

nt [See details](#)

Query ID

lc|Query_13045

Description

gnl|SRA|SRR11119762.182596220.2 182596220 (Biological)

Molecule type

dna

Query Length

151

Other reports

[Distance tree of results](#)

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage Columns

Show

1000

select all

0 sequences selected

GenBank

Graphics

Distance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/>	PREDICTED: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X3, mRNA	279	279	100%	2e-71	100.00%	XM_015150059.2
<input type="checkbox"/>	PREDICTED: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X2, mRNA	279	279	100%	2e-71	100.00%	XM_015150058.2
<input type="checkbox"/>	PREDICTED: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X1, mRNA	279	279	100%	2e-71	100.00%	XM_015150057.2
<input type="checkbox"/>	Papio anubis anubis NIPSNAP1 protein (NIPSNAP1), gene, partial cds, and merlin (NF2) and S-100/CaBP-type calcium binding domain	279	5181	100%	2e-71	100.00%	AH012454.2
<input type="checkbox"/>	PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X3, mRNA	279	279	100%	2e-71	100.00%	XM_015457315.1
<input type="checkbox"/>	PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X2, mRNA	279	279	100%	2e-71	100.00%	XM_005567555.2
<input type="checkbox"/>	PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X1, mRNA	279	279	100%	2e-71	100.00%	XM_005567554.2
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-146I12 from chromosome 6, complete sequence	279	3318	100%	2e-71	100.00%	AC241602.2
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-138D20 from chromosome 13, complete sequence	279	681	100%	2e-71	100.00%	AC239463.3
<input type="checkbox"/>	Macaca mulatta BAC CH250-74N24 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Library) complet	279	6774	100%	2e-71	100.00%	AC204493.6
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-257N12 from chromosome 6, complete sequence	274	6168	100%	9e-70	99.34%	AC241575.3
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-518J9 from chromosome 6, complete sequence	274	4718	100%	9e-70	99.34%	AC241458.2
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-124F12 from chromosome 6, complete sequence	274	2252	100%	9e-70	99.34%	AC241509.3
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-175F15 from chromosome 6, complete sequence	274	4253	100%	9e-70	99.34%	AC241845.3
<input type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-57Q21 from chromosome 6, complete sequence	274	2432	100%	9e-70	99.34%	AC241469.3

Fig.3C Presence of Primate-derived mRNA reads in [SRX7756766](#) confirms the Primate origin of these reads.

☒ select all 100 sequences selected

GraphicsDistance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	SRX7756769	278	278	0%	9e-69	100.00%	SRA-SRR11119759.99831231.2
<input checked="" type="checkbox"/>	SRX7756769	278	278	0%	9e-69	100.00%	SRA-SRR11119759.99831231.1
<input checked="" type="checkbox"/>	SRX7756769	278	4814	1%	9e-69	100.00%	SRA-SRR11119759.88019245.2
<input checked="" type="checkbox"/>	SRX7756769	278	5178	2%	9e-69	100.00%	SRA-SRR11119759.82130976.2
<input checked="" type="checkbox"/>	SRX7756769	278	278	0%	9e-69	100.00%	SRA-SRR11119759.70689253.2
<input checked="" type="checkbox"/>	SRX7756769	278	278	0%	9e-69	100.00%	SRA-SRR11119759.70689253.1
<input checked="" type="checkbox"/>	SRX7756769	278	278	0%	9e-69	100.00%	SRA-SRR11119759.57405658.2
<input checked="" type="checkbox"/>	SRX7756769	278	278	0%	9e-69	100.00%	SRA-SRR11119759.57405658.1

[AC073210.8](#)
Homo sapiens BAC clone RP11-460N20 from 7, complete seq ...
nucleic acid
203396

Fig.4a Similarly, [SRX7756769](#) contained large amount of reads that are 100% full-length matches to Human genomic DNA.

☐ select all 0 sequences selected

GenBankGraphicsDistance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/>	Homo sapiens chromosome 22 clone ABC11_00047178300_E22_complete_sequence	278	456	100%	6e-71	100.00%	AC279316.1
<input type="checkbox"/>	Homo sapiens actin related protein 2 pseudogene (LOC284441) on chromosome 19	278	278	100%	6e-71	100.00%	NG_022927.2
<input type="checkbox"/>	Homo sapiens TBC1 domain containing kinase (TBCK1) RefSeqGene on chromosome 4	278	2140	100%	6e-71	100.00%	NG_034057.3
<input type="checkbox"/>	Homo sapiens chromosome 15 clone VMRC59-280I06_complete_sequence	278	2291	100%	6e-71	100.00%	AC279072.1
<input type="checkbox"/>	Homo sapiens chromosome 2 clone VMRC59-389K09_complete_sequence	278	3905	100%	6e-71	100.00%	AC279037.1
<input type="checkbox"/>	Homo sapiens chromosome 15 clone VMRC59-359A02_complete_sequence	278	3589	100%	6e-71	100.00%	AC278991.1
<input type="checkbox"/>	Homo sapiens chromosome 16 clone VMRC59-453R14_complete_sequence	278	2239	100%	6e-71	100.00%	AC278975.1

Descriptiongnl|SRR11119759.88019245.2 88019245 (Biological)

Molecule typedna

Query Length150

Other reports[Distance tree of results](#) [MSA viewer](#) ?

Query IDlcl|Query_61480

Descriptiongnl|SRR11119759.70689253.1 70689253 (Biological)

Molecule typedna

Query Length150

Other reports[Distance tree of results](#) ?

Percent Identity to

E value to

Query Coverage to

FilterReset

DescriptionsGraphic SummaryAlignmentsTaxonomy

Sequences producing significant alignmentsDownloadManage ColumnsShow1000?

☒ select all 41 sequences selected

GenBankGraphicsDistance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	PREDICTED: Pan paniscus endogenous retrovirus group 3 member 1 envelope (ERV3-1) mRNA	278	278	100%	7e-71	100.00%	XM_034964253.1
<input checked="" type="checkbox"/>	Homo sapiens endogenous retrovirus group 3 member 1 envelope (ERV3-1) transcript variant 1 mRNA	278	278	100%	7e-71	100.00%	NM_001007253.4
<input checked="" type="checkbox"/>	PREDICTED: Pan troglodytes endogenous retrovirus group 3 member 1 envelope (ERV3-1) transcript variant X2 mRNA	278	278	100%	7e-71	100.00%	XM_024357564.1
<input checked="" type="checkbox"/>	PREDICTED: Pan troglodytes endogenous retrovirus group 3 member 1 envelope (ERV3-1) transcript variant X1 mRNA	278	278	100%	7e-71	100.00%	XM_016956774.2
<input checked="" type="checkbox"/>	Synthetic construct DNA clone: pFN21AE1261_Homo sapiens ERV3 gene for endogenous retroviral sequence 3 without stop cod	278	278	100%	7e-71	100.00%	AB590999.1
<input checked="" type="checkbox"/>	Homo sapiens cDNA FLJ60255 complete cds, highly similar to HERV-R_Tg21.2 provirus ancestral Env polyprotein precursor	278	278	100%	7e-71	100.00%	AK295189.1
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-623C19 from chromosome 7, complete sequence	278	278	100%	7e-71	100.00%	AC184799.2
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-565C10 from chromosome 7, complete sequence	278	278	100%	7e-71	100.00%	AC148313.3
<input checked="" type="checkbox"/>	Pan troglodytes endogenous retrovirus group 3 member 1 envelope (ERV3-1) mRNA	278	278	100%	7e-71	100.00%	NM_001135588.1
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-460N20 from 7, complete sequence	278	278	100%	7e-71	100.00%	AC073210.8
<input checked="" type="checkbox"/>	Homo sapiens cDNA FLJ23884 f1s clone LNG13819, highly similar to Human endogenous retrovirus ERV3, pol-env-3LTR region	278	278	100%	7e-71	100.00%	AK074464.1
<input checked="" type="checkbox"/>	Human endogenous retrovirus ERV3, pol-env-3LTR region	278	278	100%	7e-71	100.00%	M12140.1
<input checked="" type="checkbox"/>	Nomascus leucogenys endogenous retrovirus group 3 member 1 envelope (ERV3-1) mRNA	272	272	100%	3e-69	99.33%	NM_001308194.1
<input checked="" type="checkbox"/>	Gibbon endogenous retrovirus proviral envR gene for R envelope protein	272	272	100%	3e-69	99.33%	AJ862653.1
<input checked="" type="checkbox"/>	PREDICTED: Ptilocobus leucosceles endogenous retrovirus group 3 member 1 envelope (ERV3-1) transcript variant X1 mRNA	267	267	100%	2e-67	98.67%	XM_023196289.2
<input checked="" type="checkbox"/>	PREDICTED: Papio anubis endogenous retrovirus group 3 member 1 envelope (ERV3-1) transcript variant X2 mRNA	267	267	100%	2e-67	98.67%	XM_031662056.1
<input checked="" type="checkbox"/>	PREDICTED: Papio anubis endogenous retrovirus group 3 member 1 envelope (ERV3-1) transcript variant X1 mRNA	267	267	100%	2e-67	98.67%	XM_017956680.3
<input checked="" type="checkbox"/>	PREDICTED: Rhinopithecus roxellana endogenous retrovirus group 3 member 1 envelope (ERV3-1) transcript variant X2 mRNA	267	267	100%	2e-67	98.67%	XM_030932630.1
<input checked="" type="checkbox"/>	PREDICTED: Rhinopithecus roxellana endogenous retrovirus group 3 member 1 envelope (ERV3-1) transcript variant X1 mRNA	267	267	100%	2e-67	98.67%	XM_030932629.1

Fig.4b A BLAST analysis on reads sampled from the 100% hit results confirmed that it was found only in humans. Once again confirming human origin.

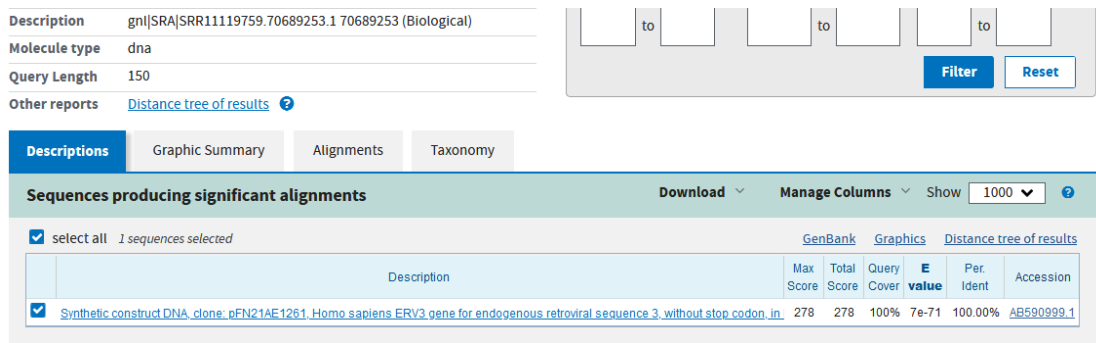


Fig.4c The sequence have no matches outside of Primates.

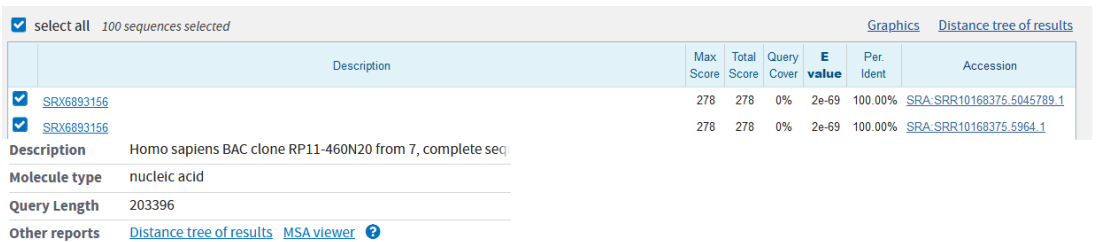


Fig.5a [SRX6893156](#) also returned 100% matched results from the human Genome.

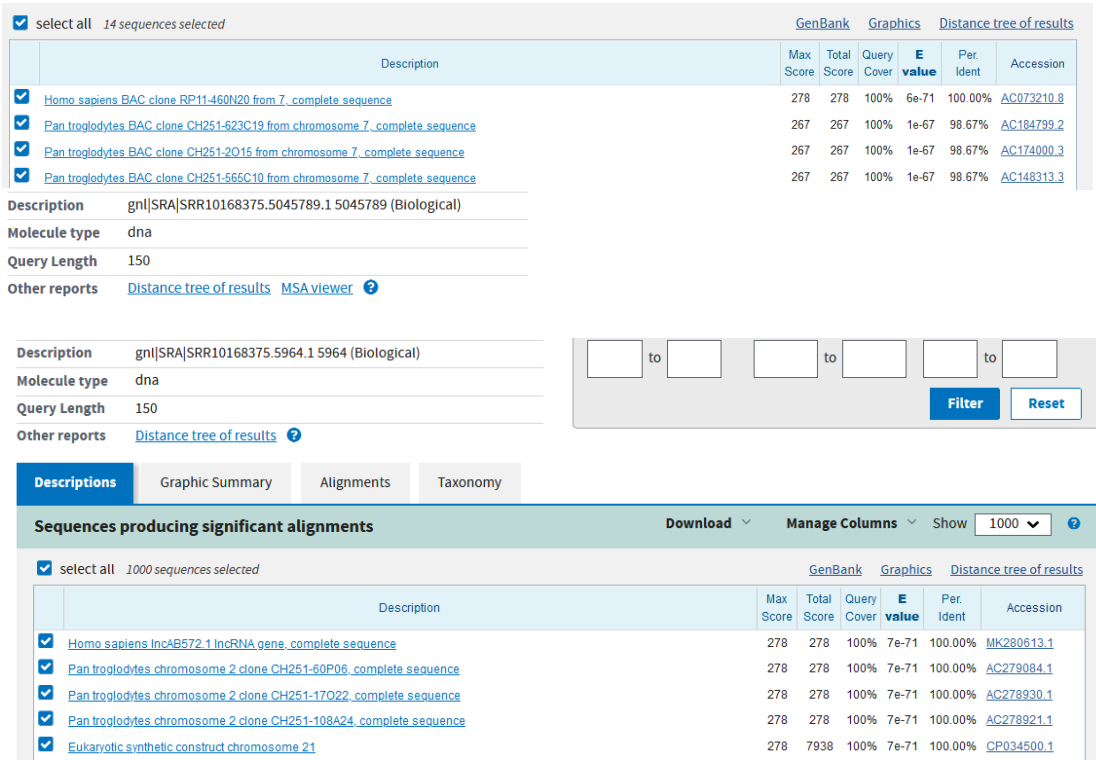


Fig.5b BLAST search on the result returned 100% match only found in humans. Confirming origin in human-derived material.

Description

gnl[SRA]SRR10168375.5964.1.5964 (Biological)

Molecule type

dna

Query Length

150

Other reports

[Distance tree of results](#)

to

to

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage Columns

Show

1000

select all

14 sequences selected

GenBank

Graphics

Distance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 21	278	7938	100%	7e-71	100.00%	CP034500.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	267	14570	100%	1e-67	98.67%	CP034516.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 16	267	10333	100%	1e-67	98.67%	CP034494.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 15	267	9021	100%	1e-67	98.67%	CP034493.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	267	14570	100%	1e-67	98.67%	CP034491.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 18	261	15047	100%	7e-66	98.00%	CP034496.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 17	261	6545	100%	7e-66	98.00%	CP034495.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 20	219	7949	98%	4e-53	93.79%	CP034499.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 19	209	3521	96%	3e-50	93.10%	CP034522.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 19	209	3766	96%	3e-50	93.10%	CP034497.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 22	207	2291	96%	9e-50	92.47%	CP034501.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 14	207	13851	96%	9e-50	92.47%	CP034492.1
<input checked="" type="checkbox"/>	Gossypium hirsutum clone NBRI_GE27093 microsatellite sequence	189	189	96%	3e-44	90.41%	JX591845.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome Y	180	2399	96%	2e-41	89.44%	CP034510.1

Fig.5c BLAST result of the sequences in question revealed that it is not found outside of Primates.

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"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.17339580.1
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.17013625.2
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.17013625.1
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.16930714.2
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.16930714.1
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.15267479.2
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.15267479.1
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.13985702.2
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.13985702.1
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.13353822.2
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.13353822.1
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.11109740.1
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.9343845.2
<input checked="" type="checkbox"/>	SRX6893155	278	278	0%	2e-69	100.00%	SRA_SRR10168376.9232549.2

Description

Homo sapiens BAC clone RP11-460N20 from 7, complete seq ...

Molecule type

nucleic acid

Query Length

203396

Other reports

[Distance tree of results](#)

[MSA viewer](#)

Fig.6a Similarly, BLAST research on [SRX6893155](#) gives large number of full length 100% matches to the human genome.

☒ select all 57 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Homo sapiens FOSMID clone ABC13-48840700E15 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	AC242196.4
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-340I24 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	AC185242.2
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-623C19 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	AC184799.2
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-114G16 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	AC183835.2
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-2015 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	AC174000.3
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-479O9 from 7, complete sequence	278	278	100%	6e-71	100.00%	AC073107.7
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-565C10 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	AC148313.3
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-460N20 from 7, complete sequence	278	278	100%	6e-71	100.00%	AC073210.8
<input checked="" type="checkbox"/>	PREDICTED: Cebus capucinus imitator small integral membrane protein 11A (SMIM11A), transcript variant X6, mRNA	87.9	87.9	49%	1e-13	88.00%	XM_017526193.1

Description gn|SRA|SRR10168376.15267479.2 15267479 (Biological)
Molecule type dna
Query Length 150
Other reports [Distance tree of results](#) [MSA viewer](#) ?

Description gn|SRA|SRR10168376.13985702.1 13985702 (Biological)
Molecule type dna
Query Length 150
Other reports [Distance tree of results](#) ?

to to to

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Manage Columns Show 1000 ?

☒ select all 87 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-460N20 from 7, complete sequence	278	278	100%	7e-71	100.00%	AC073210.8
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-623C19 from chromosome 7, complete sequence	272	272	100%	3e-69	99.33%	AC184799.2
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-565C10 from chromosome 7, complete sequence	272	272	100%	3e-69	99.33%	AC148313.3
<input checked="" type="checkbox"/>	Macaca mulatta BAC CH250-206B6 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Library) con	217	217	100%	2e-52	92.67%	AC210125.6
<input checked="" type="checkbox"/>	Homo sapiens tripartite motif containing 24 (TRIM24), RefSeqGene on chromosome 7	182	182	99%	6e-42	88.59%	NG_023286.1
<input checked="" type="checkbox"/>	Homo sapiens chromosome 7 clone RP11-199L18, complete sequence	182	182	99%	6e-42	88.59%	AC013429.12
<input checked="" type="checkbox"/>	Homo sapiens chromosome 7 clone RP11-258C24, complete sequence	182	182	99%	6e-42	88.59%	AC008265.15
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome Y	176	353	99%	3e-40	87.92%	CP034510.1

☒ select all 3 sequences selected

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome Y	176	353	99%	3e-40	87.92%	CP034510.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 14	143	143	99%	3e-30	84.00%	CP034492.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 16	111	111	72%	7e-21	85.45%	CP034494.1

Fig.6b The results, when put through BLAST, confirms that the 100% matches are in fact derived from a Hominid origin.

Description Homo sapiens BAC clone RP11-460N20 from 7, complete seq ...
Molecule type nucleic acid
Query Length 203396
Other reports [Distance tree of results](#) [MSA viewer](#) ?

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

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"/>	SRX6893153	278	278	0%	2e-69	100.00%	SRA_SRR10168376.1832954.1

Fig.7a [SRX6893153](#) have also returned 100% match full-length read on this tiny part of the human genome.

Description

gnl[SRA|SRR10168378.1832954.1 1832954 (Biological)

Molecule type

dna

Query Length

150

Other reports

[Distance tree of results](#)
[MSA viewer](#)

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage Columns

Show 1000

select all 170 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 FOSMID clone ABC18-862111 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	AC245205.1
<input checked="" type="checkbox"/>	Homo sapiens FOSMID clone ABC13-48840700E15 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	AC242196.4
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-460N20 from 7, complete sequence	278	278	100%	6e-71	100.00%	AC073210.8
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-487D11 from chromosome 7, complete sequence	272	272	100%	3e-69	99.33%	AC182733.3

select all 7 sequences selected

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

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 15	211	211	98%	7e-51	92.11%	CP034493.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 16	206	1214	98%	3e-49	91.45%	CP034494.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	200	200	98%	2e-47	90.79%	CP034516.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 21	200	401	98%	2e-47	90.79%	CP034500.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	200	200	98%	2e-47	90.79%	CP034491.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 18	195	195	96%	7e-46	90.60%	CP034496.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 17	195	195	98%	7e-46	90.13%	CP034495.1

Fig.7b Similarly, the read is only found in humans—indicating the Homo Sapiens Trace result is accurate.

Description

Homo sapiens BAC clone RP11-450O3 from 7, complete sequence

Molecule type

nucleic acid

Query Length

195834

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

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

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	SRX6893154	278	556	0%	1e-69	100.00%	SRA:SRR10168377.16302266.1
<input checked="" type="checkbox"/>	SRX6893154	278	556	0%	1e-69	100.00%	SRA:SRR10168377.12014841.2
<input checked="" type="checkbox"/>	SRX6893154	278	556	0%	1e-69	100.00%	SRA:SRR10168377.12014841.1

Fig.8a Reads from the Human PMS1 gene is recovered from [SRX6893154](#) with a query sequence only 195834bp in length.

Description	gn SRA SRR10168377.16302266.1 16302266 (Biological)					
Molecule type	dna					
Query Length	150					
Other reports	Distance tree of results					

Filter

Reset

Descriptions	Graphic Summary	Alignments	Taxonomy
--------------	-----------------	------------	----------

Sequences producing significant alignments

Download Manage Columns Show 1000

☒ select all 311 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 PMS1 homolog 2, mismatch repair system component pseudogene 8 (PMS2P8) on chromosome 7	278	278	100%	7e-71	100.00%	NG_006447.3
<input checked="" type="checkbox"/>	Homo sapiens PMS1 homolog 2, mismatch repair system component pseudogene 10 (PMS2P10) on chromosome 7	278	278	100%	7e-71	100.00%	NG_023454.4
<input checked="" type="checkbox"/>	Homo sapiens PMS1 homolog 2, mismatch repair system component pseudogene 6 (PMS2P6) on chromosome 7	278	278	100%	7e-71	100.00%	NG_006449.3
<input checked="" type="checkbox"/>	Homo sapiens BAC clone CH17-264B6 from chromosome 7, complete sequence	278	1112	100%	7e-71	100.00%	AC211476.5
<input checked="" type="checkbox"/>	Homo sapiens BAC clone CH17-220H16 from chromosome 7, complete sequence	278	1112	100%	7e-71	100.00%	AC211491.5
<input checked="" type="checkbox"/>	Homo sapiens FOSMID clone ABC10-45515000F15 from chromosome 7, complete sequence	278	556	100%	7e-71	100.00%	AC244146.2
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-45003 from 7, complete sequence	278	556	100%	7e-71	100.00%	AC105418.5
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-396K3 from 7, complete sequence	278	556	100%	7e-71	100.00%	AC006995.5
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-313P13 from 7, complete sequence	278	556	100%	7e-71	100.00%	AC005488.2
<input checked="" type="checkbox"/>	Homo sapiens chromosome 7 clone VMRC53-89F05, complete sequence	272	272	100%	3e-69	99.33%	AC278394.1
<input checked="" type="checkbox"/>	Homo sapiens chromosome 7 clone VMRC62-404M06, complete sequence	272	272	100%	3e-69	99.33%	AC278331.1
<input checked="" type="checkbox"/>	Homo sapiens chromosome 7 clone CH17-61K4, complete sequence	272	272	100%	3e-69	99.33%	AC270699.1
<input checked="" type="checkbox"/>	Homo sapiens PMS8 mRNA (yeast mismatch repair gene PMS1 homologue), partial cds (C-terminal region)	248	248	100%	6e-62	96.67%	D38503.1
<input checked="" type="checkbox"/>	Homo sapiens PMS1 homolog 2, mismatch repair system component (PMS2), RefSeqGene (LRG_161) on chromosome 7	244	244	100%	7e-61	96.00%	NG_008466.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 06a	154	154	84%	1e-33	88.89%	CP050586.1
<input checked="" type="checkbox"/>	Canis lupus familiaris breed Labrador retriever chromosome 06b	154	154	84%	1e-33	88.89%	CP050622.1
<input checked="" type="checkbox"/>	Pipistrellus pipistrellus genome assembly chromosome 5	139	139	86%	3e-29	86.15%	LR862361.1
<input checked="" type="checkbox"/>	Synthetic construct Homo sapiens clone ccsbBroadEn_14772 PMS2 gene, encodes complete protein	126	126	49%	3e-25	97.30%	KJ905275.1
<input checked="" type="checkbox"/>	PREDICTED: Zallophus californianus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X3, mRNA	124	124	54%	9e-25	93.90%	XM_027610028.2
<input checked="" type="checkbox"/>	PREDICTED: Desmodus rotundus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X2, mRNA	121	121	52%	1e-23	94.87%	XM_024576931.1
<input checked="" type="checkbox"/>	PREDICTED: Desmodus rotundus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X1, mRNA	121	121	52%	1e-23	94.87%	XM_024576930.1
<input checked="" type="checkbox"/>	PREDICTED: Myotis lucifugus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X4, mRNA	121	121	52%	1e-23	94.87%	XM_023761457.1
<input checked="" type="checkbox"/>	PREDICTED: Myotis lucifugus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X3, mRNA	121	121	52%	1e-23	94.87%	XM_023761456.1

Fig.8b This PMS1 read is only found in Humans. This is clearly a contaminant from a hominid origin.

Description	Homo sapiens BAC clone RP11-611L7 from 7, complete sequence					
Molecule type	nucleic acid					
Query Length	173967					
Other reports	Distance tree of results MSA viewer					

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"/>	SRX6893139	278	278	0%	3e-69	100.00%	SRA-SRR10168392.39544030.1
<input checked="" type="checkbox"/>	SRX6893139	278	278	0%	3e-69	100.00%	SRA-SRR10168392.28917809.1
<input checked="" type="checkbox"/>	SRX6893139	278	278	0%	3e-69	100.00%	SRA-SRR10168392.14357888.1
<input checked="" type="checkbox"/>	SRX6893139	278	278	0%	3e-69	100.00%	SRA-SRR10168392.2548655.2

Fig.9a Similarly, multiple 100% match Full length reads were obtained from [SRX6893139](#). As this query sequence is only 173967 nucleotides in length, the real extent of Human-derived contamination is also extremely severe.

Description

gnl|SRA|SRR10168392.28917809.1 28917809 (Biological)

Molecule type

dna

Query Length

150

Other reports

[Distance tree of results](#)
[MSA viewer](#)

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

Manage Columns

Show 1000

select all 66 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 zinc finger protein 316 (ZNF316) .mRNA	278	278	100%	6e-71	100.00%	NM_001278559.2
<input checked="" type="checkbox"/>	PREDICTED: Homo sapiens zinc finger protein 316 (ZNF316) .transcript variant X3 .mRNA	278	278	100%	6e-71	100.00%	XM_024446619.1
<input checked="" type="checkbox"/>	PREDICTED: Homo sapiens zinc finger protein 316 (ZNF316) .transcript variant X2 .mRNA	278	278	100%	6e-71	100.00%	XM_024446618.1
<input checked="" type="checkbox"/>	PREDICTED: Homo sapiens zinc finger protein 316 (ZNF316) .transcript variant X1 .mRNA	278	278	100%	6e-71	100.00%	XM_006715630.4
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-611L7 from 7 .complete sequence	278	278	100%	6e-71	100.00%	AC073343.6
<input checked="" type="checkbox"/>	PREDICTED: Pongo abelii zinc finger protein 316 (ZNF316) .mRNA	272	272	100%	3e-69	99.33%	XM_024250011.1
<input checked="" type="checkbox"/>	PREDICTED: Oryzoterpous afer afer zinc finger protein 316 (ZNF316) .mRNA	150	150	100%	2e-32	84.67%	XM_007942750.1
<input checked="" type="checkbox"/>	PREDICTED: Miniopertus natalensis zinc finger protein 853 (ZNF853) .mRNA	145	145	100%	7e-31	84.00%	XM_016213621.1
<input checked="" type="checkbox"/>	PREDICTED: Ochotona princeps zinc finger protein 316 (ZNF316) .mRNA	145	145	98%	7e-31	84.35%	XM_012930995.1
<input checked="" type="checkbox"/>	Pipistrellus pipistrellus genome assembly .chromosome_5	145	145	100%	7e-31	84.11%	LR862361.1

Fig.9b Examining these reads revealed that they are only found in humans and apes. This is therefore also clear evidence that there are Human/Hominid-derived contamination in [SRX6893139](#).

Description

Homo sapiens chromosome 9, clone hRPK.202_H_3, comple ...

Molecule type

nucleic acid

Query Length

187174

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"/>	SRX6893157	278	278	0%	8e-70	100.00%	SRA:SRR10168374.7906491.2

Fig.10a One read is also recovered from [SRX6893157](#). From a query sequence only 187174nt in length.

Description

gnl|SRA|SRR10168374.7906491.2 7906491 (Biological)

Molecule type

dna

Query Length

150

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"/>	PREDICTED: Homo sapiens formin binding protein 1 (FNBP1) .transcript variant X13 .mRNA	278	278	100%	6e-71	100.00%	XM_005251824.2
<input checked="" type="checkbox"/>	PREDICTED: Homo sapiens formin binding protein 1 (FNBP1) .transcript variant X4 .mRNA	278	278	100%	6e-71	100.00%	XM_011518402.1
<input checked="" type="checkbox"/>	PREDICTED: Homo sapiens formin binding protein 1 (FNBP1) .transcript variant X3 .mRNA	278	278	100%	6e-71	100.00%	XM_011518401.1
<input checked="" type="checkbox"/>	Homo sapiens formin binding protein 1 (FNBP1) .RefSeqGene on chromosome 9	278	278	100%	6e-71	100.00%	NG_033946.1
<input checked="" type="checkbox"/>	Homo sapiens cDNA FLJ13619.11s .clone PLACE1010926 .weakly similar to HYPOTHETICAL 72.2 KD PROTEIN C12C2.05C IN CHROMOSOM	278	278	100%	6e-71	100.00%	AK023681.1
<input checked="" type="checkbox"/>	Human DNA sequence from clone RP11-138E2 on chromosome 9q34.11-34.3 .complete sequence	278	278	100%	6e-71	100.00%	AL136141.13
<input checked="" type="checkbox"/>	Homo sapiens formin-binding protein 17 (FBP17) .mRNA .partial cds	278	278	100%	6e-71	100.00%	AF265550.1
<input checked="" type="checkbox"/>	Homo sapiens chromosome 9 .clone hRPK.202_H_3 .complete sequence	278	278	100%	6e-71	100.00%	AC006241.1
<input checked="" type="checkbox"/>	Homo sapiens KIAA0554 mRNA for KIAA0554 protein	278	278	100%	6e-71	100.00%	AB011126.1
<input checked="" type="checkbox"/>	PREDICTED: Nomascus leucogenys formin binding protein 1 (FNBP1) .transcript variant X18 .mRNA	272	272	100%	3e-69	99.33%	XM_030818029.1
<input checked="" type="checkbox"/>	PREDICTED: Nomascus leucogenys formin binding protein 1 (FNBP1) .transcript variant X17 .mRNA	272	272	100%	3e-69	99.33%	XM_030818028.1
<input checked="" type="checkbox"/>	PREDICTED: Nomascus leucogenys formin binding protein 1 (FNBP1) .transcript variant X16 .mRNA	272	272	100%	3e-69	99.33%	XM_030818027.1

Fig.10b This particular sequence is only found in humans—indicating that even the [SRX6893157](#) dataset was contaminated by material of human origin.

DescriptionChlorocebus aethiops BAC clone CH252-276C1 from chromo: ...
Molecule typenucleic acid
Query Length160189
Other reportsDistance tree of resultsMSA viewer

Percent Identityto
E valueto
Query Coverage to
FilterReset

DescriptionsGraphic SummaryAlignments

Sequences producing significant alignmentsDownloadManage ColumnsShow100

select all 100 sequences selectedGraphicsDistance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	SRX7756765	279	774	0%	4e-69	100.00%	SRA_SRR11119763.129105044.1
<input checked="" type="checkbox"/>	SRX7756765	279	375	0%	4e-69	100.00%	SRA_SRR11119763.106754018.1
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	SRA_SRR11119763.75715814.1
<input checked="" type="checkbox"/>	SRX7756765	279	510	0%	4e-69	100.00%	SRA_SRR11119763.47541416.1
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	SRA_SRR11119763.40703161.1
<input checked="" type="checkbox"/>	SRX7756765	279	768	0%	4e-69	100.00%	SRA_SRR11119763.38948039.1

Fig.11a The presence of Reads from Somatic Chlorocebus aethiops in [SRX7756765](#) confirms the identity of the Cercopithecinae reads there.

Descriptiongnl[SRA|SRR11119763.106754018.1 106754018 (Biological)
Molecule typedna
Query Length151
Other reportsDistance tree of results

to to to
FilterReset

DescriptionsGraphic SummaryAlignmentsTaxonomy

Sequences producing significant alignmentsDownloadManage ColumnsShow1000

select all 1000 sequences selectedGenBankGraphicsDistance tree of results

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-276C1 from chromosome 6, complete sequence	279	375	100%	2e-71	100.00%	AC241496.3
<input checked="" type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-445L10 from chromosome unknown, complete sequence	279	690	100%	2e-71	100.00%	AC238953.2
<input checked="" type="checkbox"/>	Macaca mulatta Y Chr BAC CH250-11J13 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC)	279	1513	100%	2e-71	100.00%	AC240710.4
<input checked="" type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-371E9 from chromosome 5, complete sequence	279	1124	100%	2e-71	100.00%	AC239563.3
<input checked="" type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-461K13 from chromosome 13, complete sequence	279	629	100%	2e-71	100.00%	AC239354.3
<input checked="" type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-138D20 from chromosome 13, complete sequence	279	493	100%	2e-71	100.00%	AC239463.3
<input checked="" type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-417L1 from chromosome 6, complete sequence	279	472	100%	2e-71	100.00%	AC239275.3
<input checked="" type="checkbox"/>	Chlorocebus aethiops BAC clone CH252-62B18 from chromosome 4, complete sequence	279	510	100%	2e-71	100.00%	AC239442.2

<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 18	268	1.545e+05	100%	4e-68	100.00%	CP034496.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 16	268	1.008e+05	100%	4e-68	98.68%	CP034494.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 19	263	47303	100%	2e-66	98.01%	CP034522.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	263	1.935e+05	100%	2e-66	99.31%	CP034516.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome Y	263	1.175e+05	100%	2e-66	99.31%	CP034510.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 21	263	55262	100%	2e-66	98.01%	CP034500.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 20	263	88148	100%	2e-66	99.31%	CP034499.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 19	263	47481	100%	2e-66	98.01%	CP034497.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 17	263	74732	100%	2e-66	98.01%	CP034495.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 15	263	1.384e+05	100%	2e-66	98.01%	CP034493.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 14	263	1.910e+05	100%	2e-66	99.31%	CP034492.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	263	1.935e+05	100%	2e-66	99.31%	CP034491.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 22	257	22814	100%	9e-65	97.35%	CP034501.1
<input checked="" type="checkbox"/>	Human gammaherpesvirus 4 isolate HKNPC80, partial genome	248	248	100%	5e-62	96.03%	MH590571.1
<input checked="" type="checkbox"/>	Human gammaherpesvirus 4 isolate HKHD40, partial genome	248	248	100%	5e-62	96.03%	MH590409.1
<input checked="" type="checkbox"/>	Uncultured bacterium clone LMOABA39ZH04FM1 genomic sequence	246	246	96%	2e-61	97.24%	EU058958.1
<input checked="" type="checkbox"/>	Synthetic construct Pan troglodytes LINE-1L1 1P1 retrotransposon tagged with EGFP sequence	241	241	100%	9e-60	95.36%	KF661301.1
<input checked="" type="checkbox"/>	Uncultured organism clone VC1C968TR genomic sequence	241	241	96%	9e-60	96.55%	GQ879596.1

Fig.11b the sequences from the BLAST hits indicate that they were unique to the family Cercopithecinae. Confirming Primate origin.

Analyzing the extent of contamination.

As the Specific BLAST analysis confirmed significant level of Human-derived contamination in all samples positive for SARS-CoV-2 related Coronaviruses, The TRACE result can therefore be

trusted for the analysis on the extent of contamination.

The 32nt Krona Trace system is used for elucidating the ratio of different taxa within a sample. As Specific BLAST analysis confirmed the significant presence of Human and Primate derived Genetic material--The most basal group of primates detected in all Coronavirus-positive samples belong to Catarrhini—or Humans, Apes and Old-World Monkeys. Therefore, Trace classification results that can be classified into sister nodes of Catarrhini should be considered as Contamination by Primate-derived material.

Since Catarrhini is under Simiiformes; Haplorrhini; Primates; Euarchonta; Euarchontoglires and Manis is under Pholidota; Laurasiatheria, If a read is TRACed down to Catarrhini, it can not be from a Pangolin, and it will have to be from a Primate-derived source—Contamination by material from the lab.

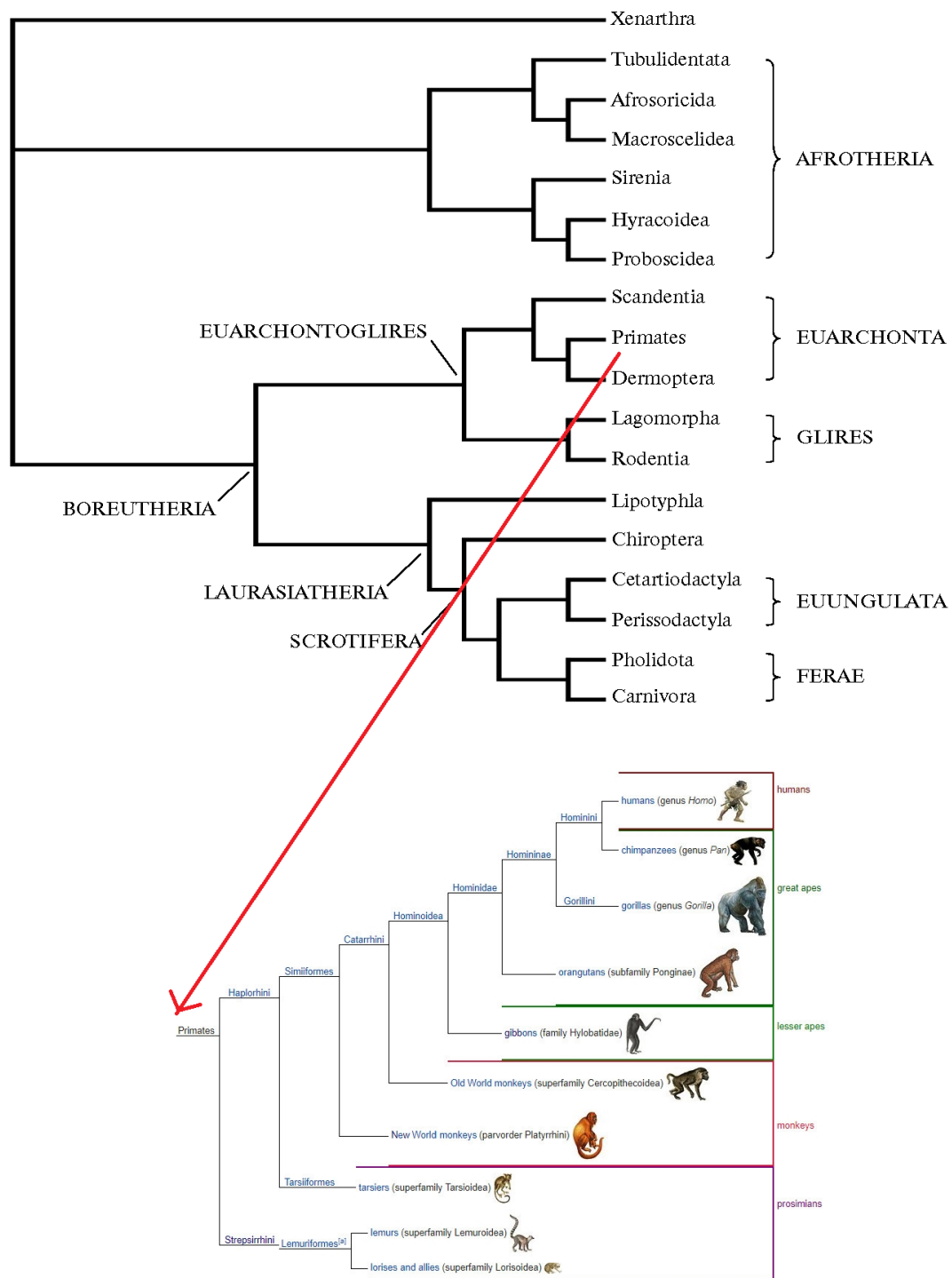


Fig. 12 Family tree of mammals, Including the position and classification of Primates in the lineage of Mammalia.

Table 3a Ratios of Hominid-traced reads to Pangolin-traced reads in the SRA datasets that contained reads of the GD- Pangolin-CoV sequence, and had Hominid reads.

Accession and date	Primate classification and total traced Kbps	Total traced Kbps to Manis Javanica (Pangolin)	Ratio of Primate to Pangolin	Virus classification and amount of reads by Kbps
SRX7756769 18-Feb-2020	Homo sapiens 5457929	15401134	0.35	Bat SARS-like coronavirus 2Kbp Wuhan seafood market pneumonia virus 2Kbp
SRX6893139 20-Sep-2019	Homo sapiens 491120	5301351	0.0926	Pangolin coronavirus 2Kbp
SRX6893157 20-Sep-2019	Catarrhini 644546	1889448	0.34	N/D***
SRX6893156 20-Sep-2019	Homo sapiens 81948	4765461	0.01719	Pangolin coronavirus 2Kbp
SRX6893155 20-Sep-2019	Homininae 3534150	525801	6.7214	Pangolin coronavirus 5Kbp
SRX6893154 20-Sep-2019	Hominoidea 356003	2232008	0.159	Pangolin coronavirus 154Kbp
SRX6893153 20-Sep-2019	Homo sapiens 162180	3110158	0.05214	Pangolin coronavirus 41Kbp

***: No trace result on Coronaviruses, despite claimed reads from [3]

Table 3b Ratios of Primate-traced reads to Coronavirus-traced reads in the SRA datasets that contained reads claimed to be traced to of the GD- Pangolin-CoV sequence, and lacked Hominid reads.

Accession and date	Primate classification and reads (in Kbp)	Virus classification and reads	Ratio of virus reads to Primate reads
SRX7756766 18-Feb-2020	Cercopithecidae 3116; BLAST to Macaca Mulatta	Betacoronavirus 2Kbp **	0.000642
SRX7756762 18-Feb-2020	Catarrhini 2831; BLAST to Chlorocebus sabaeus	Nidovirales 0Kbp Claimed 10x150bp reads	0.000530
SRX7756765 22-Apr-2020	Cercopithecinae 11339 BLAST to Chlorocebus Aethiops	N/D***	N/A
SRX7732094 15-Feb-2020	N/A*	Pangolin coronavirus	N/A*

*: No non-coronavirus reads available in the dataset with a total of 2,633 reads, making analysis impossible.

**: No claimed reads from [2]

***: Claimed 8 reads from [10]

DISCUSSIONS

The extent of contamination in the pangolin sequencing datasets

As the samples were supposed to be pangolin lung tissue, which will neither contact with nor be contaminated by non-pangolin derived mammalian tissues when still inside the animal, any non-pangolin mammalian reads within such a dataset can only be introduced to the sequencing process after the sample itself have been taken and brought into a lab.

As the classification Catarrhini itself is phylogenetically very deep down the Primate line which is itself distinguished from the Pangolin line at a very basal node (Boreoeutheria), and since we have already confirmed that the Primate line in PRJNA573298 traces mostly to humans by using Specific BLAST analysis, ([SRX6893157](#), the only one of the claimed coronavirus read dataset that gives a classification just down to Catarrhini, contained 213 full length 100% matches to the Human Mitochondrial reference genome alone, which is only 16569 bp in length. All other datasets gives definitive TRACE mapping to Homo Sapiens and contained distinct 100% matched reads to even very small parts of the Human genome.), We can deduce the extent of contamination of the PRJNA573298 dataset by Primate-related materials as from a minimum of 1.6% to as high as 87% by sample mass—using the ratio of Primate reads to Pangolin reads on TRACE. Such high level of contamination with Primate-derived material is unacceptable for a sample that was supposed to be Lung tissue. And therefore, the virome data of such samples in PRJNA573298 no longer reflects the original virome of the animal, and an potential “novel” reads from these contaminated samples may have been from in-lab contamination instead.

Deducing the dynamic of contamination in PRJNA607174

Of all 7 PRJNA607174 datasets, only [SRX7756769](#) and [SRX7756762](#) is claimed by Xiao et. Al to contain SARS-CoV-2-like reads. However, TRACE results revealed low level of contamination by Cercopithecidae (Old World Monkey) reads across all the samples. In particular, the [SRX7756762](#) dataset contained definitive mappings to *Chlorocebus sabaeus*, or African Green Monkey, while [SRX7756766](#) which contained 2Kbp unclaimed reads of Betacoronaviruses on TRACE, contained 100% full-length definitive mappings to *Macaca Mulatta* that may also be mapped to *Chlorocebus Aethiops* and *Homo Sapiens*.

[SRX7756769](#) genetically resembles other samples in PRJNA573298, in both the kind of contamination and the extent of contamination. It contained an large excess of homo sapiens reads in levels similar to the contaminated samples in PRJNA573298.

From the method section of Lam et.al, we knew that they have performed Virus isolation using VERO E6 cells—Species *Chlorocebus Sabaeus* on one of the samples that have a positive PCR test for coronaviruses. The low level of contamination by Cercopithecidae-related reads in all the samples in PRJNA607174 except for [SRX7756769](#) itself support the possibility that [SRX7756769](#) is the first sample to be sequenced, and it happens before the lab begun using VERO E6 cells in the experiment. They then isolated the virus from the contaminated [SRX7756769](#) in VERO E6 cells, characterized it but did not sequence it, and this cell culture material then contaminated [SRX7756762](#) and possibly [SRX7756766](#), resulting the 10 reads in [SRX7756762](#) and the 2Kb bat coronavirus reads in [SRX7756766](#).

The exact nature of [SRX7732094](#) needs to be further scrutinized.

The P2S dataset, [SRX7732094](#), displays very unusual property when compared to other Datasets under the same BioProject. It is the only dataset with all Non-coronavirus reads being filtered out, and contained too little spots for it to be an ILLUMINA NextSeq 550 run. Furthermore, it was the only dataset that did not contain metadata with either an isolation source or a Library prep procedure, other than “This dataset contains coronavirus-like sequence reads, based on BLAST search.”

Such a strange designation and the fact of the dataset being heavily filtered, Raises problems on whether such a dataset is an actual BioSample at all. If this sample is really as claimed by Lam et. Al, Why the dataset have to be put through such heavy filtering when the other sequencing runs was clearly not filtered as severely as this dataset? Why there was no BioSample metadata on either Biomaterial provider, Source Tissue or Collector when all other Sequencing runs clearly provided such metadata information?

Unless the complete, unfiltered sequencing reads are made available on [SRX7732094](#), and the rest of [PRJNA606875](#), this Dataset can not be considered to be a real, reliable sample, and it must be excluded as “evidence” of a SARS-CoV-2-like virus infecting

pangolins in Guangdong, 2019.

Table 4 Sequencing runs in PRJNA696875, Accession number, BioSample, Content and designation

Accession number and date	Size	Non-Coronavirus reads?	Source Tissue Provider and Collected by	Virus Designation: GD or GX?	Design
SRX7732094 15-Feb-2020	2,633	No	N/A	GD	This dataset contains coronavirus-like sequence reads, based on BLAST search.
SRX7732093 15-Feb-2020	470,344	Yes	Intestine Yanling Hu Wuchun Cao	GX	NEBNext Ultra II DNA Library Prep Kit, paired sequencing data has been integrated.
SRX7732092 15-Feb-2020	340,661	Yes	Lung Yanling Hu Wuchun Cao	GX	NEBNext Ultra II DNA Library Prep Kit, paired sequencing data has been integrated.
SRX7732091 15-Feb-2020	416,659	Yes	Intestine Yanling Hu Wuchun Cao	GX	NEBNext Ultra II DNA Library Prep Kit, paired sequencing data has been integrated.
SRX7732090 15-Feb-2020	520,254	Yes	Lung Yanling Hu Wuchun Cao	GX	NEBNext Ultra II DNA Library Prep Kit, paired sequencing data has been

					integrated.
SRX7732089 15-Feb-2020	19,607,536	Yes	Blood Yanling Hu Wuchun Cao	GX	Ion Total RNA-Seq Kit v2
SRX7732088 15-Feb-2020	4,550,437	Yes	lung and intestine Yanling Hu Wuchun Cao	GX	Ion Total RNA-Seq Kit v2

By closely examining the P2V dataset, SRX7732088, which claimed to be a culture sample in VERO E6 cells, Chlorocebus Sabaeus, the exact viral load in-culture when compared to Cellular mRNA can be deduced by dividing the total identifiable coronavirus signal to the total identifiable Primate signal within the dataset, 6943Kbp/451932Kbp, which correspond to 0.01536:1 Viral RNA to Cellular RNA.

This places the viral loads on the other datasets with Coronavirus-like reads from GD well within the threshold expected from cell culture contamination of the sequencing samples—including the samples in PRJNA607174.

Potential breach of data availability statement by Xiao et al.[2]

Sequence data that support the findings of this study have been deposited in GISAID with the accession numbers EPI_ISL_410721. Raw data of RNAseq are available from the NCBI SRA under the study accession number PRJNA607174.

Fig 13. The Data Availability Statement of Xiao et al.

In the Data availability statement, the “Raw data of RNAseq” are clearly stated to be deposited under PRJNA607174. However, only 2 of the “Extended Data Table S3” datasets actually matches the datasets deposited on PRJNA607174. The other 7 datasets were completely unavailable. And the actual deposited datasets on PRJNA607174 does not match what have been claimed by Extended Data Table S3. As the RNA-seq Raw data was stated to be available within PRJNA607174, the failure to publish all the claimed data constitute a breach of the Data Availability statement on the article. Unless such datasets are published and independently examined, All such claimed reads from the strangely unpublished datasets can not be trusted as evidence of a SARS-CoV-2-like virus infecting pangolins in Guangdong, 2019.

Identifying the Etiological agent of the Guangdong 2019 incident.

By using an approach of both SRA TRACE analysis and specific BLAST Analysis, We have uncovered the fact that all samples that does not Contain confirmed Human-derived material, also lacked Claimed reads of a SARS-CoV-2 like virus that can be confirmed using NCBI Trace. All samples with claimed or traced reads of Coronaviruses in general, contained confirmed primate reads with the lowest common phylogenetic node Catarrhini. Samples that does not give a TRACE result on primate-derived material all lacked identifiable or claimed coronavirus reads.

This strongly imply that the Coronavirus-like reads are associated with human/Primate-sourced contamination material.

Most importantly, of all dead pangolins being sampled in the studies, only 9 out of a total of 29

Analyzable samples/datasets contained TRACEd or Claimed Coronavirus reads—despite all dead pangolins displayed similar symptoms in captivity. This imply that the alleged pangolin coronavirus is not the Etiological agent of the death of the pangolins being sampled in the studies. This is further supported by the fact that 4 out of 10 lung samples in PRJNA573298 and 4 out of 7 lung samples in PRJNA607174 lacked any claimed or TRACEd coronavirus reads—despite the same symptoms displayed and similar date of death.

In order to establish the Etiological agent of the dead pangolins in the single GuangDone Accident that leads to the sampling and studies. A full virome TRACE analysis is conducted on the available samples for the determining of the exact etiological agent.

Extended Data Table S1

Full virome TRACE results of all Analyzable datasets of the GD pangolin incident

	Mammarenavirus	Nairoviridae	Murine respirovirus	Flaviviridae	Nidovirales	Rubulavirus	Nonanavirus	Peribunyavi	Amigovirus	Siphoviridae	Siphoviridae	Paheavir
SRX6893158	Yes	Yes	No	No	No	No	Yes	No	Yes	Yes	No	No
SRX6893157	Yes	Yes	No	No	Claimed	No	No	Yes	No	No	No	No
SRX6893156	No	No	Yes	Yes	Yes	No	No	No	Yes	No	No	Yes
SRX6893155	No	No	Yes	No	Yes	No	No	No	No	No	No	No
SRX6893154	No	No	Yes	No	Yes	No	No	No	No	No	No	No
SRX6893153	No	No	Yes	Yes	Yes	No	No	No	Yes	No	No	No
SRX6893152	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No	No
SRX6893151	Yes	Yes	No	Yes	No	No	No	Yes	Yes	No	No	No
SRX6893150	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	No
SRX6893149	Yes	Yes	No	No	No	No	No	No	No	No	Yes	No
SRX6893148	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No
SRX6893147	Yes	Yes	"Respirovirus"	Yes	No	No	Yes	No	Yes	No	No	No
SRX6893146	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
SRX6893145	Yes	Yes	No	No	No	No	No	No	No	No	No	No
SRX6893144	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
SRX6893143	Yes	Yes	No	No	No	No	No	No	No	No	No	No
SRX6893142	Yes	Yes	No	No	No	No	No	Yes	Yes	No	No	No
SRX6893141	Yes	Yes	No	Yes	No	No	No	No	No	No	No	No
SRX6893140	Yes	Yes	Yes	No	No	No	No	Yes	No	No	No	No
SRX6893139	No	No	Yes	No	Yes	No	No	No	No	No	No	No
SRX6893138	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No	No
SRX7756766	No	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No
SRX7756765	No	No	Yes	No	No	Yes	No	No	No	No	No	No
SRX7756764	No	No	Yes	No	No	Yes	No	No	No	No	No	No
SRX7756763	No	No	Yes	No	No	Yes	No	No	No	No	No	No
SRX7756762	No	No	Yes	No	Claimed	Yes	No	No	No	No	No	No
SRX7756761	No	No	Yes	No	No	Yes	No	No	No	No	No	No
SRX7756769	No	No	Yes	Yes	Yes	No	No	No	No	No	No	No

A full Virome TRACE result suggest all the dead pangolins were infected by either Mammarenaviruses or Murine Respirovirus, or both. Including both samples that contained Claimed or TRACEd Coronavirus reads and the samples that didn't.

Murine Respirovirus and Mammarenaviruses co-infect 7 out of 29 Available Analyzable datasets, while none of the 29 datasets lacked both—indicating that both viruses were prevalent in the location where the pangolins were captive at The Guangdong Wildlife Rescue Center.

Symptoms of Murine Respirovirus in animals resembles that of SARS-CoV-2 in humans—It forms massive Syncytia in Eukaryotic cells, suppresses the immune system and causes secondary bacterial infections. The virus causes necrosis of Lung tissue in 5 days, with similar inflammation and immunopathological effects in the lung tissues of infected animals [5]—creating the histopathological effect as reported by Xiao et al.

It should be worth pointing out that the only examined lung tissues were examined by Xiao et al. And all Lung tissue samples examined by Xiao et.al contained Reads from the Murine Respirovirus.

Similarly, Mammarenaviruses are also known to cause multi organ, lethal[7] infections, characterized by endothelial pathology and swelling of internal organs. [6] All of which were Symptoms reported in the incident. As these samples were not examined Histopathologically by either the authors of [4] nor by any of the authors of any other article who have used the

datasets/samples, leaving the only mean of elucidating the cause of death being the observed symptoms and the coarse examination of the organs during sampling. Mammarenavirus infection therefore remains the most likely cause of death of the Murine Respirivirus Negative samples in the available datasets.

Is the “GD pangolin CoV” really a virus of the pangolin?

The only examination of the binding affinity of the GD pangolin CoV RBD to different animal receptors was done by Xiao et al [2], which performed molecular dynamic simulation of the RBD docking to the Human ACE2 receptor, The Civet ACE2 receptor and the pangolin ACE2 receptor. If the RBD of GD pangolin CoV in deed evolved in pangolins, we should expect the binding affinity of the RBD toward the pangolin ACE2 receptor to be the highest binding affinity returned from the examination.

However, neither the GD pangolin CoV RBD, nor the RBD of SARS-CoV-2 which is highly similar, produced a higher binding affinity to the pangolin ACE2 receptor than to the human ACE2 receptor, and both binds the Human ACE2 receptor with the highest affinity across all 3 animal species (Human, Civet, Pangolin) examined.

This fact argues strongly against the RBD residues of the GD pangolin CoV being evolved in pangolins, and instead favoring the RBD and the virus being the result of a passage experiment of a possible virus of pangolin origin (The GX/P2V virus was isolated and passaged in VERO E6 cells during it’s collection in 2017) in Primate-derived cell lines.

There are only 2 locations of Biological sample storage in GuangDong, the Guangdong Institute of Applied Biological Resources and the China National GeneBank.

As all Credible (Non-filtered and contained analyzable Non-Coronavirus reads) samples were collected in a single incident from the Guangdong Wildlife Rescue Center[1][4][2], which the initial sample collection and storage was carried out by the Guangdong Institute of Applied Biological Resources[4], this experimental culture likely contaminated the GD pangolin samples during their initial collection or Storage, Either by the lab worker doing the initial sampling, or during their storage in the facility.

Epidemiology analysis of SARS-CoV-2 and related viruses argues strongly against the existence of a Coronavirus with the claimed RBD residues and sequence similarity in or near the Guangdong Wildlife Rescue Center at the time and date of the incident and the collection of the samples.

The earliest collection date of the GD pangolin CoV available, MP789, GenBank MT084071.1, is displayed at 29 March 2019.

Since the original location of the animals and samples in question was inside the Guangdong Wildlife Rescue Center which is neither a certified Biosafety Laboratory nor possessed adequate PPE when handling the animals, from the Simulation results by Xiao et al[2] and the observed

high human transmissibility of SARS-CoV-2 which had a very similar RBD, Should the GD pangolin CoV genuinely exists at that date and within the unprotected GuangDong Wildlife Rescue Center, It would almost certainly infect one to multiple On-site workers (Rescue workers which lacked either the Biosafety training or the adequate PPEs required to handle tissues or animals infected with a virus as characterized by the GD pangolin CoV papers) in the GuangDong Wildlife Rescue Center, and caused a SARS-level epidemic in GuangDong 2013 beginning in or around April 2019. However, no such epidemic was recorded, nor there have been any virus that genetically resembled the GD pangolin CoV sequence (which is only 90% similar to SARS-CoV-2) being isolated in humans anywhere in the world even till today.

Nor there is a possibility that the current SARS-CoV-2 pandemic may have stemmed from the 29 March incident with the GD pangolin CoV, since the estimated time of divergence between the current SARS-CoV-2 genome to the GD pangolin CoV Genome was estimated to be at least 100 years ago , ranging from 1851 [1730,1958] to 1877 [1746,1986] [8], for a genome that is only 90% similar to SARS-CoV-2 and possessed significant difference in the sequence and composition of the viral proteins they encodes.

As the Earliest time of discovery and the incident on the GD pangolin CoV is no earlier than the beginning of Year 2019, The time between the incident and the first isolate of SARS-CoV-2 is far too short for GD pangolin CoV incident to be involved in the formation of the current SARS-CoV-2 pandemic, since even the neutral sites on the RBD itself would have taken more than 19.8 years to drift/evolve into what we seen today on the actual SARS-CoV-2 genome. [9]

Homo Sapiens reads are also found in SRX7756766, SRX7756765 and SRX7756762

In addition to Chlorocebus Spp. Indicative of VERO E6 cells, we have also found trace amount of reads uniquely matched to Hominidae within SRX7756766, SRX7756765 and SRX7756762. The presence of such reads may indicate that the Coronavirus-like reads from within such dataset were the result of index-hopping from a more highly contaminated original sample dataset, such as SRX7756769, of which the Homo Sapiens reads within such datasets may have index hopped into SRX7756766, SRX7756765 and SRX7756762, alongside with extremely low level of Coronavirus-related reads. Alternatively, these Homo Sapiens reads may represent the Homo Sapiens host sequence from the original sample which were left in the cell culture medium after inoculation of the VERO E6 cells as indicated by the virus isolation procedure performed by Xiao et al[2]. The presence of low levels of Homo Sapiens sequences within these datasets also confirms the origin of the inoculum into the VERO E6 cells as being samples that had an extremely large amount of Homo Sapiens genetic material within them.

Job Title

AF228730:Homo sapiens chromosome 8 clones...

RID

ZAYT406301R

Search expires on 01-07 15:56 pm

Download All

Program

BLASTN

Citation

Database

SRA

See details

Query ID

AC006457.4

Description

Homo sapiens PAC clone RP5-1193P9 from 7p11.2-q11.2, con ...

Molecule type

nucleic acid

Query Length

134187

Other reports

Distance tree of results

MSA viewer

Descriptions

Graphic Summary

Alignments

Sequences producing significant alignments

Download

New

Manage columns

Show

100

☒ select all

100 sequences selected

Graphics

Distance tree of results

New

MSA Viewer

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession					
<input checked="" type="checkbox"/>	SRX7756766	279	964	0%	4e-69	100.00%	151	SRA_SRR11119762.119033805.2					
<input checked="" type="checkbox"/>	SRX7756766	279	552	0%	4e-69	100.00%	151	SRA_SRR11119762.33937308.1					
<input checked="" type="checkbox"/>	SRX7756766	279	992	0%	4e-69	100.00%	151	SRA_SRR11119762.27294164.2					
<input checked="" type="checkbox"/>	SRX7756766	279	666	0%	4e-69	100.00%	151	SRA_SRR11119762.27294164.1					
<input checked="" type="checkbox"/>	SRX7756766	274	548	0%	2e-67	99.34%	151	SRA_SRR11119762.33937308.2					
<input checked="" type="checkbox"/>	SRX7756766	252	705	0%	9e-61	96.69%	151	SRA_SRR11119762.87190953.2					
<input checked="" type="checkbox"/>	SRX7756766	252	964	0%	9e-61	96.69%	151	SRA_SRR11119762.65743002.2					
<input checked="" type="checkbox"/>	SRX7756766	252	705	0%	9e-61	96.69%	151	SRA_SRR11119762.65743002.1					
<input checked="" type="checkbox"/>	SRX7756766	246	675	0%	4e-59	96.03%	151	SRA_SRR11119762.267695997.2					
<input checked="" type="checkbox"/>	SRX7756766	246	686	0%	4e-59	96.03%	151	SRA_SRR11119762.87190953.1					
<input checked="" type="checkbox"/>	SRX7756766	241	905	0%	2e-57	95.36%	151	SRA_SRR11119762.267695997.1					
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13					eukaryotic synt...	279	1.710e+05	100%	2e-71	100.00%	96089878	CP034491.1
<input checked="" type="checkbox"/>	Homo sapiens glypican 5 (GPC5) RefSeqGene on chromosome 13					Homo sapiens	279	4892	100%	2e-71	100.00%	1475557	NC_009370.2

Your search is limited to records that exclude: Homininae (taxid:207598)

Job Title

2 sequences (gnl|SRA|SRR11119762.27294164.1...

RID

ZBABPT58016

Search expires on 01-07 19:13 pm

Download All

Results for

1:lc|Query_20116 gnl|SRA|SRR11119762.27294164.1 27294164 (Biologica

Program

BLASTN

Citation

Database

nt

See details

Query ID

lc|Query_20117

Description

gnl|SRA|SRR11119762.27294164.2 27294164 (Biological)

Molecule type

dna

Query Length

151

Other reports

Distance tree of results

MSA viewer

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download

New

Select columns

Show

1000

☒ select all

1000 sequences selected

GenBank

Graphics

Distance tree of results

New

MSA Viewer

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	eukaryotic synt...	279	2.436e+05	100%	2e-71	100.00%	96089878	CP034516.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 22	eukaryotic synt...	279	29397	100%	2e-71	100.00%	35194566	CP034501.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 18	eukaryotic synt...	279	1.956e+05	100%	2e-71	100.00%	84899006	CP034496.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 15	eukaryotic synt...	279	1.731e+05	100%	2e-71	100.00%	82521392	CP034493.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	eukaryotic synt...	279	2.436e+05	100%	2e-71	100.00%	96089878	CP034491.1
<input checked="" type="checkbox"/>	Pongo abelii BAC clone CH276-289G9 from chromosome unknown_complete sequence	Pongo abelii	279	279	100%	2e-71	100.00%	221923	AC215315.3
<input checked="" type="checkbox"/>	Pongo abelii BAC clone CH276-62H12 from chromosome unknown_complete sequence	Pongo abelii	279	559	100%	2e-71	100.00%	69319	AC190147.2
<input checked="" type="checkbox"/>	Pongo abelii chromosome 7 clone CH276-122D18_complete sequence	Pongo abelii	274	537	100%	1e-69	99.34%	218515	AC278951.1
<input checked="" type="checkbox"/>	Cloning vector pTLD39_complete sequence	Cloning vector p...	274	274	100%	1e-69	99.34%	76343	MG711910.1
<input checked="" type="checkbox"/>	Cloning vector pTLD67_complete sequence	Cloning vector p...	274	274	100%	1e-69	99.34%	72952	MG711909.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 19	eukaryotic synt...	274	66262	100%	1e-69	99.34%	64242768	CP034522.1

Filter Results

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

>gnl|SRA|SRR11119762.27294164.1 27294164 (Biological)

CCTCCGTCGTCCATGTGATCTCTCAATTGCCACCTATGAGTGAGAAATATGCGGTG

TTTGGTTTTTTGTTCTTCCGATAGTTTACTGAGAATGATGATTTCCAATTTTCATCATGT

CCCTACAAAGGACATGAACTCATCATTTTTT

>gnl|SRA|SRR11119762.27294164.2 27294164 (Biological)

AACCAAGCCAAATGTCCAACAATGATAGACTGGATTAGAGAAATGTGGACATATACACC

ATGGAATACTATGCAGCCATAAAAAATGATGAGTTTCATGTCCTTTGTAGGACATGGATG

AAATTGGAATCATCAATTCAGTAAACTAT

Fig 14: In addition to Chlorocebus Aethiops, reads uniquely matched to Hominidae have been found within SRX7756766.

Job Title

AC018449:Homo sapiens 3 BAC RP11-26G10 (Roswell..

RID

ZDA099D901R [Search expires on 01-08 13:19 pm](#) [Download All](#) [▼](#)

Program

BLASTN [Citation](#) [▼](#)

Database

SRA [See details](#) [▼](#)

Query ID

AC018449.11

Description

Homo sapiens 3 BAC RP11-26G10 (Roswell Park Cancer Institi ...

Molecule type

nucleic acid

Query Length

157426

Other reports

[Distance tree of results](#) [MSA viewer](#) [?](#)

Descriptions

Graphic Summary

Alignments

Sequences producing significant alignments

Download [▼](#) [New](#) [Manage columns](#) [▼](#) Show

100 [▼](#) [?](#)

☒ select all 100 sequences selected

[Graphics](#) [Distance tree of results](#) [New](#) [MSA Viewer](#)

	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA-SRR11119763.183685474.2
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA-SRR11119763.183685474.1
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA-SRR11119763.173512342.2
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA-SRR11119763.173512342.1
<input checked="" type="checkbox"/>	SRX7756765	279	1021	0%	4e-69	100.00%	151	SRA-SRR11119763.171368621.2
<input checked="" type="checkbox"/>	SRX7756765	279	1031	0%	4e-69	100.00%	151	SRA-SRR11119763.171368621.1
<input checked="" type="checkbox"/>	SRX7756765	279	992	0%	4e-69	100.00%	151	SRA-SRR11119763.152058533.2
<input checked="" type="checkbox"/>	SRX7756765	279	975	0%	4e-69	100.00%	151	SRA-SRR11119763.152058533.1
<input checked="" type="checkbox"/>	SRX7756765	279	959	0%	4e-69	100.00%	151	SRA-SRR11119763.136417907.2
<input checked="" type="checkbox"/>	SRX7756765	279	701	0%	4e-69	100.00%	151	SRA-SRR11119763.136417907.1
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA-SRR11119763.84993302.2
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA-SRR11119763.84993302.1
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA-SRR11119763.77679299.2
<input checked="" type="checkbox"/>	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA-SRR11119763.77679299.1

Job Title

2 sequences (gnl|SRA|SRR11119763.171368621.1...

RID

ZDVkJ9V7013 [Search expires on 01-08 18:20 pm](#) [Download All](#) [▼](#)

Results for

1:lc|Query_8150 gnl|SRA|SRR11119763.171368621.1 171368621 (Biologic [▼](#)

Program

BLASTN [Citation](#) [▼](#)

Database

nt [See details](#) [▼](#)

Query ID

lc|Query_8150

Description

gnl|SRA|SRR11119763.171368621.1 171368621 (Biological)

Molecule type

dna

Query Length

151

Other reports

[Distance tree of results](#) [MSA viewer](#) [?](#)

Descriptions

Graphic Summary

Alignments

Taxonomy

Sequences producing significant alignments

Download [▼](#) [New](#) [Select columns](#) [▼](#) Show

1000 [▼](#) [?](#)

☒ select all 1000 sequences selected

[GenBank](#) [Graphics](#) [Distance tree of results](#) [New](#) [MSA Viewer](#)

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Homo sapiens 3 BAC RP11-26G10 (Roswell Park Cancer Institute Human BAC Library) complete sequ...	Homo sapiens	279	1031	100%	2e-71	100.00%	157426	AC018449.11
<input checked="" type="checkbox"/>	Homo sapiens chromosome 5 clone CTB-25P15 complete sequence	Homo sapiens	278	506	99%	8e-71	100.00%	174158	AC011399.6
<input checked="" type="checkbox"/>	Homo sapiens adenosine kinase (ADK) RefSeqGene on chromosome 10	Homo sapiens	272	732	99%	4e-69	99.33%	565119	NG_020484.2
<input checked="" type="checkbox"/>	Homo sapiens neuroigin 1 (NLGN1) RefSeqGene on chromosome 3	Homo sapiens	272	742	99%	4e-69	99.33%	905473	NG_046919.2
<input checked="" type="checkbox"/>	Pongo abelii BAC clone CH276-312H8 from chromosome 15 complete sequence	Pongo abelii	272	777	99%	4e-69	99.33%	196619	AC221006.3
<input checked="" type="checkbox"/>	Pongo abelii BAC clone CH276-98A16 from chromosome 1 complete sequence	Pongo abelii	272	1032	99%	4e-69	99.33%	196248	AC205911.3
<input checked="" type="checkbox"/>	Pongo abelii chromosome UNK clone CH276-22H12 complete sequence	Pongo abelii	272	272	99%	4e-69	99.33%	225183	AC186760.1

Filter Results

Percent Identity

to

E value

to

Query Coverage

to

Filter

Reset

>gnl|SRA|SRR11119763.171368621.1 171368621 (Biological)
CTTGTGATTTTGTACATTGATTTGTATCCTGAACCTTTGCTGAAGTTGCTTATCAGCT
TAAGGAGATTTTGGGCTGAGACAAATGGGGTTTCTAGATATACAAATCATGCTCTGCAA
ACAGGGACAATTTGACTTCCTCTTTTCCTAG

>gnl|SRA|SRR11119763.171368621.2 171368621 (Biological)
AGGAAATAAAGGGTATTCAACTAGGAAAAGAGGAAGTCAAAATGTCCTGTTTGACAGCG
ACATGATTGTATATCTAGAAAAACCCATTGTCTCAGCCCAAAATCTCCTTAAGCTGATAA
GCAACTTCAGCAAAGTTTCAGGATACAAAT

Filter 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

Query ID	AP024011.1
Description	Homo sapiens DNA, sequence_id: unplaced_0526
Molecule type	nucleic acid
Query Length	13084
Other reports	Distance tree of results MSA viewer ?

Descriptions	Graphic Summary		Alignments					
Sequences producing significant alignments								
Download <div>▼</div> New Manage columns <div>▼</div> Show <div>100</div> <div>▼</div> ?								
<input checked="" type="checkbox"/> select all 2 sequences selected Graphics Distance tree of results New MSA Viewer								
<div><input checked="" type="checkbox"/></div>	Description <div>▼</div>	Max Score	Total Score	Query Cover <div>▼</div>	E value <div>▼</div>	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	SRX7756762	279	8344	44%	4e-70	100.00%	151	SRA_SRR11119766.44541545.2
<input checked="" type="checkbox"/>	SRX7756762	279	8680	44%	4e-70	100.00%	151	SRA_SRR11119766.44541545.1

>gnl|SRA|SRR11119766.44541545.1 44541545 (Biological)
AGAAAAGGAATATCTTCGTATAAAACCGACAGATCATTTCTCAGAACTGCTCTGTG
ATGTGTGCTTCAACTCACAGAGTTTAACTTTTCTTTTCATTTCAGCAGTTTGGAAACACT
CTGTTTGTAAAGTCTGCAAGTGGATATCTTG
>gnl|SRA|SRR11119766.44541545.2 44541545 (Biological)
GAAGGCTCTAAGAGGCCAAGATATCCACTTGCAGACTTTACAACAGAGTGTTTCCAAA
CTGCTGATGAAGAAAGAAATTAACCTCTGTGAGTTGAACGCACACATCAGAGCAGTT
TCTGAGGATGATTCCTCTCGGTTTATACG

Job Title	2 sequences (gnl SRA SRR1119766.44541545.1...
RID	ZNE2JD1H013 Search expires on 01-11 15:18 pm Download All ▼
Results for	1:lc Query_63282 gnl SRA SRR11119766.44541545.1 44541545 (Biologica ▼
Program	BLASTN ? Citation ▼
Database	nt See details ▼
Query ID	lc Query_63282
Description	gnl SRA SRR11119766.44541545.1 44541545 (Biological)
Molecule type	dna
Query Length	151
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to

E value to

Query Coverage to

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

Descriptions	Graphic Summary	Alignments	Taxonomy						
Sequences producing significant alignments									
Download ▼ New Select columns ▼ Show 1000 ▼ ?									
<input checked="" type="checkbox"/> select all 464 sequences selected GenBank Graphics Distance tree of results New MSA Viewer									
<input checked="" type="checkbox"/>	Description ▼	Scientific Name	Max Score	Total Score	Query Cover	E value	Per Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Homo sapiens DNA, sequence_id: unplaced_0526	Homo sapiens	279	8680	100%	2e-71	100.00%	13084	AP024011.1
<input checked="" type="checkbox"/>	Homo sapiens DNA, sequence_id: unplaced_0092	Homo sapiens	279	16564	100%	2e-71	100.00%	92273	AP023577.1
<input checked="" type="checkbox"/>	Human clone ps16 DNA	Homo sapiens	279	279	100%	2e-71	100.00%	336	X60718.1
<input checked="" type="checkbox"/>	Homo sapiens DNA, sequence_id: unplaced_0210	Homo sapiens	270	19297	100%	1e-68	98.69%	51519	AP023695.1
<input checked="" type="checkbox"/>	Homo sapiens DNA, sequence_id: unplaced_0016	Homo sapiens	268	4425	100%	5e-68	98.68%	380611	AP023501.1
<input checked="" type="checkbox"/>	Homo sapiens DNA, sequence_id: unplaced_0231	Homo sapiens	263	29291	100%	2e-66	98.01%	46435	AP023716.1
<input checked="" type="checkbox"/>	Homo sapiens DNA, sequence_id: unplaced_0494	Homo sapiens	252	10155	100%	5e-63	96.69%	15862	AP023979.1

i Your search is limited to records that exclude: Homo (taxid:9605)	
Job Title	2 sequences (gnl SRA SRR1119766.44541545.1...
RID	ZNGGJ9BZ013 Search expires on 01-11 16:00 pm Download All ▼
Results for	1:lc Query_49712 gnl SRA SRR11119766.44541545.1 44541545 (Biologica ▼
Program	BLASTN ? Citation ▼
Database	nt See details ▼
Query ID	lc Query_49712
Description	gnl SRA SRR11119766.44541545.1 44541545 (Biological)
Molecule type	dna
Query Length	151
Other reports	Distance tree of results MSA viewer ?

Filter Results

Organism only top 20 will appear ☐ exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity to

E value to

Query Coverage to

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

Descriptions	Graphic Summary	Alignments	Taxonomy						
Sequences producing significant alignments									
Download ▼ New Select columns ▼ Show 1000 ?									
<input checked="" type="checkbox"/> select all 116 sequences selected GenBank Graphics Distance tree of results New MSA Viewer									
<input checked="" type="checkbox"/>	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 19	eukaryotic synth...	246	4866	100%	2e-61	96.03%	64242768	CP034522.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 19	eukaryotic synth...	246	4941	100%	2e-61	96.03%	64242768	CP034497.1
<input checked="" type="checkbox"/>	Ralstonia solanacearum genome assembly 9 genomes_chromosome_1	Ralstonia solana...	222	439	97%	4e-54	93.88%	5416545	LN899819.1
<input checked="" type="checkbox"/>	Chimpanzee aliphoid repetitive sequence, monomers 1 and 2, clone C-alpha-R1(1020), 15-208 II	Pan troglodytes	217	217	100%	2e-52	92.72%	340	L01727.1
<input checked="" type="checkbox"/>	Chimpanzee (clone C-alpha-R1(1020), 15-164 II) aliphoid repeat sequence	Pan troglodytes	217	217	100%	2e-52	92.72%	340	M97599.1

Fig.15: Reads uniquely matched to Homo Sapiens Alpha Satellite DNA in SRX7756762.

Conclusions

The Extreme lack of transparency and the sheer level of contamination from the original samples, the lack of epidemiological evidence of it's existence at the location of it's collection, and the receptor binding affinity of the Viral RBD itself indicating it as not being evolved nor adapted in pangolins, all strongly argue against the existence of a SARS-CoV-2 like virus infecting pangolins captive in GuangDong at 2019.

Moreover, it suggests that the GD pangolin CoV exists only as a culture in Primate-derived cells within the lab/facility used for the initial collection and/or storage of the samples of the pangolins in question, raising important issues on the serial passage Gain-Of-Function research of viral pathogens.

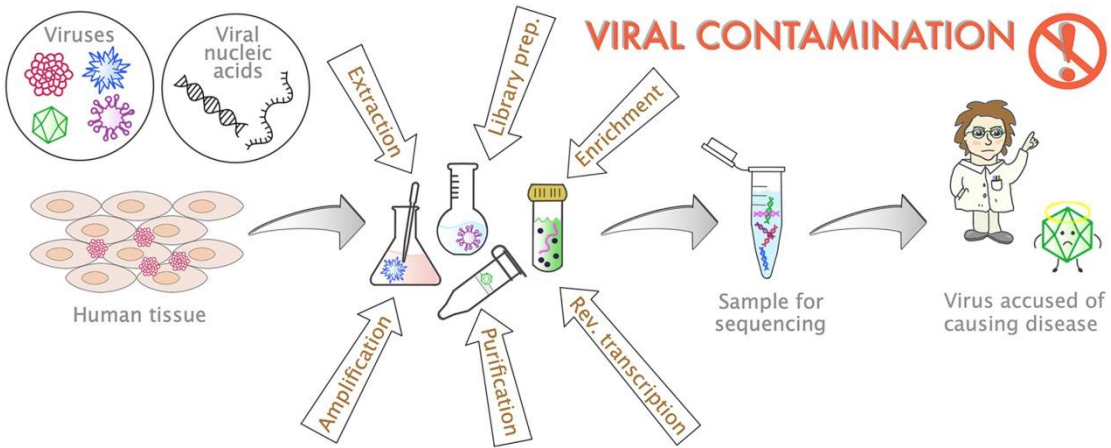


Figure 17. A cartoon diagram of contamination in sequencing experiment leading to false results and false “discoveries”.

Note as in 2020/7/23

A recent Dataset, seemingly unrelated to the Xiao et.al Nature dataset, [SRX8582289](#), appeared under [PRJNA607174](#). This dataset seems to be newly sequenced, and it was not referred in [2].

Table S2: TRACE analysis result of the [SRX8582289](#) dataset.

Accession number and registration date	Primary Mammalian Trace results and percentage	Primate-related results in Krona and read size by Kbp	Identification of “Coronaviridae” as by Trace and total read size
SRX8582289 22-Jun-2020	Manis javanica: 43.52%	Catarrhini 98913	Pangolin coronavirus 792

Nevertheless, in-depth analysis revealed significant amount of contamination from the Human genome, with ratio of Virus to cell=0.8%.

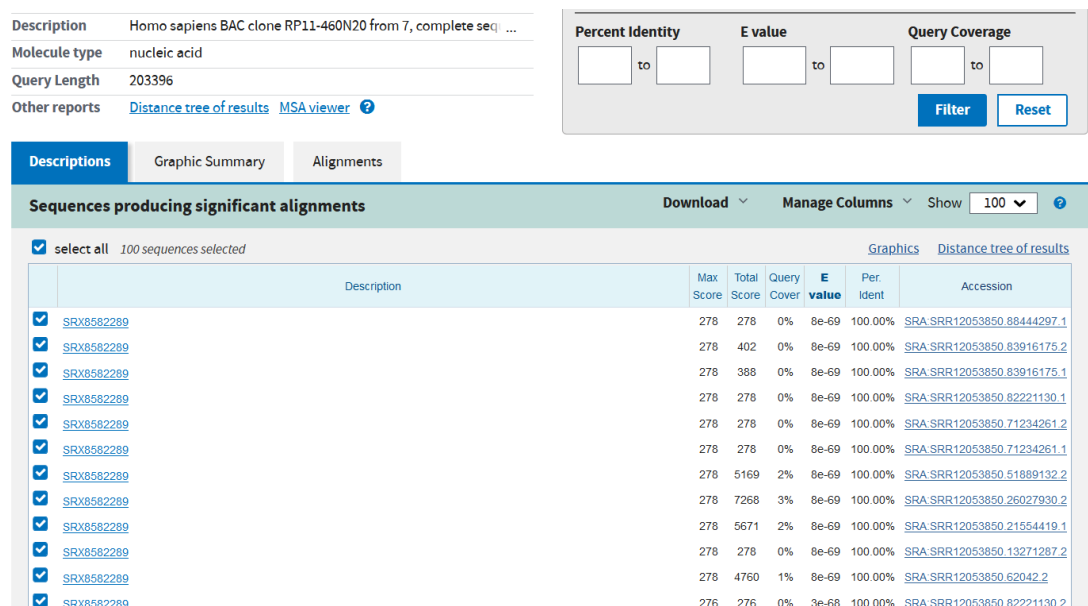


Figure S1A: Some BLAST hits out of a human Somatic BAC clone.

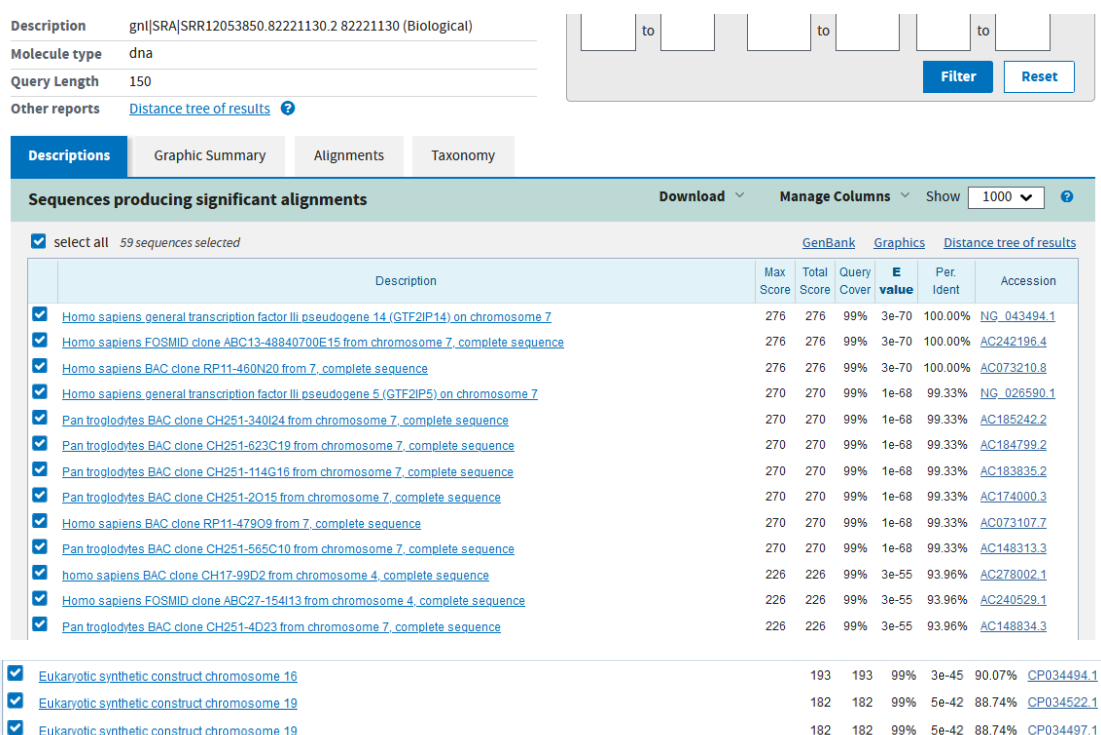


Fig. S1B: BLAST results returned only Homo Sapiens as 100% match. This indicate that the listed Catarrhini reads come from Homo Sapiens.

The significance of this particular dataset is yet unknown.

Note as in 2020/12/26

Two Recent SRAs, [SRX9714436](#) and [SRX9714921](#), were recently deposited by the Guangdong Institute of Applied Biological Resources with a listed DOI connection to 10.1371/journal.ppat.1008421 [1]. Both samples have a depositor of LinMao Li, 2020-12-21 the same time as the specified BioProject registration date. Only one of the SRAs contained

significant amount of Coronavirus-related reads.

Table S3: TRACE analysis result of [SRX9714436](#) and [SRX9714921](#).

Accession number and registration date	Primary Mammalian Trace results and percentage	Primate-related results in Krona and read size by Kbp	Identification of “Coronaviridae” as by Trace and total read size
SRX9714436	Manis javanica: 3.14% Homo sapiens: 0.04%	Homo sapiens 12332	Pangolin coronavirus 3
SRX9714921	Homo sapiens: 0.15%	Homo sapiens 9923	N/D

As expected by TRACE results, Reads that are 100% full-length uniquely matched to Homo Sapiens were obtained from [SRX9714436](#) and [SRX9714921](#).

>gnl|SRA|SRR13285085.43 43
CTATACAAACCCCATGACACGAGTTTACCTATGTAAACAACTTCA

☒ select all 100 sequences selected

Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Homo sapiens BAC clone RP11-488C22 from 2, complete sequence	human	93.5	93.5	100%	5e-16	100.00%	165351	AC019109.9
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-617M1 from chromosome unknown, complete sequence	chimpanzee	87.9	87.9	100%	2e-14	98.00%	188195	AC183921.2
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-564M21 from chromosome unknown, complete sequence	chimpanzee	87.9	87.9	100%	2e-14	98.00%	201789	AC160021.3
<input checked="" type="checkbox"/> Homo sapiens chromosome 5, BAC clone 89K19 (LBNL H179), complete sequence	human	87.9	87.9	100%	2e-14	98.00%	110312	AC005215.2
<input checked="" type="checkbox"/> Homo sapiens chromosome 5 clone CTD-2332G20, complete sequence	human	87.9	87.9	94%	2e-14	100.00%	146437	AC010489.4
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome Y	eukaryotic syn...	84.2	168	90%	3e-13	100.00%	64558752	CP034510.1
<input checked="" type="checkbox"/> Pongo abelli BAC clone CH276-222M18 from chromosome 8, complete sequence	Sumatran ora...	84.2	84.2	90%	3e-13	100.00%	196788	AC206339.2

Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome Y	eukaryotic synthetic construct	84.2	168	90%	3e-13	100.00%	64558752	CP034510.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 15	eukaryotic synthetic construct	80.5	80.5	92%	4e-12	97.83%	82521392	CP034493.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 14	eukaryotic synthetic construct	76.8	151	94%	5e-11	95.74%	88289540	CP034492.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 18	eukaryotic synthetic construct	75.0	210	100%	2e-10	95.65%	84899006	CP034496.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	96089878	CP034516.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 16	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	98200793	CP034494.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	96089878	CP034491.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 20	eukaryotic synthetic construct	69.4	138	86%	8e-09	95.35%	68480253	CP034499.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 19	eukaryotic synthetic construct	62.1	62.1	84%	1e-06	92.86%	64242768	CP034497.1

Fig.S2A: Homo Sapiens 100% full-length Unique matched read obtained from [SRX9714436](#)

>gnl|SRA|SRR13285570.36 36
ATAGGGAAGTGTGGTACCAAGGAGCAATATTCATACAGCAACCGAAG

Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> Human DNA sequence from clone XX-DSH1_29E11, complete sequence	human	93.5	93.5	100%	5e-16	100.00%	112322	CU041292.6
<input checked="" type="checkbox"/> Human DNA sequence from clone RP11-535B18 on chromosome 9, complete sequence	human	93.5	93.5	100%	5e-16	100.00%	126815	AL354931.13

Fig.S2B: Homo Sapiens 100% full-length Unique matched read obtained from [SRX9714921](#)

A provided .fastq file was also found in [SRX9714436](#). Analysis suggest significant presence of Homo Sapiens reads within this fastq file, similar to that of the Run itself.

NDX550397_RUO:309:H3FKWBGXH:1:11101:21446:1055 16 AC019109.9 87488 255
 50M * 0 0 TGAAGGTTTGTACATAGGTAAACTCGTGCATGGGGGTTTGTGTATAG *
 NH:i:1 AS:i:50 NM:i:0

	Description	Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	Homo sapiens BAC clone RP11-488C22 from 2, complete sequence	human	93.5	93.5	100%	5e-16	100.00%	165351	AC019109.9
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-617M1 from chromosome unknown, complete sequence	chimpanzee	87.9	87.9	100%	2e-14	98.00%	188195	AC183921.2
<input checked="" type="checkbox"/>	Pan troglodytes BAC clone CH251-564M21 from chromosome unknown, complete sequence	chimpanzee	87.9	87.9	100%	2e-14	98.00%	201789	AC160021.3
<input checked="" type="checkbox"/>	Homo sapiens chromosome 5, BAC clone 89K19 (LBNL H179), complete sequence	human	87.9	87.9	100%	2e-14	98.00%	110312	AC005215.2
<input checked="" type="checkbox"/>	Homo sapiens chromosome 5 clone CTD-2332G20, complete sequence	human	87.9	87.9	94%	2e-14	100.00%	146437	AC010489.4
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome Y	eukaryotic syn...	84.2	168	90%	3e-13	100.00%	64558752	CP034510.1
<input checked="" type="checkbox"/>	Pongo abelii BAC clone CH276-222M18 from chromosome 8, complete sequence	Sumatran ora...	84.2	84.2	90%	3e-13	100.00%	196788	AC206339.2
<input checked="" type="checkbox"/>	Human DNA sequence from clone RP11-987D21 on chromosome X, complete sequence	human	84.2	84.2	96%	3e-13	97.92%	55442	BX119919.5
<input checked="" type="checkbox"/>	PREDICTED: Callithrix jacchus uncharacterized LOC118154814 (LOC118154814), ncRNA	white-buffed-e...	82.4	82.4	94%	1e-12	97.87%	2719	XR_004745075.1
<input checked="" type="checkbox"/>	Homo sapiens solute carrier family 26 member 3 (SLC26A3), RefSeqGene (LRG_683) on chromos...	human	82.4	82.4	100%	1e-12	96.00%	44767	NG_008046.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome Y	eukaryotic synthetic construct	84.2	168	90%	3e-13	100.00%	64558752	CP034510.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 15	eukaryotic synthetic construct	80.5	80.5	92%	4e-12	97.83%	82521392	CP034493.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 14	eukaryotic synthetic construct	76.8	151	94%	5e-11	95.74%	88289540	CP034492.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 18	eukaryotic synthetic construct	75.0	210	100%	2e-10	95.65%	84899006	CP034496.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	96089878	CP034516.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 16	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	98200793	CP034494.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	96089878	CP034491.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 20	eukaryotic synthetic construct	69.4	138	86%	8e-09	95.35%	68480253	CP034499.1
<input checked="" type="checkbox"/>	Eukaryotic synthetic construct chromosome 19	eukaryotic synthetic construct	62.1	62.1	84%	1e-06	92.86%	64242768	CP034497.1

Fig.S3: BLAST result of the read **NDX550397_RUO:309:H3FKWBGXH:1:11101:21446:1055 TGAAGGTTTGTACATAGGTAAACTCGTGCATGGGGGTTTGTGTATAG** within the provided fastq file. The read is 100% full-length uniquely matched to Homo Sapiens.

REFERENCES

[1] Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)?

Ping Liu ,
 Jing-Zhe Jiang ,
 Xiu-Feng Wan,
 Yan Hua,
 Linmiao Li,
 Jiabin Zhou,
 Xiaohu Wang,
 Fanghui Hou,
 Jing Chen,
 Jiejian Zou,
 Jinping Chen

Published: May 14, 2020
<https://doi.org/10.1371/journal.ppat.1008421>

[2] Xiao, K., Zhai, J., Feng, Y. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2313-x>
 [3] Lam, T.T., Shum, M.H., Zhu, H. *et al.* Identifying SARS-CoV-2 related coronaviruses in Malayan

pangolins. *Nature* (2020). <https://doi.org/10.1038/s41586-020-2169-0>

[4] Liu, P.; Chen, W.; Chen, J.-P. Viral Metagenomics Revealed Sendai Virus and Coronavirus Infection of Malayan Pangolins (*Manis javanica*). *Viruses* **2019**, *11*, 979.

[5] Inducible epithelial resistance improves survival of Sendai virus pneumonia in mice by both inactivating virus and preventing CD8+ T cell-mediated immunopathology

S. Wali, J. R. Flores, A.M. Jaramillo, D. L. Goldblatt, J. Pantaleón García, M. J. Tuvim, B. F. Dickey, S. E. Evans

doi: <https://doi.org/10.1101/2020.01.30.917195>

[6] Jorlan Fernandes, Renata Carvalho de Oliveira, Alexandro Guterres, Débora Ferreira Barreto-Vieira, Ana Claudia Pereira Terças, Bernardo Rodrigues Teixeira, Marcos Alexandre Nunes da Silva, Gabriela Cardoso Caldas, Janice Mery Chicarino de Oliveira Coelho, Ortrud Monika Barth, Paulo Sergio D'Andrea, Cibele Rodrigues Bonvicino, Elba Regina Sampaio de Lemos,

Detection of Latino virus (Arenaviridae: Mammarenavirus) naturally infecting *Calomys callidus*,

Acta Tropica,

Volume 179,

2018,

Pages 17-24,

ISSN 0001-706X,

<https://doi.org/10.1016/j.actatropica.2017.12.003>.

(<http://www.sciencedirect.com/science/article/pii/S0001706X17311749>)

[7] Hemorrhagic Fever-Causing Arenaviruses: Lethal Pathogens and Potent Immune Suppressors

Morgan E. Brisse^{1,2} and Hinh Ly^{2,*}

[8] Evolutionary origins of the SARS - CoV - 2 sarbecovirus lineage responsible for the COVID-19 pandemic Maciej F Boni^{1*} , Philippe Lemey^{2*} , Xiaowei Jiang³, Tommy Tsan-Yuk Lam⁴, Blair Perry⁵, Todd Castoe⁵, Andrew Rambaut⁶ and David L Robertson⁷

[9] Xiaolu Tang, Changcheng Wu, Xiang Li, Yuhe Song, Xinmin Yao, Xinkai Wu, Yuange Duan, Hong Zhang, Yirong Wang, Zhaohui Qian, Jie Cui, Jian Lu, On the origin and continuing evolution of SARS-CoV-2, *National Science Review*, , nwaa036, <https://doi.org/10.1093/nsr/nwaa036>

[10] SARS-CoV-2-like viruses from captive Guangdong pangolins generate circular RNAs

Alexandre Hassanin 1 Huw Jones 2 Anne Ropiquet 2