

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

**Daoyu Zhang.**

## **ABSTRACT**

Recently, there were much hype about reports of SARS-like coronaviruses being found in samples of Malayan pangolins (*Manis Javanica*) collected in Guangdong 2019, which appeared to possess nearly identical RBD's to the SARS-CoV-2 coronavirus. Prominent journals cited these discoveries to claim that pangolins may be 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 potentially contaminated. Here we also report that the presences of these 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 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.

## Genome alignment and analysis of non-NCBI datasets

If online BLAST analysis was unavailable to a dataset used, the dataset (as fastq) is downloaded and the sequences are aligned with the human genome using BowTie2[20] or MagicBLAST[21]. The Alignment results (if present) are then verified for their uniqueness to the human genome as per the Specific BLAST analysis method listed above.

## Sequence assembly

Sequence assembly of PRJNA610466 short read data was performed using MEGAHIT[16] on paired-end sequencing data with default parameters.

# 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 in the NCBI SRA. 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?
<a href="#">SRX6893158</a>	16,491,648	

<a href="#">SRX6893157</a>	9,275,501	Lung12 [3] SRR10168374
<a href="#">SRX6893156</a>	22,220,187	Lung11 [1]
<a href="#">SRX6893155</a>	18,067,615	Lung09 [1] [3] SRR10168376
<a href="#">SRX6893154</a>	16,414,925	Lung08 [1] [3] [4] SRR10168377
<a href="#">SRX6893153</a>	19,045,923	Lung07 [1] [3] [4] SRR10168378
<a href="#">SRX6893152</a>	13,527,964	
<a href="#">SRX6893151</a>	16,068,654	
<a href="#">SRX6893150</a>	12,967,281	
<a href="#">SRX6893149</a>	12,590,769	
<a href="#">SRX6893148</a>	15,273,939	
<a href="#">SRX6893147</a>	15,975,904	
<a href="#">SRX6893146</a>	19,038,817	
<a href="#">SRX6893145</a>	19,055,973	
<a href="#">SRX6893144</a>	15,350,468	
<a href="#">SRX6893143</a>	11,527,782	
<a href="#">SRX6893142</a>	20,045,443	
<a href="#">SRX6893141</a>	18,903,834	
<a href="#">SRX6893140</a>	19,986,780	
<a href="#">SRX6893139</a>	39,738,679	Lung02 [3] SRR10168392
<a href="#">SRX6893138</a>	22,900,426	
<a href="#">SRX7756769</a>	107,267,359 PRJNA607174**	M1[2]***
<a href="#">SRX7756766</a>	273,651,431 PRJNA607174**	
<a href="#">SRX7756765</a>	196,761,202 PRJNA607174**	
<a href="#">SRX7756764</a>	222,286,763 PRJNA607174**	
<a href="#">SRX7756763</a>	212,161,250 PRJNA607174**	
<a href="#">SRX7756762</a>	232,433,120 PRJNA607174**	M6[2]***
<a href="#">SRX7756761</a>	113,900,941 PRJNA607174**	
<a href="#">SRX7732094</a>	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	<b>Not available</b>
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	<b>Not available</b>
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
<a href="#">SRX6893158</a> 20-Sep-2019	Manis javanica: <b>14.66%</b>	N/D	N/D
<a href="#">SRX6893157</a> 20-Sep-2019	Boreoeutheria: <b>1.24%</b>	Catarrhini 644546	N/D***
<a href="#">SRX6893156</a> 20-Sep-2019	Manis javanica: <b>7.51%</b> Homo sapiens: <b>0.03%</b>	Homo sapiens 81948	Pangolin coronavirus 2Kbp
<a href="#">SRX6893155</a> 20-Sep-2019	Homo sapiens: <b>0.37%</b>	Homininae 3534150	Pangolin coronavirus 5Kbp
<a href="#">SRX6893154</a> 20-Sep-2019	Homo sapiens: <b>0.02%</b>	Hominoidea 356003	Pangolin coronavirus 154Kbp
<a href="#">SRX6893153</a> 20-Sep-2019	Homo sapiens: <b>0.01%</b>	Homo sapiens 162180	Pangolin coronavirus 41Kbp

<a href="#">SRX6893152</a> 20-Sep-2019	Manis javanica: <b>2.87%</b> Euarchontoglires: <b>1.37%</b>	N/D	N/D
<a href="#">SRX6893151</a> 20-Sep-2019	Manis javanica: <b>7.47%</b>	N/D	N/D
<a href="#">SRX6893150</a> 20-Sep-2019	Boreoeutheria: <b>1.91%</b>	N/D	N/D
<a href="#">SRX6893149</a> 20-Sep-2019	Manis javanica: <b>1%</b>	Simiiformes 313069	N/D
<a href="#">SRX6893148</a> 20-Sep-2019	Manis javanica: <b>0.4%</b>	Catarrhini 194320	N/D
<a href="#">SRX6893147</a> 20-Sep-2019	Manis javanica: <b>2.71%</b>	Catarrhini 69937	N/D
<a href="#">SRX6893146</a> 20-Sep-2019	Boreoeutheria: <b>1.72%</b>	Hominoidea 231755	N/D
<a href="#">SRX6893145</a> 20-Sep-2019	Homininae: <b>0.27%</b> Manis javanica: <b>1.01%</b>	Homininae 2536765	N/D
<a href="#">SRX6893144</a> 20-Sep-2019	Manis javanica: <b>0.62%</b>	Hominoidea 166628	N/D
<a href="#">SRX6893143</a> 20-Sep-2019	Manis javanica: <b>1.63%</b>	N/D	N/D
<a href="#">SRX6893142</a> 20-Sep-2019	Manis javanica: <b>1.28%</b>	Simiiformes 57084	N/D
<a href="#">SRX6893141</a> 20-Sep-2019	Boreoeutheria: <b>1.41%</b>	N/D	N/D
<a href="#">SRX6893140</a> 20-Sep-2019	Boreoeutheria: <b>1.56%</b>	N/D	N/D
<a href="#">SRX6893139</a> 20-Sep-2019	Homo sapiens: <b>0.01%</b>	Homo sapiens 491120	Pangolin coronavirus 2Kbp
<a href="#">SRX6893138</a> 20-Sep-2019	Boreoeutheria: <b>1.67%</b>	Homininae 2761176	N/D
<a href="#">SRX7756769</a> 18-Feb-2020	Homo sapiens: <b>0.03%</b>	Homo sapiens 5457929	Bat SARS-like coronavirus 2Kbp Wuhan seafood market pneumonia virus 2Kbp
<a href="#">SRX7756766</a> 18-Feb-2020	Manis javanica: <b>78.6%</b>	Cercopithecidae 3116	Betacoronavirus 2Kbp**
<a href="#">SRX7756765</a> 18-Feb-2020	Manis javanica: <b>87.17%</b>	Cercopithecinae 11339	N/D****
<a href="#">SRX7756764</a> 18-Feb-2020	Manis javanica: <b>48.39%</b>	Cercopithecidae 22600	N/D
<a href="#">SRX7756763</a> 18-Feb-2020	Manis javanica: <b>94.95%</b>	Cercopithecidae 5076	N/D

<a href="#">SRX7756762</a> 18-Feb-2020	Manis javanica: <b>95.37%</b>	Catarrhini* 2831	Nidovirales 0Kbp
<a href="#">SRX7756761</a> 18-Feb-2020	Manis javanica: <b>13.63%</b>	Chlorocebus sabaesus 498506	N/D
<a href="#">SRX7732094</a> 15-Feb-2020	N/A***	N/A	Pangolin coronavirus***

\*: Chlorocebus Sabaesus

\*\* : 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 coronavirus-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"/> <a href="#">SRX7756762</a>	279	1047	0%	4e-68	100.00%	<a href="#">SRA-SRR11119766.160125840.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756762</a>	279	1366	0%	4e-68	100.00%	<a href="#">SRA-SRR11119766.138036805.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756762</a>	279	967	0%	4e-68	100.00%	<a href="#">SRA-SRR11119766.101239747.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756762</a>	279	1624	0%	4e-68	100.00%	<a href="#">SRA-SRR11119766.46413326.2</a>

Chlorocebus sabaesus 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"/> <a href="#">Macaca mulatta isolate Rh22777_6890-1b major histocompatibility complex genomic sequence</a>	279	279	100%	2e-71	100.00%	<a href="#">KT332833.1</a>	
<input checked="" type="checkbox"/> <a href="#">Macaca mulatta isolate Rh22335_5775-3 major histocompatibility complex genomic sequence</a>	279	279	100%	2e-71	100.00%	<a href="#">KT332608.1</a>	
<input checked="" type="checkbox"/> <a href="#">Macaca mulatta isolate Rh22335_5725-2 major histocompatibility complex genomic sequence</a>	279	279	100%	2e-71	100.00%	<a href="#">KT332521.1</a>	
<input checked="" type="checkbox"/> <a href="#">Macaca mulatta isolate Rh22335_5702-1a major histocompatibility complex genomic sequence</a>	279	279	100%	2e-71	100.00%	<a href="#">KT332463.1</a>	

>gn|SRA|SRR11119766.160125840.2 160125840 (Biological)  
TAAATCCTTTGGCTAATATACCCAGTAAATGGGATGGCTGGGTCATATGGTACATCTAGTTCT  
AGATCCTTGAGGAATCGCCATACCTGTTTCCCAATAATGGTTGAAGTAGTTTACAATCCAC  
CAACAGTGTAAAGTGTTCCTATTTCTCCAC

<input checked="" type="checkbox"/> <a href="#">Uncultured organism clone VC1C968TR genomic sequence</a>	230	230	100%	2e-56	94.04%	<a href="#">GQ879596.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct Pan troglodytes LINE-1L_1PT retrotransposon tagged with EGFP sequence</a>	219	219	100%	4e-53	92.72%	<a href="#">KF861301.1</a>
<input checked="" type="checkbox"/> <a href="#">Human artificial chromosome vector 21HAC4 DNA, isolated from the long arm, clone: YAC/BAC#26-2</a>	219	219	100%	4e-53	92.72%	<a href="#">AB553834.1</a>
<input checked="" type="checkbox"/> <a href="#">Mus musculus NOD-derived CD11c +ve dendritic cells cDNA, RIKEN full-length enriched library, clone: F630221F08 product un</a>	204	204	86%	1e-48	94.66%	<a href="#">AK171052.1</a>
<input checked="" type="checkbox"/> <a href="#">Mus musculus bone marrow macrophage cDNA, RIKEN full-length enriched library, clone: G530008A19 product hypothetical g</a>	204	204	86%	1e-48	94.66%	<a href="#">AK149653.1</a>
<input checked="" type="checkbox"/> <a href="#">Ralstonia solanacearum genome assembly 9 genomes, chromosome - V</a>	202	202	100%	4e-48	90.73%	<a href="#">LN899823.1</a>
<input checked="" type="checkbox"/> <a href="#">Canis lupus familiaris breed Labrador retriever chromosome 06a</a>	154	3044	98%	1e-33	86.43%	<a href="#">CP050586.1</a>
<input checked="" type="checkbox"/> <a href="#">Canis lupus familiaris breed Labrador retriever chromosome 04a</a>	154	4569	100%	1e-33	85.23%	<a href="#">CP050572.1</a>
<input checked="" type="checkbox"/> <a href="#">Canis lupus familiaris breed Labrador retriever chromosome 06b</a>	154	3042	98%	1e-33	86.43%	<a href="#">CP050622.1</a>

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"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.269072261.2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.255768440.2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.255768440.1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.255318754.2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.254520929.1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	6344	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.251645135.1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.234036838.2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.211208832.2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	9108	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.199583624.1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.198110623.2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.196936636.2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.196936636.1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.133631622.2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	279	0%	5e-67	100.00%	<a href="#">SRA_SRR11119762.108819247.2</a>

**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.

<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta Major Histocompatibility Complex BAC MMU370O02_complete sequence</a>	279	279	100%	2e-71	100.00%	<a href="#">AC148706.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta Major Histocompatibility Complex BAC MMU122H23_complete sequence</a>	279	521	100%	2e-71	100.00%	<a href="#">AC148677.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta Major Histocompatibility Complex BAC MMU065H09_complete sequence</a>	279	279	100%	2e-71	100.00%	<a href="#">AC148671.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta Major Histocompatibility Complex BAC MMU038L02_complete sequence</a>	279	279	100%	2e-71	100.00%	<a href="#">AC148668.1</a>
<input checked="" type="checkbox"/>	<a href="#">Papio anubis clone rp41-22m16_complete sequence</a>	279	554	100%	2e-71	100.00%	<a href="#">AC113268.8</a>
<input checked="" type="checkbox"/>	<a href="#">Papio anubis clone rp41-280n2_complete sequence</a>	279	465	100%	2e-71	100.00%	<a href="#">AC091778.13</a>
<input checked="" type="checkbox"/>	<a href="#">Papio anubis clone rp41-5m22_complete sequence</a>	279	279	100%	2e-71	100.00%	<a href="#">AC136143.4</a>
<input checked="" type="checkbox"/>	<a href="#">Papio anubis clone rp41-192i11_complete sequence</a>	279	526	100%	2e-71	100.00%	<a href="#">AC091671.28</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta chromosome 9 CH250-18D2_complete sequence</a>	279	548	100%	2e-71	100.00%	<a href="#">CT573219.3</a>
<input checked="" type="checkbox"/>	<a href="#">Rhesus Macaque CHR4 BAC CH250-23P16 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Libr</a>	279	279	100%	2e-71	100.00%	<a href="#">AC169807.2</a>
<input checked="" type="checkbox"/>	<a href="#">Rhesus Macaque CHR4 BAC CH250-476F18 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Libr</a>	279	279	100%	2e-71	100.00%	<a href="#">AC171646.5</a>
<input checked="" type="checkbox"/>	<a href="#">Chlorocebus aethiops BAC clone CH252-163P9 from chromosome 5_complete sequence</a>	278	556	99%	7e-71	100.00%	<a href="#">AC239684.4</a>
<input checked="" type="checkbox"/>	<a href="#">MACACA MULATTA BAC clone CH250-192J17 from chromosome unknown_complete sequence</a>	278	552	100%	7e-71	100.00%	<a href="#">AC215693.3</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta isolate Rh22335_5702-1a major histocompatibility complex genomic sequence</a>	274	274	100%	9e-70	99.34%	<a href="#">KT332463.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta isolate Rh9_6570-3 major histocompatibility complex genomic sequence</a>	274	274	100%	9e-70	99.34%	<a href="#">KT331777.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta isolate Rh9_6550-1b major histocompatibility complex genomic sequence</a>	274	274	100%	9e-70	99.34%	<a href="#">KT331733.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta isolate Rh9_6526-2 major histocompatibility complex genomic sequence</a>	274	274	100%	9e-70	99.34%	<a href="#">KT331675.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta isolate Rh18665_5547-1b major histocompatibility complex genomic sequence</a>	274	274	100%	9e-70	99.34%	<a href="#">KT329509.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta isolate Rh23717 clone 4777 major histocompatibility complex-B genomic sequence</a>	274	274	100%	9e-70	99.34%	<a href="#">KJ913523.1</a>
<input checked="" type="checkbox"/>	<a href="#">Macaca mulatta isolate Rh23108 clone 4769-2 major histocompatibility complex-B genomic sequence</a>	274	274	100%	9e-70	99.34%	<a href="#">KJ913420.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 18</a>	202	4103	96%	4e-48	91.78%	<a href="#">CP034496.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 19</a>	196	2020	96%	2e-46	92.09%	<a href="#">CP034522.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 19</a>	196	2130	96%	2e-46	92.09%	<a href="#">CP034497.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 16</a>	196	3654	96%	2e-46	91.61%	<a href="#">CP034494.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 15</a>	196	4994	97%	2e-46	92.09%	<a href="#">CP034493.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 14</a>	196	7731	96%	2e-46	92.09%	<a href="#">CP034492.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 13</a>	191	7988	97%	9e-45	90.91%	<a href="#">CP034516.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome Y</a>	191	2781	96%	9e-45	91.37%	<a href="#">CP034510.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 20</a>	191	4517	96%	9e-45	91.37%	<a href="#">CP034499.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 13</a>	191	7988	97%	9e-45	90.91%	<a href="#">CP034491.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 21</a>	185	1053	96%	4e-43	90.65%	<a href="#">CP034500.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 17</a>	185	2450	96%	4e-43	90.65%	<a href="#">CP034495.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 22</a>	183	933	96%	2e-42	90.58%	<a href="#">CP034501.1</a>

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

Results for 
  
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](#)

Organism 
  
Type common name, binomial, taxid or group name
  
[+ Add organism](#)
  
Percent Identity  to 
  
E value  to 
  
Query Coverage  to 
  
[Filter](#) [Reset](#)

**Descriptions** | Graphic Summary | Alignments | Taxonomy

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

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%	<a href="#">XM_015150059.2</a>
<input type="checkbox"/> PREDICTED: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X2_mRNA	279	279	100%	2e-71	100.00%	<a href="#">XM_015150058.2</a>
<input type="checkbox"/> PREDICTED: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X1_mRNA	279	279	100%	2e-71	100.00%	<a href="#">XM_015150057.2</a>
<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%	<a href="#">AH012454.2</a>
<input type="checkbox"/> PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X3_mRNA	279	279	100%	2e-71	100.00%	<a href="#">XM_015457315.1</a>
<input type="checkbox"/> PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X2_mRNA	279	279	100%	2e-71	100.00%	<a href="#">XM_005567555.2</a>
<input type="checkbox"/> PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X1_mRNA	279	279	100%	2e-71	100.00%	<a href="#">XM_005567554.2</a>
<input type="checkbox"/> Chlorocebus aethiops BAC clone CH252-146112 from chromosome 6, complete sequence	279	3318	100%	2e-71	100.00%	<a href="#">AC241602.2</a>
<input type="checkbox"/> Chlorocebus aethiops BAC clone CH252-138D20 from chromosome 13, complete sequence	279	681	100%	2e-71	100.00%	<a href="#">AC239463.3</a>
<input type="checkbox"/> Macaca mulatta BAC CH250-74N24 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Library) complete sequence	279	6774	100%	2e-71	100.00%	<a href="#">AC204493.6</a>
<input type="checkbox"/> Chlorocebus aethiops BAC clone CH252-257N12 from chromosome 6, complete sequence	274	6168	100%	9e-70	99.34%	<a href="#">AC241575.3</a>
<input type="checkbox"/> Chlorocebus aethiops BAC clone CH252-518J9 from chromosome 6, complete sequence	274	4718	100%	9e-70	99.34%	<a href="#">AC241458.2</a>
<input type="checkbox"/> Chlorocebus aethiops BAC clone CH252-124F12 from chromosome 6, complete sequence	274	2252	100%	9e-70	99.34%	<a href="#">AC241509.3</a>
<input type="checkbox"/> Chlorocebus aethiops BAC clone CH252-175F15 from chromosome 6, complete sequence	274	4253	100%	9e-70	99.34%	<a href="#">AC241845.3</a>
<input type="checkbox"/> Chlorocebus aethiops BAC clone CH252-57Q21 from chromosome 6, complete sequence	274	2432	100%	9e-70	99.34%	<a href="#">AC241469.3</a>

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

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"/> <a href="#">SRX7756769</a>	278	278	0%	9e-69	100.00%	<a href="#">SRA_SRR11119759.99831231.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756769</a>	278	278	0%	9e-69	100.00%	<a href="#">SRA_SRR11119759.99831231.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756769</a>	278	4814	1%	9e-69	100.00%	<a href="#">SRA_SRR11119759.88019245.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756769</a>	278	5178	2%	9e-69	100.00%	<a href="#">SRA_SRR11119759.82130976.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756769</a>	278	278	0%	9e-69	100.00%	<a href="#">SRA_SRR11119759.70689253.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756769</a>	278	278	0%	9e-69	100.00%	<a href="#">SRA_SRR11119759.70689253.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756769</a>	278	278	0%	9e-69	100.00%	<a href="#">SRA_SRR11119759.57405658.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756769</a>	278	278	0%	9e-69	100.00%	<a href="#">SRA_SRR11119759.57405658.1</a>

[AC073210.8](#)  
Homo sapiens BAC clone RP11-460N20 from 7, complete sequence  
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 [GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input type="checkbox"/> Homo sapiens chromosome 22 clone ABC11_000047178300_E22, complete sequence	278	456	100%	6e-71	100.00%	<a href="#">AC279316.1</a>
<input type="checkbox"/> Homo sapiens actin related protein 2 pseudogene (LOC284441) on chromosome 19	278	278	100%	6e-71	100.00%	<a href="#">NG_022927.2</a>
<input type="checkbox"/> Homo sapiens TBC1 domain containing kinase (TBCK), RefSeqGene on chromosome 4	278	2140	100%	6e-71	100.00%	<a href="#">NG_034057.3</a>
<input type="checkbox"/> Homo sapiens chromosome 15 clone VMRC59-280I06, complete sequence	278	2291	100%	6e-71	100.00%	<a href="#">AC279072.1</a>
<input type="checkbox"/> Homo sapiens chromosome 2 clone VMRC59-389K09, complete sequence	278	3905	100%	6e-71	100.00%	<a href="#">AC279037.1</a>
<input type="checkbox"/> Homo sapiens chromosome 15 clone VMRC59-359A02, complete sequence	278	3589	100%	6e-71	100.00%	<a href="#">AC278991.1</a>
<input type="checkbox"/> Homo sapiens chromosome 16 clone VMRC59-453B14, complete sequence	278	2239	100%	6e-71	100.00%	<a href="#">AC278975.1</a>

Description **gnl|SRA|SRR11119759.88019245.2.88019245 (Biological)**  
Molecule type **dna**  
Query Length **150**  
Other reports [Distance tree of results](#) [MSA viewer](#)



Query ID: Icl|Query\_61480  
 Description: gnl|SRA|SRR11119759.70689253.1 70689253 (Biological)  
 Molecule type: dna  
 Query Length: 150

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 41 sequences selected GenBank Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus endogenous retrovirus group 3 member 1_envelope (ERV3-1)_mRNA	278	278	100%	7e-71	100.00%	<a href="#">XM_034964253.1</a>
<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%	<a href="#">NM_001007253.4</a>
<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%	<a href="#">XM_024357564.1</a>
<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%	<a href="#">XM_016956774.2</a>
<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%	<a href="#">AB590999.1</a>
<input checked="" type="checkbox"/> Homo sapiens cDNA FLJ60255 complete cds_highly similar to HERV-R_7q21.2 provirus ancestral Env polyprotein precursor	278	278	100%	7e-71	100.00%	<a href="#">AK295189.1</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-623C19 from chromosome 7_complete sequence	278	278	100%	7e-71	100.00%	<a href="#">AC184799.2</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-565C10 from chromosome 7_complete sequence	278	278	100%	7e-71	100.00%	<a href="#">AC148313.3</a>
<input checked="" type="checkbox"/> Pan troglodytes endogenous retrovirus group 3 member 1_envelope (ERV3-1)_mRNA	278	278	100%	7e-71	100.00%	<a href="#">NM_001135588.1</a>
<input checked="" type="checkbox"/> Homo sapiens BAC clone RP11-46N20 from 7_complete sequence	278	278	100%	7e-71	100.00%	<a href="#">AC073210.8</a>
<input checked="" type="checkbox"/> Homo sapiens cDNA FLJ23884 fis_clone LNG13819_highly similar to Human endogenous retrovirus ERV3_nol-env-3LTR region	278	278	100%	7e-71	100.00%	<a href="#">AK074464.1</a>
<input checked="" type="checkbox"/> Human endogenous retrovirus ERV3_nol-env-3LTR region	278	278	100%	7e-71	100.00%	<a href="#">M12140.1</a>
<input checked="" type="checkbox"/> Nomascus leucogenus endogenous retrovirus group 3 member 1_envelope (ERV3-1)_mRNA	272	272	100%	3e-69	99.33%	<a href="#">NM_001308194.1</a>
<input checked="" type="checkbox"/> Gibbon endogenous retrovirus proviral envR gene for R envelope protein	272	272	100%	3e-69	99.33%	<a href="#">AJ862653.1</a>
<input checked="" type="checkbox"/> PREDICTED: Ptilinopus leucosceles endogenous retrovirus group 3 member 1_envelope (ERV3-1)_transcript variant X1_mRNA	267	267	100%	2e-67	98.67%	<a href="#">XM_023196289.2</a>
<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%	<a href="#">XM_031662056.1</a>
<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%	<a href="#">XM_017956680.3</a>
<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%	<a href="#">XM_030932630.1</a>
<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%	<a href="#">XM_030932629.1</a>

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.

Description: gnl|SRA|SRR11119759.70689253.1 70689253 (Biological)  
 Molecule type: dna  
 Query Length: 150

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

Descriptions Graphic Summary Alignments Taxonomy

Sequences producing significant alignments Download Manage Columns Show 1000

select all 1 sequences selected GenBank Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> Synthetic construct DNA clone_pFN21AE1261_Homo sapiens ERV3 gene for endogenous retroviral sequence 3_without stop codon in	278	278	100%	7e-71	100.00%	<a href="#">AB590999.1</a>

Fig.4c The sequence have no matches 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"/> <a href="#">SRX6893156</a>	278	278	0%	2e-69	100.00%	<a href="#">SRA_SRR10168375.5045789.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX6893156</a>	278	278	0%	2e-69	100.00%	<a href="#">SRA_SRR10168375.5964.1</a>

Description: Homo sapiens BAC clone RP11-46N20 from 7\_complete seq  
 Molecule type: nucleic acid  
 Query Length: 203396

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

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

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"/> <a href="#">Homo sapiens BAC clone RP11-46N20 from 7_complete sequence</a>	278	278	100%	6e-71	100.00%	<a href="#">AC073210.8</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-923C19 from chromosome 7_complete sequence</a>	267	267	100%	1e-67	98.67%	<a href="#">AC184799.2</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-2015 from chromosome 7_complete sequence</a>	267	267	100%	1e-67	98.67%	<a href="#">AC174000.3</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-565C10 from chromosome 7_complete sequence</a>	267	267	100%	1e-67	98.67%	<a href="#">AC148313.3</a>

Description gnl|SRA|SRR10168375.5045789.1 5045789 (Biological)  
Molecule type dna  
Query Length 150  
Other reports [Distance tree of results](#) [MSA viewer](#)

Description gnl|SRA|SRR10168375.5964.1 5964 (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 1000 sequences selected

GenBank Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> <a href="#">Homo sapiens lncAB572.1 lncRNA gene complete sequence</a>	278	278	100%	7e-71	100.00%	<a href="#">MK280613.1</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes chromosome 2 clone CH251-60P06 complete sequence</a>	278	278	100%	7e-71	100.00%	<a href="#">AC279084.1</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes chromosome 2 clone CH251-17Q22 complete sequence</a>	278	278	100%	7e-71	100.00%	<a href="#">AC278930.1</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes chromosome 2 clone CH251-108A24 complete sequence</a>	278	278	100%	7e-71	100.00%	<a href="#">AC278921.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 21</a>	278	7938	100%	7e-71	100.00%	<a href="#">CP034500.1</a>

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

**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"/> <a href="#">Eukaryotic synthetic construct chromosome 21</a>	278	7938	100%	7e-71	100.00%	<a href="#">CP034500.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 13</a>	267	14570	100%	1e-67	98.67%	<a href="#">CP034516.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 16</a>	267	10333	100%	1e-67	98.67%	<a href="#">CP034494.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 15</a>	267	9021	100%	1e-67	98.67%	<a href="#">CP034493.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 13</a>	267	14570	100%	1e-67	98.67%	<a href="#">CP034491.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 18</a>	261	15047	100%	7e-66	98.00%	<a href="#">CP034496.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 17</a>	261	6545	100%	7e-66	98.00%	<a href="#">CP034495.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 20</a>	219	7949	98%	4e-53	93.79%	<a href="#">CP034499.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 19</a>	209	3521	96%	3e-50	93.10%	<a href="#">CP034522.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 19</a>	209	3766	96%	3e-50	93.10%	<a href="#">CP034497.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 22</a>	207	2291	96%	9e-50	92.47%	<a href="#">CP034501.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 14</a>	207	13851	96%	9e-50	92.47%	<a href="#">CP034492.1</a>
<input checked="" type="checkbox"/> <a href="#">Gossypium hirsutum clone NBRJ_GE27093 microsatellite sequence</a>	189	189	96%	3e-44	90.41%	<a href="#">JX591845.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome Y</a>	180	2399	96%	2e-41	89.44%	<a href="#">CP034510.1</a>

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%	<a href="#">SRA-SRR10168376.17339580.1</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.17013625.2</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.17013625.1</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.16930714.2</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.16930714.1</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.15267479.2</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.15267479.1</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.13985702.2</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.13985702.1</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.13353823.2</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.13353823.1</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.1109740.1</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.9343845.2</a>
<input checked="" type="checkbox"/> SRX6893155	278	278	0%	2e-69	100.00%	<a href="#">SRA-SRR10168376.9232549.2</a>

**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

[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 ABC13-48840700E15 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC242195.4</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-340I24 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC185242.2</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-623C19 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC184799.2</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-114G16 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC183835.2</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-2015 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC174000.3</a>
<input checked="" type="checkbox"/> Homo sapiens BAC clone RP11-479O9 from 7, complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC073107.7</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-565C10 from chromosome 7, complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC148313.3</a>
<input checked="" type="checkbox"/> Homo sapiens BAC clone RP11-460N20 from 7, complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC073210.8</a>
<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%	<a href="#">XM_017526193.1</a>

**Description** gnl|SRR10168376.15267479.2 15267479 (Biological)

**Molecule type** dna

**Query Length** 150

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

**Description** gnl|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

select all 87 sequences selected

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

Description	Max Score	Total Score	Query Cover	E value	Per Ident	Accession
<input checked="" type="checkbox"/> Homo sapiens BAC clone RP11-460N20 from 7, complete sequence	278	278	100%	7e-71	100.00%	<a href="#">AC073210.8</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-623C19 from chromosome 7, complete sequence	272	272	100%	3e-69	99.33%	<a href="#">AC184799.2</a>
<input checked="" type="checkbox"/> Pan troglodytes BAC clone CH251-565C10 from chromosome 7, complete sequence	272	272	100%	3e-69	99.33%	<a href="#">AC148313.3</a>
<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%	<a href="#">AC210125.6</a>
<input checked="" type="checkbox"/> Homo sapiens tripartite motif containing 24 (TRIM24), RefSeqGene on chromosome 7	182	182	99%	6e-42	88.59%	<a href="#">NG_023288.1</a>
<input checked="" type="checkbox"/> Homo sapiens chromosome 7 clone RP11-199L18, complete sequence	182	182	99%	6e-42	88.59%	<a href="#">AC013429.12</a>
<input checked="" type="checkbox"/> Homo sapiens chromosome 7 clone RP11-256C24, complete sequence	182	182	99%	6e-42	88.59%	<a href="#">AC008265.15</a>
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome Y	176	353	99%	3e-40	87.92%	<a href="#">CP034510.1</a>

select all 3 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 Y	176	353	99%	3e-40	87.92%	<a href="#">CP034510.1</a>
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 14	143	143	99%	3e-30	84.00%	<a href="#">CP034492.1</a>
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 16	111	111	72%	7e-21	85.45%	<a href="#">CP034494.1</a>

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 Graphics Distance tree of results

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_SRR10168378.1832954.1

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

Description gn|[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

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 gnl|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

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> <a href="#">Homo sapiens PMS1 homolog 2, mismatch repair system component pseudogene 8 (PMS2P8) on chromosome 7</a>	278	278	100%	7e-71	100.00%	<a href="#">NG_006447.3</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens PMS1 homolog 2, mismatch repair system component pseudogene 10 (PMS2P10) on chromosome 7</a>	278	278	100%	7e-71	100.00%	<a href="#">NG_023454.4</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens PMS1 homolog 2, mismatch repair system component pseudogene 6 (PMS2P6) on chromosome 7</a>	278	278	100%	7e-71	100.00%	<a href="#">NG_006449.3</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens BAC clone CH17-264B6 from chromosome 7, complete sequence</a>	278	1112	100%	7e-71	100.00%	<a href="#">AC211476.5</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens BAC clone CH17-220H16 from chromosome 7, complete sequence</a>	278	1112	100%	7e-71	100.00%	<a href="#">AC211491.5</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens FOSMID clone ABC10-4551500F15 from chromosome 7, complete sequence</a>	278	556	100%	7e-71	100.00%	<a href="#">AC244146.2</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens BAC clone RP11-45003 from 7, complete sequence</a>	278	556	100%	7e-71	100.00%	<a href="#">AC105418.5</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens BAC clone RP11-398K3 from 7, complete sequence</a>	278	556	100%	7e-71	100.00%	<a href="#">AC006995.5</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens BAC clone RP11-313P13 from 7, complete sequence</a>	278	556	100%	7e-71	100.00%	<a href="#">AC005488.2</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens chromosome 7 clone VMRC53-89F05, complete sequence</a>	272	272	100%	3e-69	99.33%	<a href="#">AC278394.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens chromosome 7 clone VMRC62-404M06, complete sequence</a>	272	272	100%	3e-69	99.33%	<a href="#">AC278331.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens chromosome 7 clone CH17-61K4, complete sequence</a>	272	272	100%	3e-69	99.33%	<a href="#">AC270699.1</a>

<input checked="" type="checkbox"/> <a href="#">Homo sapiens PMS8 mRNA (yeast mismatch repair gene PMS1 homologues), partial cds (C-terminal region)</a>	248	248	100%	6e-62	96.67%	<a href="#">D38503.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens PMS1 homolog 2, mismatch repair system component (PMS2), RefSeqGene (LRG_161) on chromosome 7</a>	244	244	100%	7e-61	96.00%	<a href="#">NG_008466.1</a>

<input checked="" type="checkbox"/> <a href="#">Canis lupus familiaris breed Labrador retriever chromosome 06a</a>	154	154	84%	1e-33	88.89%	<a href="#">CP050586.1</a>
<input checked="" type="checkbox"/> <a href="#">Canis lupus familiaris breed Labrador retriever chromosome 06b</a>	154	154	84%	1e-33	88.89%	<a href="#">CP050622.1</a>
<input checked="" type="checkbox"/> <a href="#">Pipistrellus pipistrellus genome assembly, chromosome 5</a>	139	139	86%	3e-29	86.15%	<a href="#">LR862361.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct Homo sapiens clone ccsbBroadEn_14772 PMS2 gene, encodes complete protein</a>	126	126	49%	3e-25	97.30%	<a href="#">KJ905275.1</a>
<input checked="" type="checkbox"/> <a href="#">PREDICTED: Zalophus californianus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X3, mRNA</a>	124	124	54%	9e-25	93.90%	<a href="#">XM_027610028.2</a>
<input checked="" type="checkbox"/> <a href="#">PREDICTED: Desmodus rotundus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X2, mRNA</a>	121	121	52%	1e-23	94.87%	<a href="#">XM_024576931.1</a>
<input checked="" type="checkbox"/> <a href="#">PREDICTED: Desmodus rotundus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X1, mRNA</a>	121	121	52%	1e-23	94.87%	<a href="#">XM_024576930.1</a>
<input checked="" type="checkbox"/> <a href="#">PREDICTED: Myotis lucifugus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X4, mRNA</a>	121	121	52%	1e-23	94.87%	<a href="#">XM_023761457.1</a>
<input checked="" type="checkbox"/> <a href="#">PREDICTED: Myotis lucifugus PMS1 homolog 2, mismatch repair system component (PMS2), transcript variant X3, mRNA</a>	121	121	52%	1e-23	94.87%	<a href="#">XM_023761456.1</a>

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

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> <a href="#">SRX6893139</a>	278	278	0%	3e-69	100.00%	<a href="#">SRA_SRR10168392.39544030.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX6893139</a>	278	278	0%	3e-69	100.00%	<a href="#">SRA_SRR10168392.28917809.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX6893139</a>	278	278	0%	3e-69	100.00%	<a href="#">SRA_SRR10168392.14357888.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX6893139</a>	278	278	0%	3e-69	100.00%	<a href="#">SRA_SRR10168392.2548655.2</a>

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

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

Fig.9b examining these reads revealed that they are only found in humans and apes. This is therefore also clear evidence that there is 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

**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"/> <a href="#">SRX6893157</a>	278	278	0%	8e-70	100.00%	<a href="#">SRA:SRR10168374.7906491.2</a>

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

<input checked="" type="checkbox"/> PREDICTED: Homo sapiens formin binding protein 1 (FNBP1) transcript variant X13 mRNA	278	278	100%	6e-71	100.00%	<a href="#">XM_005251824.2</a>
<input checked="" type="checkbox"/> PREDICTED: Homo sapiens formin binding protein 1 (FNBP1) transcript variant X4 mRNA	278	278	100%	6e-71	100.00%	<a href="#">XM_011518402.1</a>
<input checked="" type="checkbox"/> PREDICTED: Homo sapiens formin binding protein 1 (FNBP1) transcript variant X3 mRNA	278	278	100%	6e-71	100.00%	<a href="#">XM_011518401.1</a>
<input checked="" type="checkbox"/> Homo sapiens formin binding protein 1 (FNBP1) RefSeqGene on chromosome_9	278	278	100%	6e-71	100.00%	<a href="#">NG_033946.1</a>
<input checked="" type="checkbox"/> Homo sapiens cDNA FLJ13619 fis clone PLACE1010926 weakly similar to HYPOTHETICAL 72.2 KD PROTEIN C12C2.05C IN CHROMOS	278	278	100%	6e-71	100.00%	<a href="#">AK023681.1</a>
<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%	<a href="#">AL136141.13</a>
<input checked="" type="checkbox"/> Homo sapiens formin-binding protein 17 (FBP17) mRNA, partial cds	278	278	100%	6e-71	100.00%	<a href="#">AF265550.1</a>
<input checked="" type="checkbox"/> Homo sapiens chromosome 9 clone hRPK.202_H_3 complete sequence	278	278	100%	6e-71	100.00%	<a href="#">AC006241.1</a>
<input checked="" type="checkbox"/> Homo sapiens KIAA0554 mRNA for KIAA0554 protein	278	278	100%	6e-71	100.00%	<a href="#">AB011126.1</a>
<input checked="" type="checkbox"/> PREDICTED: Nomascus leucogenys formin binding protein 1 (FNBP1) transcript variant X18 mRNA	272	272	100%	3e-69	99.33%	<a href="#">XM_030818029.1</a>
<input checked="" type="checkbox"/> PREDICTED: Nomascus leucogenys formin binding protein 1 (FNBP1) transcript variant X17 mRNA	272	272	100%	3e-69	99.33%	<a href="#">XM_030818028.1</a>
<input checked="" type="checkbox"/> PREDICTED: Nomascus leucogenys formin binding protein 1 (FNBP1) transcript variant X16 mRNA	272	272	100%	3e-69	99.33%	<a href="#">XM_030818027.1</a>

Description [gnl\[SRA|SRR10168374.7906491.2.7906491\]](#) (Biological)  
Molecule type dna  
Query Length 150  
Other reports [Distance tree of results](#) [MSA viewer](#) ?

<input checked="" type="checkbox"/> <a href="#">Sciurus carolinensis genome assembly, chromosome_16</a>	174	174	95%	9e-40	88.81%	<a href="#">LR738606.1</a>
<input checked="" type="checkbox"/> <a href="#">Sciurus carolinensis genome assembly, chromosome_14</a>	174	174	95%	9e-40	88.81%	<a href="#">LR738604.1</a>
<input checked="" type="checkbox"/> <a href="#">Sciurus vulgaris genome assembly, chromosome_15</a>	169	169	95%	4e-38	88.11%	<a href="#">LR738626.1</a>
<input checked="" type="checkbox"/> PREDICTED: Loxodonta africana formin binding protein 1 (FNBP1) transcript variant X2 mRNA	135	135	74%	4e-28	88.50%	<a href="#">XM_010587565.2</a>
<input checked="" type="checkbox"/> PREDICTED: Loxodonta africana formin binding protein 1 (FNBP1) transcript variant X1 mRNA	135	135	74%	4e-28	88.50%	<a href="#">XM_023544839.1</a>

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

Description Chlorocebus aethiops BAC clone CH252-276C1 from chromo: ...  
Molecule type nucleic acid  
Query Length 160189  
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 Graphics Distance tree of results

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

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

Description gnl|SRA|SRR11119763.106754018.1 106754018 (Biological)  
Molecule type dna  
Query Length 151  
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 1000 sequences selected GenBank Graphics Distance tree of results

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> <a href="#">Chlorocebus aethiops BAC clone CH252-276C1 from chromosome 6, complete sequence</a>	279	375	100%	2e-71	100.00%	<a href="#">AC241496.3</a>
<input checked="" type="checkbox"/> <a href="#">Chlorocebus aethiops BAC clone CH252-445L10 from chromosome unknown, complete sequence</a>	279	690	100%	2e-71	100.00%	<a href="#">AC238953.2</a>
<input checked="" type="checkbox"/> <a href="#">Macaca mulatta Y Chr CH250-11J13 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC)</a>	279	1513	100%	2e-71	100.00%	<a href="#">AC240710.4</a>
<input checked="" type="checkbox"/> <a href="#">Chlorocebus aethiops BAC clone CH252-371E9 from chromosome 5, complete sequence</a>	279	1124	100%	2e-71	100.00%	<a href="#">AC239563.3</a>
<input checked="" type="checkbox"/> <a href="#">Chlorocebus aethiops BAC clone CH252-461K13 from chromosome 13, complete sequence</a>	279	629	100%	2e-71	100.00%	<a href="#">AC239354.3</a>
<input checked="" type="checkbox"/> <a href="#">Chlorocebus aethiops BAC clone CH252-138D20 from chromosome 13, complete sequence</a>	279	493	100%	2e-71	100.00%	<a href="#">AC239463.3</a>
<input checked="" type="checkbox"/> <a href="#">Chlorocebus aethiops BAC clone CH252-417L1 from chromosome 6, complete sequence</a>	279	472	100%	2e-71	100.00%	<a href="#">AC239275.3</a>
<input checked="" type="checkbox"/> <a href="#">Chlorocebus aethiops BAC clone CH252-62B18 from chromosome 4, complete sequence</a>	279	510	100%	2e-71	100.00%	<a href="#">AC239442.2</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 18</a>	268	1.545e+05	100%	4e-68	100.00%	<a href="#">CP034496.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 16</a>	268	1.008e+05	100%	4e-68	98.68%	<a href="#">CP034494.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 19</a>	263	47303	100%	2e-66	98.01%	<a href="#">CP034522.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 13</a>	263	1.935e+05	100%	2e-66	99.31%	<a href="#">CP034516.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome Y</a>	263	1.175e+05	100%	2e-66	99.31%	<a href="#">CP034510.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 21</a>	263	55262	100%	2e-66	98.01%	<a href="#">CP034500.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 20</a>	263	88148	100%	2e-66	99.31%	<a href="#">CP034499.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 19</a>	263	47481	100%	2e-66	98.01%	<a href="#">CP034497.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 17</a>	263	74732	100%	2e-66	98.01%	<a href="#">CP034495.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 15</a>	263	1.384e+05	100%	2e-66	98.01%	<a href="#">CP034493.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 14</a>	263	1.910e+05	100%	2e-66	99.31%	<a href="#">CP034492.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 13</a>	263	1.935e+05	100%	2e-66	99.31%	<a href="#">CP034491.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 22</a>	257	22814	100%	9e-65	97.35%	<a href="#">CP034501.1</a>
<input checked="" type="checkbox"/> <a href="#">Human gammaherpesvirus 4 isolate HKNPC60, partial genome</a>	248	248	100%	5e-62	96.03%	<a href="#">MH590571.1</a>
<input checked="" type="checkbox"/> <a href="#">Human gammaherpesvirus 4 isolate HKHD40, partial genome</a>	248	248	100%	5e-62	96.03%	<a href="#">MH590409.1</a>
<input checked="" type="checkbox"/> <a href="#">Uncultured bacterium clone LMOABA39ZH04FM1 genomic sequence</a>	246	246	96%	2e-61	97.24%	<a href="#">EU058958.1</a>
<input checked="" type="checkbox"/> <a href="#">Synthetic construct Pan troglodytes LINE-1L 1Pt retrotransposon tagged with EGFP sequence</a>	241	241	100%	9e-60	95.36%	<a href="#">KF661301.1</a>
<input checked="" type="checkbox"/> <a href="#">Uncultured organism clone VC1C968TR genomic sequence</a>	241	241	96%	9e-60	96.55%	<a href="#">GQ879596.1</a>

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.



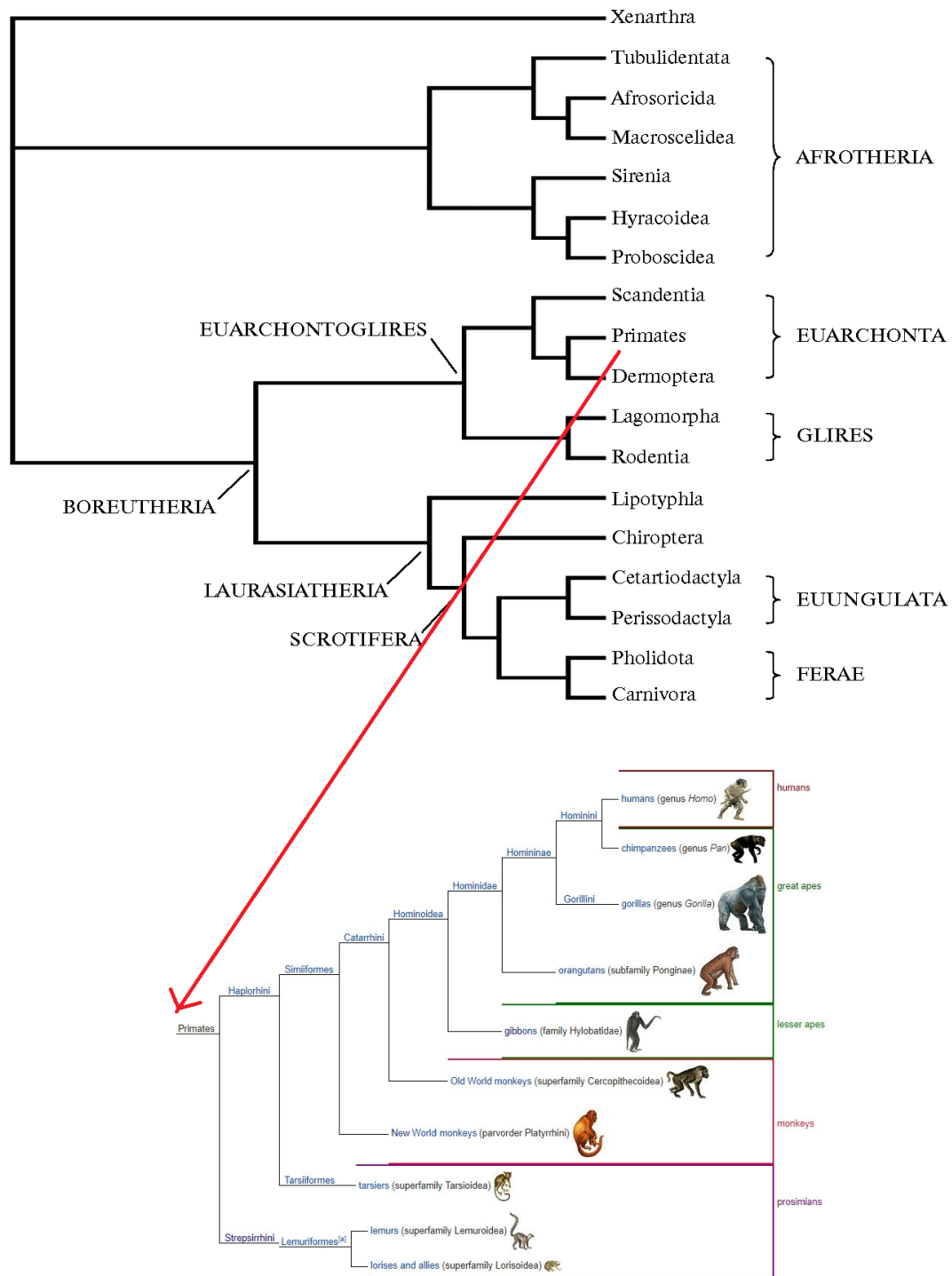


Fig. 12 Family tree of mammals, Including the position and classification of Primates in the lineage of Mammalia.

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
<a href="#">SRX7756769</a> 18-Feb-2020	Homo sapiens 5457929	15401134	0.35	Bat SARS-like coronavirus 2Kbp Wuhan seafood market pneumonia virus 2Kbp
<a href="#">SRX6893139</a> 20-Sep-2019	Homo sapiens 491120	5301351	0.0926	Pangolin coronavirus 2Kbp
<a href="#">SRX6893157</a> 20-Sep-2019	Catarrhini 644546	1889448	0.34	N/D***
<a href="#">SRX6893156</a> 20-Sep-2019	Homo sapiens 81948	4765461	0.01719	Pangolin coronavirus 2Kbp
<a href="#">SRX6893155</a> 20-Sep-2019	Homininae 3534150	525801	6.7214	Pangolin coronavirus 5Kbp
<a href="#">SRX6893154</a> 20-Sep-2019	Hominoidea 356003	2232008	0.159	Pangolin coronavirus 154Kbp
<a href="#">SRX6893153</a> 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
<a href="#">SRX7756766</a> 18-Feb-2020	Cercopithecidae 3116; BLAST to Macaca Mulatta	Betacoronavirus 2Kbp **	0.000642
<a href="#">SRX7756762</a> 18-Feb-2020	Catarrhini 2831; BLAST to Chlorocebus sabaeus	Nidovirales 0Kbp Claimed 10x150bp reads	0.000530
<a href="#">SRX7756765</a> 22-Apr-2020	Cercopithecinae 11339 BLAST to Chlorocebus Aethiops	N/D***	N/A
<a href="#">SRX7732094</a> 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 Batacoronavirus 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	Pahexavir
<a href="#">SRX6893158</a>	Yes	No	No	No	No	No	Yes	No	Yes	Yes	No	No
<a href="#">SRX6893157</a>	Yes	Yes	No	No	Claimed	No	No	Yes	No	No	No	No
<a href="#">SRX6893156</a>	No	No	Yes	Yes	Yes	No	No	No	Yes	No	No	Yes
<a href="#">SRX6893155</a>	No	No	Yes	No	Yes	No	No	No	No	No	No	No
<a href="#">SRX6893154</a>	No	No	Yes	No	Yes	No	No	No	No	No	No	No
<a href="#">SRX6893153</a>	No	No	Yes	Yes	Yes	No	No	No	Yes	No	No	No
<a href="#">SRX6893152</a>	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No	No
<a href="#">SRX6893151</a>	Yes	Yes	No	Yes	No	No	No	Yes	Yes	No	No	No
<a href="#">SRX6893150</a>	Yes	Yes	Yes	No	No	No	No	Yes	Yes	No	No	No
<a href="#">SRX6893149</a>	Yes	Yes	No	Yes	No	No	No	No	No	No	Yes	No
<a href="#">SRX6893148</a>	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No
<a href="#">SRX6893147</a>	Yes	Yes	"Respirovirus"	Yes	No	No	Yes	No	Yes	No	No	No
<a href="#">SRX6893146</a>	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No
<a href="#">SRX6893145</a>	Yes	Yes	No	No	No	No	No	No	No	No	No	No
<a href="#">SRX6893144</a>	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No
<a href="#">SRX6893143</a>	Yes	Yes	No	No	No	No	No	No	No	No	No	No
<a href="#">SRX6893142</a>	Yes	Yes	No	No	No	No	No	Yes	Yes	No	No	No
<a href="#">SRX6893141</a>	Yes	Yes	No	Yes	No	No	No	No	No	No	No	No
<a href="#">SRX6893140</a>	Yes	Yes	Yes	No	No	No	Yes	Yes	No	No	No	No
<a href="#">SRX6893139</a>	No	No	Yes	No	Yes	No	No	No	No	No	No	No
<a href="#">SRX6893138</a>	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No	No	No
<a href="#">SRX7756766</a>	No	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No
<a href="#">SRX7756765</a>	No	No	Yes	No	No	Yes	No	No	No	No	No	No
<a href="#">SRX7756764</a>	No	No	Yes	No	No	Yes	No	No	No	No	No	No
<a href="#">SRX7756763</a>	No	No	Yes	No	No	Yes	No	No	No	No	No	No
<a href="#">SRX7756762</a>	No	No	Yes	No	Claimed	Yes	No	No	No	No	No	No
<a href="#">SRX7756761</a>	No	No	Yes	No	No	Yes	No	No	No	No	No	No
<a href="#">SRX7756769</a>	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 significant amount of Homo Sapiens genetic material within them, which agrees with the hypothesis that the Coronavirus-related reads within these 3 datasets were the result of contamination by the In-lab VERO cell culture used by Xiao et al for virus isolation.

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 RPS-1193P9 from 7p11.2-q11.2, con...**  
 Molecule type **nucleic acid**  
 Query Length **134187**  
 Other reports [Distance tree of results](#) [MSA viewer](#) [?](#)

**Filter Results**

Percent Identity  to  E value  to  Query Coverage  to

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

```
>gn|SRA|SRR11119762.27294164.1 27294164 (Biological)
CTTCCCTGTCGTCATGTGATCTCTCAATGTTCCACCCTATGAGTGAAGAAATATGCGGTG
TTTGGTTTTTGTCTCTCCGATAGTTTACTGAGAATGATGATTTCCAATTTCCATCCATGT
CCCTACAAAGGACATGAACTCATCAATTTTTT

>gn|SRA|SRR11119762.27294164.2 27294164 (Biological)
AACCAAGCCAAATGTCACCAATGATAGACTGGATTAGAAATGTGGACATATACACC
ATGGAAATACATGACGCCATAAAAAATGATGAGTTTCATGCTCTTGTAGGACATGGATG
AAATGGAAATCATCTTTCAGTAAACTAT
```

**Descriptions** | Graphic Summary | Alignments

**Sequences producing significant alignments** Download ▾ [New](#) Manage columns ▾ Show  [?](#)

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"/>	<a href="#">SRX7756766</a>	279	964	0%	4e-69	100.00%	151	<a href="#">SRA_SRR11119762_119033805_2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	552	0%	4e-69	100.00%	151	<a href="#">SRA_SRR11119762_33937308_1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	992	0%	4e-69	100.00%	151	<a href="#">SRA_SRR11119762_27294164_2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	279	666	0%	4e-69	100.00%	151	<a href="#">SRA_SRR11119762_27294164_1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	274	548	0%	2e-67	99.34%	151	<a href="#">SRA_SRR11119762_33937308_2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	252	705	0%	9e-61	96.69%	151	<a href="#">SRA_SRR11119762_87190953_2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	252	964	0%	9e-61	96.69%	151	<a href="#">SRA_SRR11119762_65743002_2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	252	705	0%	9e-61	96.69%	151	<a href="#">SRA_SRR11119762_65743002_1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	246	675	0%	4e-59	96.03%	151	<a href="#">SRA_SRR11119762_267695997_2</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	246	686	0%	4e-59	96.03%	151	<a href="#">SRA_SRR11119762_87190953_1</a>
<input checked="" type="checkbox"/>	<a href="#">SRX7756766</a>	241	905	0%	2e-57	95.36%	151	<a href="#">SRA_SRR11119762_267695997_1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 13</a>							<a href="#">eukaryotic_synt... 279 1.710e+05 100% 2e-71 100.00% 96089878 CP034491.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens glypican 5 (GPC5) RefSeqGene on chromosome 13</a>							<a href="#">Homo sapiens 279 4892 100% 2e-71 100.00% 1475557 NC_009370.2</a>

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

Job Title **2 sequences (gn|SRA|SRR11119762.27294164.1...**  
 RID [ZBABPT58016](#) Search expires on 01-07 19:13 pm [Download All](#) ▾  
 Results for  ▾  
 Program **BLASTN** [Citation](#) ▾  
 Database **nt** [See details](#) ▾  
 Query ID [|Query\\_20117](#)  
 Description **gn|SRA|SRR11119762.27294164.2 27294164 (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  [?](#)

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"/>	<a href="#">Eukaryotic synthetic construct chromosome 13</a>	<a href="#">eukaryotic_synt...</a>	279	2.435e+05	100%	2e-71	100.00%	96089878	<a href="#">CP034516.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 22</a>	<a href="#">eukaryotic_synt...</a>	279	29397	100%	2e-71	100.00%	35194566	<a href="#">CP034501.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 18</a>	<a href="#">eukaryotic_synt...</a>	279	1.956e+05	100%	2e-71	100.00%	84890006	<a href="#">CP034496.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 15</a>	<a href="#">eukaryotic_synt...</a>	279	1.731e+05	100%	2e-71	100.00%	82521392	<a href="#">CP034493.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 13</a>	<a href="#">eukaryotic_synt...</a>	279	2.435e+05	100%	2e-71	100.00%	96089878	<a href="#">CP034491.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pongo abelii BAC clone CH276-289G9 from chromosome unknown_complete sequence</a>	<a href="#">Pongo abelii</a>	279	279	100%	2e-71	100.00%	221923	<a href="#">AC215315.3</a>
<input checked="" type="checkbox"/>	<a href="#">Pongo abelii BAC clone CH276-62H12 from chromosome unknown_complete sequence</a>	<a href="#">Pongo abelii</a>	279	559	100%	2e-71	100.00%	69319	<a href="#">AC190147.2</a>
<input checked="" type="checkbox"/>	<a href="#">Pongo abelii chromosome 7 clone CH276-122D18_complete sequence</a>	<a href="#">Pongo abelii</a>	274	537	100%	1e-69	99.34%	218515	<a href="#">AC278951.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cloning vector pTLD39_complete sequence</a>	<a href="#">Cloning vector p...</a>	274	274	100%	1e-69	99.34%	76343	<a href="#">MG711910.1</a>
<input checked="" type="checkbox"/>	<a href="#">Cloning vector pTLD67_complete sequence</a>	<a href="#">Cloning vector p...</a>	274	274	100%	1e-69	99.34%	72952	<a href="#">MG711909.1</a>
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome 19</a>	<a href="#">eukaryotic_synt...</a>	274	66262	100%	1e-69	99.34%	64242768	<a href="#">CP034522.1</a>

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 Instit...**

Molecule type **nucleic acid**

Query Length **157426**

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

**Filter Results**

Percent Identity  to  E value  to  Query Coverage  to

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

```
>gnl|SRA|SRR11119763.171368621.1 171368621 (Biological)
CTTGTGATTTTTGTACATTGATTTTGTATCCTGAAACTTTGCTGAAGTTGCTTATCAGCT
TAAGGAGATTTGGGCTGAGACAAATGGGGTTTCTAGATATACAATCATGTCGCTGCAA
ACAGGGACAATTTGACTTCCTCTTTTCCTAG

>gnl|SRA|SRR11119763.171368621.2 171368621 (Biological)
AGGAAATAAAGGGTATCAACTAGGAAAAGAGGAAGTCAAATGTCCTGTTTCGAGACG
ACATGATTGTATATCTAGAAAACCCCAATGTCAGCCCAAAATCTCCTTAAGCTGATAA
GCAACTTCAGCAAAGTTTCAGGATACAAAAT
```

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

**Sequences producing significant alignments** [Download](#) [New](#) [Manage columns](#) Show

select all 100 sequences selected [Graphics](#) [Distance tree of results](#) [New](#) [MSA Viewer](#)

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

Job Title **2 sequences (gnl|SRA|SRR11119763.171368621.1...**  
 RID [ZDVKJ9V7013](#) Search expires on 01-08 18:20 pm [Download All](#)

Results for

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](#)

**Filter Results**

Organism *only top 20 will appear*  exclude

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

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

**Descriptions** | [Graphic Summary](#) | [Alignments](#) | [Taxonomy](#)

**Sequences producing significant alignments** [Download](#) [New](#) [Select columns](#) Show

select all 1000 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#) [New](#) [MSA Viewer](#)

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

Fig 15: Reads uniquely matched to Homo Sapiens have been found within SRX7756765.

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 New Manage columns Show 100

select all 2 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"/> <a href="#">SRX7756762</a>	279	8344	44%	4e-70	100.00%	151	<a href="#">SRA_SRR11119766.44541545.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX7756762</a>	279	8680	44%	4e-70	100.00%	151	<a href="#">SRA_SRR11119766.44541545.1</a>

```
>gnl|SRA|SRR11119766.44541545.1 44541545 (Biological)
AGAAAAGGAARTATCTTCGTAATAAAACCCGACAGATCATCTCAGAACTGCTCTGTG
ATGTGTGCGTCAACTCAGAGATTTAACTTTTCTTTTCATTACAGATTTGGAAACACT
CTGTTGTAAAGTCTGCAAGTGGATATCTTG
>gnl|SRA|SRR11119766.44541545.2 44541545 (Biological)
GAAGGCTCTAAGAGCCAAAGATATCCACTTGCAGACTTACAAACAGAGTGTTCACAAA
CTGCTGATGAAAGAAAGATTTAACTCTGTGTGATTTGAACCGCACATCAGAGCACTT
TCTGGAATGATTTCTGCGGTTTATACG
```

Job Title 2 sequences (gnl|SRA|SRR11119766.44541545.1...  
 RID [ZNE2JD1H013](#) Search expires on 01-11 15:18 pm [Download All](#)  
 Results for 1:icl|Query\_63282 gnl|SRA|SRR11119766.44541545.1 44541545 (Biologica  
 Program BLASTN Citation  
 Database nt [See details](#)  
 Query ID icl|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

select all 464 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"/> <a href="#">Homo sapiens DNA, sequence_id: unplaced_0526</a>	<a href="#">Homo sapiens</a>	279	8680	100%	2e-71	100.00%	13084	<a href="#">AP024011.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens DNA, sequence_id: unplaced_0092</a>	<a href="#">Homo sapiens</a>	279	16564	100%	2e-71	100.00%	92273	<a href="#">AP023577.1</a>
<input checked="" type="checkbox"/> <a href="#">Human clone ps16 DNA</a>	<a href="#">Homo sapiens</a>	279	279	100%	2e-71	100.00%	336	<a href="#">X60718.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens DNA, sequence_id: unplaced_0210</a>	<a href="#">Homo sapiens</a>	270	19297	100%	1e-68	98.69%	51519	<a href="#">AP023695.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens DNA, sequence_id: unplaced_0016</a>	<a href="#">Homo sapiens</a>	268	4425	100%	5e-68	98.68%	380611	<a href="#">AP023501.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens DNA, sequence_id: unplaced_0231</a>	<a href="#">Homo sapiens</a>	263	29291	100%	2e-66	98.01%	46435	<a href="#">AP023716.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens DNA, sequence_id: unplaced_0494</a>	<a href="#">Homo sapiens</a>	252	10155	100%	5e-63	96.69%	15862	<a href="#">AP023979.1</a>

Your search is limited to records that exclude: Homo (taxid:9605)

Job Title 2 sequences (gnl|SRA|SRR11119766.44541545.1...  
 RID [ZNGGJ9BZ013](#) Search expires on 01-11 16:00 pm [Download All](#)  
 Results for 1:icl|Query\_49712 gnl|SRA|SRR11119766.44541545.1 44541545 (Biologica  
 Program BLASTN Citation  
 Database nt [See details](#)  
 Query ID icl|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

select all 116 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"/> <a href="#">Eukaryotic synthetic construct chromosome 19</a>	<a href="#">eukaryotic synth...</a>	246	4866	100%	2e-61	96.03%	64242768	<a href="#">CP034522.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 19</a>	<a href="#">eukaryotic synth...</a>	246	4941	100%	2e-61	96.03%	64242768	<a href="#">CP034497.1</a>
<input checked="" type="checkbox"/> <a href="#">Ralstonia solanacearum genome assembly 9 genomes_chromosome...</a>	<a href="#">Ralstonia solana...</a>	222	439	97%	4e-54	93.88%	5416545	<a href="#">LN899819.1</a>
<input checked="" type="checkbox"/> <a href="#">Chimpanzee aliphoid repetitive sequence, monomers 1 and 2, clone C-alpha-RI(1020), 15-208 II</a>	<a href="#">Pan.troglodytes</a>	217	217	100%	2e-52	92.72%	340	<a href="#">L01727.1</a>
<input checked="" type="checkbox"/> <a href="#">Chimpanzee (clone C-alpha-RI(1020), 15-164 II) aliphoid repeat sequence</a>	<a href="#">Pan.troglodytes</a>	217	217	100%	2e-52	92.72%	340	<a href="#">M97599.1</a>

Fig.16: Reads uniquely matched to Homo Sapiens Alpha Satellite DNA in SRX7756762.

## **The Pan-SL-CoV-GD sequences can not be found in other pangolin sequencing datasets from China.**

Recently, we are able to access and perform BLAST analysis on a large dataset of 93 pangolin samples deposited by [Hu J \*et al.\*](#), [12] located under the BioProject PRJNA529540. These samples, alongside with an older sample, SRX1319167, represent a longitudinal survey spanning from between 1990 to 2017 of Both Manis Pentadactyla and Manis Javanica from China. We could not obtain any traces of reads resembling the Pan-SL-CoV-GD sequences from these datasets. Such discovery is in agreement with the conclusion of [Hu J \*et al.\*](#), [12] which failed to find any evidence of SARS-CoV-2-like Coronaviruses within their sequencing study.

Considering that another longitudinal survey of 334 pangolins in Malaysia [13] have also failed to reveal any evidence of Coronaviruses or other potentially zoonotic viruses, the failure to isolate sequences of Coronaviruses from pangolin sequencing datasets are in good agreement that no natural infection of a Coronavirus can happen to a pangolin in the wild. This may be due to their solitary behavior [14] which keep them completely physically isolated from each other for up to 9 months for each year when the population is not in its mating season which happens from May to July. As this is longer than the time of which a pangolin could stay infected before either clearing the infection or dying in any known incidence of viral infection in captive pangolins [15], any virus species that enters a pangolin population in the wild will either be cleared or kills all of its current hosts when the population is not in or have left its mating season, resulting in the viral population to go extinct. Therefore, the absence of viral infections in pangolin populations is the normal state of such population, and any datasets that claimed viral infection of pangolins must be subjected to the highest level of scrutiny to exclude any potential presence of contamination.

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

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **009AEVJN01R** Search expires on 01-20 16:24 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796795	SRR0018583
SRX5796794	SRR0018584

Other reports [MSA viewer](#)

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

**Filter Results**  
 Percent Identity:  to  Evaluate:  to  Query Coverage:  to   
[Filter](#) [Reset](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00NSTESC01R** Search expires on 01-20 19:46 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796768	SRR0018610
SRX5796767	SRR0018611
SRX5796766	SRR0018612
SRX5796765	SRR0018613
SRX5796764	SRR0018614
SRX5796763	SRR0018615
SRX5796762	SRR0018616
SRX5796761	SRR0018617

Other reports [MSA viewer](#)

Job Title **gb|MT121216.1**  
 RID **00DAS3P501R** Search expires on 01-20 16:39 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796788	SRR0018590
SRX5796787	SRR0018591

Other reports [MSA viewer](#)

Job Title **gb|MT121216.1**  
 RID **00NSWUJX01R** Search expires on 01-20 19:46 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796760	SRR0018616
SRX5796759	SRR0018619
SRX5796758	SRR0018620
SRX5796757	SRR0018621
SRX5796756	SRR0018622

Other reports [MSA viewer](#)

Job Title **gb|MT121216.1**  
 RID **00BHGJMM01R** Search expires on 01-20 17:02 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796784	SRR0018594
SRX5796783	SRR0018595

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00C8248401R** Search expires on 01-20 17:16 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796782	SRR0018596
SRX5796781	SRR0018597
SRX5796780	SRR0018598
SRX5796779	SRR0018599

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00C33AEN01R** Search expires on 01-20 17:16 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796778	SRR0018600
SRX5796777	SRR0018601

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00G97I7101R** Search expires on 01-20 18:23 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796774	SRR0018604
SRX5796773	SRR0018605

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00D92PFR01R** Search expires on 01-20 17:58 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796776	SRR0018602
SRX5796775	SRR0018603

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00NAZGSD01R** Search expires on 01-20 19:07 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796770	SRR0018608
SRX5796769	SRR0018609

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00JVBW5501R** Search expires on 01-20 19:07 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796772	SRR0018606
SRX5796771	SRR0018607

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00RVXF4G01R** Search expires on 01-20 20:32 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796721	SRR0018657
SRX5796720	SRR0018658
SRX5796719	SRR0018659
SRX5796718	SRR0018660
SRX5796716	SRR0018662
SRX5796717	SRR0018661

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00RWSMK01R** Search expires on 01-20 20:32 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796715	SRR0018663
SRX5796714	SRR0018664
SRX5796713	SRR0018665
SRX5796712	SRR0018666
SRX5796711	SRR0018667
SRX5796710	SRR0018668

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00R66GAE01R** Search expires on 01-20 20:21 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796735	SRR0018643
SRX5796734	SRR0018644
SRX5796733	SRR0018645
SRX5796732	SRR0018646
SRX5796731	SRR0018647
SRX5796730	SRR0018648
SRX5796729	SRR0018649
SRX5796728	SRR0018650

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00SH7TX401R** Search expires on 01-20 20:44 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796709	SRR0018669
SRX5796708	SRR0018670
SRX5796707	SRR0018671
SRX5796706	SRR0018672
SRX5796705	SRR0018673
SRX5796704	SRR0018674
SRX5796703	SRR0018675

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00PFPF501R** Search expires on 01-20 20:05 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796741	SRR0018637
SRX5796740	SRR0018638
SRX5796739	SRR0018639
SRX5796738	SRR0018640
SRX5796737	SRR0018641
SRX5796736	SRR0018642

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00NV33XG01R** Search expires on 01-20 19:58 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796725	SRR0018623
SRX5796724	SRR0018624
SRX5796723	SRR0018625
SRX5796722	SRR0018626
SRX5796721	SRR0018627
SRX5796720	SRR0018628
SRX5796719	SRR0018629
SRX5796718	SRR0018630

Other reports [MSA viewer](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**  
 RID **00NVAFX01R** Search expires on 01-20 19:58 pm [Download All](#)  
 Program [Citation](#)  
 Database **SRA** [See details](#)

**SRA Blast search set information**  

SRX5796725	SRR0018623
SRX5796724	SRR0018624
SRX5796723	SRR0018625
SRX5796722	SRR0018626
SRX5796721	SRR0018627
SRX5796720	SRR0018628
SRX5796719	SRR0018629
SRX5796718	SRR0018630

Other reports [MSA viewer](#)

Fig.17a: No evidence of reads resembling the Pan-SL-CoV/GD sequences could be obtained from

PRJNA529540.

Job Title **gb|MT121216.1|**

RID [0DUJEA85401R](#) Search expires on 01-20 21:16 pm [Download All](#) ▾

Program [Citation](#) ▾

Database SRA [See details](#) ▾

SRA Blast search set information	
SRX1319167	SRR2591032

Query Length 29521

Other reports [MSA viewer](#) ⓘ

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

Fig17b: No evidence of reads resembling the Pan-SL-CoV/GD sequences could be found in SRX1319167.

### Potential malpractice associated with Chinese pangolin sequencing data

Recently, we are able to obtain 20 RNA-seq datasets for the transcriptomic sequencing of both *Manis Javanica* and *Manis Pentadactyla* skin appendage (skin, scales), deposited by the Guangdong Institute of Applied Biological Resources from a project that is Separate from the Coronavirus-related sequencing project ongoing in the People's Republic of China. Located under the BioProject accession number PRJNA610466.

[Edit Search](#) [Save Search](#) [Search Summary](#) ▾ [How to read this report?](#) [BLAST Help Videos](#) [Back to Traditional Results Page](#)

Job Title **MT121216:Pangolin coronavirus isolate MP789,...**

RID [6K8P3ES9016](#) Search expires on 04-05 19:28 pm [Download All](#) ▾

Program [Citation](#) ▾

Database SRA [See details](#) ▾

SRA Blast search set information	
SRX7911829	SRR11306687
SRX7911828	SRR11306688
SRX7911827	SRR11306689
SRX7911826	SRR11306690
SRX7911825	SRR11306691
SRX7911824	SRR11306692
SRX7911823	SRR11306693
SRX7911822	SRR11306694
SRX7911821	SRR11306695
SRX7911820	SRR11306696
SRX7911819	SRR11306697
SRX7911818	SRR11306698
SRX7911817	SRR11306699
SRX7911816	SRR11306700
SRX7911815	SRR11306701
SRX7911814	SRR11306702
SRX7911813	SRR11306703
SRX7911812	SRR11306704
SRX7911811	SRR11306705
SRX7911810	SRR11306706

**Filter Results**

Percent Identity  to  E value  to  Query Coverage  to

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

[Support Center](#)

Popular	Resources	Actions
<a href="#">PubMed</a>	<a href="#">Literature</a>	<a href="#">Submit</a>
<a href="#">PubMed Central</a>	<a href="#">Health</a>	<a href="#">Download</a>
<a href="#">Bookshelf</a>	<a href="#">Genomes</a>	<a href="#">Learn</a>
<a href="#">PubCham</a>	<a href="#">Genes</a>	<a href="#">Develop</a>

Fig 18. No evidence of the Pan-SL-CoV/GD sequences could be found within PRJNA610466. Although we did not find any evidence of Coronaviruses from these datasets, We noticed that some of the datasets contained several sequences that were associated with vectors associated with the immortalization and engineering of mammalian cells, namely sequences resembling HIV-1, Macaca Mulatta polyomavirus 1 and Human betaherpesvirus 5.







Fig.19: NCBI TRACE analysis result of SRR11306689

We downloaded the dataset with most concentrated occurrence of such sequences, SRR11306689, and performed sequence assembly using MEGAHIT[16]. Contiguous sequences with homology to these cellular engineering-associated sequences were identified using BowTie2[17] and their identities were elucidated through a combination of specific BLAST analysis and through sequence analysis using the Addgene sequence analysis tool.

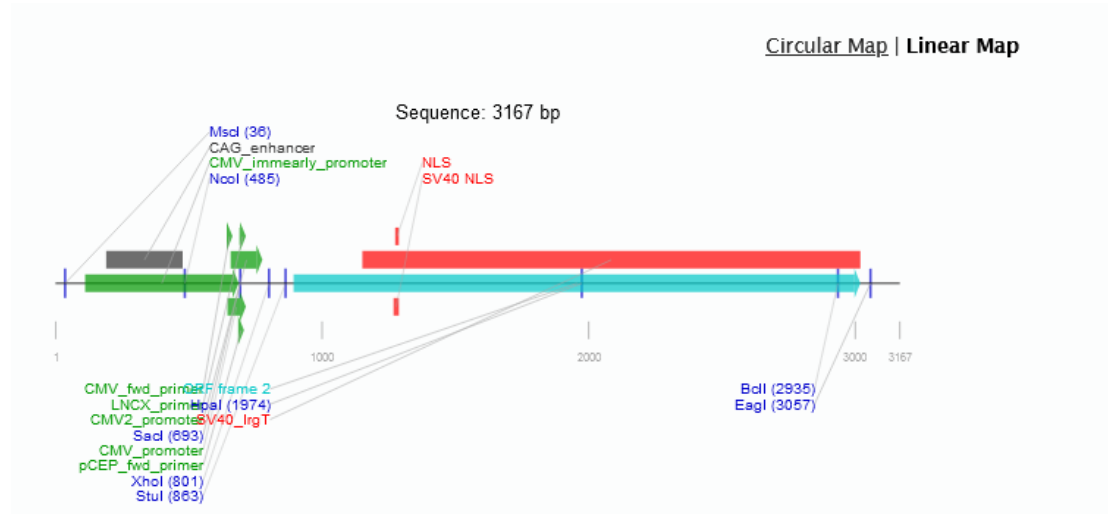


Fig.20: Addgene sequence analysis result of the single contig with detected homology to Macaca Mulatta polyomavirus 1 and Human betaherpesvirus 5.

Sequence analysis of the largest contig with homology to cellular engineering related sequences revealed a Simian Virus 40(SV40) Large T antigen (LTA) placed behind a CMV promoter. Such a sequence is normally used to immortalize cells through the oncogenic properties of the Large T antigen, normally delivered into the cells using an integrative transfection technique, such as lentiviral vectors. This sequence appear to be a partial mRNA transcript.

<input type="checkbox"/>	<a href="#">Murine retrovirus shuttle vector pZiPneoSV(TAg)_complete sequence</a>	syntheti... NA	32630	4096	4096	70%	0.0	99.78%	7020	<a href="#">Z93724.1</a>
<input type="checkbox"/>	<a href="#">Synthetic construct clone pARVA_T-Ag_complete sequence</a>	syntheti... NA	32630	3849	5119	89%	0.0	99.30%	5690	<a href="#">MF174873.1</a>
<input type="checkbox"/>	<a href="#">Mammalian expression vector pSV529HIFNG_complete sequence</a>	Mamma... NA	1945111	3502	6652	70%	0.0	99.84%	9455	<a href="#">LT727634.1</a>
<input type="checkbox"/>	<a href="#">Mammalian expression vector pSV51E6HF2_complete sequence</a>	Mamma... NA	1945103	3502	6652	70%	0.0	99.84%	9574	<a href="#">LT727623.1</a>
<input type="checkbox"/>	<a href="#">Mammalian expression vector pSV51E6HF1_complete sequence</a>	Mamma... NA	1945102	3502	6652	70%	0.0	99.84%	9472	<a href="#">LT727622.1</a>
<input checked="" type="checkbox"/>	<a href="#">Mammalian expression vector LNXCO3_complete sequence</a>	Mammalian exp...	1463	1463	25%	0.0	99.75%	7484	<a href="#">LT727330.1</a>	
<input checked="" type="checkbox"/>	<a href="#">EIAV-based lentiviral vector_complete sequence</a>	EIAV-based lent...	1461	1461	24%	0.0	100.00%	7941	<a href="#">GQ872121.1</a>	
<input checked="" type="checkbox"/>	<a href="#">Mutant Human betaherpesvirus 5 clone AD169-BAC20_complete genome</a>	Human betaher...	1448	1448	24%	0.0	100.00%	232314	<a href="#">MN920393.1</a>	
<input checked="" type="checkbox"/>	<a href="#">Mutant Human betaherpesvirus 5 clone AD169-BAC2_complete genome</a>	Human betaher...	1448	1448	24%	0.0	100.00%	233833	<a href="#">MN900952.1</a>	

Fig.21: BLAST result of this contig revealed that this sequence is associated with shuttle vectors (murine retrovirus, lentivirus) carrying the SV40 Large T antigen.

Four contigs within SRR11306689 are found with homology to Human Immunodeficiency Virus 1 (HIV-1), which upon specific BLAST analysis reveal themselves to be sequences derived from lentiviral transfer vectors.

Results for

Program **BLASTN** [Citation](#)

Database **nt** [See details](#)

Query ID **cl|Query\_64532**

Description **k141\_15599**

Molecule type **dna**

Query Length **473**

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

**Organism** only top 20 will appear

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 **500**

select all 500 sequences selected

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> PREDICTED: Manis pentadactyla coiled-coil domain containing 185 (CCDC185)_mRNA	Manis pentada...	627	627	73%	3e-175	99.14%	6791	XM_03688626.1
<input checked="" type="checkbox"/> Cloning vector sh-4FP-PPP2CA.complete.sequence	Cloning vector...	231	231	26%	3e-56	100.00%	8000	MN995874.1
<input checked="" type="checkbox"/> Cloning vector sh-PPP2CA.dna.complete.sequence	Cloning vector...	231	231	26%	3e-56	100.00%	7716	MN995873.1
<input checked="" type="checkbox"/> Cloning vector sh-ctf1.complete.sequence	Cloning vector...	231	231	26%	3e-56	100.00%	7717	MN995872.1
<input checked="" type="checkbox"/> Cloning vector sh-4FP.complete.sequence	Cloning vector...	231	231	26%	3e-56	100.00%	7716	MN995871.1
<input checked="" type="checkbox"/> Cloning vector pSF_lent1_SARS-CoV2_partial-SiEMTJ.complete.sequence	Cloning vector...	231	231	26%	3e-56	100.00%	13543	MT299805.1
<input checked="" type="checkbox"/> Cloning vector pSF_lent1_SARS-CoV2_partial-ORF1b_partial-S.complete.sequence	Cloning vector...	231	231	26%	3e-56	100.00%	13507	MT299804.1
<input checked="" type="checkbox"/> Cloning vector pSF_lent1_SARS-CoV2_partial-ORF1ab.complete.sequence	Cloning vector...	231	231	26%	3e-56	100.00%	13462	MT299803.1
<input checked="" type="checkbox"/> Cloning vector pSF_lent1_SARS-CoV2_partial-ORF1a.complete.sequence	Cloning vector...	231	231	26%	3e-56	100.00%	13408	MT299802.1

Results for

Program **BLASTN** [Citation](#)

Database **nt** [See details](#)

Query ID **cl|Query\_64533**

Description **k141\_57178**

Molecule type **dna**

Query Length **285**

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

**Organism** only top 20 will appear

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

**Filter** **Reset**

**Descriptions** Graphic Summary Alignments Taxonomy

**Sequences producing significant alignments** Download [Manage Columns](#) Show **500**

select all 0 sequences selected

Description	Scientific Name	Common Name	Taxid	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input type="checkbox"/> Cloning vector RS474_ErbB-RASERTC-CaasVP24.complete.sequence	Cloning...	NA	2577058	466	466	100%	4e-127	95.64%	16350	MK801288.1
<input type="checkbox"/> Lentiviral transfer vector FUW.tetO.complete.sequence	Lentivira...	NA	2506927	466	466	100%	4e-127	95.64%	9554	MK318529.1
<input type="checkbox"/> Cloning vector cLent1-EF1a-cCaas9-DHMT3B-E692A1a-2a-bla.complete.sequence	Cloning...	NA	2168471	466	466	100%	4e-127	95.64%	14877	MG840314.1
<input checked="" type="checkbox"/> Cloning vector cDBR.complete.sequence	Cloning vector...		448	448	100%	1e-121	94.37%	6619	FR822201.1	
<input checked="" type="checkbox"/> Cloning vector shRNA EYFP-P2A Puro.complete.sequence	Cloning vector...		444	444	100%	2e-120	94.06%	9625	MH749454.1	
<input checked="" type="checkbox"/> Human immunodeficiency virus 1 (HIV-1) RNA	Human immun...		270	270	51%	3e-68	100.00%	9781	Z11530.1	
<input checked="" type="checkbox"/> Cloning vector eAAV-EF1A-GagUtrn-FlnR-GFP-CGR3TC.complete.sequence	Cloning vector...		267	267	50%	4e-67	100.00%	7010	MT614242.1	
<input checked="" type="checkbox"/> Cloning vector eAAV-EF1A-DIO-GagUtrn-FlnR-GFP-CCR3TC.complete.sequence	Cloning vector...		267	267	50%	4e-67	100.00%	7204	MT612431.1	

Results for

Program **BLASTN** [Citation](#)

Database **nt** [See details](#)

Query ID **cl|Query\_64534**

Description **k141\_57181**

Molecule type **dna**

Query Length **300**

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

**Organism** only top 20 will appear

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

**Filter** **Reset**

**Descriptions** Graphic Summary

**Sequences producing significant alignments** Download [Manage Columns](#) Show **500**

select all 0 sequences selected

Description	Scientific Name	Common Name	Taxid	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input type="checkbox"/> Vector cLV1-TRE3G-eGFP.complete.sequence	Vector...	NA	2491125	549	549	100%	4e-152	99.67%	8536	MH325104.1
<input type="checkbox"/> Vector cLV1-TRE3G-prot.complete.sequence	Vector...	NA	2491126	549	549	100%	4e-152	99.67%	8632	MH325103.1
<input type="checkbox"/> Lentiviral transfer vector pHSCXV.complete.sequence	Lentivira...	NA	255321	520	520	100%	3e-143	98.00%	7207	AY468486.1
<input checked="" type="checkbox"/> Cloning vector cDBR.complete.sequence	Cloning vector...		484	484	100%	1e-132	95.71%	6619	FR822201.1	
<input checked="" type="checkbox"/> Cloning vector shRNA EYFP-P2A Puro.complete.sequence	Cloning vector...		479	479	100%	5e-131	95.39%	9625	MH749454.1	
<input checked="" type="checkbox"/> Eicosomal mammalian expression vector cER4.3.complete.sequence	Eicosomal mam...		285	285	51%	1e-72	100.00%	11919	MC182239.1	
<input checked="" type="checkbox"/> Vector CMV-prot.complete.sequence	Vector CMV-prot...		276	276	49%	8e-70	100.00%	9546	MH325102.1	

Results for

Program **BLASTN** [Citation](#)

Database **nt** [See details](#)

Query ID **cl|Query\_64535**

Description **k141\_9597**

Molecule type **dna**

Query Length **1040**

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

**Organism** only top 20 will appear

Type common name, binomial, taxid or group name

[+ Add organism](#)

Percent Identity  to  E value  to  Query Coverage  to

**Filter** **Reset**

**Descriptions**

**Sequences producing significant alignments** Download [Manage Columns](#) Show **500**

select all 0 sequences selected

Description	Scientific Name	Common Name	Taxid	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input type="checkbox"/> Vector cLV1-TRE3G-eGFP.complete.sequence	Vector...	NA	2491125	1507	2674	79%	0.0	99.76%	8536	MH325104.1
<input type="checkbox"/> Vector cLV1-TRE3G-prot.complete.sequence	Vector...	NA	2491126	1507	2674	79%	0.0	99.76%	8632	MH325103.1
<input type="checkbox"/> Synthetic HIV-1 clone cSF25.8.complete.sequence	syntheti...	NA	1927832	1502	2657	79%	0.0	99.64%	9723	KU521529.1
<input type="checkbox"/> Lentiviral vector cTK1940.complete.sequence	Lentivira...	NA	2249735	418	1150	43%	4e-112	99.56%	9431	MH297436.1
<input type="checkbox"/> Homo sapiens FER tyrosine kinase (FER). RefSeqGene on chromosome 5	Homo s...	human	9608	403	403	20%	1e-107	100.00%	456020	NC_011445.2

Fig.23: Specific BLAST analysis result of the 4 contigs associated with Human Immunodeficiency Virus 1 (HIV-1).

A deep analysis of these 4 contigs revealed 2 contigs spanning the vector-virus junction for common lentiviral transfer vectors, one contig displaying an Integration junction of the LTR into Manis Pentadactyla DNA and one contig displaying an integration junction of the LTR into Homo Sapiens DNA.

The presence of these fragmented sequences carrying both the payload (SV40 Large T antigen) and the vehicle (Lentiviral transfer vectors) with evidence of delivery into pangolin cells (integration junction of Vector LTR DNA into Manis Pentadactyla genomic DNA), suggesting an ongoing effort of immortalizing pangolin cells and keeping them in culture being conducted in the Guangdong Institute of Applied Biological Resources.

Indeed, the presence of Primary Fibroblast (Skin, Muscle) cells cultures from both Manis Pentadactyla and Manis Javanica and their availability to labs have been recently confirmed by two BioSamples of primary fibroblast cells “collected by Dr. Shujin Luo (Peking University, China)” placed under accession SAMN16895765 and SAMN16895764, with collection date of 20/03/2020 and 08/04/2020 respectively.

<b>Chinese pangolin</b>		<b>Sunda pangolin</b>																																																																																																																																													
Identifiers	BioSample: SAMN16895765; Sample name: Sample3814; SRA: SR5780986	Identifiers	BioSample: SAMN16895764; Sample name: Sample3820; SRA: SR5780984																																																																																																																																												
Organism	<a href="#">Manis pentadactyla</a> (Chinese pangolin) cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Placentalia; Manidae; Manis	Organism	<a href="#">Manis javanica</a> (Malayan pangolin) cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria; Placentalia; Manidae; Manis																																																																																																																																												
Package	<a href="#">Model organism or animal version 1.0</a>	Package	<a href="#">Model organism or animal version 1.0</a>																																																																																																																																												
Attributes	<table border="0"> <tr><td>strain</td><td>not collected</td></tr> <tr><td>isolate</td><td>not collected</td></tr> <tr><td>breed</td><td>not collected</td></tr> <tr><td>cultivar</td><td>not collected</td></tr> <tr><td>ecotype</td><td>not collected</td></tr> <tr><td>age</td><td>not collected</td></tr> <tr><td>development stage</td><td>not collected</td></tr> <tr><td>sex</td><td>male</td></tr> <tr><td>tissue</td><td>primary fibroblasts</td></tr> <tr><td>biomaterial provider</td><td>Polina Fereiman</td></tr> <tr><td>birth date</td><td>not collected</td></tr> <tr><td>birth location</td><td>not collected</td></tr> <tr><td>breeding history</td><td>not collected</td></tr> <tr><td>breeding method</td><td>not collected</td></tr> <tr><td>cell line</td><td>not collected</td></tr> <tr><td>cell subtype</td><td>not collected</td></tr> <tr><td>cell type</td><td>not collected</td></tr> <tr><td>collected by</td><td>not collected</td></tr> <tr><td>collection date</td><td>2020-03-20</td></tr> <tr><td>culture collection</td><td>not collected</td></tr> <tr><td>death date</td><td>not collected</td></tr> <tr><td>disease</td><td>not collected</td></tr> <tr><td>disease stage</td><td>not collected</td></tr> <tr><td>genotype</td><td>not collected</td></tr> <tr><td>geographic location</td><td>not collected</td></tr> <tr><td>growth protocol</td><td>not collected</td></tr> <tr><td>health state</td><td>n/a</td></tr> <tr><td>isolation source</td><td>not collected</td></tr> <tr><td>latitude and longitude</td><td>not collected</td></tr> <tr><td>phenotype</td><td>not collected</td></tr> <tr><td>sample type</td><td>not collected</td></tr> <tr><td>specimen voucher</td><td>not collected</td></tr> <tr><td>storage conditions</td><td>not collected</td></tr> <tr><td>stud book number</td><td>not collected</td></tr> <tr><td>treatment</td><td>not collected</td></tr> </table>	strain	not collected	isolate	not collected	breed	not collected	cultivar	not collected	ecotype	not collected	age	not collected	development stage	not collected	sex	male	tissue	primary fibroblasts	biomaterial provider	Polina Fereiman	birth date	not collected	birth location	not collected	breeding history	not collected	breeding method	not collected	cell line	not collected	cell subtype	not collected	cell type	not collected	collected by	not collected	collection date	2020-03-20	culture collection	not collected	death date	not collected	disease	not collected	disease stage	not collected	genotype	not collected	geographic location	not collected	growth protocol	not collected	health state	n/a	isolation source	not collected	latitude and longitude	not collected	phenotype	not collected	sample type	not collected	specimen voucher	not collected	storage conditions	not collected	stud book number	not collected	treatment	not collected	Attributes	<table border="0"> <tr><td>strain</td><td>not collected</td></tr> <tr><td>isolate</td><td>not collected</td></tr> <tr><td>breed</td><td>not collected</td></tr> <tr><td>cultivar</td><td>not collected</td></tr> <tr><td>ecotype</td><td>not collected</td></tr> <tr><td>age</td><td>not collected</td></tr> <tr><td>development stage</td><td>not collected</td></tr> <tr><td>sex</td><td>missing</td></tr> <tr><td>tissue</td><td>primary fibroblasts</td></tr> <tr><td>biomaterial provider</td><td>Polina Fereiman</td></tr> <tr><td>birth date</td><td>not collected</td></tr> <tr><td>birth location</td><td>not collected</td></tr> <tr><td>breeding history</td><td>not collected</td></tr> <tr><td>breeding method</td><td>not collected</td></tr> <tr><td>cell line</td><td>not collected</td></tr> <tr><td>cell subtype</td><td>not collected</td></tr> <tr><td>cell type</td><td>not collected</td></tr> <tr><td>collected by</td><td>not collected</td></tr> <tr><td>collection date</td><td>2020-04-08</td></tr> <tr><td>culture collection</td><td>not collected</td></tr> <tr><td>death date</td><td>not collected</td></tr> <tr><td>disease</td><td>not collected</td></tr> <tr><td>disease stage</td><td>not collected</td></tr> <tr><td>genotype</td><td>not collected</td></tr> <tr><td>geographic location</td><td>not collected</td></tr> <tr><td>growth protocol</td><td>not collected</td></tr> <tr><td>health state</td><td>n/a</td></tr> <tr><td>isolation source</td><td>not collected</td></tr> <tr><td>latitude and longitude</td><td>not collected</td></tr> <tr><td>phenotype</td><td>not collected</td></tr> <tr><td>sample type</td><td>not collected</td></tr> <tr><td>specimen voucher</td><td>not collected</td></tr> <tr><td>storage conditions</td><td>not collected</td></tr> <tr><td>stud book number</td><td>not collected</td></tr> <tr><td>treatment</td><td>not collected</td></tr> </table>	strain	not collected	isolate	not collected	breed	not collected	cultivar	not collected	ecotype	not collected	age	not collected	development stage	not collected	sex	missing	tissue	primary fibroblasts	biomaterial provider	Polina Fereiman	birth date	not collected	birth location	not collected	breeding history	not collected	breeding method	not collected	cell line	not collected	cell subtype	not collected	cell type	not collected	collected by	not collected	collection date	2020-04-08	culture collection	not collected	death date	not collected	disease	not collected	disease stage	not collected	genotype	not collected	geographic location	not collected	growth protocol	not collected	health state	n/a	isolation source	not collected	latitude and longitude	not collected	phenotype	not collected	sample type	not collected	specimen voucher	not collected	storage conditions	not collected	stud book number	not collected	treatment	not collected
strain	not collected																																																																																																																																														
isolate	not collected																																																																																																																																														
breed	not collected																																																																																																																																														
cultivar	not collected																																																																																																																																														
ecotype	not collected																																																																																																																																														
age	not collected																																																																																																																																														
development stage	not collected																																																																																																																																														
sex	male																																																																																																																																														
tissue	primary fibroblasts																																																																																																																																														
biomaterial provider	Polina Fereiman																																																																																																																																														
birth date	not collected																																																																																																																																														
birth location	not collected																																																																																																																																														
breeding history	not collected																																																																																																																																														
breeding method	not collected																																																																																																																																														
cell line	not collected																																																																																																																																														
cell subtype	not collected																																																																																																																																														
cell type	not collected																																																																																																																																														
collected by	not collected																																																																																																																																														
collection date	2020-03-20																																																																																																																																														
culture collection	not collected																																																																																																																																														
death date	not collected																																																																																																																																														
disease	not collected																																																																																																																																														
disease stage	not collected																																																																																																																																														
genotype	not collected																																																																																																																																														
geographic location	not collected																																																																																																																																														
growth protocol	not collected																																																																																																																																														
health state	n/a																																																																																																																																														
isolation source	not collected																																																																																																																																														
latitude and longitude	not collected																																																																																																																																														
phenotype	not collected																																																																																																																																														
sample type	not collected																																																																																																																																														
specimen voucher	not collected																																																																																																																																														
storage conditions	not collected																																																																																																																																														
stud book number	not collected																																																																																																																																														
treatment	not collected																																																																																																																																														
strain	not collected																																																																																																																																														
isolate	not collected																																																																																																																																														
breed	not collected																																																																																																																																														
cultivar	not collected																																																																																																																																														
ecotype	not collected																																																																																																																																														
age	not collected																																																																																																																																														
development stage	not collected																																																																																																																																														
sex	missing																																																																																																																																														
tissue	primary fibroblasts																																																																																																																																														
biomaterial provider	Polina Fereiman																																																																																																																																														
birth date	not collected																																																																																																																																														
birth location	not collected																																																																																																																																														
breeding history	not collected																																																																																																																																														
breeding method	not collected																																																																																																																																														
cell line	not collected																																																																																																																																														
cell subtype	not collected																																																																																																																																														
cell type	not collected																																																																																																																																														
collected by	not collected																																																																																																																																														
collection date	2020-04-08																																																																																																																																														
culture collection	not collected																																																																																																																																														
death date	not collected																																																																																																																																														
disease	not collected																																																																																																																																														
disease stage	not collected																																																																																																																																														
genotype	not collected																																																																																																																																														
geographic location	not collected																																																																																																																																														
growth protocol	not collected																																																																																																																																														
health state	n/a																																																																																																																																														
isolation source	not collected																																																																																																																																														
latitude and longitude	not collected																																																																																																																																														
phenotype	not collected																																																																																																																																														
sample type	not collected																																																																																																																																														
specimen voucher	not collected																																																																																																																																														
storage conditions	not collected																																																																																																																																														
stud book number	not collected																																																																																																																																														
treatment	not collected																																																																																																																																														
Description	not collected	Description	not collected																																																																																																																																												
BioProject	<a href="#">PRJNA612907</a> DNA Zoo Retrieve <a href="#">all samples</a> from this project	BioProject	<a href="#">PRJNA612907</a> DNA Zoo Retrieve <a href="#">all samples</a> from this project																																																																																																																																												

Figure 24: BioSample description of SAMN16895765 and SAMN16895764 as “primary fibroblast cells”.

As such a cell line could potentially be used to culture the Pan-SL-CoV/GD virus, should an isolate in VERO E6 [2] exist, these cell lines may potentially be used to contrive “novel” BioSamples and SRAs for this sequence through inoculation and serial passage of the cultured virus in order to eliminate the primate host sequences from the original samples, due to the central role of the RBD of this supposedly “wild” sequence in current publications regarding SARS-CoV-2 origin.

We therefore urge caution when adopting any short read sequencing (SRA) data or viral nucleotide sequences from pangolins with a date of deposition after the collection date for SAMN16895765 and SAMN16895764, especially after the publishing date of PRJNA610466,

24/11/2020, due to the identity of potential “tissue” samples being no longer restricted exclusively to living or dead wild animals once a primary or immortalized cell line of such a species have been established, potentially allowing malpractice when “sequencing” “new” samples from these two species.

The MEGAHIT result and the obtained vector sequences from SRR11306689 have been deposited as

Galaxy266-[Assembly\_with\_MEGAHIT\_on\_data\_260] SRR11306689.fasta

HIV-1 from SRR11306689.fasta

and

SV40 LTA+CMV from SRR11306689.fasta

## Conclusions

The Extreme lack of transparency and the sheer level of contamination from the original samples, the lack of epidemiological evidence of its existence at the location of its 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.

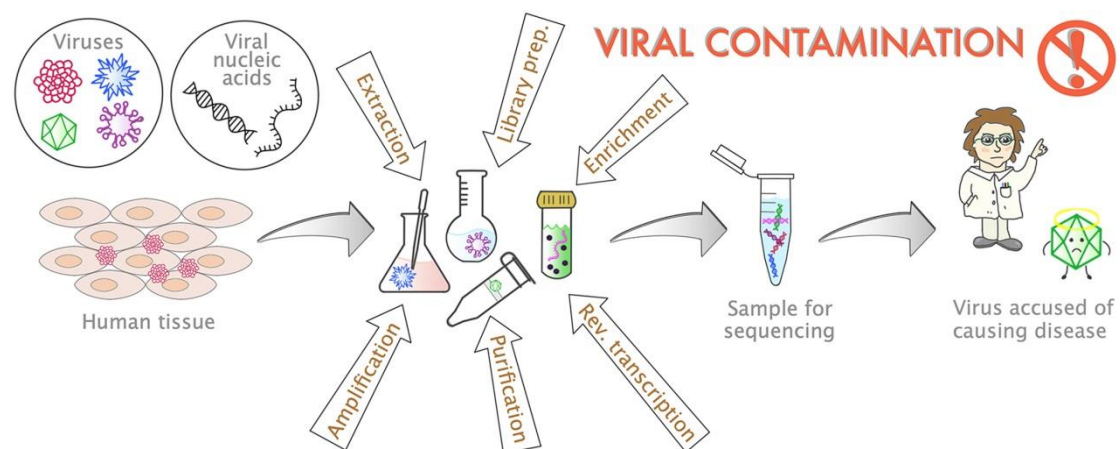


Figure 25. A cartoon diagram of contamination in sequencing experiment leading to false results and false “discoveries”.

## 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	Primary	Mammalian	Primate-related	results	Identification of
----------------------	---------	-----------	-----------------	---------	-------------------

registration date	Trace results and percentage	in Krona and read size by Kbp	“Coronaviridae” as by Trace and total read size
<a href="#">SRX8582289</a> 22-Jun-2020	Manis javanica: <b>43.52%</b>	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%.

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

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	278	0%	8e-69	100.00%	<a href="#">SRA_SRR12053850.88444297.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	402	0%	8e-69	100.00%	<a href="#">SRA_SRR12053850.83916175.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	388	0%	8e-69	100.00%	<a href="#">SRA_SRR12053850.83916175.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	278	0%	8e-69	100.00%	<a href="#">SRA_SRR12053850.82221130.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	278	0%	8e-69	100.00%	<a href="#">SRA_SRR12053850.71234261.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	278	0%	8e-69	100.00%	<a href="#">SRA_SRR12053850.71234261.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	5169	2%	8e-69	100.00%	<a href="#">SRA_SRR12053850.51889132.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	7268	3%	8e-69	100.00%	<a href="#">SRA_SRR12053850.26027930.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	5671	2%	8e-69	100.00%	<a href="#">SRA_SRR12053850.21554419.1</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	278	0%	8e-69	100.00%	<a href="#">SRA_SRR12053850.13271287.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	278	4760	1%	8e-69	100.00%	<a href="#">SRA_SRR12053850.62042.2</a>
<input checked="" type="checkbox"/> <a href="#">SRX8582289</a>	276	276	0%	3e-68	100.00%	<a href="#">SRA_SRR12053850.82221130.2</a>

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

Description gn|SRA|SRR12053850.82221130.2 82221130 (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 59 sequences selected [GenBank](#) [Graphics](#) [Distance tree of results](#)

Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
<input checked="" type="checkbox"/> <a href="#">Homo sapiens general transcription factor III pseudogene 14 (GTF2IP14) on chromosome 7</a>	276	276	99%	3e-70	100.00%	<a href="#">NG_043494.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens FOSMID clone ABC13-48840700E15 from chromosome 7, complete sequence</a>	276	276	99%	3e-70	100.00%	<a href="#">AC242196.4</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens BAC clone RP11-460N20 from 7, complete sequence</a>	276	276	99%	3e-70	100.00%	<a href="#">AC073210.8</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens general transcription factor III pseudogene 5 (GTF2IP5) on chromosome 7</a>	270	270	99%	1e-68	99.33%	<a href="#">NG_026590.1</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-340J24 from chromosome 7, complete sequence</a>	270	270	99%	1e-68	99.33%	<a href="#">AC185242.2</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-623C19 from chromosome 7, complete sequence</a>	270	270	99%	1e-68	99.33%	<a href="#">AC184799.2</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-114G16 from chromosome 7, complete sequence</a>	270	270	99%	1e-68	99.33%	<a href="#">AC183835.2</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-20I5 from chromosome 7, complete sequence</a>	270	270	99%	1e-68	99.33%	<a href="#">AC174000.3</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens BAC clone RP11-479O9 from 7, complete sequence</a>	270	270	99%	1e-68	99.33%	<a href="#">AC073107.7</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-565C10 from chromosome 7, complete sequence</a>	270	270	99%	1e-68	99.33%	<a href="#">AC148313.3</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens BAC clone CH17-99D2 from chromosome 4, complete sequence</a>	226	226	99%	3e-55	93.96%	<a href="#">AC278002.1</a>
<input checked="" type="checkbox"/> <a href="#">Homo sapiens FOSMID clone ABC27-154I13 from chromosome 4, complete sequence</a>	226	226	99%	3e-55	93.96%	<a href="#">AC240529.1</a>
<input checked="" type="checkbox"/> <a href="#">Pan troglodytes BAC clone CH251-4D23 from chromosome 7, complete sequence</a>	226	226	99%	3e-55	93.96%	<a href="#">AC148834.3</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 16</a>	193	193	99%	3e-45	90.07%	<a href="#">CP034494.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 19</a>	182	182	99%	5e-42	88.74%	<a href="#">CP034522.1</a>
<input checked="" type="checkbox"/> <a href="#">Eukaryotic synthetic construct chromosome 19</a>	182	182	99%	5e-42	88.74%	<a href="#">CP034497.1</a>

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
<a href="#">SRX9714436</a> 21-Dec-2020	Manis javanica: <b>3.14%</b> Homo sapiens: <b>0.04%</b>	<b>Homo sapiens</b> 12332	Pangolin coronavirus 3 Kbp
<a href="#">SRX9714921</a> 21-Dec-2020	Homo sapiens: <b>0.15%</b>	<b>Homo sapiens</b> 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
CTATACAACAAACCCCATGACACGAGTTTACCTATGTAACAAACCTTCA
```

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
ATAGGGAAGTGTGGTACCAAGGAGCAATATTCAATACAGCAACCCAGGAAG
```

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 using stand-alone MagicBlast[11]

suggests 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-tufted-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_883) 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.

## Note as in 2021/05/19

We have recently obtained a SRA dataset on the Chinese National GeneBank(CNGB) of a claimed “skin sample” from a Pangolin individual with claimed infection by the Pan-SL-CoV/GD genome, DOI: <http://dx.doi.org/10.26036/CNP0001573>. Analysis of the read files provided have identified significant amount of human read contamination within the dataset. The alignment result to the human genome has been deposited as CNP0001573\_human\_alignments.sam.



Job Title **Nucleotide Sequence**

RID [A9TEHE7Y013](#) Search expires on 05-20 18:17 pm [Download All](#) ▾

Program **BLASTN** [Citation](#) ▾

Database **nt** [See details](#) ▾

Query ID **lcl|Query\_19321**

Description **None**

Molecule type **dna**

Query Length **150**

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

**Filter Results**

exclude

Organism *only top 20 will appear*

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  [?](#)

select all 100 sequences selected

	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/>	<a href="#">Eukaryotic synthetic construct chromosome Y</a>	<a href="#">eukaryotic synt...</a>	278	1369	100%	1e-70	100.00%	64558752	<a href="#">CP034510.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens chromosome Y palindromes P1, P2, P3 and inverted repeat IR2 (P1-P2-P3-IR2@)</a>	<a href="#">Homo sapiens</a>	278	1112	100%	1e-70	100.00%	4538292	<a href="#">NG_004755.2</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens DNA chromosome Y nearly complete genome</a>	<a href="#">Homo sapiens</a>	278	406	100%	1e-70	100.00%	50294361	<a href="#">AP023484.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens BAC clone RP11-245K4 from Y complete sequence</a>	<a href="#">Homo sapiens</a>	278	278	100%	1e-70	100.00%	182083	<a href="#">AC007965.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens BAC clone RP11-563C13 from Y complete sequence</a>	<a href="#">Homo sapiens</a>	278	278	100%	1e-70	100.00%	184215	<a href="#">AC007322.4</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens BAC clone RP11-506M9 from Y complete sequence</a>	<a href="#">Homo sapiens</a>	278	278	100%	1e-70	100.00%	166436	<a href="#">AC016752.2</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens BAC clone RP11-427G18 from Y complete sequence</a>	<a href="#">Homo sapiens</a>	278	278	100%	1e-70	100.00%	205237	<a href="#">AC008175.2</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone CH251-412D24 from Y complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	166356	<a href="#">AC147572.4</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone CH251-506P22 from Y complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	195158	<a href="#">AC147608.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone CH251-475M5 from Y complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	220557	<a href="#">AC146198.2</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone CH251-528H6 from Y complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	169760	<a href="#">AC147361.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone CH251-548O14 from Y complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	229006	<a href="#">AC147114.2</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone CH251-113M4 from Y complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	183460	<a href="#">AC146515.2</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone RP43-5I10 from chromosome y complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	183293	<a href="#">AC142348.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes chromosome Y clone PTB-391J06 complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	171013	<a href="#">BS000677.1</a>
<input checked="" type="checkbox"/>	<a href="#">Pan troglodytes BAC clone CH251-571K7 from chromosome y complete sequence</a>	<a href="#">Pan troglodytes</a>	257	257	98%	1e-64	97.97%	171231	<a href="#">AC148933.3</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens chromosome Y palindromes P4 and P5 (P4-P5@) on chromosome Y</a>	<a href="#">Homo sapiens</a>	128	257	68%	1e-25	89.32%	1351786	<a href="#">NG_004636.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens heat shock transcription factor Y-linked 1 (HSFY1). RefSeqGene on chromosome Y</a>	<a href="#">Homo sapiens</a>	128	128	68%	1e-25	89.32%	49276	<a href="#">NG_012030.1</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens BAC clone RP11-569J3 from Y complete sequence</a>	<a href="#">Homo sapiens</a>	128	128	68%	1e-25	89.32%	156357	<a href="#">AC022486.4</a>
<input checked="" type="checkbox"/>	<a href="#">Homo sapiens BAC clone RP11-143C1 from Y complete sequence</a>	<a href="#">Homo sapiens</a>	128	128	68%	1e-25	89.32%	174082	<a href="#">AC007379.2</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Castor canadensis glycoprotein M6B (Gom6b), transcript variant X8, mRNA</a>	<a href="#">Castor canad...</a>	117	117	86%	2e-22	82.98%	4578	<a href="#">XM_020159964.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Castor canadensis glycoprotein M6B (Gom6b), transcript variant X7, mRNA</a>	<a href="#">Castor canad...</a>	117	117	86%	2e-22	82.98%	3225	<a href="#">XM_020159955.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Castor canadensis glycoprotein M6B (Gom6b), transcript variant X6, mRNA</a>	<a href="#">Castor canad...</a>	117	117	86%	2e-22	82.98%	4634	<a href="#">XM_020159948.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Castor canadensis glycoprotein M6B (Gom6b), transcript variant X5, mRNA</a>	<a href="#">Castor canad...</a>	117	117	86%	2e-22	82.98%	3033	<a href="#">XM_020159942.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Castor canadensis glycoprotein M6B (Gom6b), transcript variant X4, mRNA</a>	<a href="#">Castor canad...</a>	117	117	86%	2e-22	82.98%	3134	<a href="#">XM_020159934.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Castor canadensis glycoprotein M6B (Gom6b), transcript variant X3, mRNA</a>	<a href="#">Castor canad...</a>	117	117	86%	2e-22	82.98%	3345	<a href="#">XM_020159926.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Castor canadensis glycoprotein M6B (Gom6b), transcript variant X2, mRNA</a>	<a href="#">Castor canad...</a>	117	117	86%	2e-22	82.98%	4754	<a href="#">XM_020159918.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Castor canadensis glycoprotein M6B (Gom6b), transcript variant X1, mRNA</a>	<a href="#">Castor canad...</a>	117	117	86%	2e-22	82.98%	3275	<a href="#">XM_020159910.1</a>
<input checked="" type="checkbox"/>	<a href="#">PREDICTED: Avicanthis niloticus glycoprotein M6B (Gom6b), transcript variant X6, mRNA</a>	<a href="#">Avicanthis nilo...</a>	115	115	69%	8e-22	86.36%	2986	<a href="#">XM_034484867.1</a>

E00552:330:H2ML7CCX2:2:1207:10957:4509 16 chrY 22559655 1 150M \* 0 0  
 ATGGCTATTCAAAGTATCTCACTGTTGTTTTCCAGCAGTGGTGATGATGAAGTCCAGTTTCTAGAAGGGGAGTGCTCAACCTAATTTCTAATTTCTGGTAGCTGCAGGAATTTTTG  
 AAFATFF<JFJFATFJ<FFJFJJJF<FA<FA<FAFAAAJAA<<FA7A7A<JJJAJFA<F7J7AAJJJJFAJJJAJFJA<<JJA<J<<JFJJ<J7J<F7--FJJ<<F<<JFA<JFA<<A<FJJJAAJJJJJFJFAA AS:i:300 XS:i:300 XN:i:0  
 XM:i:0 XO:i:0 XG:i:0 NM:i:0 MD:Z:150 YT:Z:UU

Fig.S4: An example of a read aligned to the human Y chromosome from CNP0001573. This could be linked to the human lung adenocarcinoma cell line A549, which originated from a male individual and is used for experimentation with SARS-like viruses.[18][19]

## Note as in 2021/06/20

Two sequencing datasets under PRJNA607174, [SRX7756767](#) and [SRX7756768](#) have been published. These datasets are barely readable with very little sequence that resembles attributable, non-ribosomal RNA. Only one of these datasets are deemed analyzable by the NCBI TRACE tool.

Identification of reads were performed using BowTie2 against MN908947.3 and MT121216.1,



and only [SRX7756767](#) contained one read from the Pan-SL-CoV/GD sequence.

Table S4: TRACE analysis result of [SRX7756768](#) and [SRX7756767](#).

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
<a href="#">SRX7756768</a> 18-Feb-2020	Homo sapiens: <b>0.03%</b>	<b>Homo sapiens</b> 6557937	Canine coronavirus 2 2Kbp
<a href="#">SRX7756767</a> 18-Feb-2020	Homo sapiens: 8.69Mbp (BowTie2) Manis pentadactyla: 312.5Mbp (BowTie2) Mus Musculus: 12.39Kbp (BowTie2)	N/A (not analyzable)	One read matching MT121216.1 (BowTie2)

A00129:562:HWNFWDSXX:1:2267:15420:7247 16 MN908947.3 812 8 150M \* 0

O

ACTCGCTATGTCGATAACACCTACTGTGGCCCTGATGGCTACCCTCTTGAGTGCATTAAAGACTTGCTG  
GCGCGTGCTGGTAAAGCTTCTGCACCTTGTCCGAACAACCTGGACTTCTTGACACTAAGAGAGGTGTGTAC  
TGCTGCCGT

FFFFFFFFFFFFFFFFFFFFF,F:FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FFFFFFFFF:F:FFFFFFFFFFFFFFFFFFFFFFFFFFFFF  
FFF,,FFFFFFFFFFFFFFFFFFFFFFFFFFFFF:FF AS:i:-48 XN:i:0 XM:i:10  
XO:i:0 XG:i:0 NM:i:10 MD:Z:19A2T40C1T2A2A17A27A13G5A12 YT:Z:UU

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 synt...	278	278	100%	1e-70	100.00%	64242768	CP034522.1
<input checked="" type="checkbox"/> Eukaryotic synthetic construct chromosome 19	eukaryotic synt...	278	278	100%	1e-70	100.00%	64242768	CP034497.1
<input checked="" type="checkbox"/> Homo sapiens chromosome 19 open reading frame 12 (C19orf12), RefSeqGene on chromosome 19	Homo sapiens	278	278	100%	1e-70	100.00%	23904	NG_031970.2
<input checked="" type="checkbox"/> Homo sapiens isolate CHM13 chromosome 19	Homo sapiens	278	278	100%	1e-70	100.00%	61707364	CP068259.2
<input checked="" type="checkbox"/> Homo sapiens DNA chromosome 19_nearly complete genome	Homo sapiens	278	278	100%	1e-70	100.00%	59105444	AP023479.1
<input checked="" type="checkbox"/> Homo sapiens chromosome 19 clone LLNLR-232H11_complete sequence	Homo sapiens	278	278	100%	1e-70	100.00%	37770	AC010513.6
<input checked="" type="checkbox"/> PREDICTED: Pongo abelii chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant X4....	Pongo abelii	231	231	93%	8e-57	96.43%	2247	XM_009253101.2
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant XT....	Pan paniscus	152	152	54%	6e-33	100.00%	4099	XM_003816225.5
<input checked="" type="checkbox"/> PREDICTED: Pan paniscus chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant X1....	Pan paniscus	152	152	54%	6e-33	100.00%	4292	XM_003816222.4
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant X... Pan troglodytes	Pan troglodytes	152	152	54%	6e-33	100.00%	3594	XM_003953437.2
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant X... Pan troglodytes	Pan troglodytes	152	152	54%	6e-33	100.00%	3771	XM_003316249.4
<input checked="" type="checkbox"/> Homo sapiens chromosome 19 open reading frame 12 (C19orf12), transcript variant 1_mRNA	Homo sapiens	152	152	54%	6e-33	100.00%	4265	NM_001031726.4
<input checked="" type="checkbox"/> Homo sapiens chromosome 19 open reading frame 12 (C19orf12), transcript variant 5_mRNA	Homo sapiens	152	152	54%	6e-33	100.00%	4242	NM_001282929.1
<input checked="" type="checkbox"/> PREDICTED: Pan troglodytes chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant X... Pan troglodytes	Pan troglodytes	150	150	54%	2e-32	100.00%	4176	XM_009435202.3
<input checked="" type="checkbox"/> PREDICTED: Hylobates moloch chromosome unknown C19orf12 homolog (C19H19orf12), transcript... Hylobates moloch	Hylobates moloch	147	147	54%	3e-31	98.78%	3759	XM_032170324.1
<input checked="" type="checkbox"/> PREDICTED: Gorilla gorilla gorilla chromosome 19 C19orf12 homolog (C19H19orf12), transcript varia... Gorilla gorilla g...	Gorilla gorilla g...	147	147	54%	3e-31	98.78%	3598	XM_031004299.1
<input checked="" type="checkbox"/> PREDICTED: Gorilla gorilla gorilla chromosome 19 C19orf12 homolog (C19H19orf12), transcript varia... Gorilla gorilla g...	Gorilla gorilla g...	147	147	54%	3e-31	98.78%	3768	XM_031004297.1
<input checked="" type="checkbox"/> PREDICTED: Nomascus leucogenys chromosome 10 C19orf12 homolog (C10H19orf12), transcript v... Nomascus leuc...	Nomascus leuc...	147	147	54%	3e-31	98.78%	3757	XM_030820282.1
<input checked="" type="checkbox"/> PREDICTED: Pongo abelii chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant X9....	Pongo abelii	141	141	54%	1e-29	97.56%	2017	XM_009253108.2
<input checked="" type="checkbox"/> PREDICTED: Pongo abelii chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant X2....	Pongo abelii	141	141	54%	1e-29	97.56%	2187	XM_009253103.2
<input checked="" type="checkbox"/> PREDICTED: Trachypithecus francoisi chromosome unknown C19orf12 homolog (C19H19orf12), tra... Trachypithecus...	Trachypithecus...	130	130	54%	3e-26	95.12%	1637	XM_033225574.1
<input checked="" type="checkbox"/> PREDICTED: Theropithecus oeladai chromosome 19 C19orf12 homolog (C19H19orf12), transcript va... Theropithecus...	Theropithecus...	130	130	54%	3e-26	95.12%	1532	XM_025368859.1
<input checked="" type="checkbox"/> PREDICTED: Cercocebus atys chromosome unknown open reading frame_human C19orf12 (LOC10... Cercocebus atys	Cercocebus atys	130	130	54%	3e-26	95.12%	1610	XM_012043383.1
<input checked="" type="checkbox"/> PREDICTED: Papio anubis chromosome 20 C19orf12 homolog (C20H19orf12), transcript variant X5... Papio anubis	Papio anubis	124	124	54%	1e-24	93.90%	1604	XM_009194058.4
<input checked="" type="checkbox"/> PREDICTED: Macaca mulatta chromosome 19 C19orf12 homolog (C19H19orf12), transcript variant X... Macaca mulatta	Macaca mulatta	124	124	54%	1e-24	93.90%	2485	XM_015123472.2
<input checked="" type="checkbox"/> PREDICTED: Macaca nemestrina chromosome unknown C19orf12 homolog (C19H19orf12), transcri... Macaca nemes...	Macaca nemes...	124	124	54%	1e-24	93.90%	1628	XM_011751162.2

@A00129:562:HWNFWDSXX:1:1103:3875:24486  
CATGCCTACGCCACTCTGGGTGACAGCGGACGAGCGCCGGGACTCTAGTCCAGCTCTGCTGCTACTGTGTGACTTCAGCCTCTCCATGCCTCAGTTCTCCATAGGGACAAT  
GACTACTAACAGTGCCACCCCGGCCCGGG  
+  
FF

Figure S5a: Read from Homo Sapiens from SRX7756767.



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<sup>1,2</sup> and Hinh Ly<sup>2,\*</sup>

[8] Evolutionary origins of the SARS - CoV - 2sarbecovirus lineage responsible for the COVID-19 pandemic Maciej F Boni<sup>1\*</sup> , Philippe Lemey<sup>2\*</sup> , Xiaowei Jiang<sup>3</sup>, Tommy Tsan-Yuk Lam<sup>4</sup>, Blair Perry<sup>5</sup>, Todd Castoe<sup>5</sup>, Andrew Rambaut<sup>6</sup> and David L Robertson<sup>7</sup>

[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*, , nwa036, <https://doi.org/10.1093/nsr/nwaa036>

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

Alexandre Hassanin<sup>1</sup> Huw Jones<sup>2</sup> Anne Ropiquet<sup>2</sup>

[11] Boratyn, G.M., Thierry-Mieg, J., Thierry-Mieg, D. *et al.* Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* **20**, 405 (2019).

<https://doi.org/10.1186/s12859-019-2996-x>

- [12] Jing-Yang Hu, Zi-Qian Hao, Laurent Frantz, Shi-Fang Wu, Wu Chen, Yun-Fang Jiang, Hong Wu, Wei-Min Kuang, Haipeng Li, Ya-Ping Zhang, Li Yu, Genomic consequences of population decline in critically endangered pangolins and their demographic histories, *National Science Review*, Volume 7, Issue 4, April 2020, Pages 798–814, <https://doi.org/10.1093/nsr/nwaa031>
- [13] Lee, J., Hughes, T., Lee, MH. *et al.* No Evidence of Coronaviruses or Other Potentially Zoonotic Viruses in Sunda pangolins (*Manis javanica*) Entering the Wildlife Trade via Malaysia. *EcoHealth* **17**, 406–418 (2020). <https://doi.org/10.1007/s10393-020-01503-x>
- [14] Hua L, Gong S, Wang F, Li W, Ge Y, Li X, Hou F. Captive breeding of pangolins: current status, problems and future prospects. *Zookeys*. 2015 Jun 8;(507):99-114. doi: 10.3897/zookeys.507.6970. PMID: 26155072; PMCID: PMC4490220.
- [15] Wen-Hua Gao, Xian-Dan Lin, Yan-Mei Chen, Chun-Gang Xie, Zhi-Zhou Tan, Jia-Jun Zhou, Shuai Chen, Edward C Holmes, Yong-Zhen Zhang, Newly identified viral genomes in pangolins with fatal disease, *Virus Evolution*, Volume 6, Issue 1, January 2020, veaa020, <https://doi.org/10.1093/ve/veaa020>
- [16] Dinghua Li, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, Tak-Wah Lam, MEGAHit: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct *de Bruijn* graph, *Bioinformatics*, Volume 31, Issue 10, 15 May 2015, Pages 1674–1676, <https://doi.org/10.1093/bioinformatics/btv033>
- [17] Langmead B, Wilks C, Antonescu V, Charles R. [\*\*Scaling read aligners to hundreds of threads on general-purpose processors.\*\*](#) *Bioinformatics* Vol 35, Iss 3, 2019, pp 421–432.
- [18] ["A549 Cell Line: Human alveolar adenocarcinoma cell line -General Information"](#). Retrieved 3 January 2012. <http://a549.com/>
- [19] Ge, Xingyi & Li, Jia-Lu & Yang, Xinglou & Chmura, Aleksei & Zhu, Guangjian & H. Epstein, Jonathan & Mazet, Jonna & Hu, Ben & Zhang, Wei & Peng, Cheng & Zhang, Yunzhi & Luo, Chu-Ming & Tan, Bing & Wang, Ning & Zhu, Yan & Cramer, Gary & Zhang, Shu-Yi & Wang, Lin-Fa & Daszak, Peter & Shi, Zhengli. (2013). Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature*. 503. 10.1038/nature12711.
- [20] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012 Mar 4;9(4):357-9. doi: 10.1038/nmeth.1923. PMID: 22388286; PMCID: PMC3322381.
- [21] Boratyn, G.M., Thierry-Mieg, J., Thierry-Mieg, D. *et al.* Magic-BLAST, an accurate RNA-seq aligner for long and short reads. *BMC Bioinformatics* 20, 405 (2019). <https://doi.org/10.1186/s12859-019-2996-x>