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?
<u>SRX6893158</u>	16,491,648	

SRX6893157	9,275,501	Lung12 [3] SRR10168374
<u>SRX6893156</u>	22,220,187	Lung11 [1]
SRX6893155	18,067,615	Lung09 [1] [3] SRR10168376
<u>SRX6893154</u>	16,414,925	Lung08 [1] [3] [4]
		SRR10168377
<u>SRX6893153</u>	19,045,923	Lung07 [1] [3] [4]
		SRR10168378
<u>SRX6893152</u>	13,527,964	
<u>SRX6893151</u>	16,068,654	
<u>SRX6893150</u>	12,967,281	
<u>SRX6893149</u>	12,590,769	
<u>SRX6893148</u>	15,273,939	
<u>SRX6893147</u>	15,975,904	
<u>SRX6893146</u>	19,038,817	
<u>SRX6893145</u>	19,055,973	
SRX6893144	15,350,468	
<u>SRX6893143</u>	11,527,782	
SRX6893142	20,045,443	
<u>SRX6893141</u>	18,903,834	
SRX6893140	19,986,780	
<u>SRX6893139</u>	39,738,679	Lung02 [3] SRR10168392
<u>SRX6893138</u>	22,900,426	
<u>SRX7756769</u>	107,267,359 PRJNA607174**	M1[2]***
<u>SRX7756766</u>	273,651,431 PRJNA607174**	
<u>SRX7756765</u>	196,761,202 PRJNA607174**	
SRX7756764	222,286,763 PRJNA607174**	
<u>SRX7756763</u>	212,161,250 PRJNA607174**	
<u>SRX7756762</u>	232,433,120 PRJNA607174**	M6[2]***
<u>SRX7756761</u>	113,900,941 PRJNA607174**	
SRX7732094	2,633*	"P2S"[3]

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

**: All available SRA datasets from PRJNA607174

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

Article

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

				_
Sample ID	Animal species	Total reads*	No. mapped	_
M1	Malayan pangolin	107,267,359	496 ←S	RX7756769 "pangolin 9"
M2	Malayan pangolin	38,091,846	302	
МЗ	Malayan pangolin	79,477,358	14	
M4	Malayan pangolin	32,829,850	1,100	Not available
M5	Malayan pangolin	547,302,862	56	
M6	Malayan pangolin	232,433,120	¹⁰ ←S	RX7756762 "pangolin 2"
M8	Malayan pangolin	44,440,374	12	
M10	Malayan pangolin	227,801,882	0 N	ot available
Z1	Chinese pangolin	444,573,526	ο	_

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.

Accession number and	Primary Mammalian	Primate-related results	Identification of	
registration date	Trace results and	in Krona and read size	"Coronaviridae"	
	percentage	by Кbp	as by Trace and	
			total read size	
<u>SRX6893158</u>	Manis javanica: 14.66%	N/D	N/D	
20-Sep-2019				
<u>SRX6893157</u>	Boreoeutheria: 1.24%	Catarrhini 644546	N/D***	
20-Sep-2019				
SRX6893156	Manis javanica: 7.51%	Homo sapiens 81948	Pangolin	
20-Sep-2019	Homo sapiens: 0.03%		coronavirus 2Kbp	
<u>SRX6893155</u>	Homo sapiens: 0.37%	Homininae 3534150	Pangolin	
20-Sep-2019			coronavirus 5Kbp	
<u>SRX6893154</u>	Homo sapiens: 0.02%	Hominoidea 356003	Pangolin	
20-Sep-2019			coronavirus	
			154Kbp	
<u>SRX6893153</u>	Homo sapiens: 0.01%	Homo sapiens 162180	Pangolin	
20-Sep-2019			coronavirus	
			41Kbp	

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

		N/D
-	ט/א	N/D
-	N/D	
Wanis Javanica: 7.47%	N/D	N/D
Paragoutharia: 1 01%	N/D	N/D
Boreoeutnena: 1.91%	N/D	N/D
Mania invenine, 10/	Cinciliferra es 212000	
ivianis javanica: 1%	Similitormes 313069	N/D
Mania invanian 0 49/	Catarrhini 104220	
IVIANIS JAVANICA: U.4%	Catarmini 194320	N/D
Mania invaniana 2 719/	Catambini (0027	
	Catarmini 69937	N/D
Deressutheria: 1 73%	Hominoidos 2217EE	N/D
Boreoeutheria. 1.72%	HOITIITIOIUEd 231755	N/D
Hominingo: 0 27%	Hominingo 2526765	N/D
	Hominoidon 166679	N/D
ividilis javaliled. U.02 %	TOTILIOUUEd 100028	
Manis javanica: 1 62%	N/D	N/D
ivialiis javalilta. 1.03%		
Manis invanisa: 1 20%	Similformos E7094	N/D
iviallis javalilla. 1.20%	5111110111125 57064	
Boreceutheria: 1 /1%	N/D	N/D
Boreocutileria. 1.41%		
Boreceutheria: 1 56%	N/D	N/D
Borcocutiena. 1.30 /0		
Homo saniens: 0.01%	Homo saniens 491120	Pangolin
		coronavirus 2Kbp
Boreceutheria: 1.67%	Homininae 2761176	N/D
		,-
Homo sapiens: 0.03%	Homo sapiens 5457929	Bat SARS-like
		coronavirus 2Kbp
		Wuhan seafood
		market
		pneumonia virus
		2Kbp
Manis javanica: 78.6%	Cercopithecidae 3116	Betacoronavirus
		2Kbp**
Manis javanica: 87.17%	Cercopithecinae 11339	N/D****
Manis javanica: 48.39%	Cercopithecidae 22600	N/D
Manis javanica: 94.95%	Cercopithecidae 5076	N/D
	Manis javanica: 87.17% Manis javanica: 48.39%	Euarchontoglires: 1.37%Manis javanica: 7.47%N/DBoreoeutheria: 1.91%N/DManis javanica: 1%Simiiformes 313069Manis javanica: 0.4%Catarrhini 194320Manis javanica: 2.71%Catarrhini 69937Boreoeutheria: 1.72%Hominoidea 231755Homininae: 0.27% Manis javanica: 1.01%Hominoidea 166628Manis javanica: 1.63%N/DManis javanica: 1.63%N/DManis javanica: 1.63%N/DBoreoeutheria: 1.41%N/DBoreoeutheria: 1.56%N/DBoreoeutheria: 1.67%Homo sapiens 491120Boreoeutheria: 1.67%Homininae 2761176Homo sapiens: 0.03%Homo sapiens 5457929Manis javanica: 78.6%Cercopithecidae 3116Manis javanica: 48.39%Cercopithecidae 22600

SRX7756762	Manis javanica: 95.37%	Catarrhini* 2831	Nidovirales OKbp
18-Feb-2020			
<u>SRX7756761</u>	Manis javanica: 13.63%	Chlorocebus sabaeus	N/D
18-Feb-2020		498506	
SRX7732094	N/A***	N/A	Pangolin
15-Feb-2020			coronavirus***

*: Chlorocebus Sabaeus

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

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

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

Specific BLAST analysis

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

Select all 100 sequences selected							hics Distance tree of results
	Description	Max Score		Query Cover	E value	Per. Ident	Accession
	<u>SRX7756762</u>	279	1047	0%	4e-68	100.00%	SRA:SRR11119766.160125840.2
	<u>SRX7756762</u>	279	1366	0%	4e-68	100.00%	SRA:SRR11119766.138036805.1
	<u>SRX7756762</u>	279	967	0%	4e-68	100.00%	SRA:SRR11119766.101239747.1
\checkmark	SRX7756762	279	1624	0%	4e-68	100.00%	SRA:SRR11119766.46413326.2
Chlo	rocebus sabaeus isolate 1994-021 unplaced genomic sca						
dna							
1339	488						
<u>Dista</u>	nce tree of results MSA viewer 😮						

Fig.2a Specific BLAST analysis on the PRJNA607174 dataset, <u>SRX7756762</u>, 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		<u>Gen</u> B	<u>3ank</u>	<u>Grap</u>	<u>hics</u>	Distance	tree of results
Description			Total Score	Query Cover	E value	Per. Ident	Accession
Macaca mulatta isolate Rh22777_5890-1b major histocompatibility complex genomic sequence		279	279	100%	2e-71	100.00%	KT332833.1
Macaca mulatta isolate Rh22335_5775-3 major histocompatibility complex genomic sequence		279	279	100%	2e-71	100.00%	KT332608.1
Macaca mulatta isolate Rh22335_5725-2 major histocompatibility complex genomic sequence		279	279	100%	2e-71	100.00%	KT332521.1
Macaca mulatta isolate Rh22335_6702-1a major histocompatibility complex genomic sequence		279	279	100%	2e-71	100.00%	KT332463.1
IGATCCTTGAGGAATCGCCATACTGTTTTCCATAATGGTTGAACTAGTTTACAATCCCAC	230	230	1	00% 2	0.56	94.04%	GQ879596.1
	230	230		00% 2			KF661301.1
	219	219					AB553834.1
	204	204					<u>AK171052.1</u>
	204	204					<u>AK149653.1</u>
	202	202	1	00% 4	le-48	90.73%	LN899823.1
Canis lupus familiaris breed Labrador retriever chromosome 06a	154	3044	4 9	98% 1	le-33	86.43%	CP050586.1
Canis lupus familiaris breed Labrador retriever chromosome 04a	154	4569) 1	00% 1	le-33	85.23%	CP050572.1

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

	select all 100	sequences selected					Grap	hics Distance tree of results
		Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.269072261.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.255768440.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.255768440.1
	<u>SRX7756766</u>		279	279	0%	5e-67	100.00%	SRA:SRR11119762.255318754.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.254520929.1
	SRX7756766		279	6344	0%	5e-67	100.00%	SRA:SRR11119762.251645135.1
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.234036838.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.211208832.2
	SRX7756766		279	9108	0%	5e-67	100.00%	SRA:SRR11119762.199583624.1
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.198110623.2
	SRX7756766		279	279	0%	5e-67	100.00%	SRA:SRR11119762.196936636.2
	<u>SRX7756766</u>		279	279	0%	5e-67	100.00%	SRA:SRR11119762.196936636.1
	<u>SRX7756766</u>		279	279	0%	5e-67	100.00%	SRA:SRR11119762.133631622.2
	<u>SRX7756766</u>		279	279	0%	5e-67	100.00%	SRA:SRR11119762.108819247.2
Des	cription	Macaca mulatta isolate AG07107 chromosome 3 genomic sca \ldots						
Mol	ecule type	dna						
Que	ery Length	17855752						
Oth	er reports	Distance tree of results MSA viewer 😯						

Fig.3a Specific BLAST analysis of <u>SRX7756766</u> revealed large amount of 100% full-length matches with Macaca Mulatta.

1	Macaca mulatta Major Histocompatibility Complex BAC MMU370002, complete sequence	279	279	100%	2e-71	100.00%	AC148706
	Macaca mulatta Major Histocompatibility Complex BAC MMU122H23, complete sequence	279	521	100%	2e-71	100.00%	AC148677
	Macaca mulatta Major Histocompatibility Complex BAC MMU065H09, complete sequence	279	279	100%	2e-71	100.00%	AC148671
	Macaca mulatta Major Histocompatibility Complex BAC MMU038L02, complete sequence	279	279	100%	2e-71	100.00%	AC148668
	Papio anubis clone rp41-22m16, complete sequence	279	554	100%	2e-71	100.00%	AC113268
	Papio anubis clone rp41-280n2, complete sequence	279	465	100%	2e-71	100.00%	AC091778
	Papio anubis clone rp41-5m22, complete sequence	279	279	100%	2e-71	100.00%	AC136143
	Papio anubis clone rp41-192i11, complete sequence	279	526	100%	2e-71	100.00%	AC091671
	Macaca mulatta chromosome 9 CH250-18D2, complete sequence	279	548	100%	2e-71	100.00%	CT573219
	Rhesus Macaque CHR4 BAC CH250-23P16 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Librar	279	279	100%	2e-71	100.00%	AC169807
	Rhesus Macaque CHR4 BAC CH250-476F18 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Libra	279	279	100%	2e-71	100.00%	AC171646
	Chlorocebus aethiops BAC clone CH252-163P9 from chromosome 5, complete sequence	278	556	99%	7e-71	100.00%	AC239684
	MACACA MULATTA BAC clone CH250-192J17 from chromosome unknown, complete sequence	278	552	100%	7e-71	100.00%	AC21569
	Macaca mulatta isolate Rh22335_5702-1a major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	KT33246
	Macaca mulatta isolate Rh9_6570-3 major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	KT33177
	Macaca mulatta isolate Rh9_6550-1b major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	<u>KT33173</u>
	Macaca mulatta isolate Rh9_6526-2 major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	<u>KT33167</u>
	Macaca mulatta isolate Rh18665_5547-1b major histocompatibility complex genomic sequence	274	274	100%	9e-70	99.34%	KT32950
	Macaca mulatta isolate Rh23717 clone 4777 major histocompatibility complex-B genomic sequence	274	274	100%	9e-70	99.34%	KJ913523
	Macaca mulatta isolate Rh23108 clone 4769-2 major histocompatibility complex-B genomic sequence	274	274	100%	9e-70	99.34%	KJ91342
1	Eukaryotic synthetic construct chromosome 18	2	02 4	103 96	% 4e-	48 91.789	6 <u>CP0344</u>
	Eukaryotic synthetic construct chromosome 19	1	96 2	020 96	% 2e-	46 92.09%	6 <u>CP0345</u>
1	Eukaryotic synthetic construct chromosome 19	1	96 2	130 96	% 2e-	46 92.09%	6 <u>CP0344</u>
	Eukaryotic synthetic construct chromosome 16	1	96 3	654 96	% 2e-	46 91.619	6 <u>CP0344</u>
	Eukaryotic synthetic construct chromosome 15	1	96 49	994 97	% 2e-	46 92.099	6 <u>CP0344</u>
	Eukaryotic synthetic construct chromosome 14	1	96 7	731 96	% 2e-	46 92.09%	6 <u>CP0344</u>
	Eukaryotic synthetic construct chromosome 13	1	91 7	988 97	% 9e-	45 90.91%	6 <u>CP0345</u>
	Eukaryotic synthetic construct chromosome Y	1	91 2	781 96	% 9e-	45 91.379	6 <u>CP0345</u>
	Eukaryotic synthetic construct chromosome 20	1	91 4	517 96	% 9e-	45 91.379	6 <u>CP0344</u>
ľ	Eukaryotic synthetic construct chromosome 13	1	91 7	988 97	% 9e-	45 90.919	6 <u>CP0344</u>
1	Eukaryotic synthetic construct chromosome 21	1	85 1	053 96	% 4e-	43 90.65%	6 <u>CP0345</u>
·	Eukaryotic synthetic construct chromosome 17	1	85 24	450 96	% 4e-	43 90.65%	6 <u>CP0344</u>
1	Eukaryotic synthetic construct chromosome 22	1	83 9	33 96	% 2e-	42 90.589	6 <u>CP0345</u>

Fig.3b BLASTing such matches gives 1005 matches to only Primates, and with no matches outside of Primates. This indicates that <u>SRX7756766</u> also contained significant amount of material derived from primates.

Results for	2:lcl Query_13045 gnl SRA SRR11119762.182596220.2 182596220 (Biologi 🗸	Organism Only top 20		
Program	BLASTN ? <u>Citation</u> ~		e, binomial, taxid or group	name
Database	nt <u>See details</u> ~	+ Add organism		
Query ID	lcl Query_13045	Percent Identity	E value	Query Coverage
Description	gnl SRA SRR11119762.182596220.2 182596220 (Biological)	to	to	to
Molecule type	dna			
Query Length	151			Filter Reset
Other reports	Distance tree of results 🔞			
Descriptions	Graphic Summary Alignments Taxonomy			

eq	uences producing significant alignments Download Y	Ма	inage Co	olumns	⊻ S	how 1	.000 🗸 🤇
	select all 0 sequences selected			<u>nk</u> <u>G</u>			
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	PREDICTED: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X3, mRNA	279	279	100%	2e-71	100.00%	XM_015150059
	PREDICTED: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X2, mRNA	279	279	100%	2e-71	100.00%	XM_01515005
	PREDICTED: Macaca mulatta LIM domain kinase 2 (LIMK2), transcript variant X1, mRNA	279	279	100%	2e-71	100.00%	XM_01515005
	Papio anubis anubis NIPSNAP1 protein (NIPSNAP1).gene_partial cds; and merlin (NF2) and S-100/ICaBP-type calcium binding domain	279	5181	100%	2e-71	100.00%	AH012454.2
	PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X3, mRNA	279	279	100%	2e-71	100.00%	XM_01545731
	PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X2, mRNA	279	279	100%	2e-71	100.00%	XM_00556755
	PREDICTED: Macaca fascicularis LIM domain kinase 2 (LIMK2), transcript variant X1, mRNA	279	279	100%	2e-71	100.00%	XM_00556755
	Chlorocebus aethiops BAC clone CH252-146112 from chromosome 6, complete sequence	279	3318	100%	2e-71	100.00%	AC241602.2
	Chlorocebus aethiops BAC clone CH252-138D20 from chromosome 13, complete sequence	279	681	100%	2e-71	100.00%	AC239463.3
	Macaca mulatta BAC CH250-74N24 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC Library) complete	279	6774	100%	2e-71	100.00%	AC204493.6
	Chlorocebus aethiops BAC clone CH252-257N12 from chromosome 6, complete sequence	274	6168	100%	9e-70	99.34%	AC241575.3
	Chlorocebus aethiops BAC clone CH252-518J9 from chromosome 6, complete sequence	274	4718	100%	9e-70	99.34%	AC241458.2
	Chlorocebus aethiops BAC clone CH252-124F12 from chromosome 6, complete sequence	274	2252	100%	9e-70	99.34%	AC241509.3
	Chlorocebus aethiops BAC clone CH252-175F15 from chromosome 6, complete sequence	274	4253	100%	9e-70	99.34%	AC241845.3
	Chlorocebus aethiops BAC clone CH252-57021 from chromosome 6, complete sequence	274	2432	100%	9e-70	99.34%	AC241469.3

Fig.3C Presence of Primate-derived mRNA reads in <u>SRX7756766</u> confirms the Primate origin of these reads.

	select all 100 sequences selected						Grap	hics Distance tree of results
	Description	Ma Sco			Query Cover	E value	Per. Ident	Accession
	SRX7756769	27	8 2	78	0%	9e-69	100.00%	SRA:SRR11119759.99831231.2
	SRX7756769	27	8 2	78	0%	9e-69	100.00%	SRA:SRR11119759.99831231.1
	SRX7756769	27	8 4	814	1%	9e-69	100.00%	SRA:SRR11119759.88019245.2
	SRX7756769	27	85	178	2%	9e-69	100.00%	SRA:SRR11119759.82130976.2
	SRX7756769	27	8 2	78	0%	9e-69	100.00%	SRA:SRR11119759.70689253.2
	SRX7756769	27	8 2	78	0%	9e-69	100.00%	SRA:SRR11119759.70689253.1
	SRX7756769	27	8 2	78	0%	9e-69	100.00%	SRA:SRR11119759.57405658.2
≤	SRX7756769	27	8 2	78	0%	9e-69	100.00%	SRA:SRR11119759.57405658.1
AC	073210.8							
Но	mo sapiens BAC clone RP11-460N20 from 7, complete seq							
nu	icleic acid							
20	3396							

Fig.4a Similarly, <u>SRX7756769</u> contained large amount of reads that are 100% full-length matches to Human genomic DNA.

select all 0	sequences selected				i <u>k</u> <u>Gr</u>			
	Description	Ma Sci		Total Score	Query Cover	E value	Per. Ident	Accession
Homo sapien	s chromosome 22 clone ABC11_000047178300_E22, complete sequence	27	78	456	100%	6e-71	100.00%	AC279316.1
Homo sapien:	s actin related protein 2 pseudogene (LOC284441) on chromosome 19	27	78	278	100%	6e-71	100.00%	NG_022927.2
Homo sapiens	s TBC1 domain containing kinase (TBCK), RefSeqGene on chromosome 4	27	78	2140	100%	6e-71	100.00%	NG_034057.3
Homo sapiens	s chromosome 15 clone VMRC59-280106, complete sequence	27	78	2291	100%	6e-71	100.00%	AC279072.1
Homo sapiens	s chromosome 2 clone VMRC59-389K09, complete sequence	27	78	3905	100%	6e-71	100.00%	AC279037.1
Homo sapien:	s chromosome 15 clone VMRC59-359A02, complete sequence	27	78	3589	100%	6e-71	100.00%	AC278991.1
Homo sapien:	s chromosome 16 clone VMRC59-453B14, complete sequence	27	78	2239	100%	6e-71	100.00%	AC278975.1
Description	gnl SRA SRR11119759.88019245.2 88019245 (Biological)							
Molecule type	dna							
Query Length	150							
Other reports	Distance tree of results MSA viewer 🔞							

iery ID	lcl Query_61480			Percent Identity	E	value				Query Co	verage	
scription	gnl SRA SRR11119759.706	89253.1 70689253 (Biological)	to			to				to]
lecule type	dna											
ery Length	150									Filte	r Res	et
her reports	Distance tree of results											
Descriptions	Graphic Summary	Alignments	Taxonomy									
Sequences p	producing significant a	lignments		Dov	vnload 🗠	Ма	anage	Colum	ins ~	Show	1000 🗸	0
🛃 select all	41 sequences selected						<u>GenB</u>	ank .	Graphic	<u>cs Dista</u>	ince tree of re	esults
		Descr	iption			Max Score		Query Cover	E value	Per. Ident	Accessio	on
	D: Pan paniscus endogenous r	etrovirus group 3 mem	ber 1, envelope (ER	<u>3-1), mRNA</u>		278	278	100%	7e-71	100.00%	XM 0349642	2 <u>53.1</u>
Homo sapi	iens endogenous retrovirus grou	<u>ip 3 member 1, envelo</u>	pe (ERV3-1), transci	pt variant 1, mRNA		278	278	100%	7e-71	100.00%	NM 001007	253.4
	D: Pan troglodytes endogenous	retrovirus group 3 me	mber 1, envelope (E	V3-1), transcript variant X2, mR	RNA	278	278	100%	7e-71	100.00%	XM_0243575	564.1
	D: Pan troglodytes endogenous	retrovirus group 3 me	<u>mber 1, envelope (E</u>	(V3-1), transcript variant X1, mR	RNA	278	278	100%	7e-71	100.00%	XM 0169567	774.2
Synthetic co	onstruct DNA, clone: pFN21AE12	261, Homo sapiens EF	RV3 gene for endoge	nous retroviral sequence 3, with	nout stop code	278	278	100%	7e-71	100.00%	AB590999.1	
Homo sapi	iens cDNA FLJ60255 complete o	<u>ods, highly similar to H</u>	ERV-R_7q21.2 provi	us ancestral Env polyprotein pr	ecursor	278	278	100%	7e-71	100.00%	<u>AK295189.1</u>	
Pan trogloc	dytes BAC clone CH251-623C19	from chromosome 7.	complete sequence			278	278	100%	7e-71	100.00%	AC184799.2	
Pan trogloc	dytes BAC clone CH251-565C10	from chromosome 7,	complete sequence			278	278	100%	7e-71	100.00%	AC148313.3	
Pan trogloc	<u>dytes endogenous retrovirus gro</u>	up 3 member 1, envelo	<u>ope (ERV3-1), mRN/</u>			278	278	100%	7e-71	100.00%	NM_001135	588.1
Homo sapi	iens BAC clone RP11-460N20 fr	om 7, complete seque	nce			278	278	100%	7e-71	100.00%	AC073210.8	
Homo sapi	iens cDNA FLJ23884 fis, clone L	NG13819, highly simi	lar to Human endog	nous retrovirus ERV3, pol-env-3	<u>3'LTR region</u>	278	278	100%	7e-71	100.00%	<u>AK074464.1</u>	
Human end	dogenous retrovirus ERV3, pol-e	nv-3'LTR region				278	278	100%	7e-71	100.00%	<u>M12140.1</u>	
Nomascus	leucogenys endogenous retrov	irus group 3 member [.]	<u>1, envelope (ERV3-1</u>	mRNA		272	272	100%	3e-69	99.33%	<u>NM 001308</u>	194.1
Gibbon end	dogenous retrovirus proviral envi	R gene for R envelope	protein			272	272	100%	3e-69	99.33%	AJ862653.1	
	D: Piliocolobus tephrosceles en	idogenous retrovirus g	roup 3 member 1, e	velope (ERV3-1), transcript vari	iant X1, mRN/	267	267	100%	2e-67	98.67%	XM 0231962	289.2
	D: Papio anubis endogenous re	trovirus group 3 mem	ber 1. envelope (ER\	3-1), transcript variant X2, mRN	A	267	267	100%	2e-67	98.67%	XM 0316620	056.1
	D: Papio anubis endogenous re	trovirus group 3 mem	ber 1. envelope (ER\	3-1), transcript variant X1, mRN	Δ	267	267	100%	2e-67	98.67%	<u>XM_0179566</u>	<u>680.3</u>
	D: Rhinopithecus roxellana end	ogenous retrovirus gro	up 3 member 1, env	lope (ERV3-1), transcript varia	nt X2, mRNA	267	267	100%	2e-67	98.67%	XM 0309320	630.1
	D: Rhinopithecus roxellana end	ogenous retrovirus are	un 3 member 1 env	lone (ERV3-1) transcript varia	nt X1 mRNA	267	267	100%	2e-67	98 67%	XM 0309326	200.4

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 (Biologi	al)	to	to		to	
Molecule type	dna						
Query Length	150				•	ilter	Reset
Other reports	Distance tree of results						
Descriptions	Graphic Summary Alignments Tax	onomy					
Sequences p	roducing significant alignments		Download 🗡	Manage Colum	ns 🔨 Sho	ow 100	0 🗸 🛛
select all	l sequences selected			<u>GenBank</u>	Graphics I	Distance t	ree of results
	Description				tuery E over value	Per. Ident	Accession
Synthetic c	onstruct DNA, clone: pFN21AE1261, Homo sapiens ERV3 gen	for endogenous retroviral	sequence 3, without stop codon, in	278 278 1	00% 7e-71	100.00%	AB590999.1

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

•	select all 100) sequences selected					Graph	ics Distance tree of results
		Description		Total Score		E value	Per. Ident	Accession
	SRX6893156		278	278	0%	2e-69	100.00%	SRA:SRR10168375.5045789.1
<	SRX6893156		278	278	0%	2e-69	100.00%	SRA:SRR10168375.5964.1
Desc	cription	Homo sapiens BAC clone RP11-460N20 from 7, complete seq						
Mole	ecule type	nucleic acid						
Que	ry Length	203396						
Othe	er reports	Distance tree of results MSA viewer 🔞						

Fig.5a <u>SRX6893156</u> also returned 100% matched results from the human Genome.

	14 sequences selected		<u>c</u>	ienBank	<u>Grap</u>	hics	Distance	tree of result
	Description		Ma Sco		Query Cover	E value	Per. Ident	Accession
Homo sapie	ens BAC clone RP11-460N20 from 7, complete sequence		27	8 278	100%	6e-71	100.00%	AC073210.8
Pan troglod	tytes BAC clone CH251-623C19 from chromosome 7, complete sequence		26	7 267	100%	1e-67	98.67%	AC184799.2
Pan troglod	tytes BAC clone CH251-2015 from chromosome 7, complete sequence		26	7 267	100%	1e-67	98.67%	AC174000.3
Pan troglod	tytes BAC clone CH251-565C10 from chromosome 7, complete sequence		26	7 267	100%	1e-67	98.67%	AC148313.3
escription	gnl SRA SRR10168375.5045789.1 5045789 (Biological)							
olecule type	dna							
uery Length	150							
ther reports	Distance tree of results MSA viewer 😮							
Description	gnl SRA SRR10168375.5964.1 5964 (Biological)	to		to			to	
Molecule type	e dna							
Query Length	150	-					Filter	Reset
0 ah an man a mb						_		
Other reports	s <u>Distance tree of results</u>							
Descriptio	ns Graphic Summary Alignments Taxonomy	Developed of	Ma		lumne	~ 6	'how -	1000
Descriptio		Download 🗡	Ma	inage Co	lumns	~ S	ihow 🗌	1000 🗸
Description	ns Graphic Summary Alignments Taxonomy	Download 🗸	Ma	inage Co <u>GenBan</u>		→ S		1000 🗸
Description	ns Graphic Summary Alignments Taxonomy es producing significant alignments	Download 🗡	Мах	GenBan Total Qi	<u>k Gra</u>	<u>phics</u>		
Description Sequence	ns Graphic Summary Alignments Taxonomy es producing significant alignments all 2000 sequences selected	Download 🗡	Мах	GenBan Total Qi Score Ci	<u>k Gra</u> Jery	ephics E	Distanc Per. Ident	e tree of resu
Description Sequence	ns Graphic Summary Alignments Taxonomy es producing significant alignments all 1000 sequences selected Description	Download 🗡	Max Score	GenBan Total Qu Score Cu 278 10	<u>k Gra</u> uery over va 00% 7¢	ephics E ilue	Distanc Per. Ident	e tree of resu
Description Sequence Select a Homo Pantra	ns Graphic Summary Alignments Taxonomy es producing significant alignments all 1000 sequences selected Description usagiens IncAB572.1 IncRNA gene.comolete sequences	Download 🗸	Max Score 278	GenBan Total Qu Score Cu 278 10 278 10	k Gra uery over va 00% 7¢	ephics E ilue e-71 10 e-71 10	Distance Per. Ident 00.00% 00.00%	Accession
Description Sequence select a Homo Pan tre Pan tre Pan tre	ns Graphic Summary Alignments Taxonomy es producing significant alignments all 1000 sequences selected Uescription usapiens IncAB572.1 IncRNA gene.comolete sequence apploytes chromosome 2 clone CH251-60P06, complete sequence	Download 🗡	Max Score 278 278	GenBan Total Qr Score Cr 278 10 278 10 278 10 278 10	k <u>Gra</u> uery va over va 00% 7¢ 00% 7¢	ephics E ilue -71 10 -71 10	Distance Per. Ident 00.00% M 00.00% A 00.00% A	e tree of resu Accession IK280613.1 IC279084.1

Fig.5b BLAST search on the result returned 100% match only found in humans, confirming origin

in human-derived material.

scription	gnl SRA SRR10168375.596	4.1 5964 (Biologica)	to			to			to	
lecule type	dna										
ery Length	150									Filter	Reset
her reports	Distance tree of results										
Descriptions	Graphic Summary	Alignments	Taxonomy								
Sequences	producing significant a	lignments		Dow	nload 🗠	Mana	ge Colu	imns	∽ Sh	ow 100	• 00
select all	14 sequences selected					Ge	nBank	Grap	<u>hics</u>	Distance t	ree of resu
		Des	cription			Max Score	Total Score	Query Cover	E value	Per. Ident	Accessio
Eukaryotic	synthetic construct chromosome	<u>e 21</u>				278	7938	100%	7e-71	100.00%	CP03450
Eukaryotic	synthetic construct chromosome	<u>e 13</u>				267	14570	100%	1e-67	98.67%	CP03451
 Eukaryotic 	synthetic construct chromosome	<u>ə 16</u>				267	10333	100%	1e-67	98.67%	CP03449
Eukaryotic	synthetic construct chromosome	<u>e 15</u>				267	9021	100%	1e-67	98.67%	CP03449
Eukaryotic	synthetic construct chromosome	<u>e 13</u>				267	14570	100%	1e-67	98.67%	CP03449
Eukaryotic	synthetic construct chromosome	<u>e 18</u>				261	15047	100%	7e-66	98.00%	CP03449
Eukaryotic	synthetic construct chromosome	<u>e 17</u>				261	6545	100%	7e-66	98.00%	CP03449
Eukaryotic	synthetic construct chromosome	<u>e 20</u>				219	7949	98%	4e-53	93.79%	CP03449
Eukaryotic	synthetic construct chromosome	<u>e 19</u>				209	3521	96%	3e-50	93.10%	CP03452
Eukaryotic	synthetic construct chromosome	<u>e 19</u>				209	3766	96%	3e-50	93.10%	CP03449
Eukaryotic	synthetic construct chromosome	<u>a 22</u>				207	2291	96%	9e-50	92.47%	CP03450
Eukaryotic	synthetic construct chromosome	<u>e 14</u>				207	13851	96%	9e-50	92.47%	CP03449
Gossypiur	m hirsutum clone NBRI_GE2709	3 microsatellite seque	nce			189	189	96%	3e-44	90.41%	<u>JX591845</u>
_	synthetic construct chromosome					180	2399	96%		89.44%	CP03451

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

	select all	100 sequences selected					Graph	nics Distance tree of results
		Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
	SRX689315	i	278	278	0%	2e-69	100.00%	SRA:SRR10168376.17339580.1
	SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.17013625.2
	SRX689315	2	278	278	0%	2e-69	100.00%	SRA:SRR10168376.17013625.1
	SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.16930714.2
	SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.16930714.1
	SRX689315	2	278	278	0%	2e-69	100.00%	SRA:SRR10168376.15267479.2
	SRX689315	2	278	278	0%	2e-69	100.00%	SRA:SRR10168376.15267479.1
	SRX689315	2	278	278	0%	2e-69	100.00%	SRA:SRR10168376.13985702.2
	SRX689315	2	278	278	0%	2e-69	100.00%	SRA:SRR10168376.13985702.1
	SRX689315	2	278	278	0%	2e-69	100.00%	SRA:SRR10168376.13353823.2
	SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.13353823.1
	SRX689315		278	278	0%	2e-69	100.00%	SRA:SRR10168376.11109740.1
	SRX689315	1	278	278	0%	2e-69	100.00%	SRA:SRR10168376.9343845.2
	SRX689315	2	278	278	0%	2e-69	100.00%	SRA:SRR10168376.9232549.2
Descri	ption	Homo sapiens BAC clone RP11-460N20 from 7, complete seq \ldots						
Molecu	ule type	nucleic acid						
Query	Length	203396						
Other	reports	Distance tree of results MSA viewer 😮						

Fig.6a Similarly, BLAST research on <u>SRX6893155</u> gives large number of full length 100% matches to the human genome.

~	select all 57 se	quences selected					<u>GenBa</u>	<u>nk G</u>	raphics	<u>Distan</u>	ice tree of results
			Descrip	otion		Max Score		Query Cover	E value	Per. Ident	Accession
	Homo sapiens F	OSMID clone ABC13-48840700E	E15 from chromosome	7, complete sequence		278	278	100%	6e-71	100.00%	AC242196.4
	Pan troglodytes	BAC clone CH251-340l24 from c	hromosome 7, complet	te sequence		278	278	100%	6e-71	100.00%	AC185242.2
	Pan troglodytes	BAC clone CH251-623C19 from	chromosome 7, comple	ete sequence		278	278	100%	6e-71	100.00%	AC184799.2
	Pan troglodytes	BAC clone CH251-114G16 from	chromosome 7, comple	ete sequence		278	278	100%	6e-71	100.00%	AC183835.2
	Pan troglodytes	BAC clone CH251-2O15 from ch	iromosome 7 <u>, complete</u>	e sequence		278	278	100%	6e-71	100.00%	AC174000.3
	<u>Homo sapiens B</u>	AC clone RP11-479O9 from 7_c	omplete sequence			278	278	100%	6e-71	100.00%	AC073107.7
	Pan troglodytes	BAC clone CH251-565C10 from	chromosome 7, comple	ete sequence		278	278	100%	6e-71	100.00%	AC148313.3
	<u>Homo sapiens B</u>	AC clone RP11-460N20 from 7.	complete sequence			278	278	100%	6e-71	100.00%	AC073210.8
	PREDICTED: Ce	bus capucinus imitator small int	<u>egral membrane protei</u>	n 11A (SMIM11A), transcrip	<u>st variant X6, mRNA</u>	87.9	87.9	49%	1e-13	88.00%	XM_017526193.1
D	escription	gnl SRA SRR10168376.15	5267479.2 1526747	9 (Biological)							
M	olecule type	dna									
	uery Length	150									
0	ther reports	Distance tree of results	MSA viewer 🔞								
De	scription	gnl SRA SRR10168376.139	85702 1 13985702 (Biological)	to		to				
	lecule type	dna		biotogicati			10				0
	ery Length	150								Filter	Reset
_	her reports	Distance tree of results)								
	Descriptions	Graphic Summary	Alignments	Taxonomy							
		Graphic Summary	-	Taxonomy	Download 🗠	Ма	nage C	olumn	s Ý S	Show	1000 🗸 🚱
	Sequences p		-	Taxonomy	Download 🗡	Ма	nage C		s ⊻ s		1000 V ?
	Sequences p	roducing significant a	-		Download 🗠	Мах	<u>GenBa</u> Total (n <u>k G</u> i Query	raphics E		
	Sequences p	roducing significant a	lignments Descr	iption	Download 🗡	Мах	GenBar Total (Score (n <u>k Gi</u> Query Cover v	raphics E value	Distan Per. Ident	ce tree of results
	Sequences p Select all & Homo sapid	roducing significant a	lignments Descr	iption	Download 🗡	Max Score	GenBar Total (Score (278	n <u>k Gr</u> Query Cover v 100% 7	raphics E value 7e-71 1	Distant Per. Ident 00.00%	ce tree of results Accession
	Sequences p select all 6 Homo sapir Pan troglod	roducing significant a	lignments Descr om 7. complete seque	iption ince complete sequence	Download 🗸	Max Score 278	GenBar Total (Score (278 - 272 -	nk Gr Duery Cover V 100% 7	Te-71 1 Be-69	Distant Per. Ident 00.00% (99.33% (Accession
	Sequences p Select all a Homo sapir Pan troglod Pan troglod	roducing significant al 17 sequences selected ans BAC clone RP11-460N20 fm yles BAC done CH251-623C19 yles BAC done CH251-565C10	lignments Descr om 7. complete seque from chromosome 7. from chromosome 7.	tption ince complete sequence complete sequence	Download 🗸	Max Score 278 272 272	GenBar Total (Score (278 - 272 - 272 -	nk Gr Query V Cover V 100% 7 100% 3	E value 7e-71 1 3e-69 9 3e-69 9	Distant Per. Ident 00.00% (99.33% (99.33% (Accession Accession AC073210.8 AC184799.2
	Sequences p Select all & Homo sapie Pan troglod Macaca mu Homo sapie	roducing significant al 17 sequences selected ans BAC clone RP11-460N20 fm yles BAC done CH251-623C19 yles BAC done CH251-565C10	lignments Descr om 7. complete seque from chromosome 7. from chromosome 7. en's Hospital Oakland	tption Ince complete sequence complete sequence I Research Institute Rhesu		Max Score 278 272 272	GenBar Total (Score (278 - 272 - 272 - 217 -	nk Gr Duery Cover V 100% 3 100% 3 100% 2	E Palue 7e-71 1 3e-69 9 3e-69 9 2e-52 9	Distant Per. Ident 00.00% (99.33% (99	Accession AC073210.8 AC184799.2 AC148313.3
	Sequences p select all a select all a Pan troglod Anacaca mu Homo sapir Homo sapir Homo sapir	roducing significant al 7 sequences selected ans BAC clone RP11-460N20 fr des BAC clone CH251-623C19 des BAC clone CH251-565C10 lata BAC CH250-206B5 (Childr ans tripartite motif containing 24 ans chromosome 7 clone RP11	lignments Descr om 7. complete seque from chromosome 7. tfrom chromosome 7. ren's Hospital Oakland L (TRIM24), RefSeqGer -199L 18. complete se	iption ince complete sequence complete sequence I Research Institute Rhesu he on chromosome 7 guence		Max Score 278 272 272 272 217	GenBar Total 0 Score 0 278 2 272 2 272 2 217 2 217 1 182 1	nk Gi Duery Cover v 100% 3 100% 3 100% 3 99% 6	E Palue 7e-71 1 3e-69 9 2e-52 9 5e-42 8	Distant Per. Ident 99.33% (99.33% (99.33% (93.359% (38.59% (Accession Accession Ac073210.8 Ac184799.2 Ac148313.3 Ac210125.6 NG_023286.1 Ac013429.12
	Sequences p select all a select all a Anno sabi Anno sabi Anno sabi Homo sabi Anno	roducing significant al 7 sequences selected ens BAC clone RP11-460N20 fr des BAC clone CH251-623C19 des BAC clone CH251-565C10 lata BAC CH250-206B6 (Childr ens tripartite motif containing 24 ens chromosome 7 clone RP11 ens chromosome 7 clone RP11	lignments Descr om 7. complete seque from chromosome 7. thom chromosome 7. en's Hospital Oakland L(TRIM24). RefSeqGer -1991.18. complete se -256C24. complete se	iption ince complete sequence complete sequence I Research Institute Rhesu he on chromosome 7 guence		Max Score 278 272 272 272 217 182 182 182	GenBar Total (Score (278 - 272 - 272 - 217 - 182 182 182 182	nk Gr Duery V Cover V 100% 3 100% 3 100% 3 99% 6 99% 6	E Control 7e-71 1 3e-69 9 2e-52 9 3e-64 1 3e-65 1 3e-69 1 3e-62 1 3e-42 1	Distant Per. Ident 1dent 00.00% (39.33% ((39.33% ((38.59% ((38.59% ((Accession Accession Ac073210.8 Ac184799.2 Ac148313.3 Ac210125.6 NG.023286.1 Ac013429.12 Ac008265.15
	Sequences p select all a Pan troplod Pan troplod Macaca mul Homo sapir Homo sapir Homo sapir	roducing significant al 7 sequences selected ans BAC clone RP11-460N20 fr des BAC clone CH251-623C19 des BAC clone CH251-565C10 lata BAC CH250-206B5 (Childr ans tripartite motif containing 24 ans chromosome 7 clone RP11	lignments Descr om 7. complete seque from chromosome 7. thom chromosome 7. en's Hospital Oakland L(TRIM24). RefSeqGer -1991.18. complete se -256C24. complete se	iption ince complete sequence complete sequence I Research Institute Rhesu he on chromosome 7 guence		Max Score 278 272 272 217 182 182	GenBar Total (Score (278 - 272 - 272 - 217 - 182 182 182 182	nk Gr Duery V Cover V 100% 3 100% 3 100% 3 99% 6 99% 6	E Control 7e-71 1 3e-69 9 2e-52 9 3e-64 1 3e-65 1 3e-69 1 3e-62 1 3e-42 1	Distant Per. Ident 1dent 00.00% (39.33% ((39.33% ((38.59% ((38.59% ((Accession Accession Ac073210.8 Ac184799.2 Ac148313.3 Ac210125.6 NG_023286.1 Ac013429.12
	Sequences p select all a Pan troplod Pan troplod Macaca mul Homo sapir Homo sapir Homo sapir	roducing significant al 7 sequences selected ans BAC clone RP11-460N20 fr tes BAC clone CH251-622C19 tes BAC clone CH251-652C10 atta BAC CH250-206B8 (Children ans tribartite molf containing 24 ans chromosome 7 clone RP11 ans chromosome 7 clone RP11 synthetic construct chromosome	lignments Descr om 7. complete seque from chromosome 7. thom chromosome 7. en's Hospital Oakland L(TRIM24). RefSeqGer -1991.18. complete se -256C24. complete se	iption ince complete sequence complete sequence I Research Institute Rhesu he on chromosome 7 guence		Max Score 278 272 272 272 217 182 182 182	GenBar Total (Score (278 - 272 - 272 - 217 - 182 182 182 182	nk Gr Query V Cover V 100% 3 100% 3 100% 3 99% 6 99% 6 99% 6 99% 6	E Control 7e-71 1 3e-69 9 2e-52 9 3e-64 1 3e-65 1 3e-69 1 3e-62 1 3e-42 1	Distant Per. Ident 1dent 00.00% 2 99.33% 2 3 99.33% 2 3 99.33% 3 3 <th>Accession Accession Ac073210.8 Ac184799.2 Ac148313.3 Ac210125.6 NG.023286.1 Ac013429.12 Ac008265.15</th>	Accession Accession Ac073210.8 Ac184799.2 Ac148313.3 Ac210125.6 NG.023286.1 Ac013429.12 Ac008265.15
	Sequences p select all a Homo sabit Pan troglod Aacca mul Homo sabit Homo sabit Homo sabit Lomo sabit Lom	roducing significant al 7 sequences selected ans BAC clone RP11-460N20 fr tes BAC clone CH251-523C19 tes BAC clone CH251-565C10 atta BAC CH250-206B8 (Children ans tribartite molf containing 24 ans chromosome 7 clone RP11 ans chromosome 7 clone RP11 synthetic construct chromosome	lignments Descr om 7. complete seque from chromosome 7. trom chromosome 7. en's Hospital Oakland L(TRIM24), RefSeqGer -1991.18, complete se -256C24, complete se LY	iption ince complete sequence complete sequence I Research Institute Rhesu he on chromosome 7 guence		Max Score 278 272 272 272 272 272 182 182 182 182 176	GenBar Total (Score (278 - 272 - 272 - 217 - 182 - 182 - 353 - GenBar - Max To	nk Gi Query V Cover V 100% 1 100% 1 100% 2 99% 6 999% 6 999% 2 999% 2 nk Gi nk Gi	E Final Arrow 7e-71 1 3e-69 9 2e-52 9 5e-42 8 5e-42 8 3e-40 8 3e-40 8	Distant Per. Ident Ident 00.00% () 99.33% () () 92.67% () () 88.59% () () 88.59% () () 87.92% () () Distant E Per	Accession Accession
	Sequences p Select all a Pan troolod Pan troolod Macaca mul Homo sapit Homo sapit Homo sapit Select all 3 seq	roducing significant al 7 sequences selected ans BAC clone RP11-460N20 fr tes BAC clone CH251-523C19 tes BAC clone CH251-565C10 atta BAC CH250-206B8 (Children ans tribartite molf containing 24 ans chromosome 7 clone RP11 ans chromosome 7 clone RP11 synthetic construct chromosome	lignments Descr om 7. complete seque from chromosome 7. trom chromosome 7. en's Hospital Oakland L(TRIM24), RefSeqGer -1991.18, complete se -256C24, complete se LY	iption ince complete sequence complete sequence I Research Institute Rhesu- ne on chromosome 7 quence counce		Max Score 278 272 272 217 182 182 182 176	GenBar Total 0 Score 0 278 272 272 272 217 1 182 182 182 353 GenBar Transport Max Transport Score Score	nk Gr Query V Cover V 100% 1 1100% 2 999% 6 999% 6 999% 2 nk Gr nk Gr nk Gr core C	E Image: Constraint of the state of the sta	Distant Per. Ident Ident 00.00% 99.33% 2 92.67% 2 88.59% 2 88.59% 2 97.32% 2 100.00% 2<	Accession Accession Accor3210.8 Ac184799.2 Ac184799.2 Ac14813.3 Ac210125.6 NG.023286.1 Ac008265.15 CP034510.1
	Sequences p Select all & Homo saple Pan troglod Pan troglod Macaca mul Homo saple Homo saple Eukaryotic s	roducing significant al 7 sequences selected ans BAC done RP11-460N20 fm tes BAC done CH251-623C19 tes BAC done CH251-656C10 atta BAC CH260-206B8 (Child ens thoradite motif containing 24 ens chromosome 7 done RP11 ans chromosome 7 done RP11 withelic construct chromosome uences selected	lignments Descr om 7. complete seque from chromosome 7. from chromosom	iption ince complete sequence complete sequence I Research Institute Rhesu- ne on chromosome 7 quence counce		Max Score 278 272 272 272 182 182 182 182 176	GenBar Total (278 272 272 272 272 272 217 182 182 182 353 GenBar Max Tracore Score Score	nk Gr Query V Cover V 100% 7 1100% 2 999% 6 999% 6 999% 6 999% 6 999% 6 100% 2 100% 2 999% 6 100% 2 100%	E E 7e-71 1 38-69 9 38-69 9 38-69 9 38-640 8 38-40 8 asa-40 8 asa-40 8 asa-40 8 asa-40 8 asa-40 8 asa-40 8	Distant Per. Ident 1dent 00.00% 0 99.33% 0 0 99.33% 0 0 98.59% 1 0 88.59% 1 0 88.59% 1 0 97.67% 0 0 98.59% 1 0 98.59% 1 0 98.59% 1 0 98.59% 1 0 98.59% 1 0 98.59% 1 0 99.33% 1 0 99.33% 1 0 99.38 1 0 99.38 1 0 99.38 1 0 99.38 1 0 99.38 1 0 99.38 1 0 99.39 1 0 99.39 1 0 99.39 1 0 <th>Accession Accession Accor3210.8 Ac184799.2 Ac1484799.2 Ac1484733.3 Ac210125.6 NG 023286.1 Ac008265.15 CP034510.1</th>	Accession Accession Accor3210.8 Ac184799.2 Ac1484799.2 Ac1484733.3 Ac210125.6 NG 023286.1 Ac008265.15 CP034510.1
	Sequences p Select all & Homo saple Pan troolod Pan troolod Macaca mui Homo saple Homo saple Eukaryotic synth Eukaryotic synth	roducing significant al 7 sequences selected Ins BAC clone RP11-460N20 fm 4es BAC done CH251-523C19 4es BAC done CH251-565C10 atta BAC CH250-206B6 (Childr ans tribaritle molf containing 24 ans chromosome 7 clone RP11 ans chromosome 7 clone RP11 withelic construct chromosome uences selected elic construct chromosome Y	lignments Descr om 7. complete seque from chromosome 7. from chromosom	iption ince complete sequence complete sequence I Research Institute Rhesu- ne on chromosome 7 quence counce		Max Score 278 272 272 1217 182 182 182 176	GenBan Total C 278 272 272 272 272 272 217 182 182 182 353 GenBan Max Tr Trocore Str 1176 3 1143 1	nk Gr Duery V Duery V 100% T 1100% T 100% T 100% T 999% E 999% E 999% E 11k Gr 12k Gr 153 9 43 9	E E 7e-71 1 38-69 9 38-69 9 56-42 4 56-42 4 38-69 9 36-40 4 7 7 7 7 7 7 8 7 9 7 9 3 9% 3 9% 3	Distant Per. Ident 1dent 00.00% 99.33% 2 99.33% 2 98.59% 2 88.59% 2 88.59% 2 97.92% 2 Distant 2 0.000 1 0.000	Accession Accession Accor3210.8 Ac184799.2 Ac1484799.2 Ac1484733.3 Ac210125.6 NG 023286.1 Ac008265.15 CP034510.1

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

Description	Homo sapiens BAC clone R	P11 460N20 from	omploto soci							
Description	Homo sapiens BAC Clone R	111-400/020 11011	unpiece seq	Percent Identity		E va	lue			Query Coverage
Molecule type	nucleic acid			to				to		to
Query Length	203396									
Other reports	Distance tree of results M	SA viewer 🔞								Filter Reset
Descriptions	Graphic Summary	Alignments								
Sequences p	producing significant al	lignments		Dow	nload	~	Mar	nage Co	olumns	✓ Show 100 ✓
🗹 select all	100 sequences selected								Graph	nics Distance tree of results
		Description			Max Score		Query Cover	E value	Per. Ident	Accession
SRX689315	3				278	278	0%	2e-69	100.00%	SRA:SRR10168378.1832954.1

Fig.7a <u>SRX6893153</u> have also returned 100% match full-length read on this tiny part of the human genome.

escription	gnl SRA SRR10168378.18	32954.1 1832954 (Bi	ological)	Percent Identity E va	alue			Ouerv	Covera	ge	
olecule type	dna			to		0			to		
uery Length	150										
ther reports	Distance tree of results	MSA viewer 🔞						Fi	ilter	Reset	:
Descriptions	Graphic Summary	Alignments	Taxonomy								
Sequences	producing significant a	alignments		Download 🗡	Manag	e Colui	nns ~	Sho	w 100	00 🗸	0
select all	170 sequences selected				Gen	Bank	Graph	ics D	istance t	tree of res	<u>ults</u>
		De	escription		Max Score		Query Cover	E value	Per. Ident	Accessi	on
Homo sap	piens FOSMID clone ABC18-86211	1 from chromosome 7, cr	omplete sequence		278	278	100%	6e-71	100.00%	AC245205	<u>5.1</u>
Homo sag	piens FOSMID clone ABC13-48840	1700E15 from chromoson	ne 7 <u>, complete sequence</u>		278	278	100%	6e-71	100.00%	AC242196	5.4
Homo sap	piens BAC clone RP11-460N20 from	<u>n 7. complete sequence</u>			278	278	100%	6e-71	100.00%	AC073210	<u>).8</u>
Pan troglo	odytes BAC clone CH251-487D11 f	<u>from chromosome 7, con</u>	<u>nplete sequence</u>		272	272	100%	3e-69	99.33%	AC182733	3.3
select all	7 sequences selected					Gen	<u>Bank</u>	Grap	hics <u>D</u>)istance t	ree of I
		C	Description			Max Score	Total Score	Query Cover	E value	Per. Ident	Acce
Eukaryotic	synthetic construct chromoson	ne 15				211	211	98%	7e-51	92.11%	CP034
Eukaryotic	synthetic construct chromoson	<u>ne 16</u>				206	1214	98%	3e-49	91.45%	CP034
Eukaryotic	synthetic construct chromoson	<u>ne 13</u>				200	200	98%	2e-47	90.79%	CP034
Eukaryotic	synthetic construct chromoson	ne 21				200	401	98%	2e-47	90.79%	CP034
Eukaryotic	synthetic construct chromoson	ne 13				200	200	98%	2e-47	90.79%	CP034
Eukarvotic	synthetic construct chromoson	ne 18				195	195	96%	7e-46	90.60%	CP034
Eukaryotic:											

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

Description	Homo sapiens BAC clone RP11-45003 from 7, complete sequence						
Molecule type	nucleic acid						
Query Length	195834						
Other reports	Distance tree of results MSA viewer 😨						
Descriptions	Graphic Summary Alignments						
Sequences p	producing significant alignments	Downle	oad ~	м	anage	Columns	s ~ Show 100 🗸
_ `	Droducing significant alignments 100 sequences selected	Downle	oad ~	М	anage		s × Show 100 ×
		Downle Max Score	Total	M Query Cover	E		
_ `	100 sequences selected Description	Мах	Total	Query	E	<u>Gra</u> Per.	phics Distance tree of res
select all	100 sequences selected Description 54	Max Score	Total Score	Query Cover	E value	Gra Per. Ident 100.00%	phics Distance tree of resi

Fig.8a Reads from the Human PMS1 gene is recovered from <u>SRX6893154</u> with a query sequence only 195834bp in length.

Descri	ption	gnl SRA SRR10168377.163	02266.1 16302266 (Biological)]
Moleci	ule type	dna							Filter	Rese	nt
Query	Length	150									
Other	reports	Distance tree of results)								
Des	criptions	Graphic Summary	Alignments	Taxonomy							
Sec	quences p	producing significant a	lignments		Download	⊻ Ma	nage Colı	ımns 🗸	Show	1000 🗸	0
	select all	311 sequences selected					GenBank	Graphics	Dista	nce tree of re	sults
							Total Que		Per.		
			Desc	iption		Score S		er value	Ident	Accessio	a
	<u>Homo sap</u>	<u>iens PMS1 homolog 2, mismatc</u>	<u>h repair system comp</u>	onent pseudogene 8	(PMS2P8) on chromosome 7	278	278 100	% 7e-71 1	00.00%	<u>NG_006447.3</u>	3
	Homo sap	iens PMS1 homolog 2, mismatc	h repair system comp	onent pseudogene 1	0 (PMS2P10) on chromosome 7	278	278 100	% 7e-71 1	00.00%	NG 023454.4	1
	<u>Homo sap</u>	iens PMS1 homolog 2, mismate	<u>h repair system comp</u>	onentipseudogene 6	(PMS2P6) on chromosome 7	278	278 100	% 7e-71 1	00.00%	NG 006449.3	ž
	<u>Homo sap</u>	iens BAC clone CH17-264B6 fro	<u>m chromosome 7, co</u>	<u>mplete sequence</u>		278	1112 100	% 7e-71 1	00.00%	AC211476.5	
	Homo sap	iens BAC clone CH17-220H16 fi	<u>om chromosome 7, c</u>	omplete sequence		278				AC211491.5	
	<u>Homo sap</u>	iens FOSMID clone ABC10-4551	5000F15 from chrom	osome 7, complete s	equence	278	556 100	% 7e-71 1	00.00%	AC244146.2	
	<u>Homo sapi</u>	ens BAC clone RP11-45003 fro	<u>im 7, complete sequi</u>	ence		27	8 556	100% 7e-	71 100.	00% <u>AC105</u> 4	<u>418.5</u>
~	<u>Homo sapi</u>	ens BAC clone RP11-396K3 fro	m 7, complete seque	nce		27	8 556	100% 7e-	71 100.	00% <u>AC006</u>	995.5
	<u>Homo sapi</u>	ens BAC clone RP11-313P13 fi	om 7, complete sequ	ience		27	8 556	100% 7e-	71 100.	00% <u>AC0054</u>	488.2
	<u>Homo sapi</u>	ens chromosome 7 clone VMR	C53-89F05, complete	sequence		27	2 272	100% 3e-	69 99.3	33% <u>AC2783</u>	<u>394.1</u>
~	<u>Homo sapi</u>	ens chromosome 7 clone VMR	C62-404M06, comple	te sequence		27	2 272	100% 3e-	69 99.3	33% <u>AC278</u>	<u>331.1</u>
~	<u>Homo sapi</u>	ens chromosome 7 clone CH1	7-61K4, complete se	quence		27	2 272	100% 3e-	69 99.3	33% <u>AC2706</u>	<u>699.1</u>
	Homo sapie	ens PMS8 mRNA (yeast mismat	<u>ch repair gene PMS1</u>	<u>homologue), partial</u>	cds (C-terminal region)	248	248 1	00% 6e-62	96.679	6 <u>D38503.1</u>	
	Homo sapie	ens PMS1 homolog 2, mismatcl	n repair system comp	onent (PMS2), RefS	eqGene (LRG_161) on chromosome 7	244	244 1	00% 7e-61	96.009	6 <u>NG 00846</u>	<u>36.1</u>
	Canis lupus	familiaris breed Labrador retrie	ver chromosome 06a			154	154 8	4% 1e-33	88.89%	CP050586.1	
~	Canis lupus	familiaris breed Labrador retrie	ver chromosome 06b			154	154 8	4% 1e-33	88.89%	CP050622.1	
	Pipistrellus	<u>pipistrellus genome assembly, c</u>	<u>chromosome: 5</u>			139	139 8	6% 3e-29	86.15%	LR862361.1	
	<u>Synthetic co</u>	nstruct Homo sapiens clone ccs	bBroadEn_14772 PM	<u>S2 gene, encodes c</u>	omplete protein	126	126 4	9% 3e-25	97.30%	KJ905275.1	
	PREDICTED): Zalophus californianus PMS1	homolog 2, mismatch	repair system com	ionent (PMS2), transcript variant X3, mR	<u>NA</u> 124	124 5	4% 9e-25	93.90%	XM_0276100	128.2
	PREDICTED): Desmodus rotundus PMS1 ho	molog 2, mismatch re	pair system compo	nent (PMS2), transcript variant X2, mRNA	121	121 5	2% 1e-23	94.87%	XM 0245769	<u>J31.1</u>
	PREDICTED): Desmodus rotundus PMS1 ho	molog 2, mismatch re	pair system compo	nent (PMS2), transcript variant X1, mRNA	121	121 5	2% 1e-23	94.87%	XM 0245769	<u>130.1</u>
	PREDICTED): Myotis lucifugus PMS1 homolo	<u>ig 2. mismatch repair</u>	system component	PMS2), transcript variant X4, mRNA	121	121 5	2% 1e-23	94.87%	XM_0237614	57.1
✓	PREDICTED): Myotis lucifugus PMS1 homolo	<u>ig 2, mismatch repair</u>	system component	PMS2), transcript variant X3, mRNA	121	121 5	2% 1e-23	94.87%	XM 0237614	<u>456.1</u>

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	Percent Identity		E va	lue			Query Coverage
Molecule type	nucleic acid		to				to		to
Query Length	173967								
Other reports	Distance tree of results	MSA viewer 😯							Filter Reset
Descriptions	Graphic Summary	Alignments							
Sequences p	producing significant	alignments	Dov	vnload	I ~	Mai	nage Co	olumns	✓ Show 100 ✔
🗹 select all	100 sequences selected							Graph	nics Distance tree of results
		Description		Max Score		Query Cover	E value	Per. Ident	Accession
SRX689313	<u>19</u>			278	278	0%	3e-69	100.00%	SRA:SRR10168392.39544030.1
SRX689313	<u>19</u>			278	278	0%	3e-69	100.00%	SRA:SRR10168392.28917809.1
SRX689313	9			278	278	0%	3e-69	100.00%	SRA:SRR10168392.14357888.1
SRX689313	<u>19</u>			278	278	0%	3e-69	100.00%	SRA:SRR10168392.2548655.2

Fig.9a similarly, multiple 100% match Full length reads were obtained from <u>SRX6893139</u>. 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)	Percent Identity E	value		Qu	lery Cov	erage
Molecule type	dna	to		to		1	o
Query Length	150						
Other reports	Distance tree of results MSA viewer 😨					Filter	Reset
Descriptions	Graphic Summary Alignments Taxonomy						
Sequences p	producing significant alignments	Download 🐣	Manag	ge Column	s Y	Show	1000 🗸 🔞
🗹 select all	66 sequences selected		Ge	nBank <u>G</u>	raphics	<u>Distan</u>	ce tree of result
	Description			otal Query core Cover	E value	Per. Ident	Accession
Homo sapie	ens zinc finger protein 316 (ZNF316), mRNA		278	278 100%	6e-71	100.00%	NM_001278559.2
	D: Homo sapiens zinc finger protein 316 (ZNF316), transcript variant X3, mRNA		278	278 100%	6e-71	100.00%	XM_024446619.1
	D: Homo sapiens zinc finger protein 316 (ZNF316), transcript variant X2, mRNA		278	278 100%	6e-71	100.00%	XM_024446618.1
	D: Homo sapiens zinc finger protein 316 (ZNF316), transcript variant X1, mRNA		278	278 100%	6e-71	100.00%	XM_006715630.4
Homo sapie	ens BAC clone RP11-611L7 from 7, complete sequence		278	278 100%	6e-71	100.00%	AC073343.6
PREDICTED	<u>D: Pongo abelii zinc finger protein 316 (ZNF316), mRNA</u>		272	272 100%	3e-69	99.33%	XM_024250011.1
PREDICTED: C	Prycteropus afer afer zinc finger protein 316 (ZNF316), mRNA		150 1	50 100%	2e-32	84.67%	XM 00794275
PREDICTED: M	liniopterus natalensis zinc finger protein 853 (ZNF853), mRNA		145 1	45 100%	7e-31	84.00%	XM_01621362
PREDICTED: 0	<u> Ochotona princeps zinc finger protein 316 (ZNF316), mRNA</u>		145 1	45 98%	7e-31	84.35%	<u>XM_01293099</u>
Pipistrellus pip	istrellus genome assembly, chromosome: 5		145 1	45 100%	7e-31	84.11%	LR862361.1

Fig.9b examining these reads revealed that they are only found in humans and apes. This is therefore also clear evidence that there is Human/Hominid-derived contamination in **SRX6893139**.

Description	Homo sapiens chromosor	ne 9, clone hRPK.20	2_H_3, complet	Percent Identity		E val	ue			Query Co	verage	
Molecule type	nucleic acid			to				to		to		
Query Length	187174											
Other reports	Distance tree of results	ISA viewer 🔞								Filter		leset
Descriptions	Graphic Summary	Alignments										
Sequences p	producing significant a	lignments		Down	nload	~	Man	age Co	lumns	Show	100 💊	•
🗹 select all	100 sequences selected								Graphi	<u>cs Dista</u>	nce tree c	of result
		Description			Max Score		Query Cover	E value	Per. Ident		Accession	
SRX689315	7				278	278	0%	8e-70	100.00%	SRA:SRR1	0168374.7	906491

Fig.10a one read is also recovered from <u>SRX6893157</u>, from a query sequence only 187174nt in length.

	PRED	DICTED: Homo sapiens formin binding.protein 1 (FNBP1), transcript variant X13, mRNA	278	278	100%	6e-71	100.00	% <u>XM</u>	005251824	1.2
	PRED	DICTED: Homo sapiens formin binding protein 1 (FNBP1), transcript variant X4, mRNA	278	278	100%	6e-71	100.00	% <u>XM</u>	011518402	<u>11</u>
	PREC	DICTED: Homo sagiens formin binding.protein 1 (FNBP1). transcript variant X3, mRNA	278	278	100%	6e-71	100.00	% <u>XM</u>	011518401	1
	Home	<u>sapiens formin binding protein 1 (FNBP1). RefSeqGene on chromosome 9</u>	278	278	100%	6e-71	100.00	% <u>NG</u>	033946.1	
	Home	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	% <u>AK</u>	<u>023681.1</u>	
	Huma	an DNA sequence from clone RP11-138E2 on chromosome 9q34.11-34.3, complete sequence	278	278	100%	6e-71	100.00	% <u>AL1</u>	136141.13	
	Home	sapiens formin-binding protein 17 (FBP17) mRNA_partial cds	278	278	100%	6e-71	100.00	% <u>AF</u> 2	265550.1	
Solution	Home	sapiens chromosome 9, clone hRPK.202_H_3, complete seguence	278	278	100%	6e-71	100.00	% <u>AC</u>	<u>006241.1</u>	
	Home	sapiens KIAA0554 mRNA for KIAA0554 protein	278	278	100%	6e-71	100.00	% <u>AB</u>	<u>011126.1</u>	
Sec.	PRED	DICTED: Nomascus leucogenys formin binding protein 1 (ENBP1), transcript variant X18, mRNA	272	272	100%	3e-69	99.339	6 <u>XM</u>	030818029	0.1
Image:	PREC	DICTED: Nomascus leucogenys formin binding protein 1 (ENBP1), transcript variant X17, mRNA	272	272	100%	3e-69	99.339	6 <u>XM</u>	030818028	1.1
	PREC	DICTED: Nomascus leucogenys formin binding protein 1 (ENBP1), transcript variant X16, mRNA	272	272	100%	3e-69	99.339	6 <u>XM</u>	030818027	11
Descripti	ion	gnl SRA SRR10168374.7906491.2 7906491 (Biological)								
Molecule	type	dna								
Query Le	ngth	150								
Other rep	ports	Distance tree of results MSA viewer 🔞								
Sciu	irus cai	rolinensis genome assembly, chromosome: 16		17	4 17	74 9	5% 9	e-40	88 81%	LR738606.1
				17						LR738604.1
		rolinensis genome assembly, chromosome: 14								
_		<u>garis genome assembly, chromosome: 15</u>		16						LR738626.1
	DICTE	<u>:D: Loxodonta africana formin binding protein 1 (FNBP1), transcript variant X2, mRNA</u>		13	5 13	35 7	4% 4	e-28	88.50%	XM_010587565.2
	DICTE	D: Loxodonta africana formin binding protein 1 (FNBP1), transcript variant X1, mRNA		13	5 13	35 7	4% 4	e-28	88.50%	XM_023544839.1

Fig.10b this particular sequence is only found in humans—indicating that even the <u>SRX6893157</u> dataset was contaminated by material of human origin.

escription	Chlorocebus aethiops BAC clone CH252-276C1 from chromos	Percent Identity	Ev	alue		Query Coverage
olecule type	nucleic acid	to		to		to
ery Length	160189					
her reports	Distance tree of results MSA viewer 2					Filter Reset
Descriptions	Graphic Summary Alignments					
Sequences	producing significant alignments	Downl	oad ~	Manage	Columns	∽ Show 100 ✔ 🤅
_	producing significant alignments 100 sequences selected	Downl	oad ~	Manage	Columns <u>Grap</u>	
_		Downl	x Total	Manage Query E Cover value	<u>Grar</u> Per.	
_	100 sequences selected Description	Ma	x Total re Score	Query E	Grag Per. Ident	phics Distance tree of result
Select all	100 sequences selected Description T65	Ma	x Total re Score 9 774	Query E Cover value 0% 4e-69	Grag Per. Ident 100.00%	Chics Distance tree of result
Select all	100 sequences selected Description	Mi Scc 21	x Total Score 9 774 9 375	Query E Cover value 0% 4e-69 0% 4e-69	Grag Per. Ident 100.00%	Accession SRA-SRR11119763.129105044 SRA-SRR11119763.106754018
Select all	100 sequences selected Description	Ma Sco 21 21	x Total re Score 9 774 9 375 9 279	Query Cover E 0% 4e-69 0% 4e-69 0% 4e-69	Grag Per. Ident 100.00% 100.00%	Chics Distance tree of result Accession SRA.SRR11119763.129105044.
 select all SRX775670 SRX775670 SRX775670 	100 sequences selected Description	Mi Scc 23 23 27 27	x Total re Score 9 774 9 375 9 279 9 510	Query Cover E 0% 4e-69 0% 4e-69 0% 4e-69 0% 4e-69 0% 4e-69	Gran Per. Ident 100.00% 100.00% 100.00%	hics Distance tree of result Accession SRA SRR11119763 129105044 SRA SRR11119763 106754018 SRA SRR11119763 75715814 1

Fig.11a The presence of Reads from Somatic Chlorocebus aethiops in <u>SRX7756765</u> confirms the identity of the Cercopithecinae reads there.

scription	gnl SRA SRR11119763.106754018.1106754018 (Biological) to		to				to
lecule type	dna						
ery Length	151					Filter	Reset
her reports	Distance tree of results 🔞						
Descriptions	Graphic Summary Alignments Taxonomy						
Sequences	oroducing significant alignments Download	~	Manage	Colum	ins ~	Show	1000 🗸
select all	1000 sequences selected		GenE	<u>Bank</u>	<u>Graphic</u>	<u>s Dista</u>	nce tree of res
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Chloroceb	us aethiops BAC clone CH252-276C1 from chromosome 6, complete sequence	279	375	100%	2e-71	100.00%	AC241496.3
_	us aethiops BAC clone CH252-445L10 from chromosome unknown, complete sequence	279	690	100%	2e-71	100.00%	AC238953.2
_	ulatta Y Chr BAC CH250-11J13 (Children's Hospital Oakland Research Institute Rhesus macaque Adult Male BAC I	-	1513		2e-71	100.00%	
_	us aethiops BAC clone CH252-371E9 from chromosome 5, complete sequence	279	1124	100%		100.00%	AC239563.3
_	us aethiops BAC clone CH252-461K13 from chromosome 13, complete sequence	279	629	100%		100.00%	AC239354.3
_	us aethiops BAC clone CH252-138D20 from chromosome 13, complete sequence	279	493	100%			AC239463.3
_	us aethiops BAC clone CH252-417L1 from chromosome 6, complete sequence us aethiops BAC clone CH252-62B18 from chromosome 4, complete sequence	279 279	472 510	100%		100.00%	AC239275.3 AC239442.2
	nthetic construct chromosome 18 Inthetic construct chromosome 16	268 268	1.545e+05 1.008e+05	100% 100%		100.00% 98.68%	CP034496.1 CP034494.1
	nthetic construct chromosome 16						
	nthetic construct chromosome 19	263	47303 1.935e+05	100% 100%		98.01% 99.31%	CP034522.1
	nthetic construct chromosome 13	263 263	1.935e+05		2e-66	99.31%	CP034516.1 CP034510.1
	nthetic construct chromosome Y nthetic construct chromosome 21	263	55262	100%		98.01%	CP034510.1 CP034500.1
	nthetic construct chromosome 20	263	88148			99.31%	01 00 1000.1
Eukarvotic si							CP034499.1
		263	47481		2e-66	99.31% 98.01%	<u>CP034499.1</u> <u>CP034497.1</u>
Eukaryotic s	Inthelic construct chromosome 19 Inthelic construct chromosome 17		47481 74732	100%			
Eukaryotic sy	nthetic construct chromosome 19	263		100% 100%	2e-66	98.01%	CP034497.1
Eukaryotic sy Eukaryotic sy Eukaryotic sy	nthetic construct chromosome 19 nthetic construct chromosome 17	263 263	74732	100% 100%	2e-66 2e-66	98.01% 98.01%	CP034497.1 CP034495.1
Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy	nthelic construct chromosome 19 nthelic construct chromosome 17 nthelic construct chromosome 15	263 263 263	74732 1.384e+05	100% 100% 100% 100%	2e-66 2e-66 2e-66	98.01% 98.01% 98.01%	CP034497.1 CP034495.1 CP034493.1
Eukaryotic su Eukaryotic su Eukaryotic su Eukaryotic su Eukaryotic su	nthelic construct chromosome 19 nthelic construct chromosome 17 nthelic construct chromosome 15 nthelic construct chromosome 14	263 263 263 263	74732 1.384e+05 1.910e+05	100% 100% 100% 100%	2e-66 2e-66 2e-66 2e-66	98.01% 98.01% 98.01% 99.31%	CP034497.1 CP034495.1 CP034493.1 CP034492.1
Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy	nthelic construct chromosome 19 nthelic construct chromosome 17 nthelic construct chromosome 15 nthelic construct chromosome 14 nthelic construct chromosome 13	263 263 263 263 263	74732 1.384e+05 1.910e+05 1.935e+05	100% 100% 100% 100% 100%	2e-66 2e-66 2e-66 2e-66 2e-66	98.01% 98.01% 98.01% 99.31% 99.31%	CP034497.1 CP034495.1 CP034493.1 CP034492.1 CP034491.1
Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy Human gam	nthelic construct chromosome 19 nthelic construct chromosome 17 nthelic construct chromosome 15 nthelic construct chromosome 14 nthelic construct chromosome 13 nthelic construct chromosome 22	263 263 263 263 263 263 257	74732 1.384e+05 1.910e+05 1.935e+05 22814	100% 100% 100% 100% 100% 100%	2e-66 2e-66 2e-66 2e-66 2e-66 9e-65	98.01% 98.01% 98.01% 99.31% 99.31% 97.35%	CP034497.1 CP034495.1 CP034493.1 CP034492.1 CP034491.1 CP034501.1
Eukaryotic si Eukaryotic si Eukaryotic si Eukaryotic si Eukaryotic si Human gam Human gam	nthelic construct chromosome 19 nthelic construct chromosome 17 nthelic construct chromosome 15 nthelic construct chromosome 14 nthelic construct chromosome 13 nthelic construct chromosome 22 maherpesvirus 4 isolate HKNPC60, partial genome	263 263 263 263 263 263 257 248	74732 1.384e+05 1.910e+05 1.935e+05 22814 248	100% 100% 100% 100% 100% 100%	2e-66 2e-66 2e-66 2e-66 9e-65 5e-62	98.01% 98.01% 98.01% 99.31% 99.31% 97.35% 96.03%	CP034497.1 CP034495.1 CP034493.1 CP034492.1 CP034491.1 CP034501.1 MH590571.1
Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy Eukaryotic sy Human gam Human gam Uncultured b	nthetic construct chromosome 19 nthetic construct chromosome 17 nthetic construct chromosome 15 nthetic construct chromosome 14 nthetic construct chromosome 13 nthetic construct chromosome 22 maherpesvirus 4 isolate HKNPC60, partial genome maherpesvirus 4 isolate HKNPC60, partial genome	263 263 263 263 263 263 257 248 248	74732 1.384e+05 1.910e+05 1.935e+05 22814 248 248	100% 100% 100% 100% 100% 100% 100%	2e-66 2e-66 2e-66 2e-66 9e-65 5e-62 5e-62	98.01% 98.01% 99.31% 99.31% 97.35% 96.03% 96.03%	CP034497.1 CP034495.1 CP034493.1 CP034492.1 CP034492.1 CP034491.1 CP034501.1 MH590571.1 MH590409.1

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

Analyzing the extent of contamination.

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

trusted for the analysis on the extent of contamination.

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

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

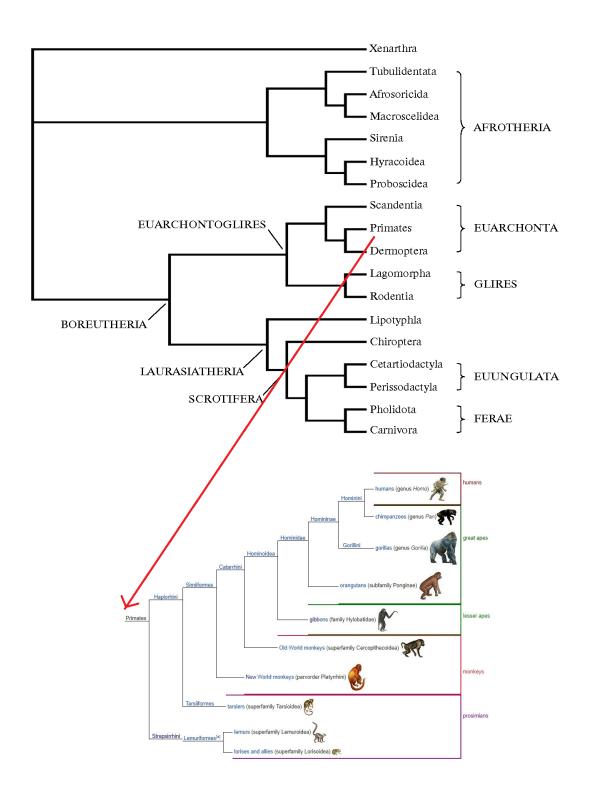


Fig. 12 Family tree of mammals, Including the position and classification of Primates in the lineage of Mammalia.

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

Accession and	Primate	Total traced Kbps	Ratio of	Virus
date	classification and	to Manis Javanica	Primate to	classification
	total traced Kbps	(Pangolin)	Pangolin	and amount of
				reads by Kbps
SRX7756769	Homo sapiens	15401134	0.35	Bat SARS-like
18-Feb-2020	5457929			coronavirus
				2Kbp
				Wuhan seafood
				market
				pneumonia
				virus 2Kbp
<u>SRX6893139</u>	Homo sapiens	5301351	0.0926	Pangolin
20-Sep-2019	491120			coronavirus
				2Kbp
<u>SRX6893157</u>	Catarrhini	1889448	0.34	N/D***
20-Sep-2019	644546			
SRX6893156	Homo sapiens	4765461	0.01719	Pangolin
20-Sep-2019	81948			coronavirus
				2Kbp
SRX6893155	Homininae	525801	6.7214	Pangolin
20-Sep-2019	3534150			coronavirus
				5Kbp
<u>SRX6893154</u>	Hominoidea	2232008	0.159	Pangolin
20-Sep-2019	356003			coronavirus
				154Kbp
<u>SRX6893153</u>	Homo sapiens	3110158	0.05214	Pangolin
20-Sep-2019	162180			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	Virus	Ratio of virus
	and reads (in Kbp)	classification and	reads to
		reads	Primate reads
<u>SRX7756766</u>	Cercopithecidae 3116;	Betacoronavirus	0.000642
18-Feb-2020	BLAST to Macaca	2Kbp **	
	Mulatta		
SRX7756762	Catarrhini 2831;	Nidovirales 0Kbp	0.000530
18-Feb-2020	BLAST to Chlorocebus	Claimed	
	sabaeus	10x150bp reads	
SRX7756765	Cercopithecinae 11339	N/D***	N/A
22-Apr-2020	BLAST to Chlorocebus		
	Aethiops		
SRX7732094	N/A*	Pangolin	N/A*
15-Feb-2020		coronavirus	

*: 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 <u>SRX7756769</u> and <u>SRX7756762</u> 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 <u>SRX7756762</u> dataset contained definitive mappings to Chlorocebus sabaeus, or African Green Monkey, while <u>SRX7756766</u> 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.

<u>SRX7756769</u> 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 <u>SRX7756769</u> itself support the possibility that <u>SRX7756769</u> 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 <u>SRX7756769</u> in VERO E6 cells, characterized it but did not sequence it, and this cell culture material then contaminated <u>SRX7756762</u> and possibly <u>SRX7756766</u>, resulting the 10 reads in <u>SRX7756762</u> and the 2Kb Batacoronavirus reads in <u>SRX7756766</u>.

The exact nature of <u>SRX7732094</u> 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 SRX7732094 15-Feb-2020	Size 2,633	Non-Coronavirus reads? No	Source Tissue Provider and Collected by N/A	Virus Designation: GD or GX? GD	Design 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

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

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 th	e 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.

	Mammarenavirus N	airoviridae	Murine respirovirus	Flaviviridae	Nidovirales	Rubulavirus	Nonanavirus	Peribunyavi	Amigovirus	Siphoviridae	Siphoviridae	Pahexavir
SRX6893158				No				No	Yes			No
SRX6893157	Yes Y		No	No	Claimed	No		Yes	No	No	No	No
SRX6893156	No	lo	Yes	Yes	Yes	No	No	No	Yes	No	No	Yes
SRX6893155	No	lo	Yes	No	Yes	No	No	No	No	No	No	No
SRX6893154	No	lo	Yes	No	Yes	No	No	No	No	No		No
SRX6893153	No	lo	Yes	Yes	Yes	No	No	No	Yes	No		No
SRX6893152	Yes Y	es	Yes	Yes	No	No	No	Yes	No	No		No
SRX6893151	Yes Y	es	No	Yes		No		Yes	Yes	No		No
SRX6893150	Yes Y		Yes	No		No		Yes	Yes	No		No
SRX6893149			No	No		No		No	No	No		No
SRX6893148	Yes Y	es	Yes	No	No	No	No	No	Yes	No	No	No
SRX6893147	Yes Y	les	"Respirovirus"	Yes	No	No	Yes	No	Yes	No	No	No
SRX6893146	Yes Y	les	Yes	No	No	No	No	No	No	No	No	No
SRX6893145	Yes Y	les	No	No			No	No	No	No		No
SRX6893144			Yes	Yes		No	No	No	No	No		No
SRX6893143	Yes Y		No	No		No		No	No	No		No
SRX6893142	Yes Y		No	No	No	No		Yes	Yes	No		No
			No	Yes		No	No	No	No	No	No	No
SRX6893140			Yes	No		No		Yes	No	No		No
SRX6893139			Yes	No		No	No	No	No	No		No
SRX6893138	Yes Y			Yes				Yes	Yes	No		No
SRX7756766				Yes		Yes	No	No	No	No		No
SRX7756765				No					No	No		No
			Yes	No		Yes			No	No		No
			Yes	No		Yes		No	No	No		No
SRX7756762	No	lo	Yes	No		Yes		No	No	No		No
SRX7756761	No N		Yes	No	No	Yes		No	No	No		No
SRX7756769	No	lo	Yes	Yes	Yes	No	No	No	No	No	No	No

Extended Data Table S1

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

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 Respirovirus 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

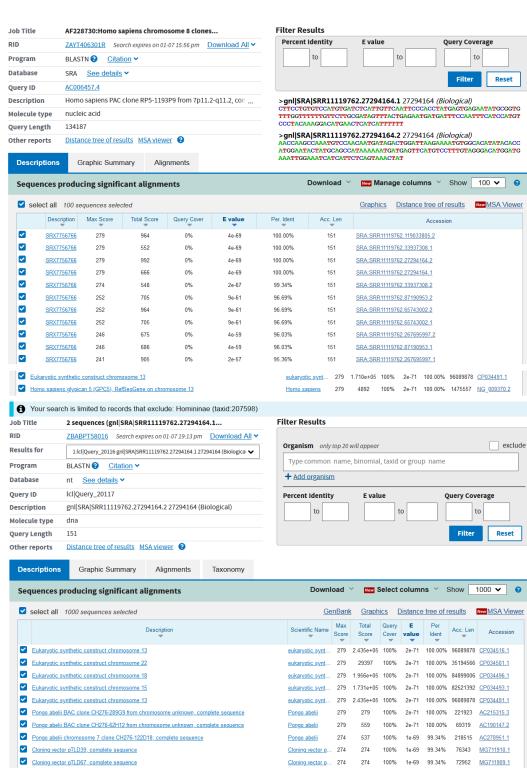
resembled the GD pangolin CoV sequence (which is only 90% similar to SARS-CoV-2) being isolated in humans anywhere in the world even till today.

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.



Eukaryotic synthetic construct chromosome 19

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

eukaryotic synt... 274 66262 100% 1e-69 99.34% 64242768 CP034522.1

Job Title	AC018449:Homo sapiens 3 BAC RP11-26G10 (Roswell	Filter Results
RID	ZDA099D901R Search expires on 01-08 13:19 pm Download All 🗸	Percent Identity
Program	BLASTN ? Citation ~	to
Database	SRA <u>See details</u> ✓	
Query ID	AC018449.11	
Description	Homo sapiens 3 BAC RP11-26G10 (Roswell Park Cancer Instit	>gnl SRA SRR111
Molecule type	nucleic acid	CTTGTGATTTTTGTA TAAGGAGATTTTGGG
Query Length	157426	ACAGGGACAATTTG/
Other reports	Distance tree of results MSA viewer 🔞	>gnl SRA SRR111 AGGAAATAAAGGGTA ACATGATTGTATATG
Descriptions	Graphic Summary Alignments	GCAACTTCAGCAAAO

Percent Identity	E value	Query C	overage	•
to	to		to	
		Filte	er	Reset

119763.171368621.1 171368621 (Biological) Acattgatittotatcctgaactittgctgaactitgcttatcaggt gctggagcaatggggtttctagatatacaatcatgtcgtctctgca acttcctcttttcctag

119763.171368621.2 171368621 (Biological) AttCAACTACGAAAGAGGAAGTCAAATTGTCCCTGTTGCAGACG CTAGAAAACCCCATTGTCTCAGCCCAAAATCTCCTTAAGCTGATAA GTTTCAGGATACAAAAT

Sequ	ences produ	cing signifi	cant alignme	ents			Download ~	New Manage columns Y Show	100 🗸 🔞
🔽 se	elect all 100 s	equences sele	cted					Graphics Distance tree of results	New MSA Viewer
	Description	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession	
	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA:SRR11119763.183685474.2	
	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA:SRR11119763.183685474.1	
	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA:SRR11119763.173512342.2	
	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA:SRR11119763.173512342.1	
	SRX7756765	279	1021	0%	4e-69	100.00%	151	SRA:SRR11119763.171368621.2	
	SRX7756765	279	1031	0%	4e-69	100.00%	151	SRA:SRR11119763.171368621.1	
	SRX7756765	279	992	0%	4e-69	100.00%	151	SRA:SRR11119763.152058533.2	
	SRX7756765	279	975	0%	4e-69	100.00%	151	SRA:SRR11119763.152058533.1	
	SRX7756765	279	959	0%	4e-69	100.00%	151	SRA:SRR11119763.136417907.2	
	SRX7756765	279	701	0%	4e-69	100.00%	151	SRA:SRR11119763.136417907.1	
	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA:SRR11119763.84993302.2	
	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA:SRR11119763.84993302.1	
	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA:SRR11119763.77679299.2	
	SRX7756765	279	279	0%	4e-69	100.00%	151	SRA:SRR11119763.77679299.1	

Job Title	2 sequences (gnl SRA SRR11119763.171368621.1
RID	ZDVKJ9V7013 Search expires on 01-08 18:20 pm Download All 🗸
Results for	1:lcl Query_8150 gnl SRA SRR11119763.171368621.1 171368621 (Biologic: 🗸
Program	BLASTN ? Citation ~
Database	nt <u>See details</u> ¥
Query ID	lcl 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	Resu	lts

Organism only top 20 w	vill appear		exclude
Type common name,	binomial, taxid or g	roup name	
+ Add organism			
Percent Identity	E value	Query Cover	age
to	to	to	
		Filter	Reset

Descriptions	Graphic Summary	Alignments	Taxonomy								
Sequences pro	oducing significant al	ignments		Down	oad ~	New S	Select o	column	is Y S	Show	1000 🗸 (
select all 1	000 sequences selected			Ge	nBank	Graph	ics D	istance	tree of r	<u>esults</u>	New MSA View
		Description		Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Homo sapiens	3 BAC RP11-26G10 (Roswell Pa	ark Cancer Institute Hu	man BAC Library) comp	lete segu Homo sapiens	279	1031	100%	2e-71	100.00%	157426	AC018449.11
Homo sapiens	chromosome 5 clone CTB-25P1	5. complete sequence		Homo sapiens	278	506	99%	8e-71	100.00%	174158	AC011399.6
Homo sapiens	adenosine kinase (ADK). RefSe	Gene on chromosome	10	Homo sapiens	272	732	99%	4e-69	99.33%	565119	NG_030484.2
Homo sapiens	neuroligin 1 (NLGN1), RefSeqGe	ne on chromosome 3		Homo sapiens	272	742	99%	4e-69	99.33%	905473	<u>NG_046919.2</u>
Pongo abelii B.	AC clone CH276-312H8 from chi	omosome 15, complet	e sequence	Pongo abelii	272	777	99%	4e-69	99.33%	196619	AC221006.3
Pongo abelii B.	AC clone CH276-98A16 from ch	omosome 1, complete	sequence	Pongo abelii	272	1032	99%	4e-69	99.33%	196248	AC205911.3
Pongo abelii cl	hromosome UNK clone CH276-2	2H12. complete seque	<u>nce</u>	Pongo abelii	272	272	99%	4e-69	99.33%	225183	AC186760.1

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

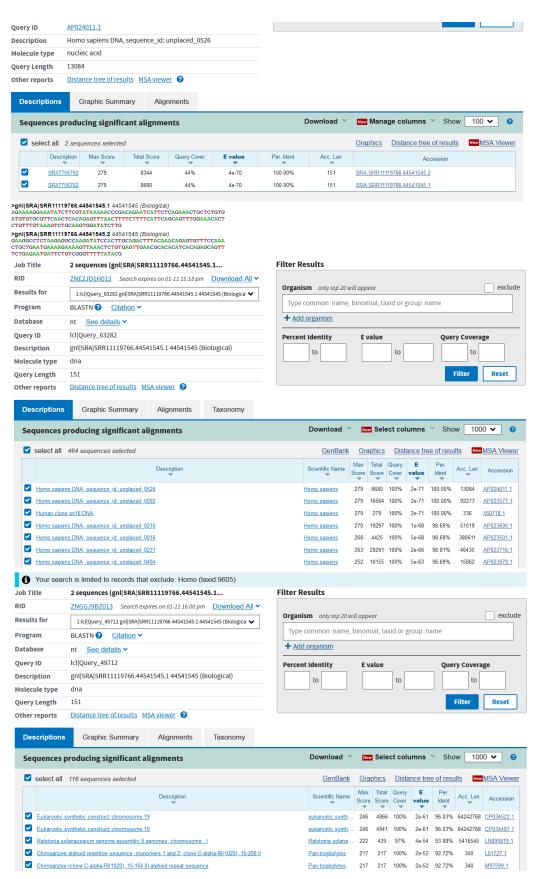


Fig.16: Reads uniquely matched to Homo Sapiens Alpha Satellite DNA in SRX7756762.

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 <u>Hu J *et al.*</u>,[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 <u>Hu J *et al.*</u>,[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 it's 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.

	Save Search S	Search Summary 🛩			this report?	BLAST Help Vide	os DBack to Traditio	nal Results Page	Job Title	MT121216:Pangoli	n coronavirus isolate M	iP789,
Job Title	MT121216:Pangolin	coronavirus isolate MP789,		Filter Result	ts				RID		arch expires on 01-20 19:46 j	pm Download All ~
RID		ch expires on 01-20 16:24 pm Downk	ad All Y	Percent Idea	ntity	E value	Query Cover	rage	Program			
Program	Citation ~			to		to	to		Database	SRA See details	*	
Database	SRA See details	2										
		-					Filter	Reset	SRA Blast se SRX5796768	arch set information	SRR9018610	
SRA Blast sea	urch set information								SRX5796767		SRR9018611	
SRX5796795		SRR9018583							SRX5796766 SRX5796765		SRR9018612 SRR9018613	
SRX5796794		SRR9018584							SRX5796764		SRR9018614	
Other reports	MSA viewer								SRX5796763		SRR9018615	
									SRX5796762 SRX5796761		SRR9018616 SRR9018617	
A No sign	nificant similarity foun	id. For reasons why, <u>click here</u>							360,3790701		3883010017	
Job Title	gb MT121216.1			Job Title	gb MT121	216.1						
RID		rch expires on 01-20 16:38 pm Downl	oad All V	RID			on 01-20 16:39 pm Do	wnload All 🗸	Job Title RID	gb[MT121216.1]	earch expires on 01-20 19.46	
Program	Citation ~			Program	O Cita				RID	Download All ~	earch expires on 01-20 19.46	(pm
Database	SRA See details	-		Database		e details ¥			Program			
									Database	SRA See details	~	
SRA Blast sea	urch set information			SRA Blast se	arch set info	ormation						
SRX5796792		SRR9018586		SRX5796788		SRR9018			SRA Blast se	arch set information	1	
SRX5796791 SRX5796790		SRR9018587 SRR9018588		SRX5796787		SRR9018	591		SRX5796760		SRR9018618	
SRX5796789		SRR9018589		Other reports	MSA viewe	er 🔞			SRX5796759 SRX5796758		SRR9018619 SRR9018620	
SRX5796793		SRR9018585							SRX5796757		SRR9018621	
A No sign	nificant similarity foun	id. For reasons why, click here		A No si	gnificant simi	ilarity found. For rea	asons why, click here		SRX5796756		SRR9018622	
		and a remain method		- C.					A No si	gnificant similarity fou	nd. For reasons why,clic	k here
Job Title		n coronavirus isolate MP789,		Job Title	gb MT1212	16.1						
RID	0DBGGGJM01R Se Download All ~	earch expires on 01-20 17:01 pm		RID	ODBHSJG80		on 01-20 17:02 pm Dov	vnload All	Job Title		n coronavirus isolate N	
Program	Citation ✓			Program	O Citati				RID		arch expires on 01-20 19:58	pm Download All ~
Database				Database	SRA See	details 🗸			Program			
Jarang26	SRA See details	<u></u>							Database	SRA See details	~	
SRA Blast see	arch set information	1		SRA Blast sea	irch set info				Ten			
SRX5796785		SRR9018593		SRX5796784 SRX5796783		SRR90185 SRR90185		-	SRA Blast se SRX5796755	arch set information	srr9018623	
SRX5796786		SRR9018592		any3146183		aRR90185			SRX5796755 SRX5796754		SRR9018623 SRR9018624	
Other reports	MSA viewer @			Other reports	MSA viewer	0			SRX5796753		SRR9018625	
									SRX5796752 SRX5796751		SRR9018626 SRR9018627	
A No sig	gnificant similarity fou	nd. For reasons why, click here		A No sig	nificant simil	arity found. For reas	sons why click here		SRX5796750		SRR9018628	
_									SRX5796749 SRX5796748		SRR9018629 SRR9018630	
Job Title	MT121216:Pangoli	n coronavirus isolate MP789,		Job Title	MT1212	16:Pangolin coron	avirus isolate MP789,		5605796746		51010010630	
RID	ODCB248401R Sea	urch expires on 01-20 17:16 pm Downl	load All ¥	RID	ODCB3A	EN01R Search expl	res on 01-20 17:16 pm	Download All	Job Title	MT121216:Pangoli	n coronavirus isolate N	4P789
Program				Program	0 <u>Ci</u>	tation ~			RID		arch expires on 01-20 19:58	
Database	SRA See details	~		Database	SRA S	See details ¥			Program			
									Database	SRA See details	~	
	arch set information				search set in					5104 000 001015	-	
SRX5796782 SRX5796781		SRR9018596 SRR9018597		SRX5796771 SRX5796771		SRR90 SRR90		_	SRA Blast se	arch set information	1	
SRX5796780		SRR9018598		380,019011		Skrau	10001		SRX5796747	aren set miorination	SRR9018631	
SRX5796779		SRR9018599		Other reports	MSA vie	wer 🔞			SRX5796746		SRR9018632	
									SRX5796744 SRX5796743		SRR9018634 SRR9018635	
A No sig	gnificant similarity four	nd. For reasons why, <u>click here</u>		A No	significant si	milarity found. For r	easons why, click here		SRX5796745		SRR9018633	
Job Title	MT121216-Pangoli	in coronavirus isolate MP789		Job Title	MT12121	16:Pangolin corona	virus isolate MP789,.		SRX5796742		SRR9018636	
RID		arch expires on 01-20 17:58 pm Down	V IA heal	RID			s on 01-20 18:23 pm		-			
Program	Citation ~	urch expires on 01-20 17:58 pm	iluau /vi +	Program		ation ¥			Job Title	MT121216:Pangoli	in coronavirus isolate N	dP789,
Database	SRA See details	-		Database		ee details ¥			RID	ODP9PPF501R Se	arch expires on 01-20 20:05	pm Download All ~
Database	SRA Obe details								Program			
SRA Blast se	arch set information	n		SRA Blast	earch set in	formation			Database	SRA See details	~	
SRX5796776			· · · · ·	SRX5796774			19604					
SRX5796775		SRR9018602				SRR901						
Other reports		SRR9018602 SRR9018603	1	SRX5796773						arch set information		
	MSA viewer 🔞				MSA view	SRR901 SRR901			SRA Blast se SRX5796741 SRX5796740	arch set information	sRR9018637 SRR9018638	
	MSA viewer			SRX5796773		SRR901 SRR901			SRX5796741 SRX5796740 SRX5796739	arch set information	SRR9018637 SRR9018638 SRR9018639	
A No si			1	SRX5796773 Other reports	MSA viev	SRR901 SRR901			SRX5796741 SRX5796740	earch set information	SRR9018637 SRR9018638	
-	gnificant similarity fou	SRR8018603	<u> </u>	SRX5796773 Other reports	MSA view	SRR901 SRR901	18605 Basons why <u>,click here</u>		SRX5796741 SRX5796740 SRX5796739 SRX5796738	arch set information	SRR9018637 SRR9018638 SRR9018639 SRR9018640	
Job Title	gnificant similarity fou MT121216:Pangol	SRR8018603 md. For reasons why,click.here in coronavirus isolate MP789,		SRX5796773 Other reports No s Job Title	MSA view significant sir MT12121	SRR901 SRR901 SRR901	18605 easons why <u>,click here</u> wirus isolate MP789,.		SRX5796741 SRX5796740 SRX5796739 SRX5796738 SRX5796737	earch set information	SRR9018637 SRR9018638 SRR9018639 SRR9018640 SRR9018640 SRR9018641	
Job Title RID	gnificant similarity fou MT121216:Pangol 0D.IVAZGS01R Se	SRR8018603	nioad All Y	SRX5796773 Other reports Job Title RID	MSA view significant sir MT12123 0DJV8W5	SRR901 SRR901 SRR901 UNIT O	18605 Basons why <u>,click here</u>		SRX5796741 SRX5796740 SRX5796739 SRX5796738 SRX5796737		SRR9018637 SRR9018638 SRR9018639 SRR9018640 SRR9018640 SRR9018641	IP789,
Job Title RID Program	MT121216:Pangol 0D.IVAZGS018 Se Citation ~	SRR9016603 Ind. For reasons why.click.here in coronavirus isolate MP789, orch expires on 01-2018.07 pm Down	nload All V	SRX5796773 Other reports Job Title RID Program	MSA view significant sir MT12123 ODJV8W5	SRR901 SRR901 milarity found. For re 16:Pangolin corona 5501R Search expire ation ~	18605 easons why <u>,click here</u> wirus isolate MP789,.		SRX5796741 SRX5796740 SRX5796739 SRX5796738 SRX5796737 SRX5796736	MT121216:Pangoli	SRR9018637 SRR9018638 SRR9018639 SRR9018640 SRR9018641 SRR9018642	
Job Title RID	gnificant similarity fou MT121216:Pangol 0D.IVAZGS01R Se	SRR9016603 Ind. For reasons why.click.here in coronavirus isolate MP789, orch expires on 01-2018.07 pm Down	nioad All Y	SRX5796773 Other reports Job Title RID	MSA view significant sir MT12123 ODJV8W5	SRR901 SRR901 SRR901 UNIT O	18605 easons why <u>,click here</u> wirus isolate MP789,.		SRX5796741 SRX5796740 SRX5796739 SRX5796739 SRX5796737 SRX5796736	MT121216:Pangoli	SRR9018637 SRR9018639 SRR9018639 SRR9018640 SRR9018641 SRR9018641 sRR9018642	
Job Title RID Program Database	Inificant similarity fou MT121216:Pangol DDIVAZGSD1R Se Citation ~ SRA See dotails	sRR9015603 md. For reasons why.click.here in coronavirus isolate MP789, arch expires on 81-2018 of pm Down V	nload All ~	SRX5796773 Other reports Job Title RID Program Database	MSA view significant sin MT12121 ODJV8W5 © <u>Citu</u> SRA <u>S</u>	SRR901 SRR901 milarity found. For re 16:Pangolin corona 5018 Search expire ation ~ co dotails ~	18605 easons why <u>,click here</u> wirus isolate MP789,.		SRX5796741 SRX5796740 SRX5796739 SRX5796738 SRX5796738 SRX5796736	MT121216:Pangoli ODPFHXN801R Se	SRR9018637 SRR9018638 SRR8018639 SRR8018639 SRR8018640 SRR8018641 SRR8018642 n coronavirus isolate M arch expires on 01-20 20:09	
Job Title RID Program Database	MT121216:Pangol 0D.IVAZGS018 Se Citation ~	sRR9015603 md. For reasons why.click.here in coronavirus isolate MP789, arch expires on 81-2018 of pm Down V	nioad Ali v	GRX5706773 Other reports Job Title RiD Program Database SRA Blast s SRX5796772	MSA view significant sin MT12121 ODJV8W5 © <u>Citu</u> SRA <u>S</u>	SRR001 SRR001 milarity found. For re 16:Pangolin corona 5018 Seorch expire abion ~ ee dotails ~ formation SRR001	easons why. <u>click here</u> wirus isolate MP789, es on 01-2019-07 pm		SRX5796741 SRX5796740 SRX5796740 SRX5796739 SRX5796738 SRX5796737 SRX5796737 SRX5796737 SRX5796737 SRX5796737 SRX5796737 SRX5796737	MT121216:Pangoli ODPEHXN801R Se Citation ~	SRR9018637 SRR9018638 SRR8018639 SRR8018639 SRR8018640 SRR8018641 SRR8018642 n coronavirus isolate M arch expires on 01-20 20:09	
Job Title RID Program Database	Inificant similarity fou MT121216:Pangol DDIVAZGSD1R Se Citation ~ SRA See dotails	sRR9015603 nd. For reasons why.click.here in coronavirus isolate MP789, arch expires on 01-20 18 07 pm Down v	nioad All ×	SRA Blast 3	MSA view significant sin MT12121 ODJV8W5 © <u>Citu</u> SRA <u>S</u>	SRR001 SRR001 SRR001 milarity found. For re- ts:Pangolin corona SSDIR Search expire ation ~ co-dotails ~ formation	easons why. <u>click here</u> wirus isolate MP789, es on 01-2019-07 pm		SRX5796741 SRX5796740 SRX5796739 SRX5796738 SRX5796738 Job Title RID Program Database SRA Blass sc	MT121216:Pangoli ODPEHXN801R Se Citation ~	SRR9019637 SRR9019638 SRR9019639 SRR9019640 SRR9019640 SRR9019641 SRR9019641 SRR9019642 V	
Job Title RID Program Database SRA Blast se SRx5796770 SRX5796769	gnificant similarity fou MT121216:Pangol DDJVAZGSD1R Se Citation ~ SRA See dotails earch set informatio	sRR9015603 nd. For reasons why.click.here in coronavirus isolate MP789, arch expires on 01-2018 of pm ov srR9015605	nioad All V	SRA Blass S SRA Stars	MSA view significant sir MT12121 ODJVBW5 © Citl SRA(<u>S</u> exearch set in	SRR001 SRR001 SRR001 millarity found. For re t6:Pangolin corona Stoll Search expire ation ~ co details ~ formation SRR001 SRR001	easons why. <u>click here</u> wirus isolate MP789, es on 01-2019-07 pm		SRX5796741 SRX5796740 SRX5796740 SRX5796738 SRX5796738 SRX5796737 SRX5796737 SRX5796737 SRX5796737 SRX5796737 SRX5796737 SRX5796737 SRX5796737 SRX5796741	MT121216:Pangoli ODPPHXN801R Se Citation ~ SRA See details	IRR0018637 SRR0018638 SRR0018638 SRR0018638 SRR0018640 SRR0018641 SRR0018642 arch expres on 01.70 20 69 V SRR0018643	
Job Title RID Program Database SRA Blast se SRX5796770	Inificant similarity fou MT121216:Pangol DDIVAZGSD1R Se Citation ~ SRA See dotails	sRR9015603 nd. For reasons why.click.here in coronavirus isolate MP789, arch expires on 01-2018 of pm ov srR9015605	nioad All ~	GRX5706773 Other reports Job Title RiD Program Database SRA Blast s SRX5796772	MSA view significant sin MT12121 ODJV8W5 © <u>Citu</u> SRA <u>S</u>	SRR001 SRR001 SRR001 millarity found. For re t6:Pangolin corona Stoll Search expire ation ~ co details ~ formation SRR001 SRR001	easons why. <u>click here</u> wirus isolate MP789, es on 01-2019-07 pm		SRX5796741 SRX5796740 SRX5796730 SRX5796739 SRX5796739 SRX5796739 SRX5796730 SRX5796730 SRX5796730 SRX5796730 SRX5796734	MT121216:Pangoli ODPPHXN801R Se Citation ~ SRA See details	SPR30116027 SPR30116023 SPR30116023 SPR30116023 SPR30116023 SPR3011602 SPR3011602 SPR3011602 SPR3011602 SPR3011602 SPR3011602 SPR3011604 SPR30116043 SPR30116043 SPR30116043	
Job Title RID Program Database SRXS796770 SRXS796769 Other reports	gntficant similarity fou MT12116:Pangol DD.WAZGSDIR Se @ Citation ~ SRA See dotails earch set informatio	sRR9015603 nd. For reasons why.click.here in coronavirus isolate MP789, arch expires on 01-2018 of pm ov srR9015605	nioad All ~	SRA Blast 1 SRX5796773 Other reports Job Title RID Program Database SRA Blast 1 SRX5796771 SRX5796771	MSA view significant sin MT12121 QDJV8WS © Citi SRA SRA search set in MSA view	SRR001 SRR001 SRR001 Milarity found. For re t6:Pangolin corona 5018 Seorch expire ation ~ oc dotails ~ formation SRR001 SRR001	easons why. <u>click here</u> wirus isolate MP789, es on 01-2019-07 pm		SRX5796741 SRX5796730 SRX5796730 SRX5796730 SRX5796738 SRX5796735 SRX5796735 SRX5796735 SRX5796735 SRX5796735 SRX5796735 SRX5796735	MT121216:Pangoli ODPPHXN801R Se Citation ~ SRA See details	ISPR3016637 ISPR3016638 ISPR3016638 ISPR3016639 ISPR3016640 ISPR3016641 ISPR3016642 ISPR3016643 ISPR3016643 ISPR3016643 ISPR3016643 ISPR3016645 ISPR3016645	
Job Title RID Program Database SRX5796770 SRX5796769 Other reports No si	gnificant similarity fou MT121216:Pangol 00.WAZGSDIR Se © Citation ~ SRA See dotais earch set informatio	ARR2015603 and. For reasons why,click.here in coronavirus isolate MP789, arch expires on 02-2018/7 pm Down arch expires on 02-2018/7 pm Down arch expires on 02-2018/7 pm Branch Stresson 5000 and. For reasons why,click.heree	nload All V	SRAS706773 Other reports Job Title RID Program Database SRAS706771 Other reports	MSA view significant sir MT12121 ODJV8W5 Cith SRA S earch set in MSA view	SRR001 SRR01 SRR01 SRR01 Milarity found. For re SSOLE Search appre and Malarity found. For re SRR001 SRR01 S	easons why,click here wirus isolate MP789, eson 01-2019/07pm eson	ownload All V	SRX5796741 SRX5796740 SRX5796730 SRX5796739 SRX5796739 SRX5796739 SRX5796730 SRX5796730 SRX5796730 SRX5796730 SRX5796734	MT121216:Pangoli ODPPHXN801R Se Citation ~ SRA See details	SPR30116027 SPR30116023 SPR30116023 SPR30116023 SPR30116023 SPR3011602 SPR3011602 SPR3011602 SPR3011602 SPR3011602 SPR3011602 SPR3011604 SPR30116043 SPR30116043 SPR30116043	
Job Title RID Program Database SRA.Blaat se SRXX5796769 Other reports Other reports Job Title	gnificant similarity fou MT321216:Pangol DDJA2CSO1B se Gladion × SRA See dotain earch set informatio MSA viewer • wgnificant similarity for MT32216:Pangolin	SRR0015603 and. For reasons why click here in coronavirus isolate MP789, srR001600 srR001600 and, For reasons why click here to coronavirus isolate MP789,		SRASTORITS Other reports Job Title RiD Program Database SRA Reast SRASTORITS Other reports Job Title	MSA view significant sir MT12122 ODJV8W5 © Citt SRA S mearch set in MSA view significant sin MT12121	SPROTO	Reasons why.click.here	ownload All V	SRAS79674 BRX379674 BRX3796749 BRX3796739 BRX3796739 BRX3796739 BRX3796736 BRX3796736 BRX3796736 SRAEBLASLSS SRAEBLASLSS SRAEBLASLSS SRAEBLASLSS SRAE706735 SRAE706735 SRAE706735 SRAE706735 SRAE706735 SRAE706735 SRAE706735 SRAE706735 SRAE706735 SRAE706735 SRAE706735 SRAE706735	MT121216:Pangoli ODPPHXN801R Se Citation ~ SRA See details	RPR00160/7 RPR00160/7 RPR00160/0 RPR00160/0 RPR00160/0 RPR00160/0 RPR00160/0 RPR00160/0 RPR00160/0 RPR0016642 SRR0016643 SRR0016644 SRR0016644 SRR0016645 SRR001664 SRR001664 SRR001664 SRR001664 SRR001664 SRR001664 SRR01664 SRR001664 SRR001664 SRR001664 SRR01664 SRR0164 SRR01664 SRR01664 SRR01664 SRR0164 SRR0164	
Job Title RID Program Database SRX5796770 SRX5796769 Other reports No si	gnificant similarity fou MT121216-Pangol OD/WAZGS01B se Catalon ~ SRA Sociotatis earch set informatio MSA viewer MT321216-Pangolir MT121216:Pangolir ODWXSSMK01B se	ARR2015603 and. For reasons why,click.here in coronavirus isolate MP789, arch expires on 02-2018/7 pm Down arch expires on 02-2018/7 pm Down arch expires on 02-2018/7 pm Branch Stresson 5000 and. For reasons why,click.heree		SRASTORITS Other reports Job Title RID Program Database SRASTBATS SRASTBATS SRASTBATS SRASTBATS SRASTBATS SRASTBATS Database No I	MSA view significant sin MT12122 ODJV8W5 © Cith SRA S MSA view significant sin MT12121 ODRVXF40	SPRODO	easons why,click here wirus isolate MP789, eson 01-2019/07pm eson	ownload All V	SRX5796741 SRX5796740 SRX5796730 SRX5796730 SRX5796737 SRX5796737 SRX5796737 Database SRX5796731 SRX5796731 SRX5796731 SRX5796731	MT121216:Pangoli ODPPHXN801R Se Citation ~ SRA See details		
Job Title RID Program Database SRA Blaat se SRAS796770 SRX5796770 Other reports Other reports Job Title RID	gnificant similarity fou MT121216:Pangol DD/AZGSDIB Se Catadon ~ SRA See datais earch set informatio MSA viewer information MSA viewer mt121216:Pangin MT121216:Pangin DOWN/SSMAPOIR Se DOWN/SSMAPOIR Se DOWN/SSMAPOIR Se	SRR0015603 and. For reasons why click here in coronavirus isolate MP789, srR001600 srR001600 and, For reasons why click here to coronavirus isolate MP789,		SRUGTPOTTA Other reports Job Title RiD Program Database SRUGTBOTTZ SRUGTBOTTZ SRUGTBOTTZ Job Title RiD Program	MSA view significant sin MT12121 ODJV9W2 Citi SRA S carch set in MSA view MSA view M	SRR010 SIGNA	Reasons why.click.here	ownload All V	Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Database Stat Blass es Stat Blass es Stat Blass es Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria Socionaria	MT121216:Pangoli ODPFHXN8018 Se Catation ~ SRA See details arch set information	IRR0018637 IRR0018637 IRR0018638 IRR0018639 IRR0018640 IRR0018641 IRR0018642 IRR0018642 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018645 IRR0018646 IRR0018646 IRR0018646	pm Download All >
Job Title RID Program Database SRX6706770 SRX6706770 SRX6706770 Other reports Other reports Job Title RID Program	gnificant similarity fou MTJ21216:Pangol DDA/25018 se Galadon × SRA See dotais earch set informatio MSA.viewer • MSA.viewer • MTJ21216:Pangolie DDB/VSSAK018 se MTJ212216:Pangolie DDB/VSSAK018 se Citation ×	sRR0015603 and. For reasons why.click.here in coronavirus isolate MP789, sRR0016609 and, For reasons why.click.here to coronavirus isolate MP789, arch expires on 01-20-20-32 pm	Noad All ~	SRASTORITS Other reports Job Title RID Program Database SRASTBATS SRASTBATS SRASTBATS SRASTBATS SRASTBATS SRASTBATS Database No I	MSA view significant sin MT12121 ODJV9W2 Citi SRA S carch set in MSA view MSA view M	SPRODO	Reasons why.click.here	ownload All V	Secondaria	MT121216:Pangoli ODPFHXISDIR Se Cataton * SRA Sen dotails arch set information MT121216:Pangoli	IRR001607 IRR001607 IRR001607 IRR0016030 IRR0016030 IRR0016030 IRR0016040 IRR0016040 IRR0016042 IRR0016042 INCONDENTION INCLUENT INCONDENTION INCONDENTION INCONDECT INR0016643 INR0016643 INR0016643 INR0016643 INR0016643 INR0016643 INR0016643 INR0016643 INR0016643 INR0016643 INR0016643 INR0016643	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blaat se SRAS796770 SRX5796770 Other reports Other reports Job Title RID	gnificant similarity fou MT121216:Pangol DD/AZGSDIB Se Catadon ~ SRA See datais earch set informatio MSA viewer information MSA viewer mt121216:Pangin MT121216:Pangin DOWN/SSMAPOIR Se DOWN/SSMAPOIR Se DOWN/SSMAPOIR Se	sRR0015603 and. For reasons why.click.here in coronavirus isolate MP789, sRR0016609 and, For reasons why.click.here to coronavirus isolate MP789, arch expires on 01-20-20-32 pm		SRUG790773 Other reports Job Title RID Program Database SRA Data SRA Database SRA Database Job Title RID Program Database No 31 Job Title RID Program	MSA view significant sin MT12222 ODJV8W5 © Cith SRA S earch set in MSA view significant sin MT12121 ODRVF44 © Cith SRA S	SREDTO	Reasons why.click.here	ownload All V	SCONTRACT SIGNOFFACT S	MT121216:Pangoli ODPFHONBOLR Se Chalton ~ SRA See details arch set information MT122126:Pangoli ODB6594E018 See	IRR0018637 IRR0018637 IRR0018638 IRR0018639 IRR0018640 IRR0018641 IRR0018642 IRR0018642 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018645 IRR0018646 IRR0018646 IRR0018646	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Stabase SRA Stabase SRA Stabase SRA Stabase No stabase	gnificant similarity fou MT11216:Pangol 9D/VAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Pangolin Vignificant similarity for MT12126:Pangolin ODWVSSMMCNIB Se Download All ~ @ Catalon ~ SRA See dotais	ARROOTS603		SRUSTORITS Other reports Other reports Dab Title RiD Program Database SRA Blass 1 SRA Blass 1 Database FID Program Database SRA Blass 1	MSA view significant sin MT12222 ODJV8W5 © Cith SRA S earch set in MSA view significant sin MT12121 ODRVF44 © Cith SRA S	ARROYOLL ARROYOL ARROYOLL ARROYOL ARROYOLL ARROYOLL ARROYOLL ARROYOLL ARROYOLL ARROYOLL ARROYOLL	essons why.click here wirus isolate MP789, s on 01-20 13-07 pm essons essons why.click here virus isolate MP789, on 01-20 20-32 pm L	ownload All V	Socomparial Socomp	MT121216:Pangoli DDFHXN8018 Se Cataton ~ SRA See dataits arch set information arch set information MT121216:Pangoli DDE503AED18 See Cataton ~	IRR0018637 IRR0018637 IRR0018630 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR0018642 IRR0018643 IRR0018644 IRR0018644 IRR0018645 IRR0018646 IRR0018646 IRR0018646 IRR0018646 IRR0018646 IRR0018646 IRR0018640 IRR018640 IRR018640 IRR018640 IRR0	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA/SNP/T0 SRA/SNP/T0 SRA/SNP/T0 NO st NO st NO st NO st NO st NO st NO st NO st SRA Blast sec	gnificant similarity fou MTJ21216:Pangol DDA/25018 se Galadon × SRA See dotais earch set informatio MSA.viewer • MSA.viewer • MTJ21216:Pangolie DDB/VSSAK018 se MTJ212216:Pangolie DDB/VSSAK018 se Citation ×	SRR0015603 arch express on 02-2018/7 pm proved to the press on 02-2018/7 pm proved to the press on 02-2018/7 pm proved to the press on 02-2018/7 pm presson to the presson swhy, click heree arch express on 02-2028 pm SRR0015605 SRR0015665 SRR0015665		SRUG790773 Other reports Job Title RID Program Database SRA Data SRA Database SRA Database Job Title RID Program Database No 31 Job Title RID Program	MSA view significant sin MT12222 ODJV8W5 © Cith SRA S earch set in MSA view significant sin MT12121 ODRVF44 © Cith SRA S	SREDTO	assons why.click here wirus isolate MP789, ss on 01-20 19-07 pm secos	ownload All V	SCONTRACT SIGNOFFACT S	MT121216:Pangoli ODPFHONBOLR Se Chalton ~ SRA See details arch set information MT122126:Pangoli ODB6594E018 See	IRR0018637 IRR0018637 IRR0018630 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR0018642 IRR0018643 IRR0018644 IRR0018644 IRR0018645 IRR0018646 IRR0018646 IRR0018646 IRR0018646 IRR0018646 IRR0018646 IRR0018640 IRR018640 IRR018640 IRR018640 IRR0	pm <u>Rownload</u> All • (P789,
Job Title RiD Program Database SR02576770 Other reports Job Title RID Program Database SRA Blast see SRA Blast see	gnificant similarity fou MT11216:Pangol 9D/VAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Pangolin Vignificant similarity for MT12126:Pangolin ODWVSSMMCNIB Se Download All ~ @ Catalon ~ SRA See dotais	SRR0015603 and. For reasons why click here in coronavirus isolate MP789, spreao 15000 and, For reasons why click here coronavirus isolate MP789, arch expires on 01-20 20.32 pm SRR00156653 SRR00156653 SRR00156654		SRUG796773 Other reports Job Title RID Program Database SRA Base t SRA Base t SRA Base t SRA Base t SRA Base t SRA Base t SRA Base t	MSA view significant sin MT12222 ODJV8W5 © Cith SRA S earch set in MSA view significant sin MT12121 ODRVF44 © Cith SRA S	SRR001 SRR01 SR01 S	assons why.click here wirus isolate MP789, ss on 01-20 19.07 pm 06005 assons why.click here assons why.click here virus isolate MP789, s on 01-20 20-32 pm 0 06057 0656 0657 0656	ownload All V	SRA Blas se Ro Program SRA Program SRA Blas se SRA BLAS SA S	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018630 IRR0018640 IRR0018641 IRR0018642 IRR0018643 IRR0018644 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018645 IRR0018646 IRR0018646 IRR0018645 IRR0018646 IRR0018646 IRR0018645 IRR0018646 IRR0018647 IRR0018648 IRR0018649 IRR0018640 IRR0018641 IRR0018641 IRR0018642 IRR0018643 IRR0018643 IRR0018649 IRR0018640 IRR0018641 IRR0018641 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IR	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA/SIM/TO SRA/SIM/TO AND Other reports No si No si	gnificant similarity fou MT11216:Pangol 9D/VAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Pangolin Vignificant similarity for MT12126:Pangolin ODWVSSMMCNIB Se Download All ~ @ Catalon ~ SRA See dotais	SRR0015603 arch express on 02-2018/7 pm proved to the press on 02-2018/7 pm proved to the press on 02-2018/7 pm proved to the press on 02-2018/7 pm presson to the presson swhy, click heree arch express on 02-2028 pm SRR0015605 SRR0015665 SRR0015665		SRA STATES	MSA view significant sin MT12222 ODJV8W5 © Cith SRA S earch set in MSA view significant sin MT12121 ODRVF44 © Cith SRA S	SPROTO	assons why.click here wirus isolate MP769, son 01-2019-07 pm 8606 8607 asons why.click here virus isolate MP769, 9005 8006 8007 8006 8007 8006 8007 8008 8009	ownload All V	SRA Blas se Ro Program SRA Program SRA Blas se SRA BLAS SA S	MT121216:Pangoli DDFHXN8018 Se Cataton ~ SRA See dataits arch set information arch set information MT121216:Pangoli DDE503AED18 See Cataton ~	IRR0018637 IRR0018637 IRR0018630 IRR0018630 IRR0018640 IRR0018641 IRR0018642 IRR0018643 IRR0018644 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018645 IRR0018646 IRR0018646 IRR0018645 IRR0018646 IRR0018646 IRR0018645 IRR0018646 IRR0018647 IRR0018648 IRR0018649 IRR0018640 IRR0018641 IRR0018641 IRR0018642 IRR0018643 IRR0018643 IRR0018649 IRR0018640 IRR0018641 IRR0018641 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IR	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015665 SRR0015665 SRR0015666 SRR001566 SRR00156 SRR0015 SRR00156 SRR00156 SRR0015 SRR00156 SRR0015 SRR00156 SRR0015 SRR00156	 IA baok 	SRUG796773 Other reports Job Title RID Program Database SRA Base t SRA Base t SRA Base t SRA Base t SRA Base t SRA Base t SRA Base t	MSA view significant sin MT12222 ODJV8W5 © Cith SRA S earch set in MSA view significant sin MT12121 ODRVF44 © Cith SRA S	ARROTOLOGY	BBBOD BBBOD BBBODS BBBODS BBBODS BBBDS BBBODS	ownload All V	Socomparial Socomparial Socomparial Socomparial Socomparial Socomparial Socomparial Socomparial Socomparial Socomparial Socomparial Database Stat Blass se Socomparial Socompa	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018630 IRR0018640 IRR0018641 IRR0018642 IRR0018642 IRR0018643 IRR0018643 IRR0018643 IRR0018643 IRR0018645 IRR0018645 IRR0018645 IRR0018650 IRR0018650 IRR0018650 IRR0018650 IRR0018650 IRR0016651 IRR0016651 IRR0016651	pm <u>Rownload</u> All • (P789,
Job Title RiD Program Database SR02576770 Grosports Job Title RiD Program Database SR02576775 SR02576775	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 Arran of a construction of a constru	It has been as a second sec	SRUGTORY SRUGTORY SRUGTORY Job Title RID Program Database SRA Blast s SRA SRA SRA SRA SRA SRA SRA SRA SRA SRA	MSA view significant sin MT12222 ODJV8W5 © Cith SRA S earch set in MSA view significant sin MT12121 ODRVF44 © Cith SRA S	SRR001 SRR01 SRR0	BBBOD BBBOD BBBODS BBBODS BBBODS BBBDS BBBODS	ownload All V	Joborner Att SIGKOPPE ATT SIGKO	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR8018637 IRR8018637 IRR8018638 IRR8018639 IRR8018640 IRR8018640 IRR8018642 IRR8018642 IRR8018643 IRR8018643 IRR8018643 IRR8018643 IRR8018643 IRR8018643 IRR8018643 IRR8018645 IRR8018646 IRR8018646 IRR8018646 IRR8018646 IRR8018645 IRR8018645 IRR8018645 IRR8018645 IRR8018645	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015665 SRR0015665 SRR0015666 SRR001566 SRR00156 SRR0015 SRR00156 SRR00156 SRR0015 SRR00156 SRR0015 SRR00156 SRR0015 SRR00156		SRUGTORY SRUGTORY SRUGTORY Job Title RID Program Database SRA Blast s SRA SRA SRA SRA SRA SRA SRA SRA SRA SRA	MSA.view MT122222 GD.VHPS95 SRA SRA <td>RR0010 RR001 RR01 RR01 RR001 RR001 RR001 RR001 RR001 RR</td> <td>BBBOD BBBOD BBBODS BBBODS BBBODS BBBDS BBBODS</td> <td>viela Ali V</td> <td>SRA Blass er SRA B</td> <td>MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details</td> <td>IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR01186463 IRR0118643 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118653</td> <td>pm <u>Rownload</u> All • (P789,</td>	RR0010 RR001 RR01 RR01 RR001 RR001 RR001 RR001 RR001 RR	BBBOD BBBOD BBBODS BBBODS BBBODS BBBDS BBBODS	viela Ali V	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR01186463 IRR0118643 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118653	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015665 SRR0015665 SRR0015666 SRR001566 SRR00156 SRR0015 SRR00156 SRR00156 SRR0015 SRR00156 SRR0015 SRR00156 SRR0015 SRR00156	 IA baok IA baok 	SRUGTPOTTO Other reports Job Title RID Program Database SRA Flace t SRA Flac	M5Aviewagenficant sir M113222 © Cell SRAI S Cell M12322 M124 Service Service M12322 M12322 SRAI S Cell SRAI S Cell SRAI S Cell M12322	SRR001 SRR001 SRR001 SRR001 SRR001 SRR001 SRR001 CORA SRR001 CORA SRR001 SRR01 SRR001	B800 Image: Control of the second secon	ventoed All ¥	I SECONDATA SECO	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0016637 IRR0016637 IRR001663 IRR0016640 IRR0016640 IRR0016640 IRR0016642 IRR0016642 IRR0016643 IRR0016643 IRR0016643 IRR0016643 IRR0016643 IRR0016643 IRR0016643 IRR0016660 IRR0016651 IRR0016632 IRR0016632 IRR0016634	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR		SRUGTORITS Other reports Job Title RID Program Database SRA Data SRA	M5Aviewagenficant sir M113222 © Cell SRAI S Cell M12322 M124 Service Service M12322 M12322 SRAI S Cell SRAI S Cell SRAI S Cell M12322	BRR001 BRR001 BRR001 BRR001 BRR001 BRR001 BRR001 COMMING SOLE Source appre Source S	assons why.click here wirus isolate MP789, ss on 01-20 13-07 pm assons why.click here	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR01186463 IRR0118643 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118653	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR		SRUG790773 Other reports Job Title RID Program Database SRUG790773 SRUG790773 Other reports Job Title RID Program Database RID Program Database SRUG797719 SRUG79777 SRUG79777 SRUG79777 SRUG7977 SRUG79777 SRUG79777 SRUG79777 SRUG79777 SRUG797777 SRUG7977777 SRUG777777 SR	MSA view MT12222 00.2005 SRA SRA SC CONTRACTOR MT12222 000000000000000000000000000000000	BRR001 BRR001 BRR001 BRR001 BRR001 BRR001 BRR001 COMMING SOLE Source appre Source S	assons why.click here wirus isolate MP789, ss on 01-20 13-07 pm assons why.click here	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR01186463 IRR0118643 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118653	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR	 IA baok IA baok 	SRASTPORTS SRASTPORTS Job Title RID Program Database SRA Flacer s SRASTPORTS SRASTPORTS Job Title RID Program Database FRA Flacer s Job Title RID Program Database	MSA view MT12222 00.2005 SRA SRA SC CONTRACTOR MT12222 000000000000000000000000000000000	Israoto espre ser of or other of the series	assons why.click here wirus isolate MP789, ss on 01-20 13-07 pm assons why.click here	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118646 IRR0118646 IRR0118645 IRR0118646 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118654	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR		SRA 5796773 Other reports Job Title RID Program Job Title RID SRA Blast t SRA 57 S S SRA 57 S SRA 57 S S SRA 57 S S SRA 57 S S SRA 57 S S S SRA 57 S S S S S S S S S S S S S S S S S S S	MSA view MT12222 00.2005 SRA SRA SC CONTRACTOR MT12222 000000000000000000000000000000000	APRODUCE	BB005 Insacons why.click.here wirus isolate MP789, Image: Son 01-20 19:07 pm Image: Son 01-20 20:03 pm Image: Son 01-20 20:04 pm Image: Son	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118646 IRR0118646 IRR0118645 IRR0118646 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118654	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR		SRUGTORITS SRUGTORITS SRUGTORITS SRUGTORITS Job Title RID Program Database SRUGTORITS SRUGTORITS Job Title RID Program Database SRUGTORITS Job Title RID Program Database SRUGTORITS SRUGTORIT	MSA view upgenficant ser an MT122122 000-0000 SRAI Se SRAI SE	Isreet of a series	Bood Bood Bood Bood Bood Bood Son 01-20 13-07 pm P Bood	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118646 IRR0118646 IRR0118645 IRR0118646 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118654	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR		SRUGTORY SRUGTO	MSA view upgenficant ser an MT122122 000-0000 SRAI Se SRAI SE	Isreen of a series	BBG05 BBG05 BBG05 BBG05 BBG05 BBG05 SS on 01-20 13-07 pm D BBG05 BBG07 BBBG05 BBBG07	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118646 IRR0118646 IRR0118645 IRR0118646 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118654	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Statuse SRA Statuse No st Job Title RID Program Database SRA Blast sea SRA Statuse SRA Statuse	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR		SRAJPORTS SRAJPORTS	MSA view upgenficant ser an MT122122 000-0000 SRAI Se SRAI SE	Israel of a serie of a	BB005 Image: Control of the second seco	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118646 IRR0118646 IRR0118645 IRR0118646 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118654	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Statuse SRA Statuse No st Job Title RID Program Database SRA Blast sea SRA Statuse SRA Statuse	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR		SRUGTORY SRUGTO	MSA view MT12212 00,04997 SRA SRA S SRA S S SRA S S SRA S S SRA S S SRA S S SRA S S S S S S S S S S S S S S S S S S S	Isreen of a series	BBBOD BBBOD exacons why.click here Wirus isolate MP789, BBBOD BBBOD	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118646 IRR0118646 IRR0118645 IRR0118646 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118654	pm <u>Rownload</u> All • (P789,
Job Title RID Program Database SRA Blast se SRA Blast se SRA STAR No sl Database No sl SRA Blast sea SRA Blast sea SRA Blast sea SRA Blast sea	gnificant similarity fou MT11216:Panglo DDVAZGSDIB Se Catalon ~ SRA See dotais MSA viewer @ MSA viewer @ MSA viewer @ MSA viewer @ MT12126:Panglin DDVXSSMMCAI @ Citation ~ SRA See dotais	SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015663 SRR0015664 SRR0015666 SRR001566 SRR0015666 SRR001566 SRR00156 SRR001566 SRR001566 SRR00156 SRR00156 SRR0015 SRR00156 SRR00156 SRR00156 SRR00156 SRR00156 SRR	 IA back 	SRUGTORY SRUGTO	MSA view MI3222	Isreen of a series	BB005 Image: Control of the second seco	ventoed All ¥	SRA Blass er SRA B	MT121216:Pangoli ODPFHXN80LR Se Chalton × SRA See details arch set information MT121216:Pangoli ODR66924E01R Set Chalton × SRA See details	IRR0018637 IRR0018637 IRR0018630 IRR0018640 IRR0018640 IRR0018641 IRR0018642 IRR01186464 IRR01186464 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118646 IRR0118646 IRR0118645 IRR0118646 IRR0118645 IRR0118645 IRR0118645 IRR0118645 IRR0118651 IRR0118651 IRR0118651 IRR0118653 IRR0118654	pm <u>Rownload</u> All • (P789,

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

PRJNA529540.

RID	0DUEA8S401R Search	expires on 01-20 21:16 pm	Download All
Program	② <u>Citation</u> ✓		
Database	SRA See details V		
SRA Blast se	arch set information		
SRX1319167	SF	RR2591032	
Query Length	29521		
Other reports	MSA viewer 😮		

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.

D rogram	6K8P3ES9016 se					
ogram		earch expires on 04-05 19:28 pm	Download All ~	Percent Identity	E value	Query Coverage
				to	to	to
atabase	SRA See detail	s v				
	or vi <u>occ uciu</u>	<u> </u>		-		Filter Reset
SRA Blast sea	rch set information	-	е			
SRX7911829		SRR11306687				
SRX7911828		SRR11306688				
SRX7911827		SRR11306689				
SRX7911826		SRR11306690				
SRX7911825		SRR11306691				
SRX7911824		SRR11306692				
SRX7911823		SRR11306693				
SRX7911822		SRR11306694				
SRX7911821		SRR11306695				
SRX7911820		SRR11306696				
SRX7911819		SRR11306697				
SRX7911818		SRR11306698				Support Center
SRX7911817		SRR11306699				•
SRX7911816		SRR11306700				
SRX7911815		SRR11306701			_	
SRX7911814		SRR11306702		Popular	Resources	Actions
SRX7911813		SRR11306703		PubMed	Literature	Submit
SRX7911812		SRR11306704				
SRX7911811		SRR11306705		PubMed Central	Health	Download
SRX7911810		SRR11306706		Bookshelf	Genomes	Learn

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.

HS14 (SRR11306689)

Metadata Analysis Reads Data access Taxonomy Analysis Unidentified reads: 7.23% Identified reads: 92.77% cellular organisms: 92.77% Eukaryota: 76.86% Opisthokonta: 76.69% Metazoa: 75.87% Boreoeutheria: 74.22% Laurasiatheria: 68.86% -Manis javanica: 64.4% Euarchontoglires: 1.72% Simiiformes: 0.92% Catarrhini: 0.87% Hominoidea: 0.79% Hominidae: 0.7% Homininae: 0.63% -Homo sapiens: 0.3% Fungi: 0.79% Viridiplantae: 0.11% Sar: < 0.01% (188 Kbp) Bacteria: 13.22% Archaea: < 0.01% (132 Kbp)</p> └ Viruses: < 0.01% (222 Kbp) Caudovirales: < 0.01% (100 Kbp) Siphoviridae: < 0.01% (62 Kbp) Tunavirinae: < 0.01% (33 Kbp) Rtpvirus: < 0.01% (29 Kbp) unclassified Rtpvirus: < 0.01% (29 Kbp) Enterobacteria phage vB_EcoS_IME542: < 0.01% (16 Kbp) Escherichia phage vB_Ecos_CEB_EC3a: < 0.01% (12 Kbp) unclassified Tunavirinae: < 0.01% (5 Kbp) Escherichia phage IMM-001: < 0.01% (5 Kbp) Pahexavirus: < 0.01% (15 Kbp) Propionibacterium phage Solid: < 0.01% (9 Kbp) Myoviridae: < 0.01% (23 Kbp)</p> Eneladusvirus: < 0.01% (13 Kbp) Cronobacter phage vB_CsaM_GAP32: < 0.01% (13 Kbp) - Tevenvirinae: < 0.01% (10 Kbp) unclassified Tevenvirinae: < 0.01% (10 Kbp) Riboviria: < 0.01% (74 Kbp) Tymovirales: < 0.01% (35 Kbp) Actinidia seed-borne latent virus: < 0.01% (35 Kbp) Potyviridae: < 0.01% (29 Kbp) Sugarcane mosaic virus: < 0.01% (29 Kbp) Negarnaviricota: < 0.01% (9 Kbp) Mononegavirales: < 0.01% (9 Kbp)</p> Filoviridae: < 0.01% (7 Kbp) Paramyxoviridae: < 0.01% (2 Kbp) Mammalian rubulavirus 5: < 0.01% (2 Kbp) Polyomaviridae: < 0.01% (18 Kbp) Macaca mulatta polyomavirus 1: < 0.01% (18 Kbp) Ortervirales: < 0.01% (16 Kbp) Orthoretrovirinae: < 0.01% (11 Kbp) Gammaretrovirus: < 0.01% (9 Kbp) RD114 retrovirus: < 0.01% (7 Kbp) Lentivirus: < 0.01% (2 Kbp) Human immunodeficiency virus 1: < 0.01% (2 Kbp) Herpesvirales: < 0.01% (6 Kbp) Herpesviridae: < 0.01% (5 Kbp) Gammaherpesvirinae: < 0.01% (3 Kbp) Human gammaherpesvirus 4: < 0.01% (3 Kbp) Betaherpesvirinae: < 0.01% (2 Kbp) Human betaherpesvirus 5: < 0.01% (2 Kbp) Poxviridae: < 0.01% (4 Kbp) Vaccinia virus GLV-1h68: < 0.01% (4 Kbp)

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.

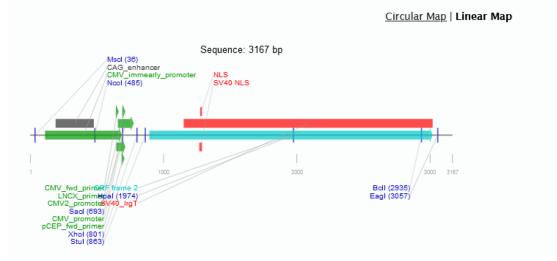


Fig.20: Addgene sequence analysis result of the single contig with detected homology to Macaca Mulatta polyomavirus 1 and Human betaherpesvirus 5.

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.

	Murine retrovirus shuttle vector pZIPneoSV(TAg), complete sequence	syntheti I	NA	<u>32630</u>	4096	4096	70%	0.0	99.78%	7020	<u>Z93724.1</u>
	Synthetic construct clone pARVA T-Ag, complete sequence	<u>syntheti</u> I	<u>AN</u>	<u>32630</u>	3849	5119	89%	0.0	99.30%	5690	MF174873.1
	Mammalian expression vector pSV529HIFNG, complete sequence	Mamma	<u>NA</u>	<u>1945111</u>	3502	6652	70%	0.0	99.84%	9455	LT727634.1
	Mammalian expression vector pSV51E6Hf2, complete sequence	Mamma	<u>NA</u>	<u>1945103</u>	3502	6652	70%	0.0	99.84%	9574	LT727623.1
	Mammalian expression vector pSV51E6Hf1, complete sequence	<u>Mamma</u>	<u>AN</u>	<u>1945102</u>	3502	6652	70%	0.0	99.84%	9472	LT727622.1
	Mammalian expression vector LNXCO3, complete sequence		Mammal	<u>an exp</u>	1463	1463	25%	0.0	99.75%	7484	LT727330.1
	EIAV-based lentiviral vector, complete sequence		EIAV-bas	ed lent	1461	1461	24%	0.0	100.00%	7941	<u>GQ872121.1</u>
~	Mutant Human betaherpesvirus 5 clone AD169-BAC20, complete genome		Human b	etaher	1448	1448	24%	0.0	100.00%	232314	MN920393.1
	Mutant Human betaherpesvirus 5 clone AD169-BAC2, complete genome		<u>Human b</u>	etaher	1448	1448	24%	0.0	100.00%	233833	MN900952.1

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 Immunodefiency Virus 1 (HIV-1), which upon specific BLAST analysis reveal themselves to be sequences derived from lentiviral transfer vectors.

Results for Program		
Program	1:Icl/Query_64532 k141_15599(473bp)	Organism only top 20 will appear Type common name, binomial, taxid or group name
	BLASTN 😧 Citation 🛩	Add organism
Database	nt <u>See details</u> ~	
Query ID	Icl Query_64532	Percent Identity E value Query Coverage
Description	k141_15599	to to to
Nolecule type	dna	Filter Reset
Query Length	473	Reset
Other reports	Distance tree of results MSA viewer @	
Descriptions	Graphic Summary Alignments Taxonomy	
Sequences p	producing significant alignments	Download 🐣 🔤 Select columns 🐣 Show 500 🕶 📀
		GenBank Graphics Distance tree of results Item MSA Viewer
Select all	500 sequences selected	GenBank Graphics Distance tree of results Mee MSA Viewer
	Description	Scientific Name Score Score Cover value Ident Len Accession
PREDICTER	D: Manis pentadactyla colled-coll domain containing 186 (CCDC186), mRNA	Manis pentada 627 627 73% 3e-175 99.14% 6791 XM 036898626.1
		Cloning vector 231 231 26% 3e-56 100.00% 8000 MN996874.1
	tor sh-AFP-PPP2CA, complete sequence tor sh-PPP2CA, dna, complete sequence	Cloning vector 231 231 26% 3e-56 100.00% 8000 MN996874.1 Cloning vector 231 231 26% 3e-56 100.00% 7716 MN996873.1
	tor sh-ctrl, complete sequence	Cloning vector 231 231 26% 3e-56 100.00% 7717 MN996872.1
	tor sh-AFP, complete sequence	Cloning vector 231 231 26% 3e-56 100.00% 7716 MN996871.1
Cloning vec	tor pSF_lentl_SARS-CoV-2_partial-S/E/M/N_complete sequence	Cloning vector 231 231 26% 3e-56 100.00% 13543 MT299805.1
Cloning vec	tor pSF_lenti_SARS-CoV-2_partial-ORF1b/partial-S_complete sequence	Cloning vector 231 231 26% 3e-56 100.00% 13507 MT299804.1
Cloning vec	tor pSF_lenti_SARS-CoV-2_partial-ORF1ab, complete sequence	Cloning vector 231 231 26% 3e-56 100.00% 13462 MT299803.1
Cloning vec	tor pSF_lenti_SARS-CoV-2_partial-ORF1a.complete sequence	Cloning vector 231 231 26% 3e-56 100.00% 13408 MT299802.1
esults for	2:Icl]Query_64533 k141_57178(285bp)	Organism only top 20 will appear
		Type common name, binomial, taxid or group name
rogram		+ Add organism
atabase	nt See details ~	Percent Identity E value Query Coverage
uery ID	Icl[Query_64533	to to to to
escription	k141_57178 dna	
lolecule type uery Length	dna	Filter Reset
ther reports	285 Distance tree of results MSA viewer ?	
aner reports	Distance lice of results IMOA VIEWEL	
Descriptions	Graphic Summary Alignments Taxonomy	
C	dealers dealers the	Download 🎽 🚾 Manage Columns 🎽 Show 500 💌 💡
Sequences p	producing significant alignments	Download * New Manage Columns * Show 500 V
select all	0 sequences selected	GenBank Graphics Distance tree of results May MSA Viewer
	Description	Scientific Common Taxid Max Total Query E Per. Acc.
	-	Name Name Iaxid Score Score Cover value Ident Len Accession
Cloning ved	tor RS474_ErbB-RASER1C-dCas9VP64, complete sequence	<u>CloningNA</u> <u>2577058</u> 466 466 100% 4e-127 95.64% 16350 <u>MK801288.1</u>
	insfer vector FUW-tetO, complete sequence	LentiviraNA 2506927 466 466 100% 4e-127 95.64% 9554 MK318529.1
Cloning yed	tor pLenti-EF1a-dCas9-DNMT3B(E697A)-2A-bla.complete sequence	CloningNA 2169471 466 466 100% 4e-127 95.64% 14877 MG840314.1
Cloning ved	tor pDBR, complete sequence	Cloning vector p 448 448 100% 1e-121 94.37% 6619 ER822201.1
	tor shRNA EYFP-P2A Puro, complete sequence	Cloning vector s 444 444 100% 2e-120 94.06% 9625 MH749464.1
	nunodeficiency virus I (HIV-1) RNA	Human immun 270 270 51% 3e-68 100.00% 9781 Z11530.1
-	tor pAAV-EF1A-Gephyrin FingR-GFP-CCR5TC, complete sequence	Cloning vector p 267 267 50% 4e-67 100.00% 7010 MT612432.1
Cloning ved	tor pAAV-EF1A-DIO-Gephyrin.FingR-GFP-CCR5TC, complete sequence	Cloning vector p 267 267 50% 4e-67 100.00% 7204 MT612431.1
Results for	3:Icl]Query_64534 k141_57181(300bp)	Organism only top 20 will appeal
Program	BLASTN 😧 Citation 🗸	Type common name, binomial, taxid or group name
Database	nt <u>See details</u> Y	+ Add organism
Query ID	Icl Query_64534	Percent Identity E value Query Coverage
	k141_57181	to to to
Description	K141_0/101	
	dna	
Molecule type		Filter
Molecule type Query Length	dna	Filter Reset
Molecule type Query Length Other reports	dna 300 Distance tree of results MSA viewer @	Filter Reset
Nolecule type Query Length	dna 300 Distance tree of results MSA viewer @	Filter Reset
Nolecule type Query Length Other reports Descriptions	dna 300 Distance tree of results MSA viewer @ Graphic Summary	
Molecule type Query Length Other reports Descriptions Sequences	dna 300 Distance tree of results MSA viewer @ Graphic Summary producing significant alignments	Download Y C Manage Columns Y Show 500 V
Molecule type Query Length Other reports Descriptions Sequences	dna 300 Distance tree of results MSA viewer @ Graphic Summary	
Molecule type Query Length Other reports Descriptions Sequences	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments Ø sequences selected	Download → Inter Manage Columns → Show 500 → GenBank Graphics Distance free of results Inter MSA Viewer Scientific Common - Max Total Query E Per, Acc
Molecule type Query Length Other reports Descriptions Sequences	dna 300 Distance tree of results MSA viewer @ Graphic Summary producing significant alignments	Download V C Manage Columns V Show 500 V P GenBank Graphics Distance tree of results MSA Viewer
Molecule type Query Length Other reports Descriptions Sequences	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments Ø sequences selected	Download ✓ IIII Manage Columns ✓ Show 500 ✓ @ GenBank Graphics Distance tree of results III MSA Viewer Scientific Common Tank Max Total Query E Pec Acc
Molecule type Query Length Other reports Descriptions Sequences select all yedor pl.V. yedor pl.V.	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XITESQ. 60FP: complete sequence XITESQ. 60FP. complete sequence XITESQ. 60FP. complete sequence	Download Control Manage Columns Show 500 ▼ ● GenBank Graphics Distance tree of results Immode MSA Viewer Scientific Common Taud Max Total Ouery E Hert Acc Accession Waterie Scientific Common Taud Score Score Value Ident Acc Accession Vetorice Na 2491125 549 549 100% 4e-152 98.7% 8632 IH1225103.1 Vetorice Na 2491125 549 549 100% 4e-152 98.7% 9632 IH1325103.1
Molecule type Query Length Other reports Descriptions Sequences select all yector pl.V. yector pl.V.	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE3G.aCFP.complete sequence	Download Control Manage Columns Show 500 ✓ Image: State and the state and t
Molecule type Query Length Other reports Descriptions Sequences Sected all Vector pLV Vector pLV Lentwiral to	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XITESQ. 60FP: complete sequence XITESQ. 60FP. complete sequence XITESQ. 60FP. complete sequence	Download Control Manage Columns Show 500 ▼ ● GenBank Gradhus Distance tree of results Immode Mask Viewer Scientific Commod Taud Max Total Ouery E Pet Accession Name Name Taud Max Total Ouery E Pet Accession Vector.e. Nd 2491125 549 549 10% 4e-152 98.7% 832 IH12251021 Vector.e. Nd 2491125 549 549 10% 4e-152 98.7% 932 IH12251021
Addecule type Query Length Dther reports Descriptions Sequences Select all Vector eLV. Vector eLV.	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description X TRESQ eGCPP complete sequence XTRESQ eGCPP complete sequence RTRESQ eGC complete sequence RTRESQ eGC complete sequence	Download Image Columns Show 500 Image GenBank Graphics Distance iree of results Image Max Total Ouery Image Image Image Max Total Ouery Image Per Acc Accession Vendore Name Total Score Score Cover Value Per Acc Accession Vendore Nd 2491125 549 549 100% 4e-152 99.67% 6536 Min2251041 Vendore Nd 2481125 549 549 100% 4e-152 99.67% 6536 Min2251041 Vendore Nd 2481125 549 549 100% 4e-152 99.67% 6536 Min2251041 Vendore Nd 2481125 549 549 100% 4e-152 99.67% 6536 Min2251041 Lemburg Nd 255321 520 520 100% 3e-143 98.09% 7207 X4684861
Wolecule type Query Length Uther reports Descriptions Sequences select all Vector tV/ Vector tV/ Vector tV/ Coming vector Coming vector Coming vector Coming vector	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments O sequences selected Elescipton XTRESG.60FP.complete sequence XTRESG.complete sequence cdor.0DER.complete sequence cdor.0DER.complete sequence	Download Image Columns Show 500 ● GenBank Graphics Distance tree of results Image Manage Columns Image Manage Image Image Image Mana
Molecule type Query Length Other reports Sequences select all Vector tV/ Vector tV/ Vector tV/ Coning use Coning use Coning use Coning use	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE30.e0FP.complete sequence XTRE30.edfP.complete sequence durghRynether sequence	Download Image Manage Columns Show 500 Image Manage GenBant Graphics Distance tree of results Image Mission Image M
Wolecule type Query Length Dther reports Sequences select all yestor et/y yestor et/y yestor et/y select all yestor et/y select all g yestor et/y yestor et/y yestor et/y yestor et/y yestor et/y select all g yestor et/y select all g yestor et/y select all g yestor et/y select all g yestor et/y yestor et/y yestor et/y select all g yestor et/y select g yestor et/y select all g yestor et/y select all g yestor et/y select all g yestor et/yestor et/yestor et/yestor g yestor et/yestor et/yestor et/yestor et/yestor g yestor et/yestor et/yestor et/yestor g yestor et/yestor et/yestor et/yestor et/yestor et/yestor g yestor et/yestor e	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Usescripton XXTEE30.60FP.complete sequence XXTEE30.60FP.complete sequence data sRNA.EYP.PP2A.Purg.complete sequence data sRNA.EYP.PP2A.Purg.complete sequence	Download Image Columns Show 500 Image GenBank Granbes Distance tree of results Image M48 Verset Scientific Common Tadi Scient Cover Value Image M48 Verset Name Tadi Scient Cover Value Image M48 Verset Verder.e. Na 2491125 549 549 100% 44-152 96.7% 630 Imt221041 Verder.e. Né 2491125 549 549 100% 44-152 96.7% 630 Imt221041 Verder.e. Né 25521 550 500 100% 5413 90.0% 707 V468661 Classing werder a 44 440 100% 16-13 95.3% 925 Imt784641 Existemal man 26 25 154 16-70 100.00% 1918 M2023231 Verder Childrond 26 27 276 494 8-70 100.00% 1918 M2023231 Verder Childrond
Adecule type Duery Length Dther reports Descriptions Sequences Select all Select all Select all Sel	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XLTRE30.aGFP: complete assuence KLTRE30.aGFP: complete assuence der cOBR.complete assuence	Download Control Control Show 500 Image GenBank Graphics Distance tree of results Control MAC Mace
Wolecule type Duery Length Dther reports Sequences select all Wedder (U.V. Vedder (U.V. Vedder (U.V. Sciences Select all Select all Select all	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE3G act Complete assumce ATRE3G act Complete assumce ATRE3G act Complete assumce complete assumce complet	Download Image Columns Show 500 Image GenBank Granbes Distance tree of results Image M48 Verset Scientific Common Tadi Scient Cover Value Image M48 Verset Name Tadi Scient Cover Value Image M48 Verset Verder.e. Na 2491125 549 549 100% 44-152 96.7% 630 Imt221041 Verder.e. Né 2491125 549 549 100% 44-152 96.7% 630 Imt221041 Verder.e. Né 25521 550 500 100% 5413 90.0% 707 V468661 Classing werder a 44 440 100% 16-13 95.3% 925 Imt784641 Existemal man 26 25 154 16-70 100.00% 1918 M2023231 Verder Childrond 26 27 276 494 8-70 100.00% 1918 M2023231 Verder Childrond
Adecule type Duery Length Other reports Sequences select all yeader clux yeader clux yeader clux claning xe claning xe xe xe xe xe xe xe xe xe xe xe xe xe x	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments <i>O sequences selected</i> Description XTRE3G 60FP complete sequence XTRE3G 60FP complete sequence sequences selected description Complete sequence description Complete sequence description descrip	Download Control Control Show 500 Image GenBank Graphics Distance tree of results Control MAC Mace
Addecule type Duery Length Other reports Sequences select all year and Lendrat du C. Coning us C. Coning us C	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments O sequences selected Descripton XTRE36.6FP.complete sequence XTRE36.6FP.complete sequence XTRE36.0FP.complete sequence data SRNA EYFP-P2A Puro.complete sequence data SRNA EYFP-P2	Download × Image Columns × Show 500 ✓ Image Columns × Show Source Columns × S
Wolecule type Query Length Dther reports Sequences Select all Westor KW Westor KW Results for Program Database Query ID Description	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description X TRESG soft Complete assuence MARSIG and complete assuence STRESG soft complete assuence tansfar vedor cHSCAW, complete assuence data SNRA EYPP-P2A Puro, complete assuence data SNRA EYPP-P2A	Download × Image Columns × Show 500 ▼ Image Columns × Show
Molecule type Query Length Dther reports Sequences select all wedar ELX Wedar ELX Catalogue Cata	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE30 GCFP complete assummers XTRE30 GCFP complete assummers XTRE30 GCFP complete assummers ALTRE30 GCFP complete assummers ALTRE30 GCFP complete assummers ALTRE30 GCFP complete assummers ALTRE30 GCFP complete assummers Complete assummers ALTRE30 GCFP complete assummers ALTRE30 GCFP c	Download ∨ Image Columns ∨ Show 500 ∨ Image Columns ∨ Gentlank Graphics Distance tree of results Image Miss Miss Viewer Name Taid Max Taid Over value Image Miss Miss Viewer Name Taid Max Taid Over value Image Miss Miss Miss Miss Miss Miss Miss Mis
Molecule type Query Length Other reports Sequences Select all Wedar (UX Wedar (UX	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Descripton XTRE30.60FP.complete sequence XTRE30.60FP.complete sequence ATRE30.60FP.complete sequence dors SREALAYTP-P2A.Pure.complete sequence dors SREALAYTP	Download × Image Columns × Show 500 ▼ Image Columns × Show
Molecule type Query Length Dther reports Sequences select all Vector eVX Vector eVX	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE30 GCFP complete assummers XTRE30 GCFP complete assummers XTRE30 GCFP complete assummers ALTRE30 GCFP complete assummers ALTRE30 GCFP complete assummers ALTRE30 GCFP complete assummers ALTRE30 GCFP complete assummers Complete assummers ALTRE30 GCFP complete assummers ALTRE30 GCFP c	Download × Image Columns × Show 500 ▼ Image Columns × Show
Molecule type Query Length Dther reports Sequences select all Vector eVX Vector eVX	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE3G act Complete assumme KTRE3G act Complete assumme KTRE3G act Complete assumme to SRE3G act Complete assumme complete assumme Kerd Complete assumme (art COBB, complete assumme complete assumme (art COBB, complete assu	Download × Image Columns × Show 500 ▼ Image Columns × Show
Molecule type Query Length Dther reports Sequences Select all Select all Sele	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE3G.GCFP.complete assumes KTRE3G.GCFP.complete a	Download ✓ Image Columns ✓ Show 500 ✓ € Gendlant Graphics Distance tree of results Image 144 Versue Name Taution Graphics Distance tree of results Image 144 Versue Name Taution Graphics Distance tree of results Image 144 Versue Vector.e. Nd 2491125 549 549 100% 44-152 9677% 830 MH221041 Vector.e. Nd 2491125 549 549 100% 44-152 9677% 830 MH221041 Vector.e. Nd 2491125 549 549 100% 54-139 9302 MH28451041 Vector.e. Nd 2491125 549 549 100% 54-139 932 MH28451 Cloning vector.g 44 44 100% 14-122 9577% 649 6495 100 1
Adecule type Duery Length Other reports Sequences select all select all selec	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE3G act Complete assumme KTRE3G act Complete assumme KTRE3G act Complete assumme to SRE3G act Complete assumme complete assumme Kerd Complete assumme (art COBB, complete assumme complete assumme (art COBB, complete assu	Download ➤ Image Columns ➤ Show 500 ▼ Image Columns ➤ Show
Molecule type Query Length Dther reports Sequences Select all Website table Select all Website table Calanaxe C	dna 300 Distance tree of results MSA viewer C Graphic Summary producing significant alignments 0 sequences selected Cesciption XTRE30 soft: complete sequence XTRE30 soft: complete sequence Cesciption XTRE30 soft: complete sequence Cesciption Cescipti	Download V Cardbank Grandva Delance tree of result Conditional Viewer Name Cardbank Grandva Delance tree of result Conditional Viewer Name Tand Kas Tand Conditional Viewer E Fet Acc Accession Viedrace-1k6 2491125 549 549 100% 4+152 957% 6536 Hit2S210141 Viedrace-1k6 2491125 549 549 100% 4+152 957% 6536 Hit2S210141 Viedrace-1k6 2491125 549 549 100% 4+152 957% 6536 Hit2S210141 Viedrace-1k6 2491125 549 549 100% 5401 Hit2S210141 Classovardarca 49 440 100% 1+122 957% 1019 Mit2S220111 Classovardarca
Molecule type Query Length Dther reports Sequences Select all Website table Select all Website table Calanaxe C	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE3G.GCFP.complete assumes KTRE3G.GCFP.complete a	Download V Cardbank Grandva Delance tree of result Conditional Viewer Name Cardbank Grandva Delance tree of result Conditional Viewer Name Tand Kas Tand Conditional Viewer E Fet Acc Accession Viedrace-1k6 2491125 549 549 100% 4+152 957% 6536 Hit2S210141 Viedrace-1k6 2491125 549 549 100% 4+152 957% 6536 Hit2S210141 Viedrace-1k6 2491125 549 549 100% 4+152 957% 6536 Hit2S210141 Viedrace-1k6 2491125 549 549 100% 5401 Hit2S210141 Classovardarca 49 440 100% 1+122 957% 1019 Mit2S220111 Classovardarca
Molecule type Query Length Dther reports Sequences Select all Website table Select all Website table Calanaxe C	dna 300 Distance tree of results MSA viewer C Graphic Summary producing significant alignments 0 sequences selected Cesciption XTRE30 soft: complete sequence XTRE30 soft: complete sequence Cesciption XTRE30 soft: complete sequence Cesciption Cescipti	Download × Image Columns × Show 500 ▼ Image Columns × Show 500 ▼ Image Columns × Show 500 ▼ Image Columns × Show Soo ▼ Image Columns × Soo × Image Columns × Show Soo ×
Wolecule type Query Length Other reports Sequences select all Wetch (UK) Wetch (UK) Sequences Sequences Control (Control	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE30.60FP.comolete assuence ATRE30.60FP.comolete assuence sequences selected Comolete assuence ctor sOBR.comolete assuence ctor sOBR.com	Download Image Manage Columns Show 500 Image Manage Columns Show 500 Image Manage Columns Image
Molecule type Query Length Dther reports Sequences Select all Wester table Wester table Wester table Select all Wester table Description Description Sequences Sequences Sequences	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description X TREBG eGFP complete sequence RTREBG eGFP complete sequence AttrEBG eGFP complete sequence tota shPIA EYPPP2 Purp complete sequence data shPIA EYPPP2 Purp Complete sequence (edf cOBF, complete sequence AttrEBG eGFP complete sequence Description XTREBG eGFP complete sequence	Download Image Columns Show 500 Image Columns Gendlank Granthas Distance tree of results Image Mage Image Columns Image Mage
Molecule type Query Length Dther reports Sequences Select all Wether the select Select all Control of the select Select all Control of the select Control	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description XTRE3G.GCFP.complete assumces XTRE3G.GCFP.complete assumces XTRE3G.GCFP.complete assumces AttrE3G.GCFP.complete assumces der GDBR.complete assumces der GDBR.complete assumces AttrE3G.GCFP.complete assumces der GDBR.complete assumces der GDRR.complete a	Download Image Columns Show 500 Image Columns Gerifland Graphics Detance tree of results Image Columns Image Colum
Sequences Select all S	dna 300 Distance tree of results MSA viewer Graphic Summary producing significant alignments 0 sequences selected Description X TREBG eGFP complete sequence RTREBG eGFP complete sequence AttrEBG eGFP complete sequence tota shPIA EYPPP2 Purp complete sequence data shPIA EYPPP2 Purp Complete sequence (edf cOBF, complete sequence AttrEBG eGFP complete sequence Description XTREBG eGFP complete sequence	Download Image Columns Show 500 Image Columns Gendlank Granthas Distance free of results Image Mage Image Columns Image Mage

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.

Chinese pa	ngolin				
Identifiers	BioSample: SAMN1689	15765; Sample name: Sample3814; SRA: SRS7809866	Sunda pang		
Organism	Manis pentadactyla (Ch	I-\	Identifiers	BioSample: SAMN1685	85764; Sample name: Sample3826; SRA: SRS7809864
organism	cellular organisms; Eukary	imee Parquenin) Kuz Opishokonai, Metazoa; Eumetazoa; Bilateria: Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Sarcopterygii, Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Laurasiatheria;	Organism		n pangolin) ota: Opisithkonta: Metazoa: Eumetazoa: Bilateria: Deuterostomia: Ohordata: Craniata: Vertebrata: Grathostomata: Sarcopterygil: Diprotetrapodomorpha: Tetrapoda; Amnica: Nammalia: Theria: Eureheria: Borneeutheria: Laurasiatheria:
Package	Model organism or anin	nal: version 1.0			
Attributes	strain	not collected	Package	Model organism or anin	nal: version 1.0
100000	isolate	not collected	Attributes	strain	not collected
	breed	not collected		isolate	not collected
	cultivar	not collected		breed	not collected
	ecotype	not collected		cultivar	not collected
	age	not collected		ecotype	not collected
	development stage	not collected		age	not collected
	sex	male		development stage	not collected
	tissue	primary fibroblasts		sex	missing
	biomaterial provider	Polina Perelman		tissue	primary fibroblasts
	birth date	not collected		biomaterial provider	Polina Perelman
	birth location	not collected		birth date	not collected
	breeding history	not collected		birth location	not collected
	breeding method	not collected		breeding history	not collected
	cell line	not collected		breeding method	not collected
	cell subtype	not collected		cell line	not collected
	cell type	not collected		cell subtype	not collected
	collected by	not collected		cell type	not collected
	collection date	2020-03-20		collected by	not collected
	culture collection	not collected		collection date	2020-04-08
	death date	not collected		culture collection	not collected
	disease	not collected		death date	not collected
	disease stage	not collected		disease	not collected
	genotype	not collected		disease stage	not collected
	geographic location	not collected		genotype	not collected
	growth protocol	not collected		geographic location	not collected
	health state	n/a		growth protocol	not collected
	isolation source	not collected		health state	n/a
	latitude and longitude	not collected		isolation source	not collected
	phenotype	not collected		latitude and longitude	
	sample type	not collected		phenotype	not collected
	specimen voucher	not collected		sample type	not collected
	storage conditions	not collected		specimen voucher	not collected
	stud book number	not collected		storage conditions	not collected
	treatment	not collected		stud book number	not collected
Description	not collected			treatment	not collected
Description	not contected		Description	not collected	
BioProject	PRJNA512907 DNA Zoo				
	Retrieve all samples fro	m this project	BioProject	PRJNA512907 DNA Zor Retrieve all samples fro	
				Revieve all samples tro	ur nus biolect

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.fa

and

SV40 LTA+CMV from SRR11306689.fa

Conclusions

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

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

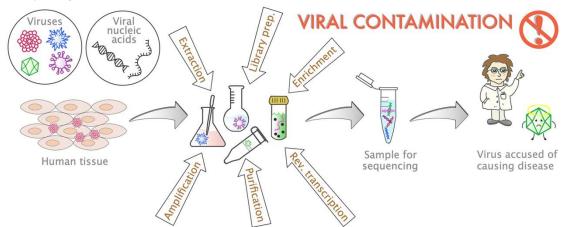


Figure 25. A cartoon diagram of contamination in sequencing experiment leading to false results and false "discoveries".

Note as in 2020/7/23

A recent Dataset, seemingly unrelated to the Xiao et.al Nature dataset, <u>SRX8582289</u>, appeared under <u>PRJNA607174</u>. 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	in Krona and read size	"Coronaviridae"
	percentage	by Кbp	as by Trace and
			total read size
<u>SRX8582289</u>	Manis javanica: 43.52%	Catarrhini 98913	Pangolin
22-Jun-2020			coronavirus 792

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

escription	Homo sapiens BAC clone	RP11-460N20 from 7, complete sequ	Percent Identity	E va	alue			Query Coverage		
olecule type	nucleic acid		to			to		to		
uery Length	203396									
ther reports	Distance tree of results	ISA viewer 😮						Filter Reset		
Descriptions	Graphic Summary	Alignments								
Sequences	producing significant a	lignments	Downlo	ad ~	Ma	nage C	olumns	✓ Show 100 ✔		
🗹 select all	100 sequences selected						<u>Grap</u>	hics Distance tree of result		
		Description	Ma Sco		Query Cover	E value	Per. Ident	Accession		
SRX85822	<u>89</u>		27	8 278	0%	8e-69	100.00%	SRA:SRR12053850.88444297.		
SRX85822	<u>89</u>		27	8 402	0%	8e-69	100.00%	SRA:SRR12053850.83916175.		
SRX85822	<u>89</u>		27	8 388	0%	8e-69	100.00%	SRA:SRR12053850.83916175		
SRX85822	<u>89</u>		27	8 278	0%	8e-69	100.00%	SRA:SRR12053850.82221130.		
SRX85822	<u>89</u>		27	8 278	0%	8e-69	100.00%	SRA:SRR12053850.71234261		
SRX85822	<u>89</u>		27	8 278	0%	8e-69	100.00%	SRA:SRR12053850.71234261		
SRX85822	<u>89</u>		27	8 5169	2%	8e-69	100.00%	SRA:SRR12053850.51889132		
SRX85822	<u>89</u>		27	8 7268	3%	8e-69	100.00%	SRA:SRR12053850.26027930		
SRX85822	<u>89</u>		27	8 5671	2%	8e-69	100.00%	SRA:SRR12053850.21554419		
SRX85822	<u>89</u>		27	8 278	0%	8e-69	100.00%	SRA:SRR12053850.13271287		
SRX85822	<u>89</u>		27	8 4760	1%	8e-69	100.00%	SRA:SRR12053850.62042.2		

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

escription	gnl SRA SRR12053850.8222	21130.2 82221130 (Biological)		to		to				to
olecule type	dna							L			
uery Length	150									Filte	Reset
her reports	Distance tree of results										
Descriptions	Graphic Summary	Alignments	Taxonomy								
Sequences pr	roducing significant al	lignments			Download >	М	anage	Colum	ns ~	Show	1000 🗸 🤇
select all 55	9 sequences selected						<u>GenB</u>	ank _	Graphic	<u>s Dista</u>	nce tree of resul
		Desa	iption			Max Score	Total Score	Query Cover	E value	Per. Ident	Accession
Homo sapier	ns general transcription factor I	<u>lli pseudogene 14 (G1</u>	F2IP14) on chromosome	7		276	276	99%	3e-70	100.00%	NG_043494.1
Homo sapier	ns FOSMID clone ABC13-4884	0700E15 from chrom	osome 7, complete sequ	ence		276	276	99%	3e-70	100.00%	AC242196.4
Homo sapier	ns BAC clone RP11-460N20 fro	om 7, complete seque	nce			276	276	99%	3e-70	100.00%	AC073210.8
Homo sapier	ns general transcription factor I	<u>lli pseudogene 5 (GTF</u>	2IP5) on chromosome 7			270	270	99%	1e-68	99.33%	NG 026590.1
Pan troglody	tes BAC clone CH251-340l24 fi	rom chromosome 7, o	complete sequence			270	270	99%	1e-68	99.33%	AC185242.2
Pan troglody	tes BAC clone CH251-623C19	from chromosome 7,	complete sequence			270	270	99%	1e-68	99.33%	AC184799.2
Pan troglody	tes BAC clone CH251-114G16	from chromosome 7,	complete sequence			270	270	99%	1e-68	99.33%	AC183835.2
Pan troglody	tes BAC clone CH251-2O15 fro	om chromosome 7, co	mplete sequence			270	270	99%	1e-68	99.33%	AC174000.3
Homo sapier	ns BAC clone RP11-47909 from	<u>m 7, complete sequer</u>	ice			270	270	99%	1e-68	99.33%	AC073107.7
Pan troglody	tes BAC clone CH251-565C10	from chromosome 7,	complete sequence			270	270	99%	1e-68	99.33%	AC148313.3
homo sapier	ns BAC clone CH17-99D2 from	i chromosome 4, com	plete sequence			226	226	99%	3e-55	93.96%	AC278002.1
Homo sapier	ns FOSMID clone ABC27-15411	13 from chromosome	4, complete sequence			226	226	99%	3e-55	93.96%	AC240529.1
Pan troglody	tes BAC clone CH251-4D23 fro	om chromosome 7, co	mplete sequence			226	226	99%	3e-55	93.96%	AC148834.3
Eukaryotic synthe	etic construct chromosome 10	<u>6</u>					193	193	99%	3e-45	90.07% <u>CP034</u>
Eukaryotic synth	etic construct chromosome 19	9					182	182	99%	5e-42	88.74% <u>CP034</u>
Eukarvotic synth	etic construct chromosome 19	9					182	182	99%	5e-42	88.74% CP034

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, <u>SRX9714436</u> and <u>SRX9714921</u>, 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	Primary Mammalian	Primate-related results	Identification of
registration date	Trace results and	in Krona and read size	"Coronaviridae"
	percentage	by Кbp	as by Trace and
			total read size
SRX9714436	Manis javanica: 3.14%	Homo sapiens 12332	Pangolin
	Homo sapiens: 0.04%		coronavirus 3
<u>SRX9714921</u>	Homo sapiens: 0.15%	Homo sapiens 9923	N/D

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

>gnl|SRA|SRR13285085.43 43

CTATACAACAAACCCCCCATGACACGAGTTTACCTATGTAACAAACCTTCA

select all 100 sequences selected					GenB	ank	Graphics	Distance	e tree of results
Description		Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Homo sapiens BAC clone RP11-488C22 from 2, complete sequence	<u>e</u>	<u>human</u>	93.5	93.5	100%	5e-16	100.009	6 165351	AC019109.9
Pan troglodytes BAC clone CH251-617M1 from chromosome unkno	wn, complete sequence	<u>chimpanzee</u>	87.9	87.9	100%	2e-14	98.00%	188195	AC183921.2
Pan troglodytes BAC clone CH251-564M21 from chromosome unkn	iown, complete sequence	<u>chimpanzee</u>	87.9	87.9	100%	2e-14	98.00%	201789	AC160021.3
Homo sapiens chromosome 5, BAC clone 89K19 (LBNL H179), cor	nplete sequence	<u>human</u>	87.9	87.9	100%	2e-14	98.00%	110312	AC005215.2
Homo sapiens chromosome 5 clone CTD-2332G20, complete sequ	Jence	<u>human</u>	87.9	87.9	94%	2e-14	100.009	6 146437	AC010489.4
Eukaryotic synthetic construct chromosome Y		eukaryotic syn	84.2	168	90%	3e-13	100.009	64558752	CP034510.1
Pongo abelii BAC clone CH276-222M18 from chromosome 8, comp	<u>plete sequence</u>	Sumatran ora	84.2	84.2	90%	3e-13	100.009	6 196788	AC206339.2
Description	Common Na	ime			otal Q	uery		Per. Acc.	Len Accession

	Description	Common Name	Score	Score	Cover	value •	Ident	Acc. Len	Accession
	Eukaryotic synthetic construct chromosome Y	eukaryotic synthetic construct	84.2	168	90%	3e-13	100.00%	64558752	CP034510.1
≤	Eukaryotic synthetic construct chromosome 15	eukaryotic synthetic construct	80.5	80.5	92%	4e-12	97.83%	82521392	CP034493.1
	Eukaryotic synthetic construct chromosome 14	eukaryotic synthetic construct	76.8	151	94%	5e-11	95.74%	88289540	CP034492.1
	Eukaryotic synthetic construct chromosome 18	eukaryotic synthetic construct	75.0	210	100%	2e-10	95.65%	84899006	CP034496.1
	Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	96089878	CP034516.1
	Eukaryotic synthetic construct chromosome 16	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	98200793	CP034494.1
	Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct	71.3	71.3	94%	2e-09	93.62%	96089878	CP034491.1
	Eukaryotic synthetic construct chromosome 20	eukaryotic synthetic construct	69.4	138	86%	8e-09	95.35%	68480253	CP034499.1
✓	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 ATAGGGAAGTGTGGTACCAAGGAGCAATATTCAATACAGCAACCAGGAAG

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

Fig.S2B: Homo Sapiens 100% full-length Unique matched read obtained from <u>SRX9714921</u> A provided .fastq file was also found in <u>SRX9714436</u>. 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
--	----	------------	-------	-----

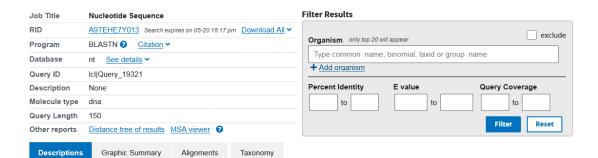
50M * 0 0 TGAAGGTTTGTTACATAGG	TAAACTCGTGTCATGGGGGGTTTGTTGTATAG *
-------------------------------	------------------------------------

Description		Common Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	/	Accession
Homo sapiens BAC clone RP11-488C22 from 2, complete	e sequence	human	93.5	93.5	100%	5e-16	100.00%	165351	AC0	<u>19109.9</u>
Pan troglodytes BAC clone CH251-617M1 from chromoso	<u>me unknown, complete sequence</u>	<u>chimpanzee</u>	87.9	87.9	100%	2e-14	98.00%	188195	AC1	<u>83921.2</u>
Pan troglodytes BAC clone CH251-564M21 from chromos	ome unknown, complete sequence	<u>chimpanzee</u>	87.9	87.9	100%	2e-14	98.00%	201789	AC1	<u>60021.3</u>
Homo sapiens chromosome 5, BAC clone 89K19 (LBNL	H179), complete sequence	<u>human</u>	87.9	87.9	100%	2e-14	98.00%	110312	AC0	05215.2
Homo sapiens chromosome 5 clone CTD-2332G20, com	plete sequence	<u>human</u>	87.9	87.9	94%	2e-14	100.00%	146437	AC0	<u>10489.4</u>
Eukaryotic synthetic construct chromosome Y		eukaryotic syn	84.2	168	90%	3e-13	100.00%	64558752	CP0	<u>34510.1</u>
Pongo abelii BAC clone CH276-222M18 from chromosom	ie 8, complete sequence	Sumatran ora	84.2	84.2	90%	3e-13	100.00%	196788	AC2	<u>06339.2</u>
Human DNA sequence from clone RP11-987D21 on chro	<u>mosome X, complete sequence</u>	<u>human</u>	84.2	84.2	96%	3e-13	97.92%	55442	<u>BX1</u>	<u>19919.5</u>
PREDICTED: Callithrix jacchus uncharacterized LOC1181	54814 (LOC118154814), ncRNA	white-tufted-e	82.4	82.4	94%	1e-12	97.87%	2719	<u>XR</u>	0047450
Homo sapiens solute carrier family 26 member 3 (SLC26	A3), RefSeqGene (LRG_683) on chromos.	<u>human</u>	82.4	82.4	100%	1e-12	96.00%	44767	<u>NG</u>	008046.1
Eukaryotic synthetic construct chromosome Y	eukaryotic synthetic construct		4	34.2	168 9	90% 3	e-13 100	.00% 6455	8752	CP0345
Eukaryotic synthetic construct chromosome 15	eukaryotic synthetic construct		4	30.5	80.5 9	92% 4	e-12 97.	83% 8252	1392	<u>CP0344</u>
Eukaryotic synthetic construct chromosome 14	eukaryotic synthetic construct		1	76.8	151 9	94% 5	ie-11 95	74% 8828	9540	<u>CP0344</u>
Eukaryotic synthetic construct chromosome 18	eukaryotic synthetic construct		1	75.0	210 1	00% 2	e-10 95	65% 8489	9006	<u>CP0344</u>
Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct		1	71.3	71.3 9	94% 2	e-09 93	62% 9608	9878	<u>CP0345</u>
Eukaryotic synthetic construct chromosome 16	eukaryotic synthetic construct		1	71.3	71.3 9	94% 2	e-09 93	62% 9820	0793	<u>CP0344</u>
Eukaryotic synthetic construct chromosome 13	eukaryotic synthetic construct		1	71.3	71.3 9	94% 2	e-09 93	62% 9608	9878	<u>CP0344</u>
Eukaryotic synthetic construct chromosome 20	eukaryotic synthetic construct			59.4	138 8	86% 8	e-09 95	35% 6848	0253	<u>CP0344</u>
Eukaryotic synthetic construct chromosome 19	eukaryotic synthetic construct			621	621 8	34% 1	e-06 92	86% 6424	2768	CP0344

Fig.S3: BLAST result of the read NDX550397_RUO:309:H3FKWBGXH:1:11101:21446:1055 TGAAGGTTTGTTACATAGGTAAACTCGTGTCATGGGGGGTTTGTTGTATAG 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: <u>http://dx.doi.org/10.26036/CNP0001573</u>. 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.



~	select all 100 sequences selected	GenBank Graphics				Distar	nce tree d	New MSA View	
	Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
2	Eukaryotic synthetic construct chromosome Y	eukaryotic synt	278	1369	100%	1e-70	100.00%	64558752	CP034510.1
1	Homo sapiens chromosome Y palindromes P1, P2, P3 and inverted repeat IR2 (P1-P2-P3-IR2@)	Homo sapiens	278	1112	100%	1e-70	100.00%	4538292	NG 004755.2
	Homo sapiens DNA, chromosome Y, nearly complete genome	Homo sapiens	278	406	100%	1e-70	100.00%	50294361	AP023484.1
1	Homo sapiens BAC clone RP11-245K4 from Y, complete sequence	<u>Homo sapiens</u>	278	278	100%	1e-70	100.00%	182083	AC007965.3
1	Homo sapiens BAC clone RP11-553C13 from Y, complete sequence	<u>Homo sapiens</u>	278	278	100%	1e-70	100.00%	184215	AC007322.4
1	Homo sapiens BAC clone RP11-506M9 from Y, complete sequence	<u>Homo sapiens</u>	278	278	100%	1e-70	100.00%	166436	AC016752.2
1	Homo sapiens BAC clone RP11-427G18 from Y complete sequence	<u>Homo sapiens</u>	278	278	100%	1e-70	100.00%	205237	AC008175.2
1	Pan troglodytes BAC clone CH251-412D24 from Y, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	166356	AC147572.4
1	Pan troglodytes BAC clone CH251-506P22 from Y, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	195158	AC147608.1
1	Pan troglodytes BAC clone CH251-475M5 from Y, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	220557	AC146198.2
•	Pan troglodytes BAC clone CH251-528H6 from Y, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	169760	AC147361.1
2	Pan troglodytes BAC clone CH251-548014 from Y, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	229006	AC147114.2
•	Pan troglodytes BAC clone CH251-113M4 from Y, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	183460	AC146515.2
•	Pan troglodytes BAC clone RP43-5I10 from chromosome y, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	183293	AC142348.1
•	Pan troglodytes chromosome Y clone: PTB-391J06, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	171013	<u>BS000677.1</u>
2	Pan troglodytes BAC clone CH251-571K7 from chromosome y, complete sequence	Pan troglodytes	257	257	98%	1e-64	97.97%	171231	AC148933.3
•	Homo sapiens chromosome Y palindromes P4 and P5 (P4-P5@) on chromosome Y	<u>Homo sapiens</u>	128	257	68%	1e-25	89.32%	1351786	NG_004636.1
2	Homo sapiens heat shock transcription factor Y-linked 1 (HSFY1), RefSeqGene on chromosome Y	<u>Homo sapiens</u>	128	128	68%	1e-25	89.32%	49276	NG_012030.1
1	Homo sapiens BAC clone RP11-569J3 from Y complete sequence	<u>Homo sapiens</u>	128	128	68%	1e-25	89.32%	156357	AC022486.4
•	Homo sapiens BAC clone RP11-143C1 from Y, complete sequence	<u>Homo sapiens</u>	128	128	68%	1e-25	89.32%	174082	AC007379.2
1	PREDICTED: Castor canadensis glycoprotein M6B (Gpm6b), transcript variant X8, mRNA	Castor canade	117	117	86%	2e-22	82.98%	4578	XM_02015996
•	PREDICTED: Castor canadensis glycoprotein M6B (Gpm6b), transcript variant X7, mRNA	Castor canade	117	117	86%	2e-22	82.98%	3225	XM_02015995
	PREDICTED: Castor canadensis glycoprotein M6B (Gpm6b), transcript variant X6, mRNA	Castor canade	117	117	86%	2e-22	82.98%	4634	XM 02015994
2	PREDICTED: Castor canadensis glycoprotein M6B (Gpm6b), transcript variant X5, mRNA	Castor canade	117	117	86%	2e-22	82.98%	3033	XM_02015994
•	PREDICTED: Castor canadensis glycoprotein M6B (Gpm6b), transcript variant X4, mRNA	Castor canade	117	117	86%	2e-22	82.98%	3134	XM_02015993
	PREDICTED: Castor canadensis glycoprotein M6B (Gpm6b), transcript variant X3, mRNA	Castor canade	117	117	86%	2e-22	82.98%	3345	XM_02015992
1	PREDICTED: Castor canadensis glycoprotein M6B (Gpm6b), transcript variant X2, mRNA	Castor canade	117	117	86%	2e-22	82.98%	4754	XM_02015991
1	PREDICTED: Castor canadensis glycoprotein M6B (Gpm6b), transcript variant X1, mRNA	Castor canade	117	117	86%	2e-22	82.98%	3275	XM_02015991
1	PREDICTED: Arvicanthis niloticus glycoprotein M6B (Gpm6b), transcript variant X6, mRNA	Arvicanthis nilo	115	115	69%	8e-22	86.36%	2986	XM_03448486

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]

REFERENCES

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

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

Published: May 14, 2020

https://doi.org/10.1371/journal.ppat.1008421

[2] Xiao, K., Zhai, J., Feng, Y. *et al.* Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* (2020). <u>https://doi.org/10.1038/s41586-020-2313-x</u>

[3] Lam, T.T., Shum, M.H., Zhu, H. *et al.* Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* (2020). <u>https://doi.org/10.1038/s41586-020-2169-0</u>

[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. Brisse1,2 and Hinh Ly2,*

[8] Evolutionary origins of the SARS - CoV - 2sarbecovirus lineage responsible for the COVID-19 pandemicMaciej F Boni1*, Philippe Lemey2*, Xiaowei Jiang3, Tommy Tsan-Yuk Lam4, Blair Perry5, Todd Castoe5, Andrew Rambaut6 and David L Robertson7

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

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

Alexandre Hassanin 1 Huw Jones 2 Anne Ropiquet 2

[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, <u>https://doi.org/10.1093/nsr/nwaa031</u>

[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, <u>https://doi.org/10.1093/ve/veaa020</u>

[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. <u>Scaling read aligners to hundreds of threads</u> on general-purpose processors. <u>Bioinformatics</u> Vol 35, Iss 3, 2019, pp 421–432.

[18] <u>"A549 Cell Line: Human alveolar adenocarcinoma cell line -General Information"</u>. Retrieved
 3 January 2012. <u>http://a549.com/</u>

[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 & Crameri, 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