OPEN ACCESS Freely available online

Journal of Cell Signaling

Research Article

Identification *Sus scrofa* and *Mus musculus* as Potential Parasitifers of SARS-CoV-2 *via* Phylogenetic and Homologous Recombination Analysis

Fhang Z[1*], Li W[2], Xu H[2,3], He Z[3]

[1]Research Center of Integrative Medicine, School of Basic Medical Sciences, Guangzhou University of Chinese Medicine; Guangzhou 510405, China; [2]Shenzhen Samii Medical Center, Shenzhen 518000, China; [3]Department of Pharmacology, School of Medicine, Shenzhen University, Shenzhen 518000, China

ABSTRACT

Wuhan Huanan seafood wholesale market was highly suspected as the original place for the outbreak of SARS-CoV-2 previously. Most study focus on the wild animals being sold in the market, neglecting the livestock around the city Wuhan could be also reasonable suspect zero. Implementing phylogenetic and recombination analysis, both porcine and murine coronavirus may attend the reorganization and evolution of SARS-CoV-2. Overall, it's the first time to illustrate that swine and mice would be the probable reservoir for the SARS-CoV-2.

Keywords: SARS-CoV-2; PEDV; Human coronavirus HKU; SADS

INTRODUCTION

A novel coronavirus, SARS-CoV-2, was recently reported in the city Wuhan, causing severe respiratory diseases as well as epidemic all around the China. The inflection point of confirmed cases didn't occur until the February 2020, since the first case was hospitalized on the 12th of December 2019 [1]. By the end of 27th February, Chinese mainland reported 399 new confirmed cases of novel coronavirus infection, bringing the total to 78,630.29 new fatalities daily were reported and the cumulative fatalities were up to 2747. Meanwhile, outside the Chinese mainland, more than 800 cases have been confirmed in Asia, like Japan, Singapore, Thailand and South Korea, Europe, like Germany, France and the Americas. It's consensus that the SARS-CoV-2 origin from bats [2], but one or two intermediate host in nature is deemed as mediating human infection *via* gradually adapting the mechanism of transcription and translation in human body. However, the exact putative parent of SARS-CoV-2 remains unconcerned. As a matter of fact, either wild animal or rear livestock deserve the suspicion. In the past ten years, there were more than 300 strains of Porcine Epidemic Diarrhea Virus (PEDV) had been reported in China (Data collected from https://www.ncbi.nlm.nih.gov/labs/virus/). Hubei used to be the severely afflicted area. Therefore, we assume that SARS-CoV-2 correlated with swine to some extent. Furthermore, the result of phylogenetic and recombination analysis indicated that both *Sus scrofa* and *Mus musculus* could be ideal reservoirs for SARS-CoV-2.

MATERIALS AND METHODS

Sequence data collection

More than 300 genome sequences were obtained and analyzed from GenBank, including the novel SARS-CoV-2 (MN908947). Related alphacoronavirus hosting in *Sus scrofa* (n=163) and betacoronavirus hosting in *Mus musculus* (n=15) from China were refined and downloaded from NCBI virus database (Table 1). Here, the mitochondrial genes represent the whole genome of potential host. Clustalx1.83 was applied to align the sequences. Filtering the unrelated genome sequences, the remaining sequences were applied for further analysis (Tables 1 and 2). Coronavirus hosting in *Sus scrofa* from different regions of China were calculated (Table 3).

Phylogenetic analysis

MEGA X was applied to construct the phylogenetic trees using the Neighbor-Joining method [3]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches [4].

Correspondence to: Fhang Z, Research Center of Integrative Medicine, School of Basic Medical Sciences, Guangzhou University of Chinese Medicine, Guangzhou 510405, China, E-mail: zhangfengxue@gzucm.edu.cn

Received: February 28, 2020; **Accepted:** March 13, 2020; **Published:** March 20, 2020

Citation: Fhang Z, Xu H, Li W, He Z (2020) Identification *Sus scrofa* and *Mus musculus* as Potential Parasitifers Of SARS-Cov-2 *via* Phylogenetic and Homologous Recombination Analysis. J Cell Signal. 5:205. DOI:10.35248/2576-1471.20.5.205

Table 1: The information of genomic analysis for phylogenetic analysis.

NO	Gene Bank accession	Reference strains name	Species
1	MN908947	SARS-CoV-2	*Homo sapiens*
2	KF686346.1	Human_coronavirus_HKU1-2010	*Homo sapiens*
3	NC_006577.2	coronavirus_HKU1	*Homo sapiens*
4	MG428704.1	Human_coronavirus_NL63	*Homo sapiens*
5	MN996532.1	Bat_coronavirus_RaTG13	*Rhinolophus sinicus*
6	EF203064.1	Bat_coronavirus_HKU2	*Rhinolophus sinicus*
7	KY073747.1	Bat_coronavirus_BtKY229E-1	*Rhinolophus sinicus*
8	FJ647225	MHV-A59	*Mus musculus*
9	AC_000192.1	MHV JHM	*Mus musculus*
10	KY419105	PHEV _USA-15TOSU0582	*Sus scrofa*
11	MH697599	PEAV_GDS04_P12	*Sus scrofa*
12	NC_039208	Porcine_coronavirus_HKU15	*Sus scrofa*
13	JX501318.1	CH/HBQX/10	*Sus scrofa*
14	JX188454.1	AJ1102	*Sus scrofa*
15	JQ239429.1	CH1	*Sus scrofa*
16	KF453512.1	HB121229	*Sus scrofa*
17	JX501317.1	CH/CY/12	*Sus scrofa*
18	JX163294.1	HBMC2012	*Sus scrofa*
19	MK644602.1	L6-HB2017	*Sus scrofa*
20	MK644605.1	T10-HB2018	*Sus scrofa*
21	MK644604.1	N7-GD2017	*Sus scrofa*
22	MK584552.1	AJ1102 (F12)	*Sus scrofa*
23	MN037494	WHLL	*Sus scrofa*
24	MK644603.1	M3-SX2017	*Sus scrofa*
25	MF807952	B5-HB2017	*Sus scrofa*
26	MH748550.1	PEDV JS-A	*Sus scrofa*
27	MH708243.1	H11-SD2017	*Sus scrofa*
28	MF807951.1	C3-HB2017	*Sus scrofa*
29	MK138516.1	V7-HB2018	*Sus scrofa*
30	MK644601.1	G2-HE2017	*Sus scrofa*
31	KT021227	YN1	*Sus scrofa*
32	KT021233.1	YN200	*Sus scrofa*
33	KT021230.1	YN60	*Sus scrofa*
34	KT021229.1	YN30	*Sus scrofa*
35	KT021228.1	YN15	*Sus scrofa*
36	KT021232.1	YN144	*Sus scrofa*
37	KT021231.1	YN90	*Sus scrofa*
38	KX499468	TGEV	*Sus scrofa*

Abbreviations: PEDV: Porcine Epidemic Diarrhea Virus, MHV: Mouse Hepatitis Virus, PHEV: Porcine Hemagglutinating Encephalomyelitis Virus, PEAV: Porcine Enteric Alpha Coronavirus, TGEV: Transmissible Gastroenteritis Virus.

The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The complete genome of SARS-CoV-2 and other candidate genome would be aligned before implementing of phylogenetic analysis.

Synonymous codon usage analysis

Exerting the relative synonymous codon usage (RSCU) bias to estimate the potential host of SARS-CoV-2Table 4). Whole coding sequences of genomic were utilized for identifying the host among different viruses [5]. The RSCU were calculated with Codon W1.4.2. The heat map clustering of RSCU was relized *via* applying MeV 4.9.0. The homologous analysis was executive *via* analyzing the pairwise distance of mitochondrial genes from the potential host animal. The pairwise distances were computed with MEGA X in the bootstrap test (1000 replicates) to evaluate the potential hosts [6].

Genome homologous recombination analysis

The genome sequences of bat RaTG13, murine JHM, Human HKU1, PEDV H11-SD2017 and PEDV YN15 were obtained from NCBI virus database https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/. Potential recombination events of SARS-CoV-2 were implemented with Recombination Detection Program v4 (RDP4) firstlythen the Simplot (version 3.5.1) was utilized to verify the possible breakpoints. The method (RDP, GENECONV, Boot Scan, maximum chi square, Chimera, SISCAN, Phylpro, LARD, and 3SEQ) were taken into the

RDP4 analysis. Potential recombination events were characterized with similarity plots, accompanying with possible region of recombination.

Table 2: The information of potential host sequence.

NO	Gene Bank accession	Reference strains name	Species
1	MN908947	SARS-CoV-2 Sus_scrofa_domesticus_mitochondrion _complete_genome	*Homo sapiens*
2	NC_012095.1	Mus_musculus_mitochondrion_ complete_genome	*Sus scrofa*
3	NC_005089.1	Mustela_putorius_voucher_ROM1171 43_mitochondrion_complete_genome	*Mus musculus*
4	NC_020638.1	*Marmota flaviventris* mitochondrion complete_genome	*Mustela putorius*
5	NC_042243.1	_Homo_sapiens_mitochondrion _complete_genome	*Marmota flaviventris*
6	NC_012920.1	_Manis_pentadactyla_mitochondrion _complete_genome	*Homo sapiens*
7	JN411577.1	Rhinolophus_ sinicus _mitochondrion	*Manis pentadactyla*
8	NC_005434.1	_complete_genome	*Rhinolophus sinicus*

Table 3: 362 PEDV genomes distribution in China from 2010~2020. Note: the data were calculated from NCBI virus database:https://www.ncbi.nlm.nih.gov/labs/virus/.

NO.	Region	Amount of PEDV related strains (2010~2020)
1	Guangdong	93
2	Taiwan	40
3	Jiangsu	30
4	Hubei	21
5	Zhejiang	18
6	Sichuan	17
7	Shandong	16
8	Hebei	16
9	Henan	16
10	Beijing	13
11	Fujian	13
12	Shanghai	11
13	Jiangxi	10
14	Anhui	8
15	Gansu	8
16	Heilongjiang	8
17	Guangxi	6
18	Yunnan	5
19	Hunan	4
20	Jilin	4
21	Shanxi	2
22	Shaanxi	1
23	Guizhou	1
24	Hainan	1

RESULTS

Geographical hot map

The distribution of PEDV with Geographical hot map indicated the risk of pandemic of porcine epidemic diarrhea in different regions [7] (Figure 1). In the past ten years, there were more than 300 strains of PEDV had been reported in China. Guangdong province ranked top with 93 strains, being followed by Taiwan with 40 strains. It's fourthly with 21 strains when it comes to the amount of PEDV in Hubei. Higher amounts means higher risk being suffered from occasional cross species epidemics of coronavirus. Therefore, we believe that the outbreak of SARS-CoV-2 has the relevance with porcine coronavirus to some extent.

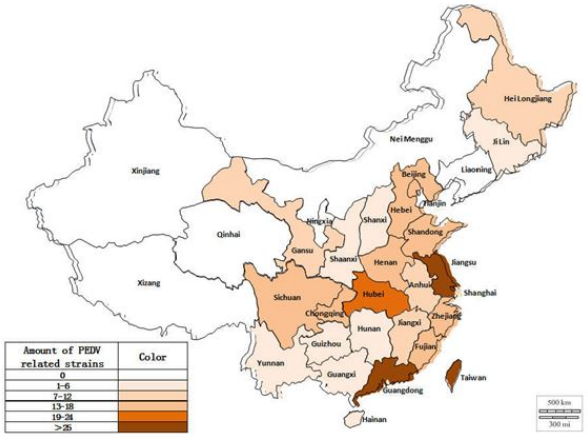

Figure 1: The hot map of 362 PEDV genomes distribution in China.

Note that the Porcine Epidemic Diarrhea Virus (PEDV) strains locating accurately indifferent provinces were collected from NCBI virus database between 2010-2020 (Table 3). The depth of the color represents the amount of PEDV related strains in distinctive regions. Map was obtained from www.d-maps.com.

The RSCU has been widely utilized to analyze the affiliation between virus and potential parasitifer [5]. The output from RSCU heat map basing on the enclidean distance inferred that both *Sus scrofa* and *Mus musculus* have similar synonymous codon usage bias with SARS-CoV-2 (Figure 2A). The enclidean distance between *Sus scrofa* and SARS-CoV-2 is minimums, hinting that SARS-CoV-2 could effectively use porcine translation machinery better than that of other animals, suggesting that the epidemic SARS-CoV-2 might originate from swine. The result of pairwise distance also supports the above judgments. Swine, mouse and SARS-CoV-2 deserves the shorter pairwise distance as compared to those of snake, marmota, mink, bat and human (Figure 2B). Pairwise Distance between the SARS-CoV-2 and *Sus scrofais* is 14.89, as for *Mus musculus* is 14.94. It indicated that SARS-CoV-2 may origin from both swine and mouse to some extent.

The intermediate animal reservoirs of the SARS-CoV-2 point to the pork and mouse

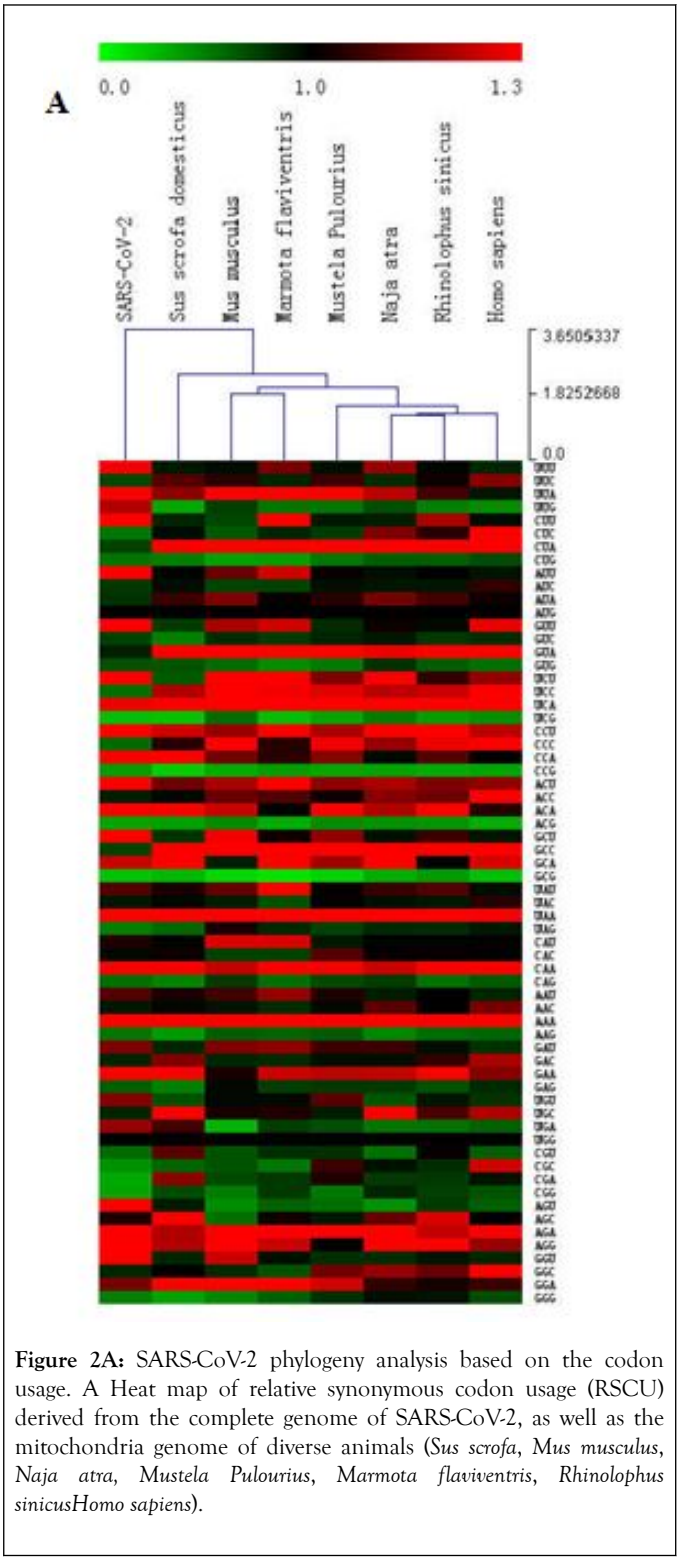

Figure 2A: SARS-CoV-2 phylogeny analysis based on the codon usage. A Heat map of relative synonymous codon usage (RSCU) derived from the complete genome of SARS-CoV-2, as well as the mitochondria genome of diverse animals (*Sus scrofa*, *Mus musculus*, *Naja atra*, *Mustela Pulourius*, *Marmota flaviventris*, *Rhinolophus sinicusHomo sapiens*).

Pairwise Distance	A	B	C	D	E	F	G	H
A	—	—	—	—	—	—	—	—
B	14.89	—	—	—	—	—	—	—
C	14.94	13.34	—	—	—	—	—	—
D	16.95	16.13	18.90	—	—	—	—	—
E	17.04	12.12	6.14	13.31	—	—	—	—
F	17.19	9.96	14.83	19.16	14.64	—	—	—
G	20.70	19.14	19.79	13.72	15.20	13.91	—	—
H	21.15	15.39	13.66	14.60	11.56	16.33	15.24	—

A: MN908947.3_Wuhan_seafood_market_pneumonia_virus_isolate_Wuhan-Hu-1_complete_genome;
B: NC_012095.1_Sus_scrofa_domesticus_mitochondrion_complete_genome;
C: NC_005089.1_Mus_musculus_mitochondrion_complete_genome;
D: NC_020638.1_Mustela_putorius_voucher_ROM117143_mitochondrion_complete_genome6;
E: NC_005434.1_Rhinolophus_sinicus_mitochondrion_complete_genome
F: NC_012920.1_Homo_sapiens_mitochondrion_complete_genome;
G: NC_042243.1_Marmota_flaviventris_mitochondrion_complete_genome;
H: EU579522.1_Bungarus_multicinctus_mitochondrion_complete_genome;

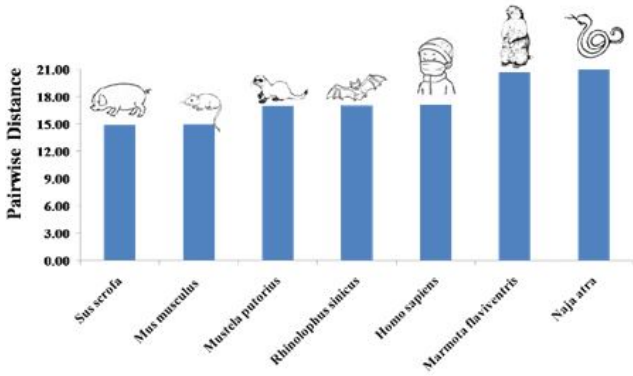

Figure 2B: Homology analysis between theSARS-CoV-2 and different animal species (Table 2), Pairwise Distance was applied to evaluate the homology as compared with SARS-CoV-2 *via* MEGA-X.

Phylogenetic analysis

The 21 PEDV strains were filtered from 2010~2020 in Hubei based on NCBI virus database. Phylogenetic analysis of the coronavirus (HKU15, PHEV, NL63, H11-SD2017, 229E, HKU2, TGEV et al) derived from varied species (Bat, Swine, Human), representing the sister lineage to SARS-CoV-2 with 99% bootstrap support (Figure 3) What's more, The RSCU depicted that related coronavirus have similar synonymous codon usage bias with SARS-CoV-2 (Figure 4). The other PEDV strains obtained from Hubei also showed the closely phylogenetic correlation with SARS-CoV-2. Overall, the close phylogenetic relationship to *Sus scrofa* provides evidence for bat-swine axis, being one of the origins of SARS-CoV-2.

Phylogeny-based geographical dissection of 21 PEDV strains (Brown box) derives from Hubei. The information of the 21 PEDV strains is shown in Table 3.

Meanwhile, the result of RSCU also presented that both SARS-CoV-2 and related coronavirus lean towards to having similar synonymous codon usage bias (Figure 4 and Table 5). Therefore, the origination of SARS-CoV-2 could be further focus on the coronavirus isolating from pork and mice. Particularly, the relationship with PEDV H11-SD2017 and PEDV YN15 needed to be further lucubrated.

Table 4: The RSCU analysis of the preferred codons derived from SARS-CoV-2 complete geneome and potential host mitochondrion complete genome.

Amino acid	Codon	A	B	C	D	E	F	G	H
Phe	UUU	1.29	0.9	0.95	1.18	0.93	1.14	0.98	0.84
	UUC	0.71	1.1	1.05	0.82	1.07	0.86	1.02	1.16
Leu	UUA	1.57	1.17	2.13	1.22	1.37	1.57	1.08	0.91
	UUG	1.21	0.34	0.75	0.69	0.53	0.54	0.49	0.44
	CUU	1.41	0.85	0.72	0.88	0.9	1.3	1.2	0.95
	CUC	0.53	0.95	0.65	1.15	0.76	0.84	1.05	1.34
	CUA	0.74	2.18	1.39	1.44	1.91	1.38	1.56	1.65
	CUG	0.54	0.5	0.36	0.63	0.53	0.38	0.62	0.7
Ile	AUU	1.43	0.98	1.12	0.94	1.01	1.25	1	0.92
	AUC	0.77	0.95	0.73	0.92	0.94	0.74	0.92	1.06
	AUA	0.8	1.07	1.15	1.14	1.05	1.01	1.08	1.02
Met	AUG	1	1	1	1	1	1	1	1

Val	GUU	1.7	0.76	1.2	1.02	0.83	1.25	0.96	1.28
	GUC	0.74	0.49	0.82	0.9	0.86	0.76	0.78	0.81
	GUA	0.88	2.09	1.4	1.28	1.79	1.54	1.62	1.34
	GUG	0.68	0.66	0.58	0.8	0.52	0.45	0.63	0.56
Ser	UCU	1.41	0.64	1.49	1.4	1.15	1.49	1.06	1.18
	UCC	0.56	1.21	1.5	1.22	1.27	1.29	1.24	1.33
	UCA	1.28	1.65	1.45	1.32	1.77	1.34	1.3	1.44
	UCG	0.25	0.23	0.57	0.51	0.35	0.25	0.38	0.43
Pro	CCU	1.47	1.25	1.18	1.51	1.2	1.52	1.3	1.22
	CCC	0.58	1.05	1.36	1.18	1.29	1.05	1.28	1.44
	CCA	1.58	1.5	1.14	0.96	1.16	1.05	1.1	0.99
	CCG	0.38	0.19	0.32	0.36	0.34	0.38	0.32	0.35
Thr	ACU	1.3	1.13	1.2	1.2	1.17	1.59	1.15	1.17
	ACC	0.88	0.97	1.12	1.16	1.02	1.1	1.13	1.46
	ACA	1.47	1.54	1.24	1.21	1.37	0.98	1.32	1.05
	ACG	0.35	0.37	0.45	0.43	0.44	0.32	0.41	0.32
Ala	GCU	1.8	0.79	1.29	0.93	1.17	1.02	1.07	0.92
	GCC	0.73	1.37	1.68	1.4	1.46	1.37	1.55	1.57
	GCA	1.24	1.55	0.88	1.35	1.19	1.5	0.98	1.26
	GCG	0.23	0.28	0.15	0.32	0.17	0.11	0.39	0.25
Tyr	UAU	1.1	1.03	1.12	1.07	1.01	1.32	1.1	0.95
	UAC	0.9	0.97	0.88	0.93	0.99	0.68	0.9	1.05
TER	UAA	1.33	1.33	1.69	1.61	1.58	1.41	1.63	1.5
	UAG	0.5	0.59	1.03	0.84	0.73	0.84	0.83	0.9
His	CAU	1.04	0.98	1.26	0.99	0.9	1.26	0.99	1.01
	CAC	0.96	1.02	0.74	1.01	1.1	0.74	1.01	0.99
Gln	CAA	1.43	1.52	1.23	1.24	1.29	1.44	1.45	1.36
	CAG	0.57	0.48	0.77	0.76	0.71	0.56	0.55	0.64
Asn	AAU	1.1	1.03	1.08	0.89	1.04	1.18	1	0.86
	AAC	0.9	0.97	0.92	1.11	0.96	0.82	1	1.14
Lys	AAA	1.47	1.6	1.37	1.53	1.36	1.4	1.42	1.4

	AAG	0.53	0.4	0.63	0.47	0.64	0.6	0.58	0.6
Asp	GAU	1.15	0.85	1.15	1.07	1.07	1.17	0.94	0.8
	GAC	0.85	1.15	0.85	0.93	0.93	0.83	1.06	1.2
Glu	GAA	1.38	1.5	1.03	1.23	1.22	1.24	1.32	1.16
	GAG	0.62	0.5	0.97	0.77	0.78	0.76	0.68	0.84
Cys	UGU	1.16	0.68	0.97	0.66	1.11	0.96	0.91	0.79
	UGC	0.84	1.32	1.03	1.34	0.89	1.04	1.09	1.21
TER	UGA	1.18	1.08	0.29	0.54	0.69	0.76	0.55	0.61
Trp	UGG	1	1	1	1	1	1	1	1
Arg	CGU	0.58	1.11	0.67	0.55	0.8	0.81	1.01	0.62
	CGC	0.43	0.6	0.67	0.93	1.07	0.54	0.82	1.25
	CGA	0.33	1.17	0.71	0.77	1.04	0.78	0.78	0.92
	CGG	0.35	0.7	0.51	0.86	0.53	0.78	0.75	0.67
Ser	AGU	1.46	0.89	0.44	0.43	0.56	0.61	0.77	0.62
	AGC	1.04	1.38	0.56	1.12	0.9	1.02	1.26	1
Arg	AGA	2.98	1.21	1.8	1.52	1.55	1.84	1.23	1.36
	AGG	1.32	1.21	1.65	1.36	1.01	1.24	1.41	1.18
Gly	GGU	1.51	0.84	1.24	0.82	0.81	0.93	0.92	0.84
	GGC	0.85	0.99	0.83	1.18	1.13	0.67	1.12	1.4
	GGA	1.14	1.81	1.47	1.06	1.25	1.8	1.03	1.07
	GGG	0.5	0.36	0.46	0.94	0.81	0.61	0.92	0.69

Note:AMN908947.3_Wuhan_seafood_market_pneumonia_virus_isolate_Wuhan-Hu-1_complete_genome (SARS-CoV-2)

B: NC_012095.1_Sus_scrofa_domesticus_mitochondrion_complete_genome

CNC_005089.1_Mus_musculus_mitochondrion_complete_genome

D EU579522.1_Bungarus_multicinctus_mitochondrion_complete_genome

E: NC_020638.1_Mustela_putorius_voucher_ROM117143_mitochondrion_complete_genome 6

FNC_042243.1_Marmota_flaviventris_mitochondrion_complete_genome

GNC_005434.1_Rhinolophus_sinicus_mitochondrion_complete_genome

HNC_012920.1_Homo_sapiens_mitochondrion_complete_genome

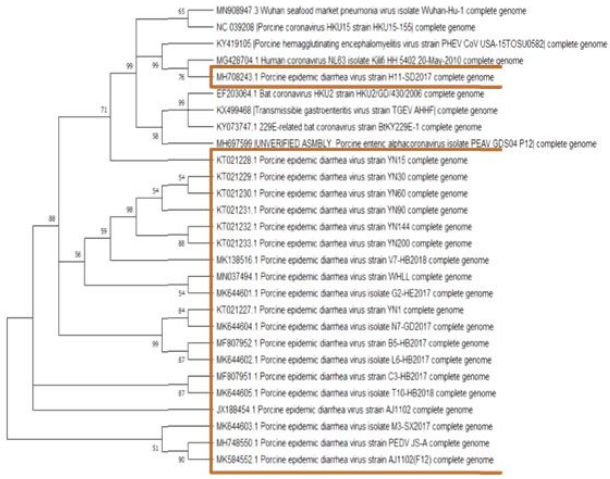

Figure 3: Phylogenetic tree of SARS-CoV-2 and related coronavirus strains. The neighbor-joining tree (bootstrap n=1,000; p-distance) was exerted to represent the evolutionary history of the taxa analyzed.

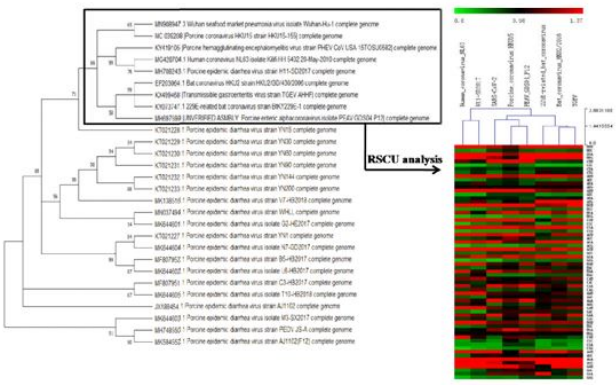

Figure 4: RSCU analysis of SARS-CoV-2 and related coronavirus. polyphyletic coronavirus. The enclidean distance was calculated to cluster the related coronavirus.

Table 5: The RSCU analysis of the preferred codons derived from SARS-CoV-2. complete geneome and related polyphyletic coronavirus complete genome

Amino acid	Codon	A	B	C	D	E	F	G	H
	UUU	1.29	1.54	1.43	1.44	1.15	1.38	1.62	1.35
Phe	UUC	0.71	0.46	0.57	0.56	0.85	0.62	0.38	0.65
	UUA	1.57	1.69	1.47	1.43	1.2	1.39	1.49	1.55
	UUG	1.21	1.65	1.72	1.68	0.93	1.48	1.64	1.64
	CUU	1.41	1.13	0.77	0.71	1.37	1.34	0.89	0.79
	CUC	0.53	0.26	0.31	0.29	0.74	0.49	0.29	0.32
	CUA	0.74	0.69	0.84	0.85	0.98	0.69	0.85	0.85
Leu	CUG	0.54	0.58	0.89	1.04	0.78	0.61	0.83	0.84
	AUU	1.43	1.73	1.43	1.37	1.23	1.66	1.44	1.3
	AUC	0.77	0.4	0.5	0.56	0.86	0.66	0.47	0.62
Ile	AUA	0.8	0.87	1.07	1.07	0.9	0.68	1.09	1.08
Met	AUG	1	1	1	1	1	1	1	1
	GUU	1.7	2.16	1.15	1.19	1.38	1.69	1.66	1.2
	GUC	0.74	0.32	0.49	0.43	0.75	0.77	0.32	0.48
	GUA	0.88	0.75	0.87	0.89	0.89	0.72	0.91	0.94
Val	GUG	0.68	0.78	1.49	1.49	0.98	0.82	1.11	1.37

	UCU	1.41	1.82	1.23	1.19	1.5	1.48	1.62	1.4
	UCC	0.56	0.38	0.62	0.52	0.7	0.73	0.41	0.55
	UCA	1.28	1.28	1.19	1.48	1.42	1.25	1.02	1.53
Ser	UCG	0.25	0.25	0.53	0.39	0.43	0.41	0.38	0.41
	CCU	1.47	2.12	1.28	0.9	1.35	1.4	1.67	1.3
	CCC	0.58	0.25	0.66	0.42	0.8	0.76	0.5	0.53
	CCA	1.58	1.34	1.36	2.08	1.48	1.28	1.46	1.71
Pro	CCG	0.38	0.29	0.7	0.6	0.38	0.56	0.37	0.47
	ACU	1.3	1.71	1.25	0.88	1.35	1.57	1.56	1.24
	ACC	0.88	0.65	0.72	0.79	1.18	0.88	0.73	0.7
	ACA	1.47	1.42	1.45	1.76	1.12	1.1	1.3	1.63
Thr	ACG	0.35	0.22	0.58	0.57	0.34	0.45	0.41	0.43
	GCU	1.8	1.99	1.4	1.23	1.37	1.46	1.77	1.31
	GCC	0.73	0.56	0.63	0.62	0.76	0.65	0.52	0.45
	GCA	1.24	1.21	1.37	1.61	1.35	1.27	1.35	1.69
Ala	GCG	0.23	0.24	0.6	0.54	0.52	0.61	0.35	0.55
	UAU	1.1	1.35	1.22	0.93	1.03	1.26	1.48	1.1
Tyr	UAC	0.9	0.65	0.78	1.07	0.97	0.74	0.52	0.9
	UAA	1.33	1.41	1.13	1.19	1.17	1.27	1.05	1.13
TER	UAG	0.5	0.71	0.67	0.67	0.49	0.48	1	0.68
	CAU	1.04	1.17	1.12	1.06	0.99	0.93	1.23	1.19
His	CAC	0.96	0.83	0.88	0.94	1.01	1.07	0.77	0.81
	CAA	1.43	1.3	1.03	1.12	1.22	1.37	0.85	1.28
Gln	CAG	0.57	0.7	0.97	0.88	0.78	0.63	1.15	0.72
	AAU	1.1	1.37	1.08	0.98	1.05	1.05	1.47	1.14
Asn	AAC	0.9	0.63	0.92	1.02	0.95	0.95	0.53	0.86
	AAA	1.47	1.19	1.11	1.04	1.16	1.36	1.03	1.11
Lys	AAG	0.53	0.81	0.89	0.96	0.84	0.64	0.97	0.89
	GAU	1.15	1.47	1.16	0.93	0.94	1.13	1.67	1.21
Asp	GAC	0.85	0.53	0.84	1.07	1.06	0.88	0.33	0.79
Glu	GAA	1.38	1.26	1.08	1.27	1.04	1.31	1.13	1.31

	GAG	0.62	0.74	0.92	0.73	0.96	0.69	0.87	0.69
	UGU	1.16	1.5	1.22	1.19	1.13	1.24	1.39	1.21
Cys	UGC	0.84	0.5	0.78	0.81	0.87	0.76	0.61	0.79
TER	UGA	1.18	0.88	1.21	1.14	1.34	1.24	0.95	1.19
Trp	UGG	1	1	1	1	1	1	1	1
	CGU	0.58	1.91	0.81	0.83	1.1	1.13	1.04	0.66
	CGC	0.43	0.37	0.49	0.38	0.95	0.67	0.66	0.45
	CGA	0.33	0.39	0.57	0.36	0.77	0.77	0.35	0.39
Arg	CGG	0.35	0.17	0.34	0.39	0.49	0.52	0.3	0.34
	AGU	1.46	1.73	1.46	1.52	0.98	1.29	1.68	1.21
Ser	AGC	1.04	0.53	0.98	0.9	0.97	0.84	0.89	0.89
	AGA	2.98	1.81	1.85	2.15	1.67	1.81	2.14	2.66
Arg	AGG	1.32	1.35	1.94	1.89	1.03	1.09	1.51	1.5
	GGU	1.51	2.75	1.55	0.95	1.35	1.17	2.16	1.54
	GGC	0.85	0.41	0.87	0.93	1.06	1.02	0.65	0.72
	GGA	1.14	0.56	0.89	1.37	1.12	1.09	0.73	1.18
Gly	GGG	0.5	0.27	0.69	0.75	0.47	0.72	0.46	0.56

Note: AMN908947Wuhan_seafood_market_pneumonia_virus_isolate_Wuhan-Hu-1 (SARS-CoV-2)

BMG428704.1Human_coronavirus_NL63

CEF203064.1Bat_coronavirus_HKU2

DKY073747.1Bat_coronavirus_ BtKY229E-1

ENC_039208 Porcine_coronavirus_HKU15

FMH697599PEAV_GDS04_P12 GMH708243.1H11-SD2017

HKX499468_Transmissible_gastroenteritis_virus_strain_TGEV

It's the first time to report that porcine and murine coronavirus may attend the reorganization of SARS-CoV-2. The RDP4 estimated the possible reorganization regions for SARS-CoV-2 (Table 1 and Figure 5A).

Furthermore, SimPlot analysis confirmed the homologous recombination of sequence similarity between SARS-CoV-2 and coronavirus from potential host. The potential recombination breakpoints (16205-16358nt) are shown in red dash lines (Figure 5B), indicating the recombination between PEDV YN15 and Murine hepatitis virus JHM occurred while SARS-CoV-2 was queried. In addition, the region between 20923 and 21181 not also reflected the recombination event taking place between PEDV H11-SD2017 and Human coronavirus HKU1 while SARS-CoV-2 was queried (Figure 5C).

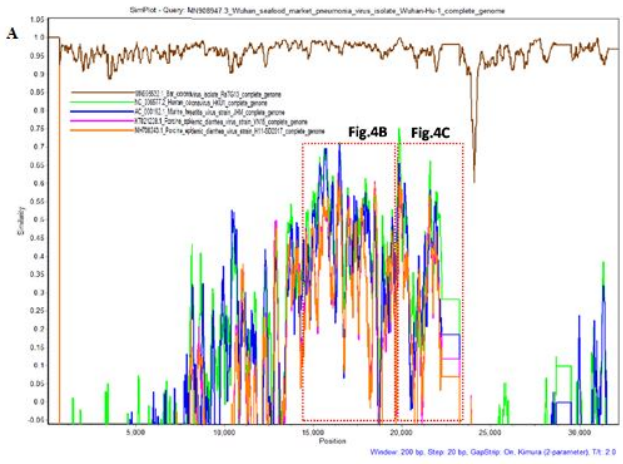

Figure 5A: Simplot analysis of varies coronavirus from potential host. The similarities to different reference sequences are indicated by different colors shown in the legend box at the top.

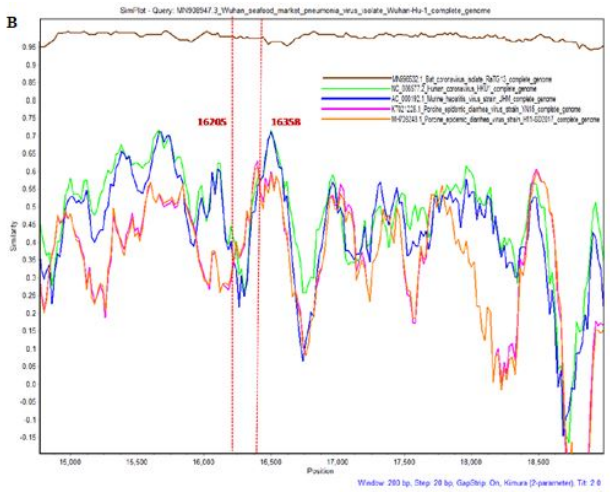

Figure 5B: Simplot analysis of varies coronavirus from potential host. The enlarged figure identify the homologous recombination region (16205-16358nt) of PEDV YN15 and Murine hepatitis strain JHM.

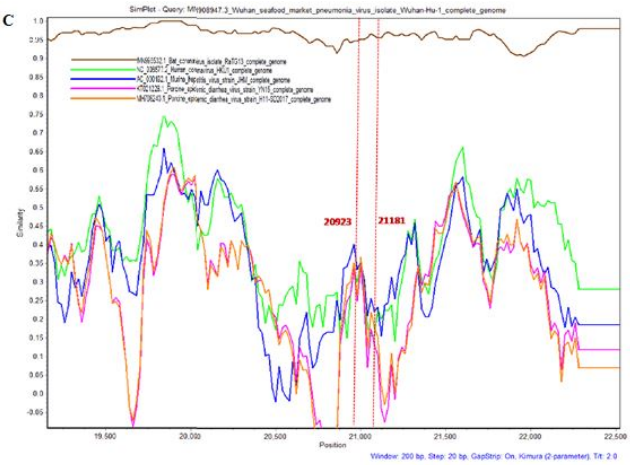

Figure 5C: Simplot analysis of varies coronavirus from potential host. The enlarged figure identify the homologous recombination region (20923-21181nt) of PEDV H11-SD2017 and Human coronavirus HKU1.

DISCUSSION

The cumulative diagnosis of SARS-CoV-2 would be thought to over 100,000 and more than two thousands cases would be died in China. It is the emergency public health crisis again since SARS 17 years ago [8].The local residences were reined with anxiety and confused during the virus outbreak, although the Chinese government takes massive actions into preventing and controlling the virus. What's worse, multiple messages could also disrupt the attention of medical workers and science researchers. The currently researches about the origin of SARS-CoV-2 mostly focus on the wildlife due to the first case was thought to be highly associated with Wuhan sea food market. There are two main points about the origin of SARS-CoV-2. Natural variation was the dominant view holding by most scholars, believing that bats and other wild animals provide virus reservoir for the novel coronavirus. But the coronavirus in bats cannot directly infect humans, one or two intermediate hosts in nature may provide the homologous recombination, enabling the coronavirus gradually adjust to the human genetic code, then survival and breeding successfully. Recently, some people ascribe the origin of SARS-CoV-2 to the modification variation in laboratory. The two most concentrated legends are the laboratory leakage and the conspiracy theory of the institute of viruses of the Chinese academy of sciences. However, no one should be blamed for creating the novel coronavirus due to the limited evidence until now. The issues also induce great concern of White House. The official requested that scientific experts "rapidly" look into the origins of the virus in order to address both the current spread and "to inform future outbreak preparation and better understand animal/human and environmental transmission aspects of coronaviruses." So it is urgent need to illuminate the origin of SARS-CoV-2 and its possible intermediate host. From our perspective, natural variation would be more reasonable for explaining the origin of SARS-CoV-2. However, the wild animal in Wuhan seafood market could not deserve all the liability because there are some proofs that patients with early infection didn't have any contact with the market [8]. Some researchers also pointed out that the Wuhan seafood market is not the only area of origin, indicating that there would be another creature act as intermediate host apart from wild animals selling in the market. In our study, the captivity animal, *Sus scrofa* (Swine), *Mus musculu* (s Mice), was highly suspected to be a critical host of SARS-CoV-2.

Our finding anatomize with natural variation. Natural variation believes that people get infected because they eat or come into contact with intermediate hosts. It was found that SARS-CoV-2 is 96% identical at the whole-genome level to the bat coronavirus [2]. What's more, some reported reveal that snake, mink, pangolins could be the potential host for SARS-CoV-2 [9].

Actually, either wild animal or rear livestock deserve the suspicion. The journal of Science also has reported that Wuhan seafood market may not be source of novel virus spreading globally because the earliest patient became ill on 1 December 2019 and had no epidemiological link to the seafood market or later cases. The official details about the first 41 hospitalized patients showed 13 of the 41 case had no link to the marketplace at all [8]. One possible hypothesis is that the cross-species transmission occurred in other space before the outbreak of Wuhan Huanan Seafood market. Previously study about fatal swine acute diarrhea syndrome (SADS) presented that SADS related coronavirus was responsible for a large-scale outbreak of fatal disease in pigs in China [7]. Here we discovered Porcine epidemic diarrhea virus (PEDV) periodicity burst in China from 2010-2020(Figure 1) [10].

Among one of the strain H11-SD2017 showed closely affiliation with SARS-CoV-2 *via* implementing relative synonymous codon usage (RSCU) and phylogenetic analysis. Swine-to-human cross species transmission may explain man (Corona Virus Disease 2019) COVID-19 patients not only suffering from severe respiratory diseases, but also with the diarrhea [8].

In the past ten years, PEDV has spread into most provinces with swine industry in China. Hubei was the first region being infected with PEDV strain CH/HBQX/10 in 2010, from then on, more and more provinces reported the presence of PEDV [10]. The spread of PEDV seems have bad broken through its geographical limitation. There were more than 360 PEDV strains were clustered into pandemic, meaning the significant natural variation took place in the spread of PEDV in different regions. According to the statistics in China from 2010~2020, up to 21 strains of PEDV had been founded in Hubei, ranking fourth in China (Figure 1). Further analysis among 21 strains of PEDV and coronavirus from varied species depicted that swine and mice could be the other parasitifer of SARS-CoV-2 (Figures 2A and 3, Figure 4). Both RDP4 and Simplot analysis help us better understand homologous recombination of SARS-CoV-2 (Table 1 and Figure 5). It verified that not only porcine coronavirus, but also the murine coronavirus attend the recombination events. Therefore, we speculate that SARS-CoV-2 May origin from the bat firstly, undergoing a series of recombination events, where pork and mice playing critical role in mediating cross species transmission. What's more, the analysis of pairwise distance also presented that *Mus musculus* was another possible host of SARS-CoV-2 (Figure 2B). Previously, CDC said that 33 of the samples were positive for the novel coronavirus nucleic acid. The positive samples were distributed among 22 stalls and 1 garbage truck in the Wuhan Huanan seafood market, so the outbreak is highly suspected to be related with wildlife trade. But how could wild animal move around the market and then leading cross infection with disparate species in different regions? One plausible phenomenon is that mice could be infected with SARS-CoV-2 firstly, then mediating the infection of other wild species in the market. So the role of mice and swine in the market deserved more attentions. It's no hard to prove with Koch's Postulate.

CONCLUSION

Eventually, the rear livestock should be deserved more notice apart from wild vertebrate creatures. It's the first time to illustrate that swine and mouse would be the probable livestock reservoir for the SARS-CoV-2. What's more, the mice around the seafood market may also involve into the chains of cross transmissions to some extent. The chains of host propagation need further clarify.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the support of the Wake Forest Baptist Comprehensive Cancer Center Cell Engineering Shared Resource, supported by the National Cancer Institute's Cancer Center Support Grant award number P30CA012197. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute.

AUTHOR CONTRIBUTIONS

Conceived the experiments: Fengxue Zhang; Performed and data analysis: Xiaopeng Hu; Design the experiments: Zhendan He; Article modification: Weixin Li

CONFLICT OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020; 3:1-5.
2. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;3:1-4.
3. Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4:406-425.
4. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 1985;39:783-791.
5. Chen LL, Ou HY, Zhang R, Zhang CT. ZCURVE_CoV: A new system to recognize protein coding genes in coronavirus genomes, and its applications in analyzing SARS-CoV genomes. Biochem Biophys Res Commun. 2003;307:382-388.
6. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30:2725-2729.
7. Zhou P, Fan H, Lan T, Yang XL, Shi WF, Zhang W, et al. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. Nature.2018;556:255-258.
8. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. The Lancet. 2020;395:470-473.
9. Ji W, Wang W, Zhao X, Zai J, Li X. Cross-species transmission of the newly identified coronavirus 2019-nCoV. J Med Virol. 2020;92:433.
10. Sun M, Ma J, Wang Y, Wang M, Song W, Zhang W, et al. Genomic and epidemiological characteristics provide new insights into the phylogeographical and spatiotemporal spread of porcine epidemic diarrhea virus in Asia. J Clin Microbiol. 2015;53:1484-1492.