PNIE PROCEEDINGS OF THE NATIONAL INSTITUTE OF ECOLOGY OF THE REPUBLIC OF KOREA

Original Article

PNIE 2024;5(4):134-152
https://doi.org/10.22920/PNIE.2024.5.4.134
pISSN 2765-2203, eISSN 2765-2211

Host-specific Mutation Patterns and Replication Characteristics of Shaan Virus of Bat Origin by Host Switching

Ji Yeong Noh[1,2a], Hye Kwon Kim[1,2*]

[1]Department of Biological Sciences and Biotechnology, College of Natural Science, Chungbuk National University, Cheongju, Republic of Korea
[2]Department of Microbiology, College of Natural Science, Chungbuk National University, Cheongju, Republic of Korea

ABSTRACT

Viruses mutate and form quasispecies within the host, from which the next generation is selected. Studying the selected mutations of an animal virus across different host species can aid in predicting how the virus might evolve to manage future potential pandemics in humans. In this study, both high-passaged (p44) and low-passaged (p4) Shaan viruses of bat origin in MARC-145 cell lines, derived from African green monkey kidney, were passaged once into human A549, HEK-293, and HRT-18 cell lines as different host species models. High-throughput sequencing data showed that there were distinct selected mutations and single nucleotide variants (SNVs) between the low- and high-passaged Shaan viruses in MARC-145 cells. After host switching, these mutation patterns were consistently observed among the same host cells in triplicate experiments, suggesting a host cell-specific selection pattern for the progeny of Shaan viruses. Notably, HRT-18 cells of colorectal adenocarcinoma origin produced more unique selected mutations and unique SNVs. While Shaan virus-specific transcripts associated with the N gene were most abundant in MARC-145, A549, and HEK-293 cells, those associated with the M gene were most abundant in HRT-18 cells. After host switching, the relative viral RNA levels in the HRT-18 cells were significantly higher than in A549 and HEK-293 cells. Thus, this study provides evidence of host-specific mutation patterns by cell type and host cells in the evolution of Shaan virus.

Keywords: Bat, Shaan virus, Evolution, Host switching

Introduction

Viruses are obligate parasites that require hosts for their replication. As they replicate within their hosts, viruses mutate and develop quasispecies, selected for the next generation. While most viral mutations are deleterious, a sufficiently large population of variants can confer some with the capacity for interspecies transmission or the evasion of virus-specific immune responses (Acevedo *et al.*, 2014; McCrone & Lauring, 2018).

Received June 21, 2024; **Revised** August 7, 2024; **Accepted** August 8, 2024

***Corresponding author:** Hye Kwon Kim
e-mail khk1329@chungbuk.ac.kr
https://orcid.org/0000-0003-3458-3403

[a]Current affiliation: Dr. Ji Yeong Noh is currently work at Division of Emerging Infectious Diseases, Bureau of Infectious Disease Diagnosis Control, Korea Disease Control and Prevention Agency, Cheongju, South Korea.

DNA viruses typically have mutation rates ranging from 10^{-8} to 10^{-6} substitutions per nucleotide site per cell infection (s/n/c), while RNA viruses exhibit higher mutation rates between 10^{-6} and 10^{-4} s/n/c (Peck & Lauring, 2018). High mutation rates allow viruses to adapt and survive in specific host environments through selected mutations. The pressures that select for the next generation of viruses and their mutants are varied and include factors such as antiviral agents, immune responses, and host switching, all of which influence viral fitness within the environment (Wargo & Kurath, 2012).

Host switching is relatively rare, as viruses typically infect a narrow range of host species, though this spectrum varies among different viruses (Rothenburg & Brennan, 2020). Nevertheless, as observed with the recently emergent severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which is linked to the intermediate horseshoe bat (*Rhinolophus affinis*) and the Malayan pangolin (*Manis javanica*), viruses can switch hosts, potentially leading to pandemics in new host species (Lam *et al*., 2020; Zhou *et al*., 2020). Consequently, studying viral mutations across different host species can offer insights into viral evolution and help prepare for future pandemics.

In this study, we used the Shaan virus isolate (B16-40), a novel species from the family *Paramyxoviridae* originating from bats (Noh *et al*., 2018). Considering that recent viruses affecting humans have originated from bats, studying a novel paramyxovirus from bats, previously unknown to infect humans, is crucial for understanding its mutation patterns and potential to infect humans. In this context, both high- and low-passaged Shaan viruses in MARC-145 cell lines derived from African green monkey kidney (Kim *et al*., 1993) were introduced to A549, HEK-293, and HRT-18 cell lines (human origin). These cell lines served as models for different host species. Utilizing population-based analysis, which is important for studying viral evolution due to the quasispecies nature of viruses (Geoghegan & Holmes, 2018), we examined mutations in the high- and low-passaged Shaan viruses from both the original and new host environments using high-throughput sequencing in triplicate.

Materials and Methods

Cells and viruses

MARC-145 cell lines, derived from the kidney of the African green monkey, were cultured in Dulbecco's Modified Eagle Medium (cat. no. 11995; Gibco, NY, USA), supplemented with antibiotic-antimycotic (100 IU/mL penicillin, 100 µg/mL streptomycin, and 0.25 µg/mL amphotericin B) (cat. no. 15240; Gibco, NY, USA) and 10% fetal bovine serum (cat. no. SV30207; Hyclone, NY, USA). Human-origin HEK-293, A549, and HRT-18 cell lines from kidney, lung, and colon, respectively, were cultured in MEM (cat. no. 10370; Gibco, NY, USA) and RPMI1640 (cat. no. A10491; Gibco, NY, USA) media, supplemented with 10% fetal bovine serum and antibiotic-antimycotic. These cells were maintained at a constant 37°C and 5% CO_2 in an incubator.

The Shaan virus isolate (B16-40) (GenBank accession no. MG230624.1) was propagated in MARC-145 cells. For amplification, 1×10^5/mL of MARC-145 cells were seeded in tissue culture flasks using Dulbecco's Modified Eagle Medium supplemented with 5% fetal bovine serum and incubated overnight. Following this, a MARC-145 cell monolayer was infected with the Shaan virus (B16-40) at a multiplicity of infection (MOI) of 0.1, with the cytopathic effect observed in the infected cells five days post-inoculation. The virus, harvested by freeze-thawing the infected cells, was then used to infect a fresh monolayer of MARC-145 cells for further passages.

Preparation of low-passaged and high-passaged Shaan virus in MARC-145 cells

The original isolate of Shaan virus (B16-40) (GenBank accession no. MG230624.1) was passaged in MARC-145 cells, as depicted in Fig. 1. Low-passaged virus was obtained by serially passing the virus four times in MARC-145 cells, while the high-passaged virus was derived from 44 serial passages in the same cell line. Viral titers of both, low- and high-passaged viruses, were quantified using the tissue culture infectious dose 50 ($TCID_{50}$) method as described by the Reed-Muench formula.

Host switching of low-passaged and high-passaged Shaan virus

The low-passaged (p4) and high-passaged (p44) Shaan virus samples from MARC-145 cells were inoculated in triplicate into various host cells, specifically HEK-293, A549, and HRT-18 cell lines, originating from human kidney, lung, and colon respectively, as indicated in Fig. 1. Each virus was introduced into these cell lines with a multiplicity of infection (MOI) of 0.5 and allowed to infect for 2 hours. Post-infection, cells were washed with DPBS (pH 7.4) (cat. no. LB001-02; Welgene Inc., Daegu, Korea) and were further incubated with the respective maintenance medium for 24 hours, following protocols established by Lim *et al*., (2023). Virus harvesting was carried out by freeze-thawing the infected cells, and the collected viruses were then utilized for further quantification and sequencing analyses.

Viral RNA quantification

Total RNA was extracted from harvested samples using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Viral RNA was quantified using two real-time RT-PCR methods that targeted total viral RNAs and the viral antigenome. For total viral RNA quantification,

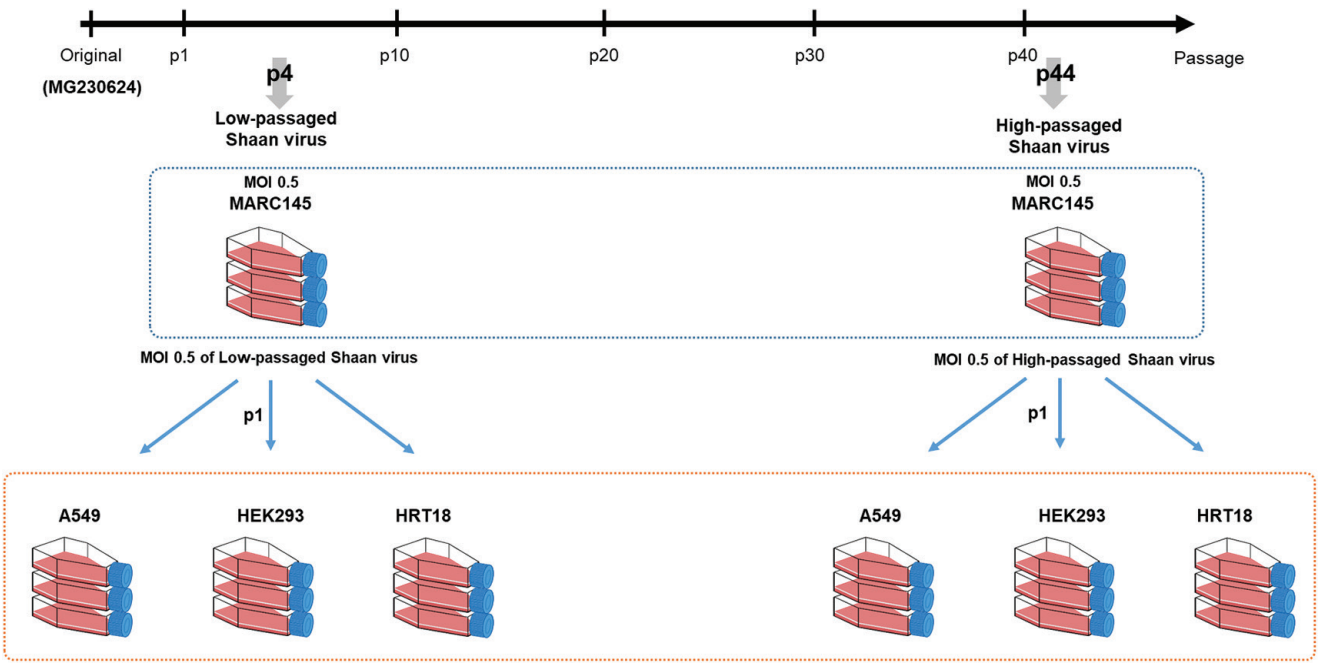


Fig. 1. High-passaged and low-passaged Shaan virus passage scheme in A549, HEK293, and HRT18 cell lines. The original Shaan virus isolate was serially passaged in the MARC145 cell line. Both high-passaged and low-passaged Shaan virus cultures were passaged at an MOI of 0.5 in triplicate across each cell line. A sample from passage 1 of both the high-passaged and low-passaged Shaan virus was subjected to high-throughput sequencing analysis.

cDNAs were synthesized using random hexamers and tested through real-time PCR with a Shaan virus detection primer, which targeted the small hydrophobic (SH) protein region (Supplementary Table 5). Antigenome quantification was performed using a primer designed specifically for the intergenic region between the nucleocapsid and phosphoprotein, suitable for reverse transcription. This antigenome was quantified in real-time PCR using an antigenome-specific primer (Supplementary Table 5). Viral RNA and antigenome levels were normalized to the monkey (for MARC-145 cells) and human GAPDH genes as controls (Das *et al*., 2000; Shi *et al*., 2016). Cell lysates from all Shaan virus-infected cells were inoculated into MARC-145 cells, and infectious virion titers were quantified as tissue culture infectious dose 50 ($TCID_{50}$) using the Reed-Muench method.

Sequencing and analysis

RNA was extracted from infected cells harboring low-passaged (p4) and high-passaged (p44) Shaan virus using TRIzol LS Reagent (Thermo Fisher Scientific Inc.), according to the manufacturer's manual. RNA samples were analyzed by Macrogen, Inc. (Seoul, Republic of Korea) for high-throughput sequencing and quality assessment. Subsequently, total cDNA libraries, sourced from RNA that met quality standards, were prepared using the Illumina TruSeq Strand Total RNA LT kit according to the manufacturer's instructions. The samples were sequenced on a NovaSeq 6,000 platform.

Data quality of the raw sequences was verified using FastQCv0.11.7. Subsequently, adapter sequences were excised using BBDuk adaptor and quality trimming version 38.84 (Bushnell, 2014). The sequences of the trimmed reads from both low- and high-passaged viruses were aligned to the Shaan virus sequence (GenBank accession no. MG230624.1) using the Geneious Prime Bowtie 2 plugin version 2.3.0 (Langmead & Salzberg, 2012). The expression levels of transcripts in cells infected with Shaan virus were quantified based on a reference annotation of Shaan virus (MG230624.1). Transcript abundance was calculated and normalized to the TPM value, considering read count, transcript length, and coverage depth. Three consensus genomic sequences obtained from Geneious Prime Bowtie 2 plugin version 2.3.0 (default setting) within the same group were aligned using the MAFFT v7.450 auto alignment algorithm (Katoh & Standley, 2013). The consensus genomic sequences of low- and high-passaged viruses served as reference sequences for mutation pattern and SNV analysis. The mean read depth of coverage for these genomic sequences is presented in Supplementary Table 1. Selected mutations were analyzed using the consensus genomic sequences of viruses passaged once in A549, HEK-293, and HRT-18 cells compared to Shaan

virus in MARC-145 cells using Geneious software, with analysis of synonymous and non-synonymous amino acid substitutions also performed.

For SNV analysis, trimmed reads from both low-passaged and high-passaged viruses, passaged once in A549, HEK-293, and HRT-18 cells, were mapped to the consensus genomic sequences of the original viruses. Single nucleotide polymorphisms (SNPs) and INDELs were detected at minimum variant frequencies of 0.25 or 0.05, with a 10^{-6} maximum variant p-value and a 10^{-5} minimum strand-bias p-value (exceeding 65% bias) using the Geneious software workflow, "Map reads then find variations/SNPs". The coverage and variant frequency data for the SNPs and INDELs across each experimental group are detailed in Supplementary Tables 2, 3, and 4. SNV results with coverage below 100× were omitted from the mutation analysis. Synonymous and non-synonymous amino acid substitutions at these SNVs were further examined.

Statistical analysis

All data in triplicate were subjected to a normality test using an unpaired t-test with multiple comparisons across groups. Statistical analyses were conducted using one-way ANOVA with Tukey's multiple comparison test in GraphPad Prism version 9.0. Asterisks denote p-values as follows: *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; ns, not significant.

Results

Quantitative analysis of Shaan virus replication in HEK-293, A549, and HRT-18 cells

Low- and high-passage viruses were infected into HEK-293, A549, and HRT-18 cells. At 24 hours post-infection, viral replication was compared by measuring total viral RNA, antigenomes, and infectious virions (Fig. 2). Relative viral RNA quantities in HRT-18 cells were

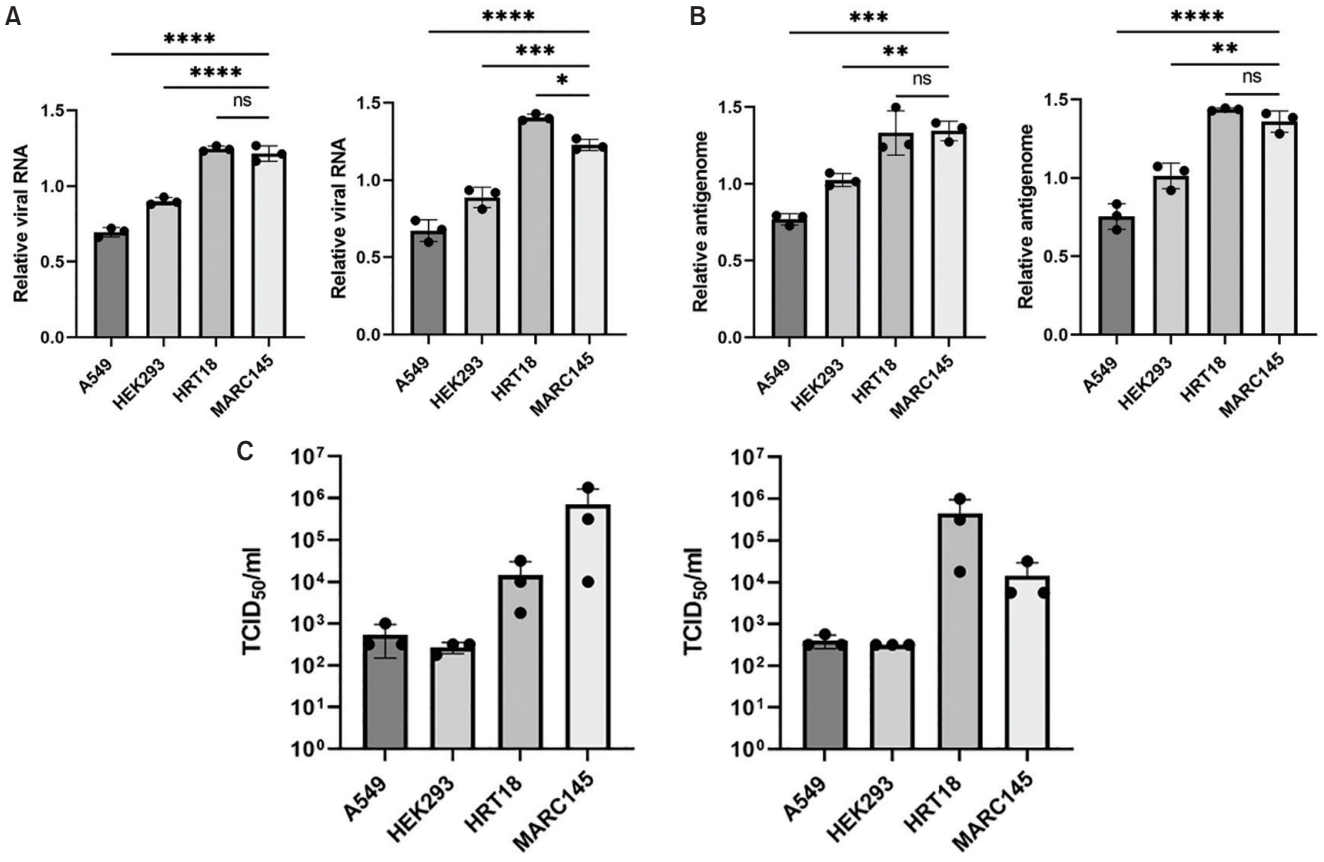


Fig. 2. Viral quantitative analysis from high-passaged and low-passaged Shaan virus passaged in A549, HEK293, HRT18, and MARC145 cell lines. (A) Viral RNAs extracted from the first passage of Shaan virus in each cell line were quantified using real-time RT-PCR targeting the SH region of the virus. (B) The viral antigenome was quantified from the total RNA of each infected cell using reverse transcription and PCR with specifically designed primers. The relative viral RNA and antigenome levels were normalized using the amount of monkey and human GAPDH genes as controls. (C) Viral titers were measured by calculating infectious virus levels (TCID50/ml) in Shaan virus-infected cell lysates. Data were analyzed using one-way ANOVA with Tukey's multiple comparison test. *p < 0.05; **p < 0.01; ***p < 0.001; ****p < 0.0001; ns, not significant.

significantly higher than those in A549 and HEK-293 cells, regardless of the Shann virus inoculum and the virus passage level. Additionally, when antigenomes produced during viral replication in host cells were measured, the relative amount of antigenomes was significantly higher in HRT-18 cells than in A549 and HEK-293 cells. Infectious virions were measured using $TCID_{50}$ in MARC-145 cells, where a significantly higher number of infectious virions was observed in HRT-18 cells infected with either low- or high-passage viruses.

Comparison of Shaan virus-specific transcripts in low- and high-passage virus-infected HEK-293, A549, and HRT-18 cells

Using high-throughput sequencing, a total of 24 sets of sequencing data were obtained, representing triplicate samples from either low- or high-passage Shaan virus, as well as those-infected HEK-293, A549, and HRT-18 cells, as documented in Supplementary Table 1. The total read counts varied from 50,285,454 to 76,298,278. The reference mapping of trimmed reads against a reference genome (GenBank accession no. MG230624.1) showed different coverage values. In HRT-18 cells, the mean read depth for mapping ranged from 2,014.9 to 2,387.8 in the low-passage Shaan virus-infected cells and from 2,177.2 to 2,384.0 in the high-passage Shaan virus-infected cells. Notably, in HEK-293 cells, the mean read depth ranged from 4,553.6 to 6,452.6 in the low-passage Shaan virus-infected cells, but only from 9.4 to 17.0 in the high-passage Shaan virus-infected cells. In A549 cells, mean read depths were 517.0-1,049.1 after low-passage virus infection and 305.0-802.1 following high-passage virus infection. The percentage rate of mapped reads with a Phred score > 30 ranged from 98.2% to 98.7% in all groups.

After reference mapping the sequenced reads, Shaan virus-specific transcripts were compared based on different viral genes. The mapped reads were normalized and expressed as transcripts per million (TPM) values (Fig. 3). Overall, transcript profiles were found to be similar to those of other paramyxoviruses and Shaan virus (Wignall-Fleming *et al.*, 2019; Lim *et al.*, 2023). No significant differences were observed in the TPM values of each gene between cells infected with low- and high-passaged Shaan virus. However, transcripts associated with the nucleocapsid (N) and matrix (M) genes varied among the host cells. Shaan virus-specific transcripts associated with the N gene were the highest, followed by those associated with the M gene in MARC-145, A549, and HEK-293 cells, regardless of the virus passage level. In HRT-18 cells, transcripts associated with the M gene were the highest, followed by those associated with the N gene.

Genomic differences between low- and high-passaged Shaan viruses in MARC-145 cells

Consensus genomic sequences of low- and high-passaged Shaan viruses were obtained from triplicate high-throughput sequencing, using the reference genome of the Shaan virus (GenBank accession no. MG230624.1) for mapping the trimmed reads. Six mutations were identified in the low-passaged Shaan virus (p4), and nine mutations

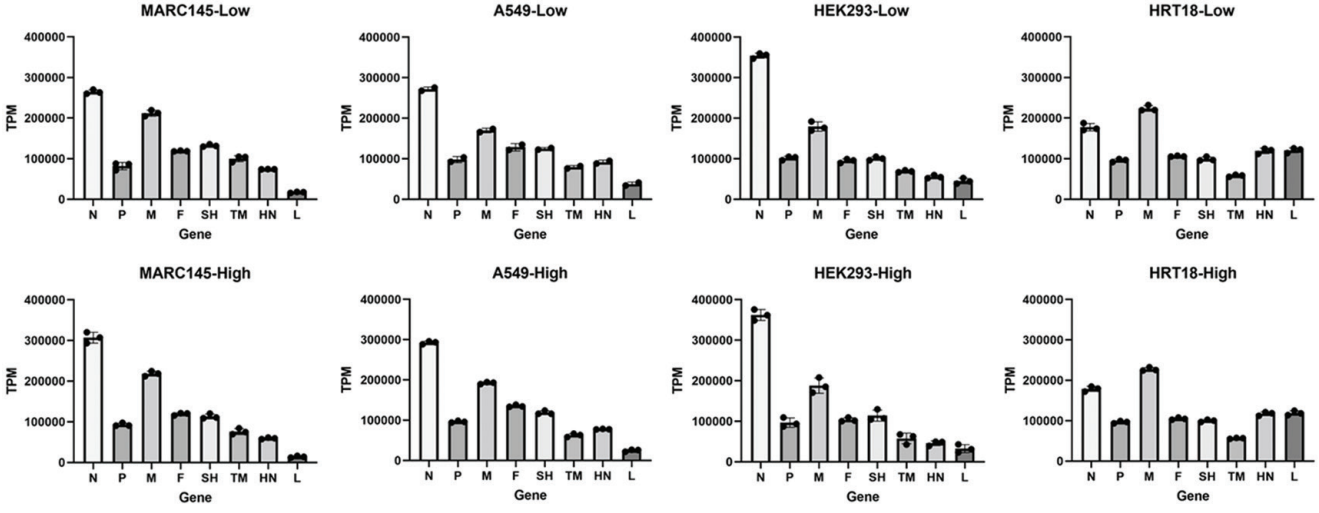


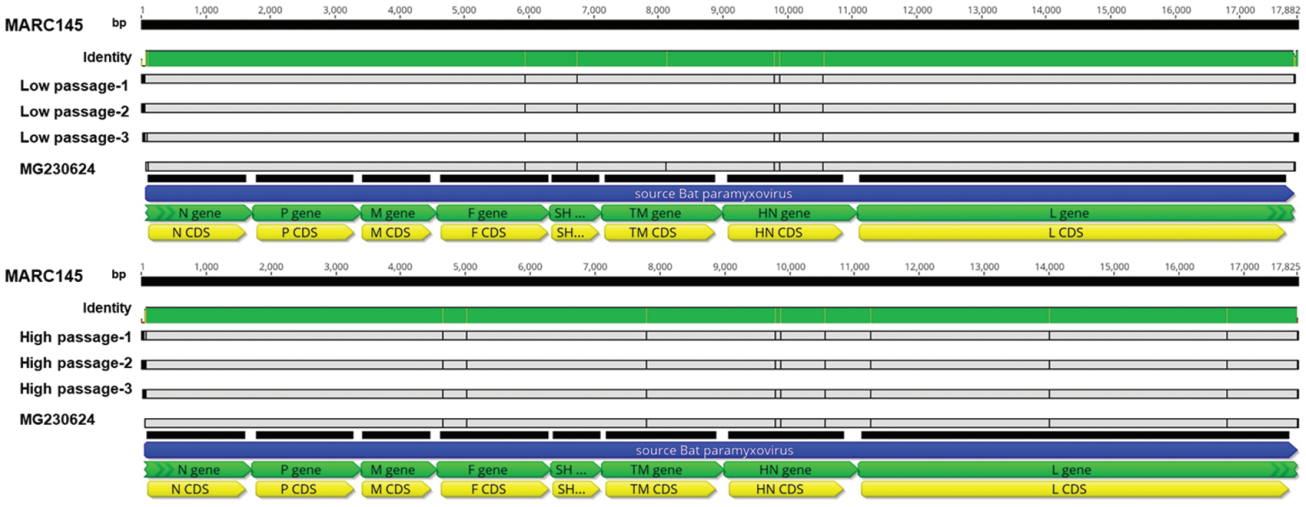
Fig. 3. Differential expression of viral transcript in MARC145, A549, HEK293, and HRT18 cells infected with Shaan virus. The expression of viral gene transcripts, based on transcripts per million mapped reads (TPM), was compared in each cell line infected with 0.5 MOI of high-passaged and low-passaged Shaan virus (passage 1). N, nucleocapsid; P, phosphoprotein; M, matrix protein; F, fusion protein; SH, small hydrophobic protein; TM, transmembrane protein; HN, attachment haemagglutinin neuraminidase glycoprotein; L, large protein.

in the high-passaged Shaan virus (p44) (Fig. 4). Three mutations in the hemaglutinin-neuraminidase (HN) gene, specifically non-synonymous substitutions C9,724T (H239Y in amino acids), C9,803T (A265V), and T10,484C (I492T), were consistently observed in both virus passages. Additionally, in the low-passaged viruses, non-synonymous substitutions A5,876T (K435N) in the fusion protein (F) gene and G6,676A (E129K) in the small hydrophobic protein (SH) gene were observed. The high-passaged virus exhibited additional non-synonymous substitutions T4,590A (F7A) in the F gene, C11,194T (S49F), and C13,936T (T964I) in the large protein (L) gene. The mutations found in the low- and high-passaged viruses were consistent across triplicate sequencing runs.

We also analyzed single nucleotide variants (SNVs) in low- and high-passaged viruses. The consensus genomic sequences we obtained were used as references for SNV analysis. SNVs found in low-passaged Shaan virus (p4) were more numerous than those in the high-passaged Shaan virus (p44) (Fig. 5). With a minimum variant frequency of 0.25, six SNVs across the F, SH, TM, HN, and L genes were identified in the low-passaged virus, compared to only one SNV in the high-passaged virus. The SNV in the high-passaged virus was characterized by the insertion of adenine (A) at an A-rich site, potentially due to a sequencing error. With a minimum variant frequency of 0.05, 15 SNVs were observed in the low-passaged virus, whereas seven SNVs were detected in the high-passaged virus.

Mutations found in infected low- and high-passaged Shaan viruses after one passage in A549, HRT-18, and HEK-293 cells

Consensus genomic sequences from low-passaged (p4) and high-passaged (p44) Shaan viruses in MARC-145 cells served as references to generate consensus genomic sequences of the viruses passaged once in A549, HRT-18, and HEK-293 cells. These consensus genomic sequences of the once-passaged viruses were compared with the original sequences of low- and high-passaged Shaan viruses. No noticeable selected mutations were found when

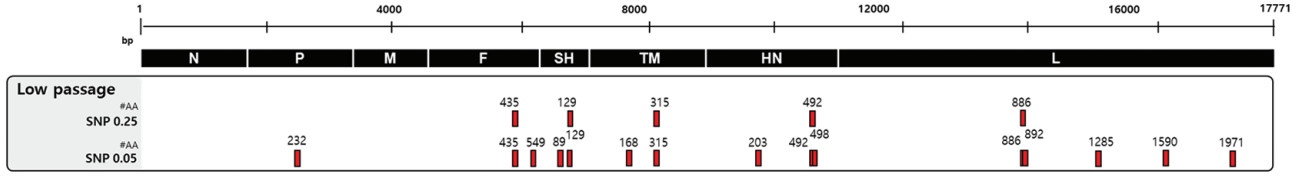


	F			SH	TM		HN			L		
#NT	4590	4964	5876	6676	7739	8057	9724	9803	10484	11194	13939	16673
#AA	7	131	435	129	209	315	239	265	492	49	964	1875
Original	**T**	**C**	**A**	**G**	**A**	**C**	**C**	**C**	**T**	**C**	**C**	**C**
Low passage	-	-	T	A	-	Y	T	T	C	-	-	-
	-	-	T	A	-	Y	T	T	C	-	-	-
	-	-	T	A	-	Y	T	T	C	-	-	-
High passage	A	T	-	-	G	-	T	T	C	T	T	T
	A	T	-	-	G	-	T	T	C	T	T	T
	A	T	-	-	G	-	T	T	C	T	T	T
Amino acid change	F>I	T>T	K>N	E>K	G>G	D>D	H>Y	A>V	I>T	S>F	T>I	Y>Y

Fig. 4. Comparison of consensus genomic sequence of original and high-passaged and low-passaged Shaan virus in MARC 145 cell. Multiple sequence alignment of consensus genomic sequences was carried out using MAFFT v7.450 integrated into Geneious Prime 2021.2.2. Selected mutations were identified in the F, SH, TM, HN, and L genes. Nucleotide and amino acid numbers (#NT, #AA) were annotated from a reference strain of bat paramyxovirus (GenBank accession no. MG230624).

low-passaged and high-passaged Shaan viruses were passaged once in A549 cells (Fig. 6). Y(C/T) 8,057T on the TM gene and A13,706R (A/G) on the L gene were observed at the SNV sites in the virus passaged in A549 cells following infection with low-passaged Shaan virus (p4); similarly, A4,590W (A/T) on the F gene was observed at the SNV sites in the virus passaged in A549 cells following infection with high-passaged Shaan virus (p44).

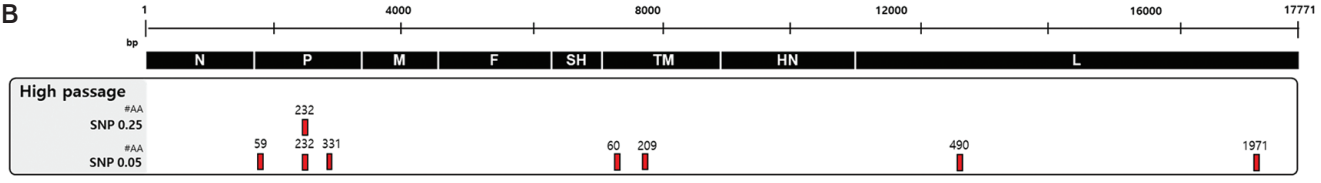
A



MARC145 Low passage

		P	F		SH		TM			HN			L				
	#NT	2415	5876	6216	6557	6676	7614	8057	8057	9618	10484	10502	13706	13723	14904	15819	16963
	#AA	232	435	549	89	129	168	315	315	203	492	498	886	892	1285	1590	1971
	Consensus	(A)7	T	G	T	A	T	Y	Y	A	C	G	A	T	T	T	(A)7
SNP 0.25	1	-	A 35.9%	-	-	G 31.8%	-	T 46.1%	C 53.9%	-	T 33.2%	-	G 33.4%	-	-	-	-
	2	-	A 34.2%	-	-	G 32.0%	-	T 48.6%	C 51.3%	-	T 34.0%	-	G 28.0%	-	-	-	-
	3	-	A 30.1%	-	-	G 27.4%	-	T 47.6%	C 52.3%	-	T 27.8%	-	G 36.1%	-	-	-	-
SNP 0.05	1	(A)8 17%	A 35.9%	A 19.4%	C 17.9%	G 31.8%	C 19.7%	T 46.1%	C 53.9%	G 14.2%	T 33.2%	A 18.6%	G 33.4%	G 12.4%	C 17.4%	A 19.4%	(A)8 6.5%
	2	-	A 34.2%	A 19.4%	C 18.3%	G 32.00%	C 19.1%	T 48.6%	C 51.3%	G 13.8%	T 34.0%	A 18.5%	G 28.0%	G 16.1%	C 17.1%	A 16.6%	-
	3	(A)8 14.6%	A 30.1%	A 16.1%	C 15.6%	G 27.4%	C 18.4%	T 47.6%	C 52.3%	G 13.1%	T 27.8%	A 15.4%	G 36.1%	G 12.7%	C 19.8%	A 18.1%	-
	Amino acid change	-	N>K	G>S	L>P	K>E	Y>H	D>D	D>D	I>M	T>I	R>K	E>G	W>G	P>P	N>K	-

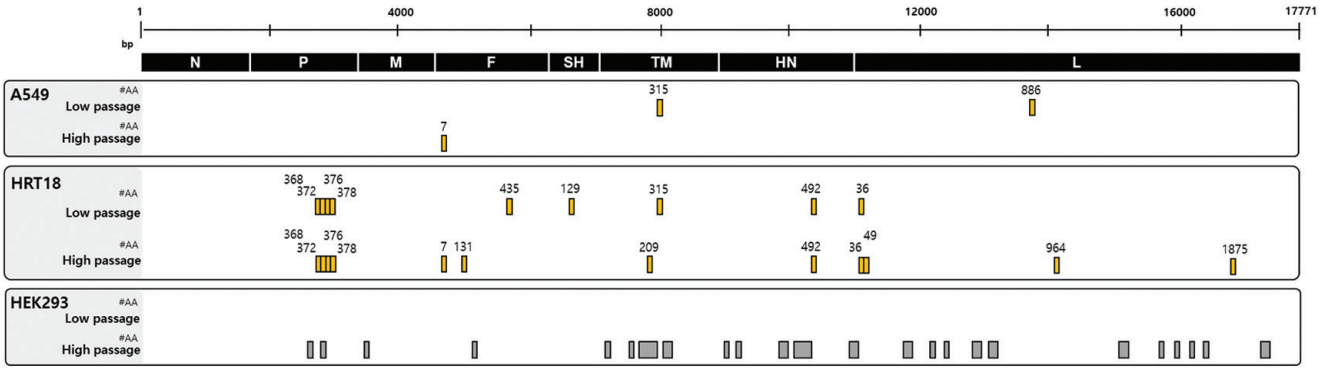
B



MARC145 High passage

		P			TM		L	
	#NT	1896	2415	2712	7292	7739	12519	16962
	#AA	59	232	331	60	209	490	1971
	Consensus	A	(A)7	T	A	G	G	(A)7
SNP 0.25	1	-	(A)8 26.8%	-	-	-	-	-
	2	-	(A)8 30.1%	-	-	-	-	-
	3	-	(A)8 25.7%	-	-	-	-	-
SNP 0.05	1	G 18.1%	(A)8 26.8%	C 5.4%	G 21.8%	A 23.4%	A 5.2%	-
	2	G 17.7%	(A)8 30.1%	-	G 17.7%	A 15.6%	-	(A)8 7.7%
	3	G 19.5%	(A)8 25.7%	C 5.1%	G 20.7%	A 21.4%	-	-
	Amino acid change	R>R	-	N>N	Q>Q	G>G	V>V	-

Fig. 5. SNVs found in high-passaged and low-passaged virus. The consensus genomic sequences of the high-passaged (B) and low-passaged (A) Shaan virus, inoculated in MARC145 cells, served as references for SNV analysis. SNVs were analyzed with a minimum variant frequency of 0.25 and 0.05 using Geneious Prime 2021.2.2. SNVs were marked with a red box and annotated with amino acid numbers at each viral gene site. The variant frequency at each SNV is expressed as a percentage value. The final row presents the synonymous or non-synonymous amino acid substitutions detected.



		P				F			SH	TM		HN	L				
	#NT	2823	2834	2847	2852	4590	4964	5876	6676	7739	8057	10484	11157	11195	13706	13940	16674
	#AA	368	372	376	378	7	131	435	129	209	315	492	36	49	886	964	1875
	Consensus	T	T	T	T	A	T	T	A	G	Y	C	A	T	A	T	T
A549	Low passage	-	-	-	-	-	-	-	-	-	T	-	-	-	R	-	-
		-	-	-	-	-	-	-	-	-	T	-	-	-	R	-	-
		-	-	-	-	-	-	-	-	-	T	-	-	-	R	-	-
	High passage	-	-	-	-	W	-	-	-	-	-	-	-	-	-	-	-
		-	-	-	-	W	-	-	-	-	-	-	-	-	-	-	-
		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
HRT18	Low passage	C	C	C	C	-	-	A	G	-	C	T	G	-	-	-	-
		C	C	C	C	-	-	A	G	-	C	T	G	-	-	-	-
		C	C	C	C	-	-	A	G	-	C	T	G	-	-	-	-
	High passage	C	C	C	C	T	C	-	-	A	-	T	G	C	-	C	C
		C	C	C	C	T	C	-	-	A	-	T	G	C	-	C	C
		C	C	C	C	T	C	-	-	A	-	T	G	C	-	C	C
	Amino acid change	N>N	I>T	P>P	F>S	I>I or F	T>T	N>K	K>E	G>G	D>D	T>I	Q>Q	F>S	E>E or G	I>T	Y>Y

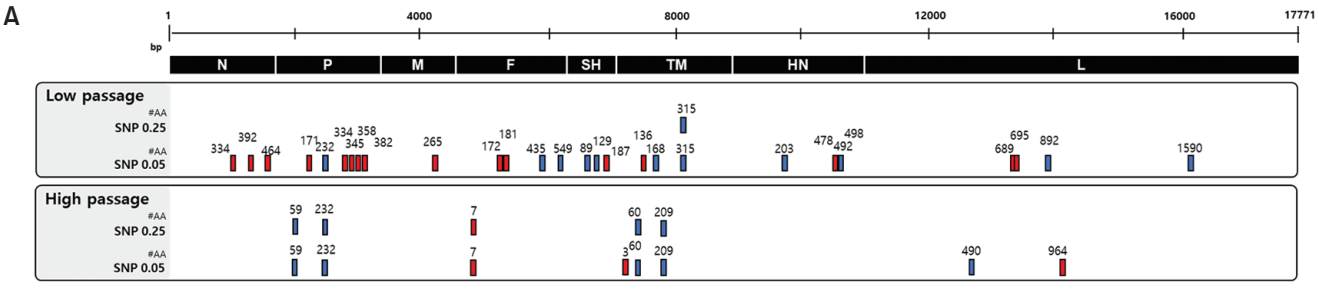
Fig. 6. Comparison of selected mutations found after one passage and in high-passaged and low-passaged Shaan virus-infected cells in human cell line. Nucleotide sequences of high-passaged and low-passaged Shaan virus genes were compared across A549, HEK293, and HRT18 cells that were passaged once. In A549 and HRT18 cells, the selected mutations are highlighted in yellow with the corresponding amino acid number at the mutation site in viral genes. In HEK293 cells, no mutations were observed in the low-passaged Shaan virus. Results from the high-passaged virus in HEK293 cells (grey box) were excluded due to the low read depth (<100×) in mutation analysis. The resulting synonymous or non-synonymous amino acid substitutions are presented in the final row.

In the case of HEK-293 cells, no selected mutations were observed following one passage of the low-passaged Shaan virus (p4). As noted in Supplementary Table 1, the low quality of the mapped reads precluded the detection of selected mutations in the viruses following one passage of the high-passaged Shaan virus (p44) in HEK-293 cells.

Selected mutations occurred more frequently in viruses passaged once in HRT-18 cells. When low-passaged Shaan virus (p4) was passed once through HRT-18 cells, nine selected mutations were detected. Among these, T2,834C (I372T) and T2,852C (F378S) in the phosphoprotein (P) gene, T5,876A (N435K) in the F gene, A6,676G (K129E) in the SH gene, and C10,484T (T492I) in the HN gene were non-synonymous substitutions. In the case of the high-passaged Shaan virus (p44) passed once through HRT-18 cells, 12 selected mutations were detected. These included consistent mutations like T2,834C (I372T) and T2,852C (F378S) on the phosphoprotein (P) gene, and C10,484T (T492I) on the HN gene from the same cells infected with low-passaged virus. Additional non-synonymous substitutions such as A4,590T (I7F) on the F gene and T11,195C (F49S) and T13,940C (I964T) on the L gene were also observed.

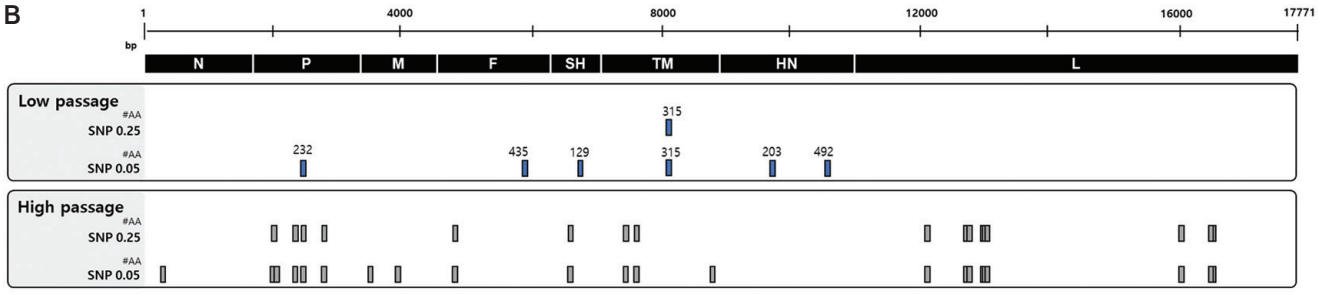
Single nucleotide variants (SNVs) found in low- and high-passaged Shaan viruses after one passage in A549, HRT18, and HEK293 cells

We also analyzed the SNVs of the viruses that were passaged once in A549, HRT-18, and HEK-293 cells with low- and high-passaged Shaan viruses. The consensus genomic sequences obtained for mutation pattern analysis served as references for the SNV analysis of each group. With a minimum variant frequency of 0.05, we found no consistent SNVs between viruses after one passage in A549 cells with low- and high-passaged Shaan viruses (Fig. 7A), except for the 2,415A insertion on the P gene. As the insertion site was in a repeated A sequence, this might be a sequencing error. Although inconsistent SNVs were observed in one sample



A549

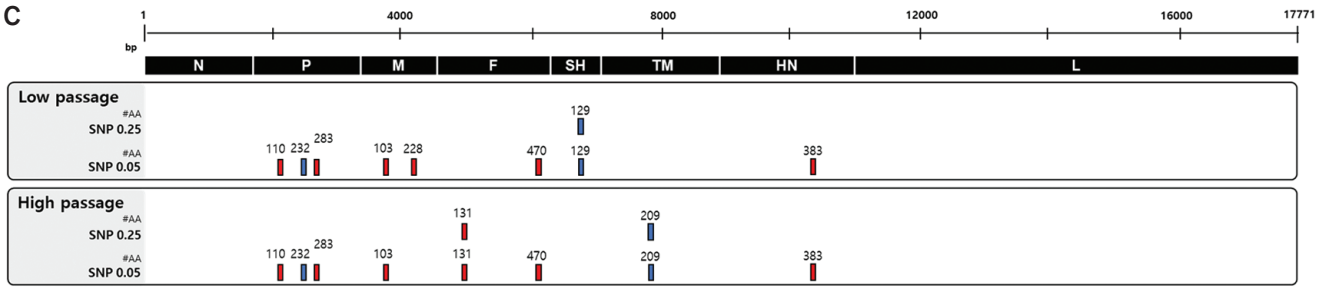
		N			P						M	F					SH			TM							HN				L					
	#NT	1045	1219	1437	1896	2248	2415	2719	2754	2793	2863	4152	4590	5085	5113	5876	6216	6557	6676	6851	7119	7292	7518	7614	7739	8057	8057	9618	10442	10484	10502	12519	13114	13133	13723	13940	15819
	#AA	334	392	464	59	177	232	334	345	358	382	265	7	172	181	435	549	89	129	187	3	60	136	168	209	315	315	203	478	492	498	490	689	695	892	964	1590
	Consensus	G	T	(A)6	A	G	(A)7	A	T	A	A	G	A	T	T	T	G	T	A	C	T	A	T	T	G	Y	Y	A	T	C	G	G	T	T	T	T	T
Low passage	1	-	-	-	-	-	(A)8 23.9%	-	-	-	-	-	-	-	-	A 17.1%	A 5.9%	C 7.5%	G 14.6%	-	-	-	-	C 8.0%	-	T 61.8%	C 38.2%	G 9.2%	-	T 14.9%	A 8.3%	-	-	-	G 7.2%	-	A 8.4%
	2	-	-	-	-	-	(A)8 20.8%	-	-	-	-	-	-	-	-	A 14.8%	-	-	G 15.6%	-	-	-	-	-	-	T 68.3%	C 31.7%	G 12.6%	-	T 17.2%	A 5.3%	-	-	-	G 11.0%	-	A 6.7%
	3	C 7.4%	C 5.6%	(A)7 6.7%		A 5.9%	-	G 11.4%	C 7.9%	G 8.6%	G 8.1%	DEL 6.1%	-	C 5.4%	C 6.2%	-	-	-	G 15.4%	T 6.0%	-	-	C 8.2%	C 10.4%	-	T 64.7%	C 35.3%	-	C 11.8%	-	-	-	C 7.4%	C 7.9%	-	-	-
High passage	1	-	-	-	G 29.1%	-	(A)8 27.5%	-	-	-	-	-	T 42.8%	-	-	-	-	-	-	-	C 18.2%	G 24.7%	-	-	A 28.7%	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	G 26.7%	-	(A)8 25.4%	-	-	-	-	-	T 41.1%	-	-	-	-	-	-	-	C 13.8%	G 34.9%	-	-	A 28.1%	-	-	-	-	-	-	-	-	-	-	C 7.1%	-
	3	-	-	-	G 26.6%	-	(A)8 31.3%	-	-	-	-	-	-	-	-	-	-	-	-	-	C 15.0%	G 28.9%	-	-	A 26.5%	-	-	-	-	-	-	A 6.4%	-	-	-	-	-
	Amino acid change	A>P	L>L	-	R>R	V>I	-	K>E	S>S	I>M	I>V	-	I>L	L>L	I>T	N>K	G>S	L>P	K>E	P>L	Y>H	Q>Q	W>R	Y>H	G>G	D>D	D>D	I>M	L>P	T>I	R>K	V>V	L>L	L>S	W>G	I>T	N>K



HEK293

		P	F	SH	TM		HN	
	#NT	2415	5876	6676	8057	8057	9618	10484
	#AA	232	435	129	315	315	203	492
	Consensus	(A)7	T	A	Y	Y	A	C
Low passage	1	(A)8 20.2%	A 11.1%	G 11.2%	T 43.1%	C 56.9%	G 8.7%	T 10.6%
	2	(A)8 18.4%	A 12.3%	G 11.8%	T 46.0%	C 53.8%	-	-
	3	(A)8 19.9%	A 11.0%	G 10.9%	T 44.5%	C 55.3%	G 8.6%	-
	Amino acid change	-	N>K	K>E	D>D	D>D	I>M	T>I

Fig. 7. SNVs found in high-passaged and low-passaged virus after one passage in human cells. The consensus genomic sequence of the high-passaged and low-passaged Shaan virus inoculated in MARC145 cells served as a reference for SNV analysis. SNVs were analyzed at minimum variant frequencies of 0.25 and 0.05 in A549 (A), HEK293 (B), and HRT18 (C) cells. The SNVs are highlighted in red with the corresponding amino acid number at the viral gene site. The variant frequency at each SNV is expressed as a percentage. The SNVs found in the original low-passaged and high-passaged Shaan virus are indicated by blue boxes. In the HEK293 cells, results from the high-passaged Shaan virus-infected group (grey box) were excluded from analysis due to low SNV coverage (<100× read depth). The resulting synonymous or non-synonymous amino acid substitutions are presented in the final row.



HRT18		P			M		F		SH	TM	HN
	#NT	2048	2415	2567	3666	4042	4964	5981	6676	7739	10158
	#AA	110	232	283	103	228	131	470	129	209	383
	Consensus	T	(A)7	C	C	T	T	G	A	G	C
Low passage	1	C 6.5%	(A)8 12.1%	A 8.2%	T 15.8%	C 5.5%	-	-	G 99.9%	-	-
	2	C 6.1%	(A)8 13.8%	-	T 15.4%	C 5.6%	-	A 16.4%	G 99.9%	-	-
	3	C 8.8%	(A)8 9.5%	A 9.5%	T 12.3%	-	-	A 14.4%	G 99.8%	-	T 8.8%
High passage	1	C 8.4%	(A)8 12.1%	-	T 15.8%	-	C 99.9%	A 14.7%	-	-	T 9.5%
	2	-	(A)8 10.5%	A 9.9%	T 13.7%	-	-	A 14.8%	-	A 100%	-
	3	C 9.1%	(A) 15.7%	A 9.0%	T 16.2%	-	C 99.9%	-	-	A 100%	T 6.0%
	Amino acid change	F>S	-	A>E	L>F	V>A	T>T	K>K	K>E	G>G	I>I

Fig. 7. Continued.

of three replicates in the low passage-virus-infected A549 cells, all the SNVs from the other low-passaged Shaan virus passaged once in A549 cells matched those found in the original low-passaged virus. In the high-passaged Shaan virus-infected group, out of a total of eight SNVs, five SNVs consistently appeared in the same positions as in the original high-passaged Shaan virus.

A similar pattern was observed in HEK-293 cells (Fig. 7B). Here, all the SNVs in the viruses from one passage of low-passaged Shaan virus in HEK-293 cells matched those in the original low-passaged virus. However, due to low coverage, we were unable to interpret the SNVs in high-passaged virus-infected HEK-293 cells.

Notably, SNVs found in the viruses after one passage in HRT-18 cells were relatively unique compared to those in the other groups (Fig. 7C). Among the eight SNVs found in the viruses from one passage of low-passaged Shaan virus, two SNVs were consistent with the original virus, while the others were HRT-18 cell-specific. Additionally, of the eight SNVs found in the viruses from one passage of the high-passaged Shaan virus, two were consistent with the original virus.

Discussion

Viruses continuously evolve in the infected host. High mutation rates allow viruses a chance to survive through the selected mutants in a given host environment. Certain host environments, such as antiviral drugs, host switching, and immune responses, act as selection factors for the next generation of viruses. Because mutations occur randomly (Luria & Delbrück, 1943; Murray, 2016), given the host environment, certain selected mutations from the random mutation pool may drive viral evolution. In this context, the serial passage of viruses in a particular cell line typically results in highly adapted viruses with changes in virulence and host specificity, which can be applied to the development of attenuated vaccine strains and animal models of human viruses (Song *et al*., 2007; Sanders *et al*., 2016; Li *et al*., 2017; Dinnon *et al*., 2020).

In this context, we investigated selected mutation patterns in a Shaan virus isolate by conducting host-switching experiments in MARC-145 cells (Noh *et al*., 2018). We prepared low-passaged (p4) and high-passaged (p44) Shaan viruses in MARC-145 cells to infect new hosts, specifically A549, HRT-18, and HEK-293 cells of human origin, and then conducted high-throughput sequencing for genetic analysis. As illustrated in Fig. 4, distinct mutations were identified between the low-passaged (p4) and high-passaged (p44) Shaan viruses in MARC-145 cells. In SNV analysis, using a minimum variant frequency of 0.25, six SNVs were detected in the F, SH, TM, HN, and L genes of the low-passaged virus, whereas only one SNV was identified in the high-passaged virus. Given that SNVs in the population are linked to selected mutations that drive evolution (Casci, 2010; Li *et al*., 2022), the high-passaged

Shaan virus (p44) may exhibit better adaptation to MARC-145 cells than the low-passaged virus (p4). Subsequently, we compared the replication and genetic changes between the highly adapted and less adapted viruses in MARC-145 cells when introduced to new host cells (A549, HRT-18, and HEK-293 cells).

We observed replication of Shaan virus in three human cell lines. Given that its receptor was identified as alpha 2,3 sialic acids and it could trigger an innate immune response in HEK293 and A549 cells (Jang *et al*., 2022; Lim *et al*., 2023), replication of Shaan virus in these cell lines is expected. Although both low- and high-passaged viruses were able to replicate and produce infectious virions after one passage in A549, HRT-18, and HEK-293 cells, the high-throughput sequencing data for generating a consensus genomic sequence was insufficient in HEK-293 cells infected with high-passaged Shaan virus (Supplementary Table 1). As high-passaged Shaan virus (p44) was better adapted to MARC-145 cells than low-passaged Shaan virus (p4), it may not replicate efficiently in HEK-293 cells. However, when the viruses successfully infected new host cells, selected mutations were observed after one passage in the new host cells, regardless of whether they were A549, HRT-18, or HEK-293 cells. When low-passaged Shaan virus (p4) was passaged, no mutations were found in HEK-293 cells and two in A549 cells, while nine selected mutations were observed in HRT-18 cells, despite all host cells being of human origin. Although reliable consensus genomic sequences from high-passaged Shaan virus-infected HEK-293 cells were not available due to low mean read depth, distinct differences in the selected mutations were evident between the high-passaged Shaan virus-infected A549 and HRT-18 cells. Interestingly, although data were derived from triplicate experiments, the mutation patterns consistently occurred in the same host cells. These findings suggest a host cell-specific selection pattern for the further evolution of Shaan viruses within a single passage, which may be linked to cell type rather than cell origin.

In this study, HRT-18 cells may serve as a unique host for Shaan viruses. Regardless of whether the Shaan virus infections were low- or high-passaged, viruses that had passed through HRT-18 exhibited more unique selected mutations than those from other cells. Furthermore, Shaan virus replication in HRT-18 cells differed from that in other cell types. As depicted in Fig. 2, Shaan viruses passaged in HRT-18 cells demonstrated significantly enhanced replication properties and yielded higher quantities of infectious virions. Additionally, the level of Shaan virus-specific transcripts associated with the N gene was highest in MARC-145, A549, and HEK-293 cells, regardless of the infection's nature, whether low- or high-passaged, and transcripts related to the M gene were most abundant in HRT-18 cells. Relative to other cells, HRT-18 cells infected with low- and high-passaged virus had higher counts of unique SNVs, with six unique SNVs each (Fig. 7). Given that SNPs in a population are linked to selected mutations for evolution (Casci, 2010; Li *et al*., 2022), it can be inferred that these unique SNVs found in viruses post HRT-18 cell passage may contribute to a greater diversity of Shaan virus variants in these cells. HRT-18 cells are a known colorectal adenocarcinoma cell line. A previous study reported increasing frequencies of novel cryptic SARS-CoV-2 lineages in feces-associated wastewater (Smyth *et al*., 2022). Thus, additional studies examining the relationship between intestinal infection-specific mutations and the emergence of future variant viruses are warranted.

In conclusion, the selected mutations of high-passaged (p44) and low-passaged Shaan virus (p4) after host switching from MARC-145 cells to A549, HRT-18, and HEK-293 cells differed from each other, even when the switched host cells were of the same human origin. Compared to the low-passaged virus, the high-passaged Shaan virus did not generate sufficient high-throughput sequencing data to produce a consensus genomic sequence in HEK-293 cells. However, when analyzing the others, mutation patterns were consistent across the same host cells in triplicate experiments, suggesting a host cell-specific selection pattern for the next generation of Shaan viruses. Among the cells in this study, HRT-18 cells, originating from colorectal adenocarcinoma, produced a more distinct mutation pattern, as well as unique SNVs, during both high- and low-passaged Shaan virus infections. Therefore, this study provides potential evidence for host-specific selective mutation patterns by cell type, as well as host cells, for Shaan virus evolution.

Data Availability

The partial genome sequence of the bat paramyxovirus isolate Bat-ParaV/B16-40 has been deposited in GenBank with the accession number MG230624. The RNA-seq raw data are available in the Gene Expression Omnibus database under the GEO Series accession number GSE201344. Additionally, raw sequence reads can be accessed under the BioProject accession number PRJNA830904 and the SRA accession number SRP371774.

Author Contributions

Conceptualization: Ji Yeong Noh, Hye Kwon Kim; **Data curation:** Ji Yeong Noh; **Formal analysis:** Ji Yeong Noh, Hye Kwon Kim; **Funding acquisition:** Hye Kwon Kim; **Investigation:** Ji Yeong Noh; **Methodology:** Ji Yeong Noh; **Project administration:** Ji Yeong Noh, Hye Kwon Kim; **Resources:** Ji Yeong Noh, Hye Kwon Kim; **Supervision:** Hye Kwon Kim; **Visualization:** Ji Yeong Noh; **Writing-original draft:** Ji Yeong Noh, Hye Kwon Kim.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This research was supported by the Young Researcher Program through the National Research Foundation of Korea (NRF), which is funded by the Ministry of Science and ICT of Korea (NRF-2020R1C1C1010440), the Bio & Medical Technology Development Program through the NRF, funded by the Ministry of Science & ICT (2021M3E5E3083402), and the Basic Science Research Program through the NRF, funded by the Ministry of Education (2020R1A6A1A06046235).

References

Acevedo, A., Brodsky, L., and Andino, R. (2014). Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*, 505, 686-690. https://doi.org/10.1038/nature12861 [PMID: 24284629 PMCID: PMC4111796]

Bushnell, B. (2014). BBMap: A fast, accurate, splice-aware aligner. In: Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States).

Casci, T. (2010). Population genetics: SNPs that come in threes. *Nat Rev Genet*. 11(1):8. https://doi.org/10.1038/nrg2725. [PMID: 20050277]

Das, H., Koizumi, T., Sugimoto, T., Chakraborty, S., Ichimura, T., Hasegawa, K., *et al.* (2000). Quantitation of Fas and Fas ligand gene expression in human ovarian, cervical and endometrial carcinomas using real-time quantitative RT-PCR. *British Journal of Cancer*, 82, 1682-1688. https://doi.org/10.1054/bjoc.2000.1118 [PMID: 10817504 PMCID: PMC2374514]

Dinnon, K.H., Leist, S.R., Schäfer, A., Edwards, C.E., Martinez, D.R., Montgomery, S.A., *et al.* (2020). A mouse-adapted model of SARS-CoV-2 to test COVID-19 countermeasures. *Nature*, 586, 560-566. https://doi.org/10.1038/s41586-020-2708-8 [PMID: 32854108 PMCID: PMC8034761]

Geoghegan, J.L., and Holmes, E.C. (2018). Evolutionary virology at 40. *Genetics*, 210, 1151-1162. https://doi.org/10.1534/genetics.118.301556 [PMID: 30523166 PMCID: PMC6283181]

Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution*, 30, 772-780. https://doi.org/10.1093/molbev/mst010 [PMID: 23329690 PMCID: PMC3603318]

Kim, H., Kwang, J., Yoon, I., Joo, H., and Frey, M. (1993). Enhanced replication of porcine reproductive and respiratory syndrome (PRRS) virus in a homogeneous subpopulation of MA-104 cell line. *Archives of Virology*, 133, 477-483. https://doi.org/10.1007/BF01313785 [PMID: 8257302]

Jang, S.S., Noh, J.Y., Kim, M.C., Lim, H.A., Song, M.S., and Kim, H.K. (2022). α2,3-linked sialic acids are the potential attachment receptor for Shaan virus infection in MARC-145 cells. *Microbiol Spectr*, 10(4), e0125622. https://doi.org/10.1128/spectrum.01256-22 [PMID: 35924912 PMCID: PMC9430483]

Lam, T.T., Jia, N., Zhang, Y.W., Shum, M.H., Jiang, J.F., Zhu, H.C., *et al.* (2020). Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature*. 583(7815):282-285. https://doi.org/10.1038/s41586-020-2169-0. Epub 2020 Mar 26. [PMID: 32218527]

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357-359. https://doi.org/10.1038/nmeth.1923 [PMID: 22388286 PMCID: PMC3322381]

Li, J., Du, P., Yang, L., Zhang, J., Song, C., Chen, D., *et al.* (2022). Two-step fitness selection for intra-host variations in SARS-CoV-2. *Cell Reports*, 38, 110205. https://doi.org/10.1016/j.celrep.2021.110205 [PMID: 34982968 PMCID: PMC8674508]

Li, K., Wohlford-Lenane, C.L., Channappanavar, R., Park, J.-E., Earnest, J.T., Bair, T.B., *et al.* (2017). Mouse-adapted MERS coronavirus causes lethal lung disease in human DPP4 knockin mice. *Proceedings of the National Academy of Sciences*, 114, E3119-E3128. https://doi.org/10.1073/pnas.1619109114 [PMID: 28348219 PMCID: PMC5393213]

Lim, H.A., Noh, J.Y., Jang, S.S., Kim, M.C., Moon, S.H., Kim, H.Y., *et al.* (2023). Innate antiviral responses against Shaan virus infection in HEK293, A549 and MARC-145 cells and limited role of viperin against Shaan virus replication. *Heliyon*, 9(12), e22597. https://doi.org/10.1016/j.heliyon.2023.e22597 [PMID: 38076073 PMCID: PMC10709056]

Luria, S.E., and Delbrück, M. (1943). Mutations of bacteria from virus sensitivity to virus resistance. *Genetics*, 28, 491. https://doi.org/10.1093/genetics/28.6.491 [PMID: 17247100 PMCID: PMC1209226]

McCrone, J.T., and Lauring, A.S. (2018). Genetic bottlenecks in intraspecies virus transmission. *Current Opinion in Virology*, 28, 20-25. https://doi.org/10.1016/j.coviro.2017.10.008 [PMID: 29107838 PMCID: PMC5835166]

Murray, A. (2016). Salvador Luria and Max Delbrück on random mutation and fluctuation tests. *Genetics*, 202, 367-368. https://doi.org/10.1534/genetics.115.186163 [PMID: 26869479 PMCID: PMC4788220]

Noh, J.Y., Jeong, D.G., Yoon, S.W., Kim, J.H., Choi, Y.G., Kang, S.Y., *et al.* (2018). Isolation and characterization of novel bat paramyxovirus B16-40 potentially belonging to the proposed genus Shaanvirus. *Sci Rep*. 8(1):12533. https://doi.org/10.1038/s41598-018-30319-7. [PMID: 30135435; PMCID: PMC6105681]

Peck, K.M., and Lauring, A.S. (2018). Complexities of viral mutation rates. *Journal of Virology*, 92, e01031-01017. https://doi.org/10.1128/JVI.01031-17 [PMID: 29720522 PMCID: PMC6026756]

Rothenburg, S., and Brennan, G. (2020). Species-specific host-virus interactions: Implications for viral host range and virulence. *Trends Microbiol*. 28(1):46-56. https://doi.org/10.1016/j.tim.2019.08.007. Epub 2019 Oct 6. [PMID: 31597598; PMCID: PMC6925338]

Sanders, B.P., de los Rios Oakes, I., van Hoek, V., Bockstal, V., Kamphuis, T., Uil, T.G., *et al.* (2016). Cold-adapted viral attenuation (CAVA): highly temperature sensitive polioviruses as novel vaccine strains for a next generation inactivated poliovirus vaccine. *PLoS Pathogens*, 12, e1005483. https://doi.org/10.1371/journal.ppat.1005483 [PMID: 27032093 PMCID: PMC4816566]

Shi, X., Zhang, X., Chang, Y., Jiang, B., Deng, R., Wang, A., *et al*. (2016). Nonstructural protein 11 (nsp11) of porcine reproductive and respiratory syndrome virus (PRRSV) promotes PRRSV infection in MARC-145 cells. *BMC Veterinary Research*, 12, 1-7. https://doi.org/10.1186/s12917-016-0717-5 [PMID: 27268206 PMCID: PMC4895886]

Smyth, D.S., Trujillo, M., Gregory, D.A., Cheung, K., Gao, A., Graham, M., *et al*. (2022). Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Nature Communications*, 13, 1-9. https://doi.org/10.1038/s41467-022-28246-3 [PMID: 35115523 PMCID: PMC8813986]

Song, D., Oh, J., Kang, B., Yang, J.S., Moon, H., Yoo, H.S., *et al*. (2007). Oral efficacy of Vero cell attenuated porcine epidemic diarrhea virus DR13 strain. *Research in Veterinary Science*, 82, 134-140. https://doi.org/10.1016/j.rvsc.2006.03.007 [PMID: 16730762 PMCID: PMC7111784]

Wargo, A.R., and Kurath, G. (2012). Viral fitness: Definitions, measurement, and current insights. *Current Opinion in Virology*, 2, 538-545. https://doi.org/10.1016/j.coviro.2012.07.007 [PMID: 22986085 PMCID: PMC7102723]

Wignall-Fleming, E.B., Hughes, D.J., Vattipally, S., Modha, S., Goodbourn, S., Davison, A.J., and Randall, R.E. (2019). Analysis of paramyxovirus transcription and replication by high-throughput sequencing. *Journal of Virology*, 93(17):e00571-19. https://doi.org/10.1128/JVI.00571-19. [PMID: 31189700; PMCID: PMC6694822]

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., *et al*. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579, 270-273. https://doi.org/10.1038/s41586-020-2012-7 [PMID:32015507 PMCID: PMC7095418]

Supplementary Materials

Supplementary Table 1. Sequencing data

Cell name	No. passage	Total reads	Trimmed reads	Mapped reads	Mean read depth (×)	The percentage of reads of phred score (>Q30)*	
						Trimmed reads	Mapped reads
MARC145	p4 (Low passage)	76155466	69476952	1630576	8280.2	98.3%	98.2%
		75634052	69000772	1256180	6322.2	98.3%	98.2%
		76298278	69709920	748856	3766.0	98.3%	98.2%
	p44 (High passage)	75536994	68748356	399789	2014.8	98.3%	98.2%
		76099468	69025618	411691	2069.4	98.3%	98.2%
		75747270	68793346	462013	2336.4	98.3%	98.2%
A549	p1 of low passage	70276644	65154680	205120	1049.1	98.4%	98.4%
		57073026	48082936	125545	624.9	98.4%	98.0%
		53316092	49406698	100752	517.0	98.4%	98.4%
	p1 of high passage	50285654	46592890	59306	305.0	98.4%	98.4%
		54227194	49983278	93271	477.5	98.1%	98.4%
		52714662	48615816	156932	802.1	98.4%	98.3%
HEK293	p1 of low passage	75796120	71982628	1031234	5434.7	98.7%	98.7%
		75707030	71871186	867148	4553.6	98.7%	98.6%
		73091564	69472950	1221750	6452.6	98.7%	98.7%
	p1 of high passage	74771938	70963538	1800	9.4	98.7%	98.7%
		76279780	72405440	3256	17.0	98.7%	98.7%
		76111894	71953910	2976	15.6	98.7%	98.7%
HRT18	p1 of low passage	58328530	53270462	465262	2387.8	98.5%	98.4%
		59752932	54840618	397711	2059.2	98.6%	98.5%
		53310344	48593456	392797	2014.9	98.5%	98.4%
	p1 of high passage	55034724	49700814	470425	2384.0	98.3%	98.2%
		61325484	55897850	431535	2217.6	98.5%	98.4%
		59167894	53474980	430264	2177.2	98.4%	98.2%
	Mean						

*Base call accuracy (if phred assigns a Q score of 30 (Q30) to a base, this is equivalent to the probabiloty of an incorrect base call 1 in 1000 times.)

Supplementary Table 2. The raw data in SNV analysis in low- and high-passaged Shaan viruses after one passage in A549 cells

	Gene	Position		Nucleotide change		Amino acid change	Coverage	Variant frequency (%)
		#NT	#AA					
A549-low passaged virus								
1	P	2415	232	(A)7→(A)8	-	-	1006	23.90%
	F	5876	435	T→A	AAT>AAA	N>K	1305	17.10%
		6216	549	G→A	GGC>AGC	G>S	1234	5.90%
	SH	6557	89	T→C	CTA>CCA	L>P	1278	7.50%
		6676	129	A→G	AAA>GAA	K>E	1454	14.60%
	TM	7614	168	T→C	TAT>CAT	Y>H	797	8.00%
		8057	315	Y→T	GAY>GAT	D>D	976	61.80%
				Y→C	GAY>GAC	D>D	976	38.20%
	HN	9618	203	A→G	ATA>ATG	I>M	1073	9.20%
		10484	492	C→T	ACT>ATT	T>I	1143	14.90%
		10502	498	G→A	AGA>AAA	R>K	1198	8.30%
	L	13723	892	T→G	TGG>GGG	W>G	2674	7.20%
		15819	1590	T→A	AAT>AAA	N>K	4770	8.40%
2	N	1045	334	G→C	GCT>CCT	A>P	176	7.40%
		1219	392	T→C	TTG>CTG	L>L	302	5.60%
		1437	464	(A)6→(A)7	-	-	298	6.70%
	P	2248	2316	G→A	GTT>ATT	V>I	237	5.90%
		2415	232	(A)7→(A)8	-	-	66	16.70%
		2706	329	A→G	GAA>GAG	E>E	95	11.60%
		2719	334	A→G	AAG>GAG	K>E	105	11.40%
		2754	345	T→C	TCT>TCC	S>S	127	7.90%
		2793	358	A→G	ATA>ATG	I>M	151	8.60%
		2863	382	A→G	ATT>GTT	I>V	149	8.10%
	M	4152	265	G→DEL	-	-	264	6.10%
		4398	347	A→G	AAG>GAG	K>E	68	5.90%
	F	5085	172	T→C	TTA>CTA	L>L	166	5.40%
		5113	181	T→C	ATA>ACA	I>T	145	6.20%
	SH	6676	129	A→G	AAA>GAA	K>E	8220	15.40%
		6851	187	C→T	CCA>CTA	P>L	184	6.00%
	TM	7518	136	T→C	TGG>CGG	W>R	110	8.20%
		7614	168	T→C	TAT>CAT	Y>H	164	10.40%
		8057	315	Y→T	GAY>GAT	D>D	119	64.70%
				Y→C	GAY>GAC	D>D	119	35.30%
		8682	524	A→G	AAA>GAA	K>E	45	22.20%
	HN	9252	82	AC→DEL	-	-	77	36.40%
		10442	478	T→C	CTA>CCA	L>P	110	11.80%
		10484	492	C→T	ACT>ATT	T>I	98	9.20%
	L	12316	423	(A)5→(A)4	-	-	36	22.20%
		13114	689	T→C	TTG>CTG	L>L	163	7.40%
		13133	695	T→C	TTA>TCA	L>S	126	7.90%

Supplementary Table 2. Continued.

	Gene	Position		Nucleotide change		Amino acid change	Coverage	Variant frequency (%)
		#NT	#AA					
A549-low passaged virus								
		13609	854	A→C	AAA>CAA	K>Q	36	33.30%
		13723	892	T→G	TGG>GGG	W>G	73	20.50%
		16158	1703	A→G	ATA>ATG	I>M	89	5.60%
		16175	1709	A→G	AAG>AGG	K>R	77	6.50%
		16178	1710	A→G	AAA>AGA	K>R	79	6.30%
		16189	1714	A→G	ATG>GTG	M>V	72	6.90%
		16229	1727	A→G	TAT>TGT	Y>C	78	11.50%
		16272	1741	A→G	GAA>GAG	E>E	57	8.80%
3	P	2415	232	(A)7→(A)8	-	-	448	20.80%
	F	5876	435	T→A	AAT>AAA	N>K	1305	14.80%
	SH	6676	129	A→G	AAA>GAA	K>E	675	15.60%
	TM	8057	315	Y→T	GAY>GAT	D>D	460	68.30%
				Y→C	GAY>GAC	D>D	460	31.70%
	HN	9618	203	A→G	ATA>ATG	I>M	610	12.60%
		10484	492	C→T	ACT>ATT	T>I	618	17.20%
		10502	498	G→A	AGA>AAA	R>K	677	5.30%
	L	13723	892	T→G	TGG>GGG	W>G	191	11.00%
		15819	1590	T→A	AAT>AAA	N>K	194	6.70%
A549-high passaged virus								
1	P	1896	59	A→G	AGA>AGG	R>R	399	29.10%
		2415	232	(A)7→(A)8	-	-	273	27.50%
	F	4590	7	A→T		I>L	285	42.80%
	TM	7119	3	T→C	TAT>CAT	Y>H	143	18.20%
		7292	60	A→G	CAA>CAG	Q>Q	154	24.70%
		7739	209	G→A	GGG>GGA	G>G	178	28.70%
	L	16666	1873	T→C	TTT>CTT	F>L	92	5.40%
		16674	1875	T→C	TAT>TAC	Y>Y	96	8.30%
2	P	1896	59	A→G	AGA>AGG	R>R	673	26.70%
		2415	232	(A)7→(A)8	-	-	469	25.40%
	F	4590	7	A→T	ATA>TTA	I>L	521	41.10%
	TM	7119	3	T→C	TAT>CAT	Y>H	188	13.80%
		7292	60	A→G	CAA>CAG	Q>Q	321	34.90%
		7739	209	G→A	GGG>GGA	G>G	295	28.10%
	L	13940	964	T→C	ATA>ACA	I>T	112	7.10%
3	P	1896	59	A→G	AGA>AGG	R>R	1107	26.60%
		2415	232	(A)7→(A)8	-	-	776	31.30%
	TM	7119	3	T→C	TAT>CAT	Y>H	407	15.00%
		7292	60	A→G	CAA>CAG	Q>Q	589	28.90%
		7739	209	G→A	GGG>GGA	G>G	509	26.50%
	L	12519	490	G→A	GTG>GTA	V>V	267	6.40%

Supplementary Table 3. The raw data in SNV analysis in low- and high-passaged Shaan viruses after one passage in HEK293 cells

	Gene	Position		Nucleotide change		Amino acid change	Coverage	Variant frequency (%)
		#NT	#AA					
HEK293-low passaged virus								
1	P	2415	232	(A)7→(A)8	-	-	5881	20.20%
	F	5876	435	T→A	AAT>AAA	N>K	4434	11.10%
	SH	6676	129	A→G	AAA>GAA	K>E	6329	11.20%
	TM	8057	315	Y→T	GAY>GAT	D>D	4128	43.10%
				Y→C	GAY>GAC	D>D	4128	56.90%
	HN	9618	203	A→G	ATA>ATG	I>M	3385	8.70%
		10484	492	C→T	ACT>ATT	T>I	3470	10.60%
2	P	2415	232	(A)7→(A)8	-	-	3895	18.40%
	F	5876	435	T→A	AAT>AAA	N>K	3662	12.30%
	SH	6676	129	A→G	AAA>GAA	K>E	4731	11.80%
	TM	8057	315	Y→T	GAY>GAT	D>D	2871	46.00%
				Y→C	GAY>GAC	D>D	2871	53.80%
3	P	2415	232	(A)7→(A)8	-	-	6435	19.90%
	F	5876	435	T→A	AAT>AAA	N>K	5244	11.00%
	SH	6676	129	A→G	AAA>GAA	K>E	7214	10.90%
	TM	8057	315	Y→T	GAY>GAT	D>D	4349	44.50%
				Y→C	GAY>GAC	D>D	4349	55.30%
	HN	9618	203	A→G	ATA>ATG	I>M	3793	8.60%
HEK293-high passaged virus								
1	N	279	78	T→C	TCT>TCC	S>S	47	6.40%
	P	1908	63	T→C	CAT>CAC	H>H	14	21.40%
		2342	208	C→T	CCC>CTC	P>L	6	33.30%
		2415	232	(A)7→(A)8	-	-	5	80.00%
	F	4590	7	A→T	ATA>TTA	I>L	12	33.30%
	HN	12113	355	(T)7→(T)8	-	-	6	33.30%
2	P	1896	59	A→G	AGA>AGG	R>R	22	36.40%
		2415	232	(A)7→(A)8	-	-	22	27.30%
	M	3851	164	(T)3→(T)4	-	-	45	8.90%
	F	4590	7	A→T	ATA>TTA	I>L	13	69.20%
	SH	6563	91	(A)6→(A)7	-	-	22	45.50%
3	P	1896	59	A→G	AGA>AGG	R>R	20	25.00%
		2415	232	(A)7→(A)8	-	-	13	23.10%
		2712	331	T→C	AAT>AAC	N>N	13	30.80%
	M	3491	44	C→T	ATA>TTA	I>L	28	10.70%
	TM	7292	60	A→G	CAA>CAG	Q>Q	15	40.00%
		7394	94	G→A	CCG>CCA	P>P	9	44.40%
		8751	547	T→C	TAC>CAC	Y>H	37	10.80%
	L	12684	545	A→G	AAA>AAG	K>K	5	80.00%
		12691	548	A→G	AGA>GGA	R>G	7	57.10%
		12919-12920	624	AA→GG	AAC>GGC	N>G	8	50.00%
		12928	627	A→G	ACA>GCA	T>A	9	44.40%
		12955-12956	636	AA→GG	AAA>GGA	K>G	9	44.40%
		12964	639	A→G	AGA>GGA	R>G	9	44.40%
		16088	1680	G→A	AGA>AAA	R>K	10	60.00%
		16490	1814	A→G	TAC>TGC	Y>C	4	50.00%
		16500	1817	A→G	TCA>TCG	S>S	5	40.00%

Supplementary Table 4. The raw data in SNV analysis in low- and high-passaged Shaan viruses after one passage in HRT18 cells

	Gene	Position		Nucleotide change		Amino acid change	Coverage	Variant frequency (%)
		#NT	#AA					
HRT18-low passaged virus								
1	P	2048	110	T→C	TTT>TCT	F>S	1424	6.50%
		2415	232	(A)7→(A)8	-	-	1410	12.10%
		2567	283	C→A	GCG>GAG	A>E	1765	8.20%
	M	3666	103	C→T	CTT>TTT	L>F	4935	15.80%
		4042	228	T→C	GTC>GCC	V>A	3945	5.50%
	SH	6676	129	A→G	AAA>GAA	K>E	2029	99.90%
2	P	2048	110	T→C	TTT>TCT	F>S	1449	6.10%
		2415	232	(A)7→(A)8	-	-	1281	13.80%
	M	3666	103	C→T	CTT>TTT	L>F	4856	15.40%
		4042	228	T→C	GTC>GCC	V>A	3715	5.60%
	F	5981	470	G→A	AAG>AAA	K>K	1935	16.40%
	SH	6676	129	A→G	AAA>GAA	K>E	1594	99.90%
3	P	2048	110	T→C	TTT>TCT	F>S	1270	8.80%
		2415	232	(A)7→(A)8	-	-	1229	9.50%
		2567	283	C→A	GCG>GAG	A>E	1673	9.50%
	M	3666	103	C→T	CTT>TTT	L>F	3984	12.30%
	F	5981	470	G→A	AAG>AAA	K>K	1444	99.80%
	HN	10158	383	C→T	ATC>ATT	I>I	1746	8.80%
HRT18-high passaged virus								
1	P	2048	110	T→C	TTT>TCT	F>S	1554	8.40%
		2415	232	(A)7→(A)8	-	-	1412	12.10%
	M	3666	103	C→T	CTT>TTT	L>F	5144	15.80%
	F	4964	131	T→C	ACT>ACC	T>T	1875	99.90%
		5981	470	G→A	AAG>AAA	K>K	2184	14.70%
	HN	10158	383	C→T	ATC>ATT	I>I	1866	9.50%
2	P	2415	232	(A)7→(A)8	-	-	1322	10.50%
		2567	283	C→A	GCG>GAG	A>E	1668	9.90%
	M	3666	103	C→T	CTT>TTT	L>F	4768	13.70%
	F	5981	470	G→A	AAG>AAA	K>K	1770	14.80%
	TM	7739	209	G→A	GGG>GGA	G>G	410	100%
3	P	2048	110	T→C	TTT>TCT	F>S	1339	9.10%
		2415	232	(A)7→(A)8	-	-	1366	15.70%
		2567	283	C→A	GCG>GAG	A>E	1745	9.00%
		3666	103	C→T	CTT>TTT	L>F	5053	16.20%
	F	4964	131	T→C	ACT>ACC	T>T	1792	99.90%
	TM	7739	209	G→A	GGG>GGA	G>G	453	100%
	HN	10158	383	C→T	ATC>ATT	I>I	1531	6.00%

Supplementary Table 5. Primer and probe of target gene detection with real time RT-PCR

Target gene	Name	Seq. (5′-3′)	Ref.
For detection of viral RNA			
Shaan virus SH gene	PVM-F	CCCAGGAGTATGGTTATCAAGTGAGG	
	PVM-R	TCCATTGGGCTCTCTTTGTTTGC	
	PVM-P	FAM-CCCATCCCAGACCAGCCACCAGACCC-TAMRA	
For detection of antigenome			
primer for cDNA synthesis	N-P cD primer	GAGCATCCAGAGACTTCCAGTGACGTTTGGCTTTGTCTG	
Shaan virus antigenom	BPV-NP-F	GGATTAACCTCAAAACCAATGC	
	BPV-tail-R	GAGCATCCAGAGACTTCCA	
	BPV-NP-P	FAM-AAAGTGACATAGGCCAAGCTC-BHQ1	
Human GAPDH gene	hGAPDH-F	GAAGGTGAAGGTCGGAGTC	Das *et al.*, 2000
	hGAPDH-R	GAAGATGGTGATGGGATTTC	
	hGAPDH-P	HEX-CAAGCTTCCCGTTCTCAGCC-BHQ1	
Monkey GAPDH gene	mGAPDH-F	TGACAACAGCCTCAAGATCG	Shi *et al.*, 2016
	mGAPDH-R	GTCTTCTGGGTGGCAGTGAT	
	mGAPDH-P	HEX-TGGAAGGACTCATGACCACA-BHQ1	