Article

Frequencies and Features of Frameshift Mutations in Microevolution and Macroevolution of Viruses

Yu-Fei Ji, Jian-Wei Shao, Ming-Hui Sun, Guo-Hui Li, Rui-Xu Chen, Huan-Yu Gong, Ji-Ming Chen*,

School of Life Science and Engineering, Foshan University, Foshan, 528225 Guangdong, China.

***Corresponding author**

Ji-Ming Chen

E-mail: jmchen@fosu.edu.cn

Abstract

Some nucleotide insertions or deletions (indels) in protein-coding open reading frames lead to frameshift mutations (FSMs) which can change amino acid sequences drastically. FSMs are widely distributed in the genomes of many organisms. However, few studies have been reported regarding frequencies of FSMs in microevolution or macroevolution. Many viruses evolve much more rapidly than cellular organisms, and they are hence suitable to investigate frequencies of FSMs in microevolution or macroevolution. In this report, we identified 667 FSMs in gene sequences of 13 virus families and each FSM changed approximately 11 amino acid residues on average. Of the FSMs, 89.21% were 2-indel compensatory FSMs, and the remaining were 1-, 3-, 4-, 5-indel FSMs. We found that FSMs usually occurred more frequently in the viruses of the same family with smaller sequence identities, and FSMs occurred in the sequences of with identities of 60.0–69.9% more frequently than in the sequences with identities of 90.0–99.9% or 80.0–89.9% by approximately 34.9 or 13.1 times on average. We also found FSMs occurred at different frequencies among genes in the same virus genome, among species in the same virus family, or among virus families (e.g., more frequently in *Coronaviridae* than in *Orthomyxoviridae*). These results suggest that FSMs are more frequent in the inter-species or macroevolution than in the intra-species or microevolution of viruses. They provide novel evidence for the hopeful monster hypothesis in evolutionary biology. They inspire researchers to investigate the roles,

frequencies, features, and functions of FSMs in other viruses and cellular organisms.

Keywords: virus; frameshift mutation; evolution; macroevolution; hopeful monster; phylogenetics

Introduction

Insertions or deletions (indels) of 3n±1 nucleotides in protein-coding open reading frames (ORFs) lead to frameshift mutations (FSMs) which can change amino acid sequences drastically (Dong, et al. 2022; Yang, et al. 2022). FSMs are widely distributed in the genomes of many organisms and can lead to phenotypic diversifications (e.g., plumage variations) and pathological changes (e.g., cancers) in organisms (Dong, et al. 2022; Yang, et al. 2022).

Few studies have been reported regarding frequencies of FSMs in microevolution or macroevolution of organisms. Many viruses evolve much more rapidly than cellular organisms, and they are hence suitable to investigate frequencies of FSMs in microevolution or macroevolution (Weaver, et al. 2016).

Viruses are the etiological agents of many infectious diseases of humans, animals, and plants, such as rabies, influenza, polio, dengue fever, and coronavirus disease 2019 (COVID-19). Although most FSMs in virus genomes likely lower fitness of viruses, some FSMs could aid viruses to adapt to new hosts,

escape from immunity, and impact viral pathogenicity and transmissibility. For example, FSMs in the genome of vaccinia virus could rapidly attenuate the virus (Senkevich, et al. 2020). FSMs in the ORF7a gene of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) could reduce the virus pathogenicity (Joonlasak, et al. 2021). An FSM in the nsp6 gene of SARS-CoV-2 could be associated with adaptation of the virus to cats, dogs, and hamsters (Bashor, et al. 2021). Two FSMs in the nsp2 gene of SARS-CoV-2 compared with SARS-CoV-1 could be associated with the virus transmissibility (Angeletti, et al. 2020). An FSM in the precure region of the genome of hepatitis B virus (HBV) could be associated with spontaneous reactivation of the virus in a chronic HBV carrier (Takakusagi, et al. 2021).

Few studies on FSMs in the evolution of viruses have been reported. A recent milestone study parsed 1,233,275 viral sequences of 10,115 genes of many RNA viruses, and found that FSMs occurred in the evolution of 385 genes on small scales (i.e., sequence identities >97%) (Park and Hahn 2021). For instance, 2 of 5,606 the hemagglutinin gene sequences (sequence identities >97%) of H1 subtype influenza virus carried FSMs, and 6 of 685 the hemagglutinin neuraminidase (HN) gene sequences of paramyxoviruses (sequence identities >97%) carried FSMs (Park and Hahn 2021).

It remains unknown about the frequencies of FSMs in the evolution of virus genomes on grander scales. It also remains unknown about any features of FSMs in the evolution of virus genomes. For example, it is unknown whether

FSMs occurred at distinct frequencies among genes in the same virus genome, among species in the same virus family, or among virus families.

In this report, we analyzed gene sequences of multiple virus families to investigate the frequencies and features of FSMs in the evolution of virus genomes on four scales and among various families or various species in the same families.

Results

FSMs in 13 virus families

We downloaded and analyzed gene sequences of 11 RNA virus families and two DNA virus families, including the newly established virus families of *Hantaviridae*, *Nairoviridae*, *Peribunyaviridae*, and *Phenuiviridae* (Laenen, et al. 2019; Garrison, et al. 2020; Hughes, et al. 2020; Sun, et al. 2022). These sequences constituted 572 fas files which harbored 358, 140, 190, and 242 groups with intra-group sequence identities 90.0–99.9%, 80.0–89.9%, 70.0–79.9%, and 60.0–69.9%, respectively. FSMs occurred in all the 13 virus families (fig. 1), and a total of 667 FSMs were identified as per sites and groups in the 572 fas files (table 1). We randomly selected 60 of the 667 FSMs and found each FSM changed 2–34 (mean ± standard deviation = 11.31 ± 6.55) amino acid residues.

FSM frequencies and sequence identities

Fig. 2 and table 1 showed that FSMs usually occurred more frequently in

virus genomes with smaller sequence identities ($P<0.01$, by the Mann-Kendall test). For example, FSMs occurred in the sequences of *Flaviviridae* with identities of 60.0–69.9% more frequently than in the sequences with identities 90.0–99.9%, 80.0–89.9%, and 70.0–79.9% by 18.7, 16.3, and 9.0 times, respectively. FSMs occurred in the sequences of *Picornaviridae* with identities of 60.0–69.9% more frequently than in the sequences with identities 90.0–99.9%, 80.0–89.9%, and 70.0–79.9% by 18.7, 7.0, and 3.3 times, respectively. Fig. 1 showed that, taken the 13 families together, FSMs occurred in the sequences of with identities of 60.0–69.9% more frequently than in the sequences with identities 90.0–99.9% and 80.0–89.9% by approximately 34.9 and 13.1 times on average, respectively.

Distribution of FSM types

Of the 667 FSMs, four (0.30%) were non-compensatory FSMs formed with only a single indel near to termination codons (fig. 3). The remaining 663 FSMs were compensatory, and 595 (89.21%) were 2-indel compensatory FSMs, 63 (9.45%) 3-indel compensatory FSMs, four (0.60%) 4-indel compensatory FSMs, and one (0.15%) 5-indel compensatory FSMs (fig. 3).

FSM frequencies and virus taxa

FSMs occurred with different frequencies among different families. For example, table 1 showed that FSMs occurred more frequently in the spike gene of coronaviruses than in the hemagglutinin gene of influenza viruses by more than 10 times ($P<0.01$). These two genes encode the viral membrane

proteins to bind to cell receptors. They both evolve more rapidly than most other genes in the viral genomes (Chen and Sun 2011). Similarly, table 1 showed that FSMs occurred more frequently in the polyprotein genes of picornaviriuses than in the polyprotein genes of flaviviruses by more than 100% when the sequence identities were 70.0–79.9% or 60.0–69.9% ($P<0.05$). These two polyprotein genes encode the viral RNA-dependent RNA polymerase. They evolve less slowly than other genes in the viral genomes.

FSMs occurred with different frequencies among different species in the same family. For example, the three FSMs in the six groups (which represented five species) of the large protein gene sequences of paramyxoviruses in table 1 with sequence identities 80.0–89.9% occurred in only one group which represented the species of *Avian metaavulavirus 2*. The FSM frequency in the large protein gene was significantly higher in *Avian metaavulavirus 2* than in other four *Paramyxoviridae* species ($P<0.01$).

FSM frequencies in the same virus genomes

We analyzed of the genomic sequences of six randomly selected species from three randomly selected virus families (fig. 4). In *Coronaviridae*, FSMs occurred more frequently in S gene than in ORF1ab and N gene by 2.4 and 3.2 times. In *Paramyxoviridae*, FSMs occurred more frequently in HN gene than in L genome by 2.2 times. In *Picornaviridae*, FSMs occurred more frequently in P1 gene than in 3D genome by 3.8 times.

FSM identification and sequence alignment

Table 2 showed that fewer FSMs could be identified if viral sequences were aligned with higher gap penalties using the software tool Mega with the MUSCLE method, likely because higher gap penalties lead to fewer indels. The table also showed that the numbers of FSMs identified using the software tool MAFFT with the mode of E-INS-I were fully consistent with the FSMs identified using MEGA with the gap penalty of –800 and more than the FSMs identified using MEGA with the gap penalty of –400 or –600 and more than the FSMs identified using the software tool MAFFT with the mode of FFT-NS-2.

Discussion

Few studies have been reported regarding FSMs in the evolution of viruses, including the viruses in the 13 families (Hughes, et al. 2010; Taubenberger and Kash 2010; Breitbart, et al. 2017; Simmonds, et al. 2017; Zell, et al. 2017; Chen, et al. 2018; Laenen, et al. 2019; Radoshitzky, et al. 2019; Rima, et al. 2019; Garrison, et al. 2020; Hughes, et al. 2020; Sun, et al. 2022; Zmasek, et al. 2022). This could be partially because researchers usually focused on virus evolution on small scales (e.g., intra-species evolution or microevolution of influenza virus or SARS-CoV-2), and FSMs are relatively rare in virus evolution on small scale, as revealed by this report and a previous report (Park and Hahn 2021). Consequently, we had the opportunity in this study to reveal frequencies and multiple features of FSMs in virus evolution on various scales for the first time.

Previous reports classified FSMs into 2-indel compensatory FSMs and non-compensatory FSMs, and one report suggested that their frequencies are similar (Park and Hahn 2021; Rogers, et al. 2021). Here we identified that, from a broader view, there were 3-, 4-, and 5-indel compensatory FSMs in virus evolution, besides 1-indel non-compensatory FSMs and 2-indel compensatory FSMs (fig. 3A). We further showed that most FSMs in virus evolution are 2-indel compensatory FSMs, even on small scales (fig. 3B).

This report demonstrated that FSMs usually occurred more frequently in virus genomes in the same family with smaller sequence identities which could correspond to inter-species or inter-genus virus evolution. This finding suggests that FSMs are more important to macroevolution or inter-species or inter-genus evolution of viruses than to microevolution or intra-species evolution of viruses.

This report also demonstrated that FSMs not always occurred more frequently in virus genomes in the same family with smaller sequence identities. This could result from sampling errors, particularly when FSMs occurred at different frequencies among species in the family.

Because FSMs can change multiple amino acid residues, they can lead to major phenotypic transformations, like virus genomic recombination, genomic reassortment, and mutations at key sites. These major phenotypic transformations could aid viruses to adapt to distinct environments or hosts, although most of them could reduce their fitness in their original

environments or hosts. Therefore, this finding provides novel evidence to support the hopeful monster hypothesis in evolutionary theory which claimed that some major evolutionary transformations have occurred in large leaps between species due to macromutations (Rogers, et al. 2021).

This report revealed that FSMs occurred at distinct frequencies among virus families or species with similar sequence identities. This suggests that different virus families or species favor different approaches or mechanisms to change their amino acid sequences. For example, influenza viruses favor nucleotide substitutions rather than FSMs, and paramyxoviruses could favor both nucleotide substitutions and FSMs, to change their amino acid sequences (table 1).

Because FSMs occurred frequently in the genomes of some species in some families, phylogenetic relationships of some viruses calculated with amino acid sequences could be somehow unreliable because one 2-indel compensatory FSM could be counted as multiple nucleotide substitutions through alignment of the relevant amino acid sequences (Park and Hahn 2021).

This report found that FSMs occurred at distinct frequencies among different genes in the same genome. This could result from the fact that different genes in the same viral genome undertake different positive selection pressure and negative selection pressure. The data of three virus families in fig. 4 all suggest that FSMs likely occurred less frequently in more conserved genes (e.g., the 3D gene of picornaviruses which encodes the viral RNA

polymerase).

Because many human viruses were caused by RNA viruses, this report analyzed more RNA virus families than DNA virus families. This report analyzed more sequences of viral RNA or DNA polymerase genes than other genes because they are shared by all the viruses in the relevant taxa.

Nucleotide sequences can be aligned correctly only at certain possibility. Consequently, the FSMs identified in this report were reliable only at certain possibility. This issue is more obvious when sequence identities were lower (Park and Hahn 2021). Therefore, we did not analyze FSM frequencies in the sequences with identities lower than 40% in this study.

To improve the alignment quality and reduce the difficulty in analyzing alignment results, each fas file in this study contained only 2–10 sequences. We found that the difference in the sequence quantity among the fas files impacted little on the FSM counting.

Table 2 suggested that the method using MAFFT with the mode of E-INS-I for identifying FSMs in this report was the most stringent, which means that all FSMs identified in this report were reliable, although some FSMs could have not been identified. This does not affect the conclusions of this report because these conclusions were based on the same FSM identification method.

This study suggests that it is valuable to investigate in the future whether the features of FSMs in virus genomes revealed by this study are applicable in cellular organisms (e.g., whether FSMs are more prevalent in human genomes

if compared with the genome of a cellular organism which is more phylogenetically distant to humans). This study also inspires researchers to investigate the roles, frequencies, and features of FSMs which have been neglected in the evolution of other viruses, particularly in their evolution on grander scales.

Together, in this report, we identified 667 FSMs in gene sequences of 13 virus families and each FSM changed approximately 11 amino acid residues on average. Most of the FSMs were 2-indel compensatory FSMs, and a few of them were 1-, 3-, 4-, 5-indel FSMs. We found that FSMs usually occurred more frequently in virus genomes of the same family with smaller sequence identities. We also found that FSMs occurred at different frequencies in different virus families, different species of the same family, and different genes of the same virus genomes. We further showed that sequence alignment methods and parameters influence substantially the counts of FSMs in viral sequences. These results suggest that FSMs are more frequent in the inter-species or macroevolution than in the intra-species or microevolution of viruses. They provide novel evidence for the hopeful monster hypothesis in evolutionary biology. They also inspire researchers to investigate FSMs in other viruses and cellular organisms.

Materials and Methods

Sequence downloading and designation. Nucleotide sequences of

complete ORFs of virus genes were downloaded from National Library of Medicine (NCBI). Sequences were designated with virus species names and sequence accession numbers.

Sequence alignment and sequence identity calculation. Nucleotide sequences were aligned using the online software tool MAFFT (version 7) with the advanced setting of E-INS-I which is very slow and recommended for <200 sequences with multiple conserved domains and long gaps (Katoh, et al. 2002). Sequence identities of nucleotide sequences were calculated using the software tool MEGA (version 11) as per pairwise deletion (Tamura, et al. 2021).

Establishment of multiple sequence files. Multiple sequence files in the fas format were established for each virus family. Each fas file contained 2–10 sequences which formed two or more groups of sequences with intra-group sequence identities of 90.0–99.9%, 80.0–89.9%, 70.0–79.9%, or 60.0–69.9% (fig. 1A). One sequence in a fas file could be shared by two or more groups (fig .1A). These fas files were aligned again using MAFFT with the mode of E-INS-I. Meanwhile, some fas files were aligned using MEGA with the MUSCLE method (Tamura, et al. 2021).

FSM frequency calculation. When two or more adjacent frameshifting indels restored the original ORF, these adjacent indels constituted compensatory FSMs. Those FSMs constituted by only one mutation were termed non-compensatory FSMs which are usually near termination codons. FSM frequencies were calculated as per FSM counts, group counts, sequence

average length, genes, and sequence identities. For example, if 90 sequences (2700-nucleotide long before alignment on average) of a virus gene constituted 12 groups with intra-group sequence identities of 80.0–89.9%, and these 12 groups contained 36 FSMs, then the FSM frequency of this gene on the scale of sequence identities of 80.0–89.9% was 36/12/2,700*5,000 = 0.56 FSMs per group per 5,000 nucleotides (FSMs/G/FTN).

Statistical tests. Differences were compared using the Chi-square test, if not specified. $P<0.05$ represented statistical significance.

Acknowledgments

This study was supported by the High-Level Talent Fund of Foshan University (No. 20210036). The funder did not play any role in this study.

References

Angeletti S, Benvenuto D, Bianchi M, Giovanetti M, Pascarella S, Ciccozzi M. 2020. COVID-2019: The role of the nsp2 and nsp3 in its pathogenesis. *J Med Virol.* 92(6):584-588.

Bashor L, Gagne RB, Bosco-Lauth AM, Bowen RA, Stenglein M, VandeWoude S. 2021. SARS-CoV-2 evolution in animals suggests mechanisms for rapid variant selection. *Proc Natl Acad Sci U S A.* 118(44): e2105253118.

Breitbart M, Delwart E, Rosario K, Segales J, Varsani A, Ictv Report C. 2017. ICTV Virus Taxonomy

Profile: Circoviridae. *J Gen Virol*. 98(8):1997-1998.

Chen J, Sun Y. 2011. Variation in the analysis of positively selected sites using nonsynonymous/synonymous rate ratios: an example using influenza virus. *PLoS One.* 6(5):e19996.

Chen R, Mukhopadhyay S, Merits A, Bolling B, Nasar F, Coffey LL, Powers A, Weaver SC, Ictv Report C. 2018. ICTV Virus Taxonomy Profile: Togaviridae. *J Gen Virol.* 99(6):761-762.

Dong MJ, Yang ZK, Yang J, Guo RQ, Xiao YY, Liu H. 2022. Gamma knife radiotherapy in a neurofibromatosis type 1 Chinese pedigrees with NF1 gene frameshift mutation: A case report. *Medicine (Baltimore)* 101(27):e29280.

Garrison AR, Alkhovsky AC, Avsic-Zupanc T, Bente DA, Bergeron E, Burt F, Di Paola N, Ergunay K, Hewson R, Kuhn JH, et al. 2020. ICTV Virus Taxonomy Profile: Nairoviridae. *J Gen Virol.* 101:798-799.

Hughes AL, Irausquin S, Friedman R. 2010. The evolutionary biology of poxviruses. *Infect Genet Evol.* 10(1):50-59.

Hughes HR, Adkins S, Alkhovskiy S, Beer M, Blair C, Calisher CH, Drebot M, Lambert AJ, de Souza WM, Marklewitz M, et al. 2020. ICTV Virus Taxonomy Profile: Peribunyaviridae. *J Gen Virol.* 101(1):1-2.

Joonlasak K, Batty EM, Kochakarn T, Panthan B, Kumpornsin K, Jiaranai P, Wangwiwatsin A, Huang A, Kotanan N, Jaru-Ampornpan P, et al. 2021. Genomic surveillance of SARS-CoV-2 in Thailand reveals mixed imported populations, a local lineage expansion and a virus with truncated ORF7a. *Virus Res.* 292:198233.

Katoh K, Misawa K Fau - Kuma K-i, Kuma K Fau - Miyata T, Miyata T. 2002. MAFFT: a novel

method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 15(14):3059-3066.

Laenen L, Vergote V, Calisher CH, Klempa B, Klingstrom J, Kuhn JH, Maes P. 2019. Hantaviridae: Current Classification and Future Perspectives. *Viruses.* 11(9):788.

Park D, Hahn Y. 2021. Rapid protein sequence evolution via compensatory frameshift is widespread in RNA virus genomes. *BMC Biotechnol*. 22(1):251.

Radoshitzky SR, Buchmeier MJ, Charrel RN, Clegg JCS, Gonzalez JJ, Gunther S, Hepojoki J, Kuhn JH, Lukashevich IS, Romanowski V, et al. 2019. ICTV Virus Taxonomy Profile: Arenaviridae. *J Gen Virol.* 100(8):1200-1201.

Rima B, Balkema-Buschmann A, Dundon WG, Duprex P, Easton A, Fouchier R, Kurath G, Lamb R, Lee B, Rota P, et al. 2019. ICTV Virus Taxonomy Profile: Paramyxoviridae. *J Gen Virol.* 100(12):1593-1594.

Rogers RA-O, Grizzard SL, Titus-McQuillan JE, Bockrath KA-O, Patel S, Wares JP, Garner JT, Moore CC. 2021. Gene family amplification facilitates adaptation in freshwater unionid bivalve Megalonaias nervosa. *Mol Ecol.* 30(5):1155-1173.

Senkevich T, Zhivkoplias E, Weisberg A, Moss B. (4.501 co-authors). 2020. Inactivation of Genes by Frameshift Mutations Provides Rapid Adaptation of an Attenuated Vaccinia Virus. *J Virol.* 94:e01053-01020.

Simmonds P, Becher P, Bukh J, Gould EA, Meyers G, Monath T, Muerhoff S, Pletnev A, Rico-Hesse R, Smith DB, et al. 2017. ICTV Virus Taxonomy Profile: Flaviviridae. *J Gen Virol.* 98(1):2-3.

Sun MH, Ji YF, Li GH, Shao JW, Chen RX, Gong HY, Chen SY, Chen JA-O. 2022. Highly adaptive Phenuiviridae with biomedical importance in multiple fields. *J Med Virol.* 94(6):2388-2401.

Takakusagi S, Takagi H, Yokoyama Y, Kizawa K, Marubashi K, Kosone T, Nagashima S, Takahashi M, Murata K, Okamoto H. 2021. Spontaneous reactivation of hepatitis B virus with a frameshift mutation in the precore region in an elderly hepatitis B virus carrier with lifestyle-related diseases. *J Clin Gastroenterol.* 14(4):1202-1210.

Tamura K, Stecher G, Kumar SA-O. 2021. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol Biol Evol.* 38(7):3022-3027.

Taubenberger JK, Kash JC. 2010. Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host Microbe.* 7(6):440-451.

Weaver SC, Denison M, Roossinck MJ, Vignuzzi M. 2016. Virus evolution : current research and future directions. Wymondham: Caister Academic Press.

Yang Y, Wang H, Li G, Liu Y, Wang C, Qiu S, Wang X, Yao J, Zhu L, He D. 2022. Using comparative genomics to detect mutations regulating plumage variations in graylag (A. anser) and swan geese (A. cygnoides). *Gene*. 834:146612.

Zell R, Delwart E, Gorbalenya AE, Hovi T, King AMQ, Knowles NJ, Lindberg AM, Pallansch MA, Palmenberg AC, Reuter G, et al. 2017. ICTV Virus Taxonomy Profile: Picornaviridae. *J Gen Virol.* 98(10):2421-2422.

Zmasek CM, Lefkowitz EJ, Niewiadomska A, Scheuermann RH. 2022. Genomic evolution of the Coronaviridae family. *Virology.* 570:123-133.

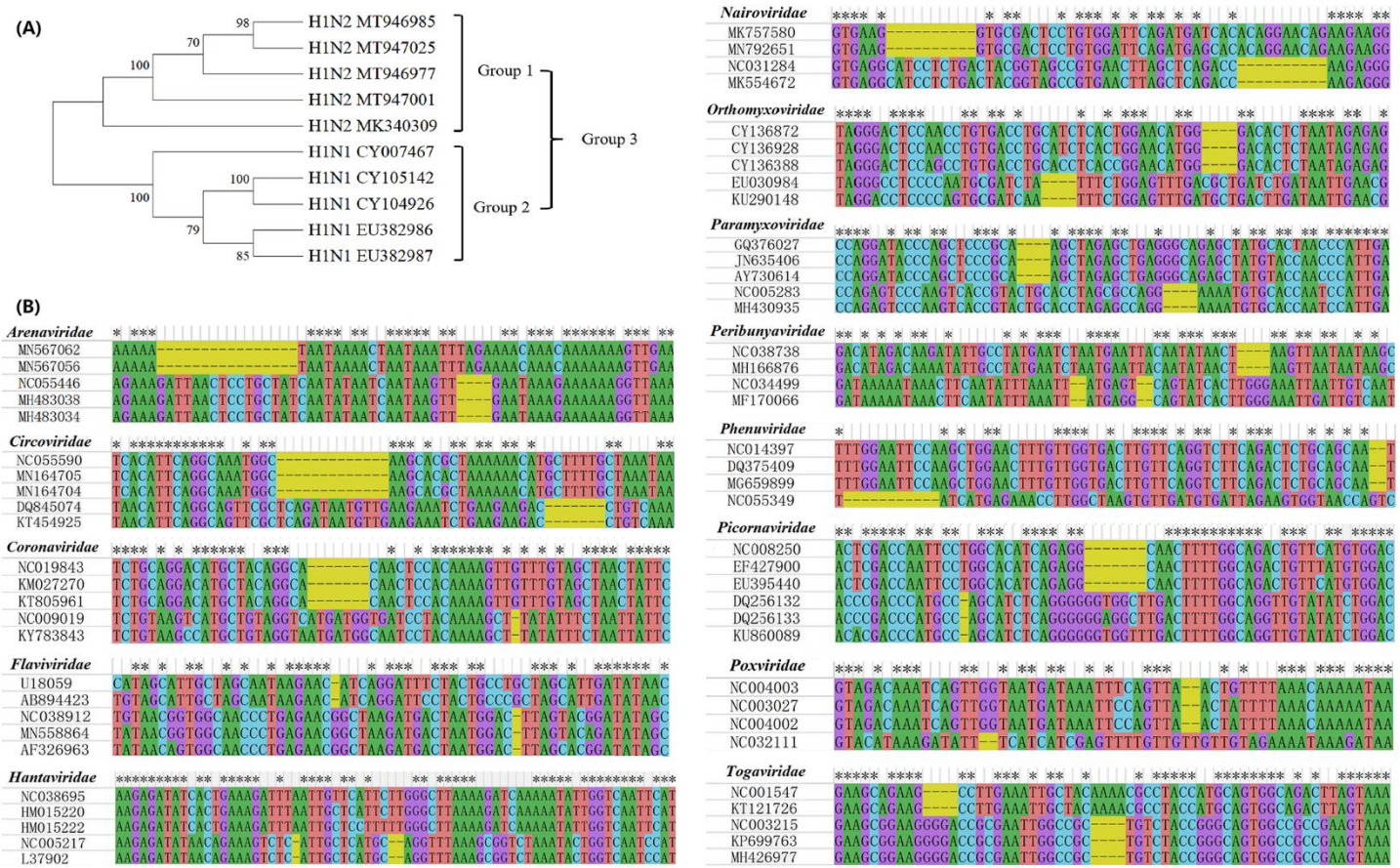

FIG. 1. Examples of groups in fas files and frameshift mutations in the genomes of 13 virus families

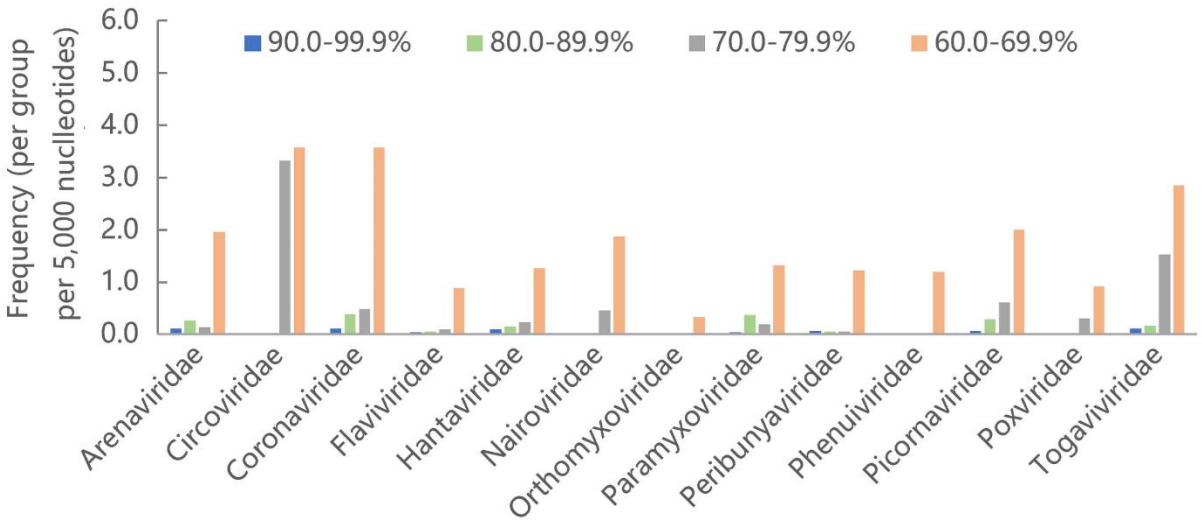

FIG. 2. Frequencies of frameshift mutations in 13 virus families among the sequences with four ranges of sequence identities

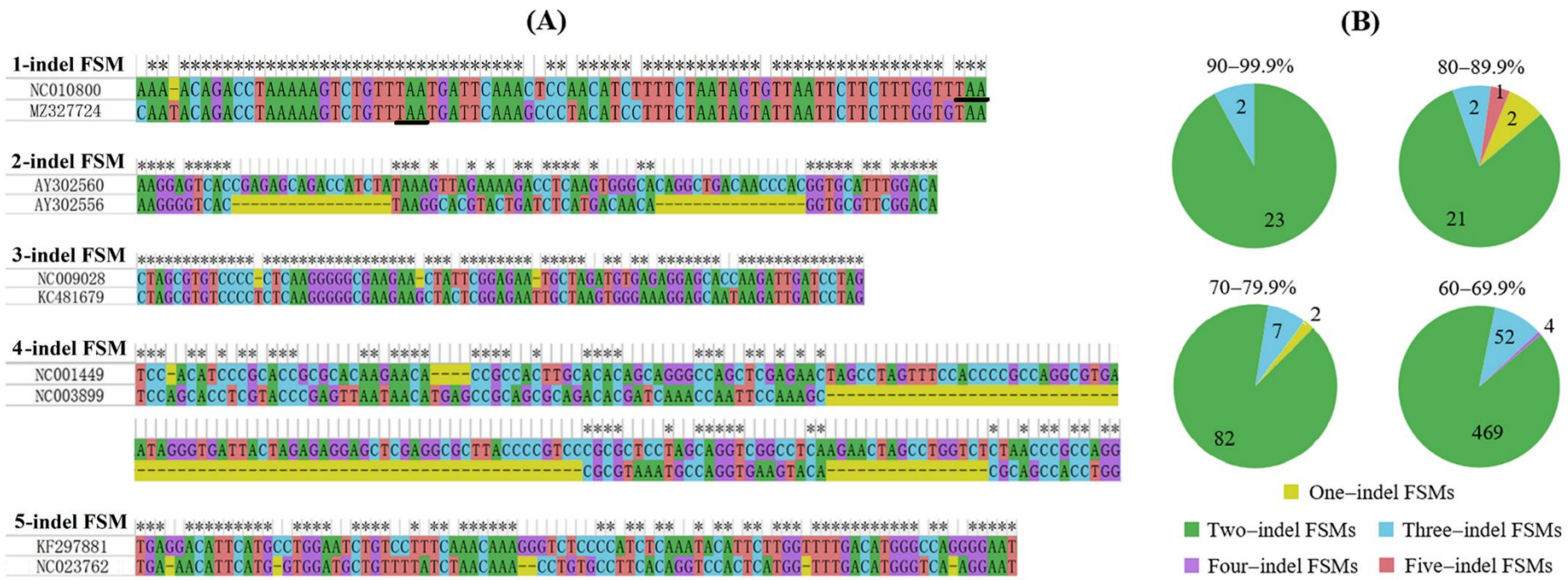

FIG. 3. Types of frameshift mutations (A) and their distribution in the frameshift mutations identified in this study on the four scales (B). The underlined TAA in panel A represented the termination codon.

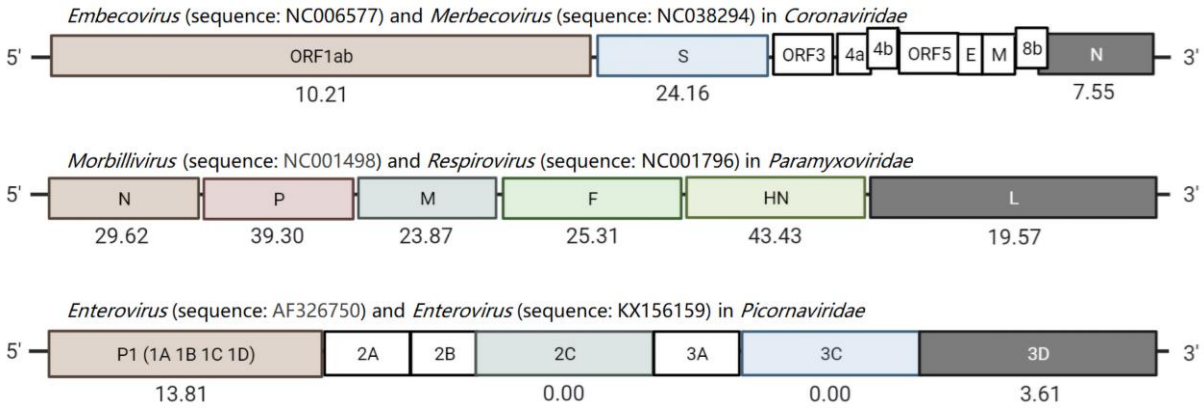

FIG. 4. FSM frequencies of some genes per 5,000 nucleotides calculated through comparing the genomic sequences of six virus species in three families. The frequencies were shown below the relevant gene names.

Table 1. Frequencies of Frameshift Mutations (FSMs) in Viral Sequences with Different Sequence Identities

Virus family	Gene	Sequence length	Sequence used	FSM counts	Frequencies as per sequence identities			
					90.0–99.9%	80.0–89.9%	70.0–79.9%	60.0–69.9%
Arenaviridae	large protein	6616	167	80	0.11 (6/41)	0.27 (5/14)	0.14 (4/22)	1.96 (65/25)
Circoviridae	capsid protein	752	76	11	0.00 (0/20)	0.00 (0/6)	3.32 (4/8)	3.58 (7/13)
Coronaviridae	spike protein	3845	156	64	0.11 (3/35)	0.39 (3/10)	0.49 (3/8)	3.58 (55/20)
Flaviviridae	polyprotein	10120	159	44	0.05 (3/31)	0.05 (1/9)	0.10 (2/10)	0.89 (38/21)
Hantaviridae	large protein	6455	85	31	0.10 (2/16)	0.15 (2/10)	0.24 (4/13)	1.27 (23/14)
Nairoviridae	large protein	11840	77	94	0.00 (0/17)	0.00 (0/10)	0.45 (14/13)	1.88 (80/18)
Orthomyxoviridae	hemagglutinin	1698	205	2	0.00 (0/43)	0.00 (0/15)	0.00 (0/11)	0.33 (2/18)
Paramyxoviridae	large protein	6696	192	48	0.04 (2/41)	0.37 (3/6)	0.20 (4/15)	1.32 (39/22)
Peribunyaviridae	large protein	6761	136	44	0.07 (3/33)	0.06 (1/13)	0.06 (2/25)	1.22 (38/23)
Phenuiviridae	large protein	6266	67	15	0.00 (0/11)	0.00 (0/5)	0.00 (0/10)	1.20 (15/10)
Picornaviridae	polyprotein	6720	207	126	0.07 (3/30)	0.29 (10/26)	0.61 (27/33)	2.00 (86/32)
Poxviridae	DNA polymerase	3014	66	7	0.00 (0/21)	0.00 (0/12)	0.30 (2/11)	0.92 (5/9)
Togaviviridae	NS protein	7425	86	101	0.11 (3/19)	0.17 (1/4)	1.53 (25/11)	2.85 (72/17)

NOTE.—The data in the parentheses after the frequencies showed relevant FSM counts versus relevant group counts.

Table 2. Counts of FSMs in Viral Sequences Aligned Using Two Software Tools

Virus genes	Sequences	Gap penalty using Mega			Using MAFFT	
		–400	–600	–800	E-INS-I	FFT-NS-2
Spike protein of coronaviruses	NC004718 NC045512	3	0	0	0	1
Large protein of nairoviruses	EU257628 NC004159	2	2	2	2	2
Polyprotein of picornnaviruses	AY843307 AY302559	8	1	1	1	1
NS protein of togaviruses	NC001449 NC038670	5	2	2	2	2
Large protein of paramyxoviruses	MH972568 NC006428	24	4	2	2	8
Large protein of arenaviruses	NC010703 NC010702	10	2	1	1	2