Article

Frameshift Mutations in the Microevolution and Macroevolution of Viruses

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

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

***Corresponding author:** E-mail: jmchen@fosu.edu.cn

ABSTRACT

Some nucleotide insertions or deletions (indels) in open reading frames of virus genes lead to frameshift mutations (FSMs) which can change amino acid sequences drastically. Most FSMs are deleterious and inhibited by natural selection, but a few FSMs could aid viruses to adapt to new hosts, escape immunity, or change viral pathogenicity or transmissibility. Surprisingly, various fundamental aspects of FSMs in virus evolution remain unknown. In this report, we identified 667 FSMs in the genomes of viruses from 13 families. Each FSM changed approximately 11.3 $\pm$ 6.6 amino acid residues. Of the 667 FSMs, 89.2% were 2-indel compensatory FSMs, and the remaining FSMs contained 1-, 3-, 4-, or 5-indels. FSMs occurred more frequently in the sequences with identities of 60.0–69.9% than in the sequences of the same viral gene with identities of 90.0–99.9%, 80.0–89.9%, and 70.0–79.9% by approximately 34.9, 13.1, and 3.1 times, respectively. These results suggested that most FSMs were fixed through the compensatory mechanism with 2 to 5 indels, and that FSMs are more important to the macroevolution (i.e., inter-species evolution including speciation) than to the microevolution (i.e., intra-species evolution) of viruses. They provided novel evidence for the hopeful monster hypothesis in evolutionary biology. Furthermore, we found that FSMs occurred at different frequencies among genes in the same virus genome, among species in the same virus family, and among virus families. Collectively, this study revealed multiple fundamental features of FSMs in virus evolution for the first time and provided novel insights into the mechanisms of speciation.

IMPORTANCE

Frameshift mutations (FSMs) can change amino acid sequences drastically, and this concept is well known to biomedical researchers. However, various fundamental aspects of FSMs, including their roles in virus evolution, remain unknown. In this report, we identified hundreds of FSMs in the genomes of viruses from a dozen of families. We found that most FSMs were fixed through the compensatory mechanism with 2 to 5 indels, and that FSMs are more important to the macroevolution (i.e., inter-species evolution) than to the microevolution (i.e., intra-species evolution) of viruses. This study revealed multiple fundamental features of FSMs in virus evolution for the first time and provided novel

evidence for the hopeful monster hypothesis in evolutionary biology pertaining to the mechanisms of speciation.

KEY WORDS

virus, frameshift mutation, evolution, macroevolution, the hopeful monster hypothesis

Introduction

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). Virus evolution involves nucleotide substitution, insertion, and deletion, besides genomic recombination, re-assortment, and rearrangement (1). Insertions or deletions (indels) of 3n±1 nucleotides in open reading frames (ORFs) of protein-coding genes lead to frameshift mutations (FSMs) (2, 3).

As FSMs can change amino acid sequences drastically, most FSMs in virus genomes could be deleterious and inhibited by natural selection. However, some FSMs could aid viruses to adapt to new hosts, escape immunity, or change the viral pathogenicity or transmissibility. For example, FSMs in the genome of vaccinia virus could rapidly attenuate the virus (4). FSMs in the ORF7a gene of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) could reduce the virus pathogenicity (5). An FSM in the nsp6 gene of SARS-CoV-2 could be associated with adaptation of the virus to cats, dogs, and hamsters (6). Two FSMs in the nsp2 gene of SARS-CoV-2 compared with SARS-CoV-1 could be associated with the virus transmissibility (7). 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 (8).

Few studies on FSMs in virus evolution 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 a small scale (i.e., among the sequences with identities of >97%) (9). For instance, two of 5,606 the hemagglutinin gene sequences with identities of >97% of H1 subtype influenza virus carried FSMs, and six of 685 the hemagglutinin neuraminidase (HN) gene sequences with identities of >97% of

paramyxoviruses carried FSMs (9).

Although frameshift mutations (FSMs) are well known to biomedical researchers, various fundamental aspects of FSMs, including their roles in virus evolution, remain unknown. In this report, we investigated some fundamental features of FSMs in virus genomes on four scales, which pertains to the microevolution (i.e., intra-species evolution) and macroevolution (i.e., inter-species evolution including speciation) of viruses.

Results

FSMs in 13 virus families. We downloaded and analyzed gene sequences of some viruses from 11 RNA virus families and two DNA virus families, including the newly established virus families of *Hantaviridae*, *Nairoviridae*, *Peribunyaviridae*, and *Phenuiviridae* (10-22). These sequences constituted 572 fas files, which harbored 358, 140, 190, and 242 groups with intra-group sequence identities of 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 between the groups were identified in the 572 fas files (Table 1). We randomly selected 60 of the 667 FSMs and found each FSM changed 11.3 ± 6.6 amino acid residues.

FSM frequencies and sequence identities. Fig. 2 and Table 1 showed that FSMs usually occurred more frequently in the sequences of the same viral gene with lower identities (P<0.01, by the Mann-Kendall test). For example, FSMs occurred 32.5, 9.2, and 7.3 times more frequently in the sequences of the spike gene of *Coronaviridae* with identities of 60.0–69.9% than in the sequences of the same gene with identities of 90.0–99.9%, 80.0–89.9%, and 70.0–79.9%, respectively. FSMs occurred 28.6, 6.9, and 3.3 times more frequently in the sequences of *Picornaviridae* with identities of 60.0–69.9% than in the sequences with identities of 90.0–99.9%, 80.0–89.9%, and 70.0–79.9%, respectively. Taken the 13 families together, FSMs occurred approximately 34.9, 13.1, and 3.1 times more frequently in the sequences with identities of 60.0–69.9% than in the sequences with identities of 90.0–99.9%, 80.0–89.9%, and 70.0–79.9%, respectively.

Distribution of FSM types. Of the 667 FSMs, four (0.3%) were non-compensatory FSMs formed with only a single indel near to termination codons (Fig. 3). The remaining 663 FSMs were compensatory, and 595 (89.2%) were 2-indel compensatory FSMs, 63 (9.4%) 3-indel compensatory FSMs, four (0.6%) 4-indel compensatory FSMs, and one (0.1%) 5-indel

compensatory FSMs (Fig. 3).

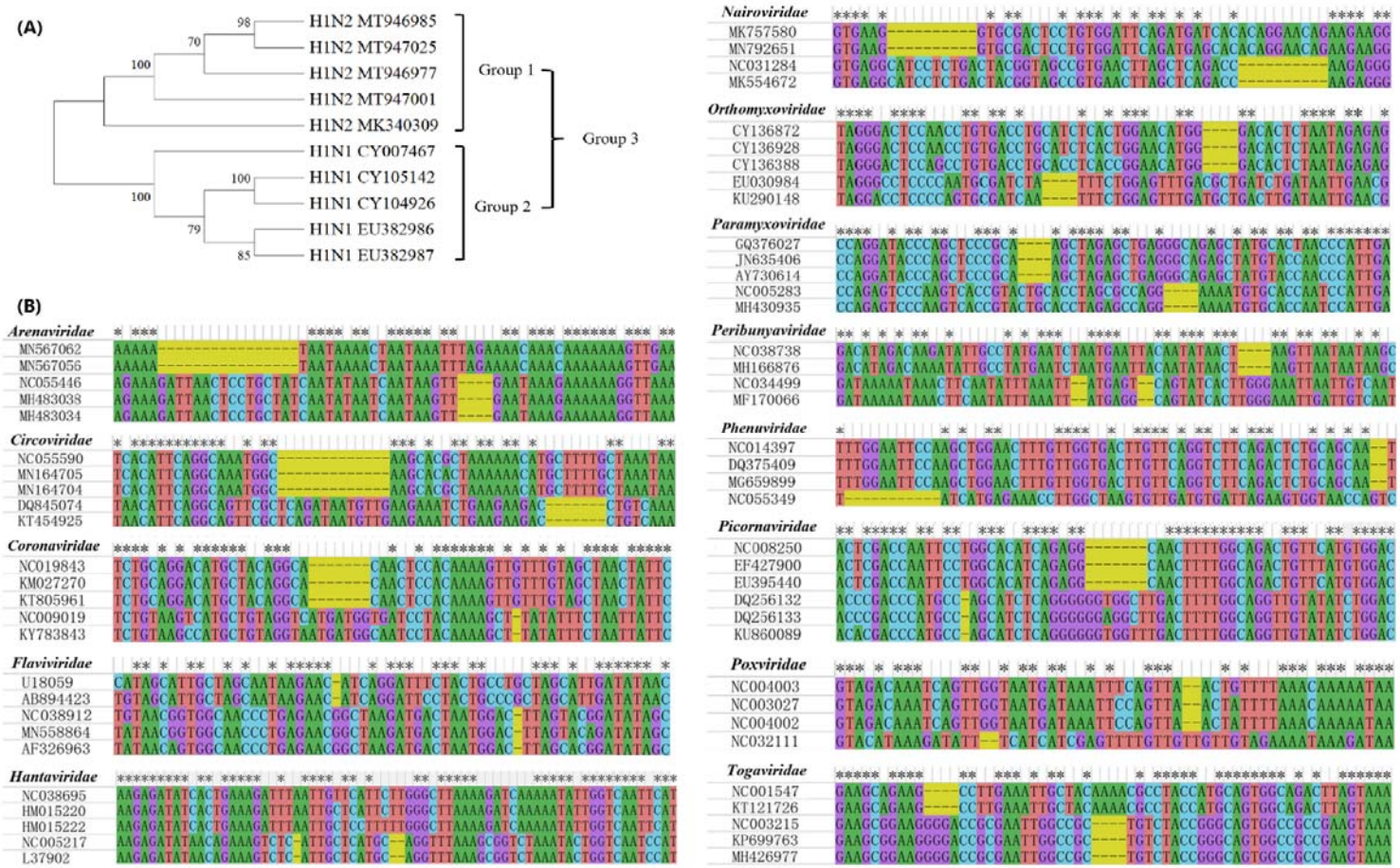

FIG 1 Examples of groups in fas files and frameshift mutations in the genomes of 13 virus families

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 the relevant FSM counts versus the relevant group counts; the frequencies were calculated per group per 5,000 nucleotides.

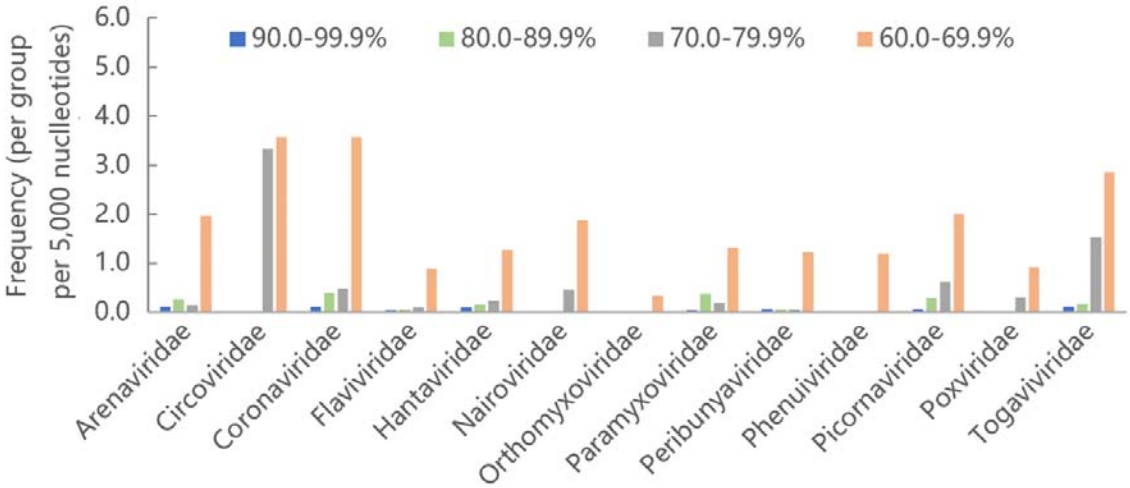

FIG 2 Frequencies of frameshift mutations in 13 virus families among the sequences with four ranges of sequence identities

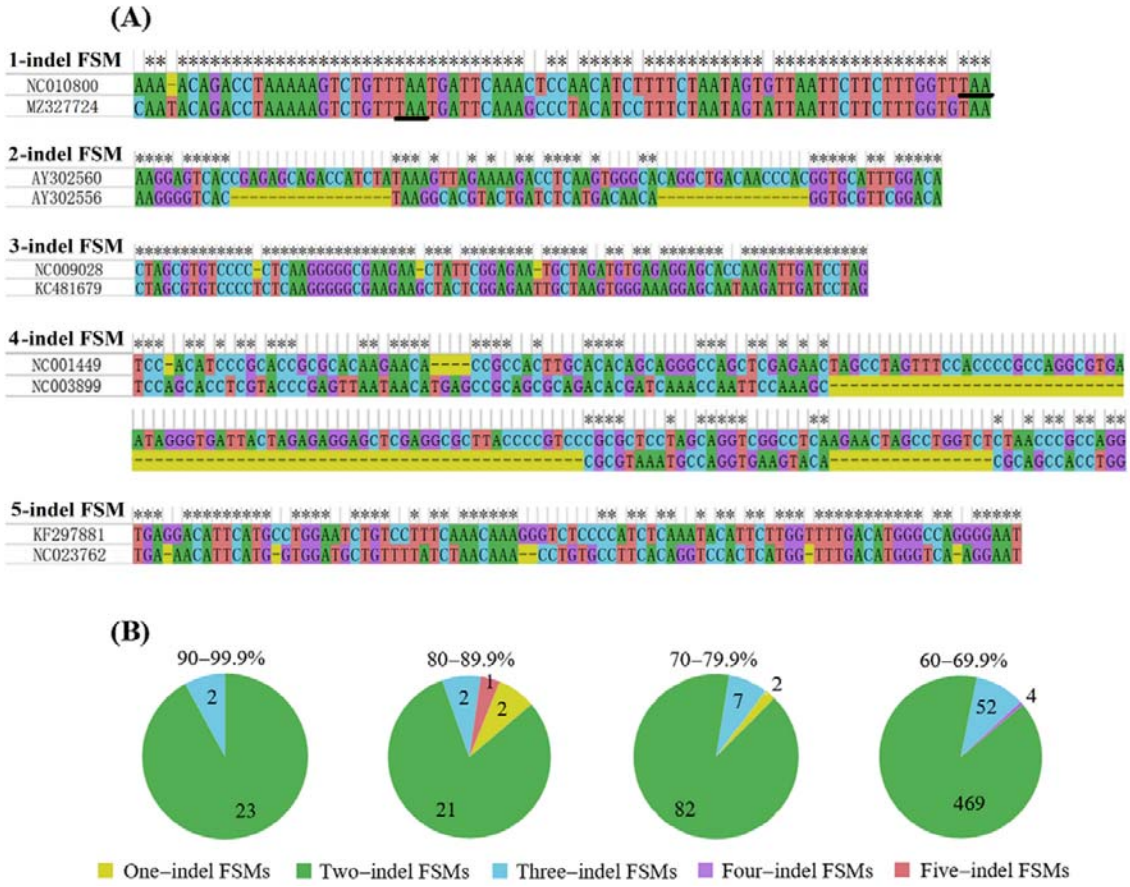

FIG 3 Types of frameshift mutations (A) and their distribution in the frameshift mutations identified in this study on the four scales of sequence identities (B). The underlined letters TAA in panel A represented termination codons.

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 (23). 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 and evolve less slowly than other genes in the viral genomes (21).

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 of 80.0–89.9% occurred in only one group which represented the species of *Avian metaavulavirus 2*. This showed that 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 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 genes by 2.4 and 3.2 times. In *Paramyxoviridae*, FSMs occurred more frequently in HN gene than in L gene by 2.2 times. In *Picornaviridae*, FSMs occurred more frequently in P1 gene than in 3D gene by 3.8 times.

FSM identification and sequence alignment. Table 2 showed that fewer FSMs could be identified if viral sequences were aligned using MEGA with higher gap penalties, likely because higher gap penalties lead to the identification of fewer indels. This table also showed that the numbers of FSMs identified using MAFFT with the mode of E-INS-I were fully consistent with the FSMs identified using MEGA with the gap penalty of –800 and less than the FSMs identified using MEGA with the gap penalty of –400 or –600 and less than the FSMs identified using MAFFT with the mode of FFT-NS-2.

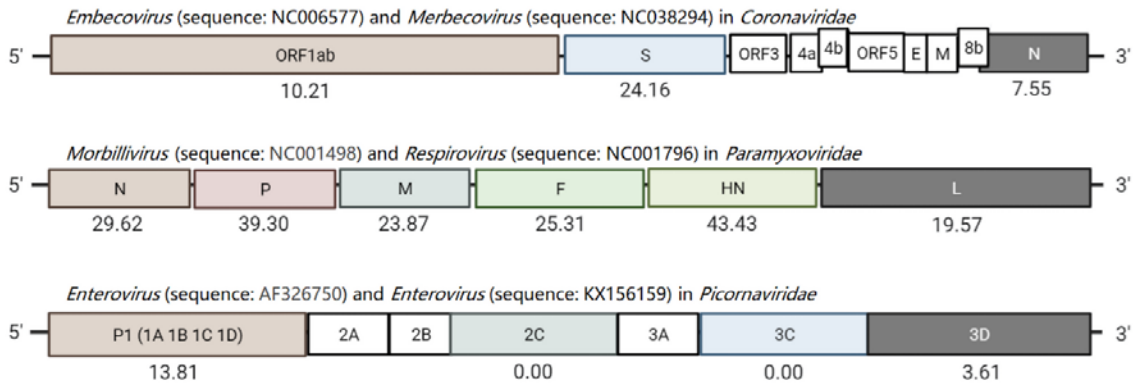

FIG 4 FSM frequencies of some genes per 5,000 nucleotides calculated through comparing the genomic sequences of six virus species in three families. The frequencies were shown below the relevant gene names.

Table 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

Discussion

Few studies have been reported regarding FSMs in the evolution of viruses, including the viruses in the 13 families investigated in this study, as indicated by the relevant reviews (10-22). This could be partially because researchers usually focus on virus evolution on small scales (e.g., intra-species evolution or microevolution of influenza viruses or SARS-CoV-2), and FSMs are relatively rare in virus evolution on small scales, as revealed by this report and a previous report (9). Consequently, we had the opportunity in this study to reveal various fundamental features of FSMs in virus evolution for the first time. We found that most FSMs were fixed through the compensatory mechanism with 2 to 5 indels, and that FSMs are important to the macroevolution (i.e., inter-species evolution including speciation) of viruses, which is interesting in virology and evolutionary biology.

Previous reports classified FSMs in virus evolution into 2-indel compensatory FSMs and non-compensatory FSMs, and one report suggested that their frequencies are similar (9, 24). 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 (Fig. 3B).

This report demonstrated that FSMs usually occurred more frequently in virus genomes in the same family with lower sequence identities which could correspond to inter-species or inter-genus virus evolution. This finding suggested 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 revealed that FSMs occurred at distinct frequencies among virus families or species with similar sequence identities. This suggested that different virus families or species likely favor different approaches or mechanisms to change their amino acid sequences. For example, influenza viruses likely favor nucleotide substitutions rather than FSMs, and paramyxoviruses could favor both nucleotide substitutions and FSMs, to change their amino acid sequences (Table 1).

Fig. 2 showed that FSMs did not always occur more frequently in virus genomes of the same family with lower sequence identities. This could result from sampling errors,

particularly when FSMs occurred at different frequencies among species in the family.

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 (9).

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 suggested 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 are 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.

Because FSMs can change multiple amino acid residues, they can lead to major phenotypic transformations (i.e., macromutations), 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 provided novel evidence to support the hopeful monster hypothesis in evolutionary theory which claims that some major evolutionary transformations between species have occurred through large leaps of macromutations (24), rather than the accumulation of small variations in Darwin's theory.

Nucleotide sequences can be aligned correctly only at a certain possibility. Consequently, the FSMs identified in this report were reliable only at a certain possibility. This issue is more obvious when sequence identities are lower (9). 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.

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 not have been identified. This does not affect the conclusions of this report because these conclusions were based on the same FSM identification method.

FSMs are widely distributed in the genomes of many organisms and can lead to phenotypic diversification (e.g., plumage variations) and pathological changes (e.g., cancers) in organisms (2, 3). However, few studies have been reported regarding the frequencies of FSMs in microevolution or macroevolution of organisms. Many viruses evolve much more rapidly than cellular organisms, and they are hence suitable to be employed to investigate various evolutionary theories and notions, including those regarding FSMs in microevolution or macroevolution. This study suggested 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. This study could inspire researchers to investigate the roles, frequencies, and features of FSMs in the evolution of other viruses and cellular organisms, particularly on grander scales.

Together, this report described multiple fundamental features of FSMs in virus evolution for the first time. It also suggested that most FSMs were fixed through the compensatory mechanism with 2 to 5 indels, and that FSMs are important to the macroevolution and speciation, which is interesting in virology and evolutionary biology.

Materials and Methods

Sequence alignment and sequence identity calculation. Viral nucleotide sequences of complete ORFs of virus genes were downloaded from National Library of Medicine (NCBI). 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 (25). Sequence identities were calculated using the software tool MEGA (version 11) as per pairwise deletion (26).

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 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 (26).

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 indel 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, and $P<0.05$ represented statistical significance, if not specified.

ACKNOWLEDGMENTS

Many viral gene sequences in GenBank constituted the vital basis of this study. The authors thank deeply all the people who produced and shared these virus sequences in GenBank. 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, and this report does not represent any views of the funder.

REFERENCES

1. Weaver SC, Denison M, Roossinck MJ, Vignuzzi M. 2016. Virus evolution : current research and future directions. Caister Academic Press, Wymondham.
2. 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 (Baltim) 101:e29280.
3. 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.

4. Senkevich T, Zhivkoplias E, Weisberg A, Moss B. 2020. Inactivation of Genes by Frameshift Mutations Provides Rapid Adaptation of an Attenuated Vaccinia Virus. J Virol 94:e01053-20.

5. Joonlasak K, Batty EM, Kochakarn T, Panthan B, Kumpornsin K, Jiaranai P, Wangwiwatsin A, Huang A, Kotanan N, Jaru-Ampornpan P, Manasatienkij W, Watthanachockchai T, Rakmanee K, Jones AR, Fernandez S, Sensorn I, Sungkanuparph S, Pasomsub E, Klungthong C, Chookajorn T, Chantratita W. 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.

6. 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.

7. 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 92:584-588.

8. 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 14:1202-1210.

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

10. Breitbart M, Delwart E, Rosario K, Segales J, Varsani A, Ictv Report C. 2017. ICTV Virus Taxonomy Profile: Circoviridae. J Gen Virol 98:1997-1998.

11. 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 99:761-762.

12. Garrison AR, Alkhovsky AC, Avsic-Zupanc T, Bente DA, Bergeron E, Burt F, Di Paola N, Ergunay K, Hewson R, Kuhn JH, Mirazimi A, Papa A, Sall AA, Spengler JR, Palacios G, Consortium IR. 2020. ICTV Virus Taxonomy Profile: Nairoviridae. J Gen Virol 101:798-799.

13. Hughes AL, Irausquin S, Friedman R. 2010. The evolutionary biology of poxviruses. Infect Genet Evol 10:50-9.

14. Hughes HR, Adkins S, Alkhovskiy S, Beer M, Blair C, Calisher CH, Drebot M, Lambert AJ, de Souza WM, Marklewitz M, Nunes MRT, Shi X, Ictv Report C. 2020. ICTV Virus Taxonomy Profile: Peribunyaviridae. J Gen Virol 101:1-2.

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

16. Radoshitzky SR, Buchmeier MJ, Charrel RN, Clegg JCS, Gonzalez JJ, Gunther S, Hepojoki J, Kuhn JH, Lukashevich IS, Romanowski V, Salvato MS, Sironi M, Stenglein MD, de la Torre JC, Ictv Report C. 2019. ICTV Virus Taxonomy Profile: Arenaviridae. J Gen Virol 100:1200-1201.

17. Rima B, Balkema-Buschmann A, Dundon WG, Duprex P, Easton A, Fouchier R, Kurath G, Lamb R, Lee B, Rota P, Wang L, Ictv Report C. 2019. ICTV Virus Taxonomy Profile: Paramyxoviridae. J Gen Virol 100:1593-1594.

18. Simmonds P, Becher P, Bukh J, Gould EA, Meyers G, Monath T, Muerhoff S, Pletnev A, Rico-Hesse R, Smith DB, Stapleton JT, Ictv Report C. 2017. ICTV Virus Taxonomy Profile: Flaviviridae. J Gen Virol 98:2-3.

19. 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 94:2388-2401.

20. Taubenberger JK, Kash JC. 2010. Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe 7:440-51.

21. Zell R, Delwart E, Gorbalenya AE, Hovi T, King AMQ, Knowles NJ, Lindberg AM, Pallansch MA, Palmenberg AC, Reuter G, Simmonds P, Skern T, Stanway G, Yamashita T, Ictv Report C. 2017. ICTV Virus Taxonomy Profile: Picornaviridae. J Gen Virol 98:2421-2422.

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

23. 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:e19996.

24. 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 30:1155-1173.

25. 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 15:3059-66.

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