Influenza A virus is enveloped with a lipid membrane (enveloped virus). Its genome consists of eight single-stranded RNA segments ranging from 0.9·10^3 to 3.2·10^3 nucleotides. The segments have the negative polarity, i.e., they cannot be directly translated by the ribosomes. Instead, they are transcribed by viral polymerase resulting in mRNAs further translated into viral proteins (Fig. 1, a and b). Transcription of some viral RNA segments is associated with mRNA splicing, while translation can start at different AUG codons and might be accompanied by the frameshift. As a result, 16 unique viral proteins are expressed via the negative-sense mechanism [1, 2]. NS is the smallest influenza virus genome segment that encodes two viral proteins: non-structural interferon-antagonist protein-1 (NS1) that counteracts the activity of the interferon system, and nuclear export protein (NEP) that regulates nuclear export of viral ribonucleoproteins in the infected cells. mRNA for the NEP protein undergoes splicing that results in the frameshift relative to the NS1 gene (Fig. 1a).

In most human influenza A virus strains, the NS segment has an extended positive-strand open reading frame (ORF) (Fig. 1a) [3-9] that was identified as a gene encoding the NSP8 protein (negative strand protein; segment 8) [4, 5]. The presence of extended reading frame containing AUG initiation and UGA termination codons and structured 5′-terminal sequence that includes IRES-like motifs [10] in the viral genomic RNA suggests that the NSP8 gene can emerge in the influenza A virus population about 100 years ago (“young” gene) and is highly evolutionary variable. The obtained experimental data suggest that NSP8 gene is expressed in the infected animals, which strengthens the concept of bipolar (ambisense) strategy of the influenza A virus genome. The high variability of the NSP8 protein suggests that the “young” NSP8 gene is in the process of functional optimization. Further accumulation of mutations may alter the functions of mature NSP8 protein and lead to the emergence of mature bipolar influenza A virus with unexpected properties that would be threatening for humans and animals.

DOI: 10.1134/S0006297920030141

Keywords: influenza virus, genome strategy, NSP gene, ambisense viruses, gene structure
influenza A viruses [5]) and stop codons in the NSP8 gene in the pandemic human influenza A virus H1N1pdm09 (see below). Moreover, the genes in the segment NS colocalize in a way that the third nucleotide in the codons of the NSP8 frame is the second and third nucleotides in the codons of the NS1 and NEP reading frames, respectively (Fig. 1a), which allows synonymous substitutions at position 3 in the codons of the NSP8 gene without affecting the NS1 gene. The ORF for the human influenza A virus NSP8 gene has emerged around 100 years ago and might be considered as a relatively new evolutionary trait. This observation was made based on phylogenetic comparison of RNA sequences of the NS segments from available human and animal influenza A viruses by using the multimetric algorithm (see Supplement to Zhirnov et al. [5]).

Still, the function of NSP8 remains unknown. Preliminary studies have demonstrated that baculoviral expression of the NSP8 gene in insect cells produced the full-size gene product that localized to the perinuclear zone [11]. In silico analysis revealed no complete homology between NSP8 and other proteins available in the databases and demonstrated a partial similarity of its amino acid sequence to the Q-motif in the DEAD-box RNA helicase from the parasitic plant fungus Ceraceosorus bombacis [12], a fragment of the protease domain in class C19 hydrolases (including ubiquitin hydrolase from the white button mushroom Agaricus bisporus var. Burnettii [13]).

Fig. 1. Gene architecture of the (a) segment 8 (NS) and (b) segment 7 (M) (GenBank accession numbers CY121118.1 and CY121121.1, respectively) of the influenza A/Aichi/2/68 (H3N2) genome. Numbers correspond to nucleotide positions, starting from the 5′-end of virion RNA. vRNA, negative-sense viral RNA; cRNA, positive-sense complementary RNA (replicative form); arrows, gene exons for NS1, NEP, and NSP8 proteins; broken line, splicing of the NEP gene mRNA.
and the transmembrane domain of the voltage-gated calcium ion channel from the unicellular eukaryote Salpingoeca rosetta [14] (Table 1 and Fig. 2). It cannot be ruled out that NSP8 has multiple functions. Its structural resemblance to other proteins indicates that NSP8 either plays the same role as these proteins or interferes with their functions or inhibits their activity in vivo. Further studies are necessary to uncover the functions of this new class of negative-strand gene products.

The discovery of a new gene in the influenza A viruses has raised a number of important questions regarding its origin, functions, and evolutionary variability. One of the essential questions is how the new positive-sense NSP8 gene has emerged in the genomic region encoding two negative-sense genes. The appearance of the ambipolar gene suggests the existence of yet unknown correspondence principle (or reverse determination rule) for the expression of ambipolar genes residing in the same region of RNA molecule. It is possible that such principle implies that a certain pre-existing gene can predetermine the emergence mechanism and the properties of a new ambipolar gene. It should be presumed that in the absence of determination mechanism, chaotic accumulation of mutations will result in the appearance of a new functional gene and its further evolutionary selection. However, the probability for such event is low, considering the ambipolar overlapping of several preexisting genes, when changes in one of them would cause changes in the coupled ambipolar genes. In this case, gene variability and selection of mutations should be interconnected in all three viral genes (NS1, NEP, and NSP8).

It is possible that other RNA segments of the influenza A virus genome also contain ambisense genes, as it might be suggested from the presence of extended ORFs in segments PB1, PB2, PA, NP, and M (Fig. 1b). The protein products of these ORFs have not been identified yet in biological systems; however, the in silico analysis points out that the proteins of the positive-sense genes have the properties of transmembrane ionic channels [4, 7]. The occurrence of ambisense genes in multiple RNA segments implies that a universal mechanism might exist in human and animal influenza A viruses that allows to implement both positive- and negative-sense genome strategies in the target cells.

Another important issue is related to the high variability of the NSP proteins in human influenza A viruses. Our data suggest that the variability of NSP is similar to that of the most variable glycoproteins, such as hemagglutinin (HA) and neuraminidase (NA), located at the virion surface and representing major target molecules for the antiviral factors of the host adaptive immunity (Table 2). The negative-sense NS1 and NEP genes in the NS segment are less variable [17]. There are two possible reasons that may account for this phenomenon. First, the NSP8 protein could be exposed on the surface of either virions or infected cells (similar to HA and NA) to be targeted by the factors of the host innate and adaptive immunity, which promotes its variability. Second, the NSP protein, as a newly emerged viral product, may be at the stage of optimization of its expression level and functions towards its viral or cell partners, which might explain its adaptation to these partners.

Table 1. Similarity between the sequences of the influenza A virus NSP8 protein and non-viral proteins

| NSP8 amino acid sequence | NSP8 region used for comparison* | Homology region in the compared protein** | Similarity*** | Function of domain in the compared protein*** | References |
|--------------------------|---------------------------------|-----------------------------------------|--------------|-----------------------------------------------|------------|
| NS segment of influenza virus A/WSN/33 (H1N1) Ac. no. U13683.1 | 1-108 (121) | ubiquitin hydrolase (UH) 479-600 (1121) | E: 6.1e-1, id.: 31%, pos.: 46% | protection of proteins from proteasomal degradation via cleaving off ubiquitin | [14] |
| NS segment of influenza virus/Moscow/328/2003 (H3N2) Ac. no. AAZ38747.1 | 8-195 (216) | α₁ subunit of voltage-gated calcium channel G1 158-321 (2453) | E: 4.6e-1, id.: 26%, pos.: 41% | transmembrane Ca²⁺ ion transport, regulation of apoptosis, autophagy, and carcinogenesis | [15, 16] |
| NS segment of influenza virus/Moscow/328/2003 (H3N2) Ac. no. AAZ38747.1 | 43-120 (216) | Q-motif in DEAD-box RNA helicase 306-383 (904) | E: 2.4e-0, id.: 33%, pos.: 47% | ATP binding and hydrolysis in active RNA helicase | [13] |

* Positions of amino acid residues in the homology region in NSP8. The total number of amino acid residues in NSP8 is shown in parentheses.

** Positions of amino acid residues in the homology region in the compared protein. The total number of amino acid residues in the compared protein is shown in parentheses. The number of amino acid residues in the protein is shown in parentheses.

*** The similarity was calculated based on the paired comparison between protein amino acid sequences with the UniProt software (Fig. 2). The expectation value (E) threshold is a statistical measure of the number of expected matches in a random database; id. and pos., percentage of identical and positive amino acid residues, respectively. Compared proteins: α₁/G1 transmembrane domain of the type 1 voltage-gated calcium ion channel from the unicellular eukaryote Salpingoeca rosetta (ID: F2U5C1); fragment of the protease domain of type C19 ubiquitin hydrolase (UH) from the white button mushroom Agaricus bisporus var. Burnettii (ID: K5Y1E3); Q-motif of the DEAD-box RNA helicase from the smut fungus Ceraceosorus bombacis (ID: A0A0P1BHD7).
amino acid sequences of the NS segments from the subtype H3N2 influenza A viruses demonstrated a large number of nonsynonymous mutations in the NSP8 gene (dN/dS > 1.5) compared to the NS1 and NEP genes (dN/dS ≤ 1) (Table 3), which is in a good agreement with the concept on the elevated variability of the NSP8 gene and the impact of factor(s) of positive selection (adaptation) involved in its evolution [18]. Accumulation of mutations in the course of evolution may result in qualitative leap in the virus functional activity, leading to the appearance of a novel subtype of influenza virus with unexpected and unpredictable properties that could pose a threat to both humans and animals. Studying the new gene and deciphering its functions linked to the newly emerged NSP8 protein will help to neutralize such threat either by specific preventive vaccination or development of pharmaceutical inhibitors against this protein or its gene.

The evolutionary dynamics of the NS segment from the human influenza virus H1N1pdm09 that caused the 2009 influenza pandemics in Mexico correlates with the predicted virus adaptive variability [19]. This strain was a triple reassortant of the human, swine, and avian viruses and has spread to humans from a natural reservoir. It carried the avian NS segment that lacked the full-size NSP gene and contained two stop codons in the NSP gene body. However, after a 10-year evolution in the human population, it lost one of the stop codons, resulting in the sequence extension to 240 bp. It is reasonable to expect that if this virus continues to circulate in the human population in the next decade, it will lose the remaining stop codon, so that the full-size NSP will appear that may acquire unexpected and life-threatening properties upon its further optimization.

Evolutionary optimization of the NSP8 gene may also affect its regulatory motifs. In particular, during more than a fifty year-long evolution, human influenza A virus subtype H3N2 has accumulated a significant number of AUG codons in its 5′ start region: from one identified in 1968 to four observed at present [5]. An increased number of AUG codons in the translation initiation region can facilitate recognition of viral RNA by translation initiation factors and ribosomes, resulting in the upregulation of its translation in the infected cells. Moreover, an increase in the number of start codons in the viral RNA template may contribute to an increased rate of defective translation of the NSP8 gene, including translation initiation at the alternative AUG codons that
results in the emergence of defective ribosomal products (DRiPs), which may be involved in the downregulation of the MHC-I-dependent antiviral immune response [20]. However, taking into consideration that expression in the baculoviral system [11] and translation in vitro [10] resulted in the synthesis of the full-size stable NSP8 protein, the defective translation of the NSP8 gene seems to be unlikely as a dominant mechanism of its expression.

The uniqueness of gene architecture of the bipolar genes deserves special attention. The existence of three or more ambisense genes in the same region of a single-stranded RNA molecule is a unique type of gene architecture that ensures the “saving” of genetic information at the RNA level and its efficient use in the synthesis of different protein products and provides the highest protein diversity per unit of genetic information in the RNA genome. The unique features of the bipolar gene architecture in the NS segment of the influenza A virus differs from the organization of ambisense gene in RNA genomes of viruses infecting animals (phleboviruses, tospoviruses, and arenaviruses) and plants (tenuiviruses) [21]. Compared to the aforementioned viruses, in which two ambipolar genes locate in different regions of RNA molecule and do not overlap, human influenza A viruses contain three overlapping ambipolar genes (the so-called stacking) within the same RNA region (Fig. 1, a and b). When bipolar genes do not overlap, their architecture is not directly coupled with their variability, and they can change independently of each other.

Although the protein product of the NSP gene has not been yet identified in biological systems, e.g., virus-infected cell lines or animal models, the data obtained confirm the possibility of expression of positive-sense viral RNA genes. In particular, the NSP8 protein can be synthesized in the in vitro translation system by mammalian ribosomes on the full-size viral RNA template [10]. Specific immune lymphocytes against this protein or its components have been found in the infected animals [17, 22, 23]. The properties of NSP8 as a “invisible protein” may be accounted by its low synthesis and/or accumulation levels, high lability, short half-life, and possible tissue-specific expression of the NSP8 gene in certain cell types containing factors necessary for the regulation of the positive-strand NSP gene expression. Further studies might help in elucidating these issues.

Finally, the discovery of the NSP8 might call for the revision of Orthomyxoviridae classification and creation of a separate genus of ambipolar influenza viruses (Ambisense Alfainfluenzavirus). However, this issue will remain open until the protein product of the NSP8 gene is unambiguously identified in biological settings. Meanwhile, the fact that the full-size NSP8 gene has been retained for more than a hundred years of evolution in the population of human influenza viruses points at its functional importance in the yet putative group of ambipolar influenza viruses. It should be noted that no stop codons altering the NSP8 gene have been identified despite pronounced evolutionary variability of the NSP8 protein, further reinforcing the idea on the biological determination of this gene.

| Table 2. Evolutionary variability of proteins of the 1968–2018 influenza A virus/H3N2 |
|---|---|---|---|---|
| Viral protein | Molecular weight (Da) | Total number of analyzed sequences | Total number of amino acid substitutions identified | Variability ratio (VR)* |
| PB1 | 86,000 | 13,215 | 23 | 0.05 |
| HA | 75,000 | 19,639 | 139 | 0.40 |
| NA | 56,000 | 17,147 | 114 | 0.40 |
| M1 | 27,000 | 16,145 | 12 | 0.09 |
| NS1 | 27,000 | 13,847 | 34 | 0.25 |
| NEP | 12,000 | 13,739 | 5 | 0.08 |
| NSP8 | 25,000 | 14,331 | 38 | 0.35 |

* Variability ratio (VR) was calculated as the number of amino acid substitutions in the proteins from the 1968–2018 human influenza virus A/H3N2 strains per year per 100 amino acid residues in each protein [17] (GenBank data). PB1, polymerase (basic) protein 1; HA, hemagglutinin; NA, neuraminidase; M1, matrix protein 1; NS1, interferon-antagonist protein-1; NEP, nuclear export protein; NSP8, negative strand protein-8.

| Table 3. A ratio of nonsynonymous vs. synonymous substitutions in the NSP8, NS1, NEP genes in the 1968–2019 influenza A virus subtype H3N2 |
|---|---|
| Viral gene | dN/dS* |
| NSP8 | 1.73 |
| NS1 | 1.25 |
| NEP | 0.23 |

* A ratio of nonsynonymous vs. synonymous substitutions (dN/dS) was calculated from the mRNAs sequences for the NSP8, NS1, and NEP proteins from the 1968–2019 human influenza A viruses/H3N2 (GenBank).
Acknowledgements. The author is thankful to Academician G. P. Georgiev (Russian Academy of Sciences, Moscow) for supporting the study and fruitful discussions and Dr. D. Anhlan (Institute of Virology, University of Munster, Germany) for assistance in analyzing NSP gene variability.

Conflict of interest. The author declares no conflict of interest.

Compliance with ethical standards. This article does not contain description of studies with human participants or animals performed by the author.

REFERENCES

1. Dou, D., Revol, R., Ostbye, H., Wang, H., and Daniels, R. (2018) Influenza A virus cell entry, replication, virion assembly and movement, Front. Immunol., 9, 1581, doi: 10.3389/fimmu.2018.01581.

2. Vasin, A. V., Temkina, O. A., Egorov, V. V., Klotchenko, S. A., Plotnikova, M. A., and Kiselev, O. I. (2014) Molecular mechanisms enhancing the proteome of influenza A viruses: an overview of recently discovered proteins, Virus Res., 185, 53-63, doi: 10.1016/j.virusres.2014.03.015.

3. Baez, M., Zazra, J. J., Elliott, R. M., Young, J. F., and Palese, P. (1981) Nucleotide sequence of the influenza A/duck/Alberta/60/76 virus RNA segment, Virology, 113, 397-402.

4. Zhirnov, O. P., Poyarkov, S. V., Vorob’eva, I. V., Safonova, O. A., Malyshev, N. A., and Klenk, H. D. (2007) Segment NS of influenza A virus contains an additional gene NSP in positive-sense orientation, Dokl. Biochem. Biophys., 414, 127-133, doi: 10.1134/s1607672907030106.

5. Zhirnov, O. P., Vorobjeva, I. V., Saphonova, O. A., Poyarkov, S. V., Ovcharenko, A. V., Anhlan, D., Malyshev, A. N., and Klenk, H. D. (2009) Structural and evolutionary characteristics of HA4, NA4, NS4 and M genes of clinical influenza A/H3N2 viruses in human and canine cells, J. Clin. Virol., 45, 322-333, doi: 10.1016/j.jcv.2009.05.030.

6. Clifford, M., Twigg, J., and Upton, C. (2009) Evidence for a novel gene associated with human influenza A viruses, Virol. J., 6, 198, doi: 10.1186/1743-422X-6-198.

7. Gong, Y. N., Chen, G. W., Chen, C. J., Kuo, R. L., and Shih, S. R. (2014) Computational analysis and mapping of novel open reading frames in influenza A viruses, PLoS One, 9, e115016, doi: 10.1371/journal.pone.0115016.

8. Yang, C. W., and Chen, M. F. (2016) Uncovering the potential pan proteomes encoded by genomic strand RNAs of influenza A viruses, PLoS One, 11, e0146936, doi: 10.1371/journal.pone.0146936.

9. Sabath, N., Morris, J. S., and Graur, D. (2011) Is there a twelfth protein-coding gene in the genome of influenza A?

A selection-based approach to the detection of overlapping genes in closely related sequences, J. Mol. Evol., 73, 305-315, doi: 10.1007/s00239-011-9477-9.

10. Zhirnov, O. P., Akulich, K. A., Lipatova, A. V., and Usachev, E. V. (2017) Negative-sense virion RNA of segment 8 (NS) of influenza A virus is able to translate in vitro a new viral protein, Dokl. Biochem. Biophys., 473, 122-127, doi: 10.1134/S1607672917020090.

11. Zhirnov, O. P., and Klenk, H. D. (2010) Integration of influenza A virus gene NSP into baculovirus and its expression in insect cells, Vopr. Virusol., 55, 4-8.

12. Cunningham, J. L., Bakshi, B. K., Lentz, P. L., and Gilliam, M. S. (1976) Two new genera of leaf-parasitic fungi (Basidiomycetidae: Brachybasiidae), Mycologia, 68, 640-654.

13. Tanner, N. K., Cordin, O., Banroques, J., Doere, M., and Linder, P. (2001) The Q motif: a newly identified motif in DEAD box helicases may regulate ATP binding and hydrolysis, Mol. Cell., 11, 127-138.

14. Quesada, V., Diaz-Perales, A., Gutierrez-Fernandez, A., Garabay, C., Cal, S., and Lopez-Otin, C. (2004) Cloning and enzymatic analysis of 22 novel human ubiquitin-specific proteases, Biochem. Biophys. Res. Commun., 314, 54-62, doi: 10.1016/j.bbrc.2003.12.050.

15. Vervliet, T., Clerix, E., Seitaj, B., Ivanova, H., Monaco, G., and Bulync, G. (2017) Modulation of Ca2+ signaling by anti-apoptotic B-cell lymphoma 2 proteins at the endoplasmic reticulum-mitochondrial interface, Front. Oncol., 7, 75, doi: 10.3389/fonc.2017.00075.

16. Lauren, A., and Martin-Caraballo, M. (2019) T-type calcium channels in cancer, Cancers, 11, 134, doi: 10.3390/cancers11020134.

17. Zhirnov, O. P., Konakova, T. E., Anhlan, D., Ludwig, S., and Isaeva, T. I. (2019) Cellular immune response in infected mice to NSP protein encoded by the negative strand NS RNA of influenza A virus, Microbiol. Ind. Res. J., 6, 10-17, doi: 10.18527/2500-2236-2019-6-1-28-36.

18. Yang, Z., and Bielawski, J. P. (2000) Statistical methods for detecting molecular adaptation, Trends Ecol. Evol., 15, 496-503, doi: 10.1016/s1098-9252(01)00094-7.

19. York, I., and Donis, R. O. (2013) The 2009 pandemic influenza virus: where did it come from, where is it now, and where is it going? Curr. Top. Microbiol. Immunol., 370, 241-257, doi: 10.1007/82_2012_221.

20. Wei, J., and Yewdell, J. W. (2018) Flu DRiPs in MHC class I immunosurveillance, Virol. Sin., 34, 162-167, doi: 10.1007/s12250-018-0061-y.

21. Nguyen, M., and Haenni, A. L. (2003) Expression strategies of ambisense viruses, Virus Res., 93, 141-150, doi: 10.1016/s0168-1702(03)00094-7.

22. Zhong, W., Reche, P. A., Lai, C. C., Reinhold, B., and Reinherz, E. L. (2003) Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire, J. Biol. Chem., 278, 45135-45144, doi: 10.1074/jbc.M307417200.

23. Hickman, H. D., Mays, J. W., Gibbs, J., Kosik, I., Magadan, J. G., Takeda, K., and Yewdell, J. W. (2018) Influenza A virus negative strand RNA is translated for CD8+ T cell immunosurveillance, J. Immunol., 201, 1222-1228, doi: 10.4049/jimmunol.1800586.