

インフルエンザA型ウイルス遺伝子の塩基置換パターン

誌名	The Japanese journal of genetics
ISSN	0021504X
著者	斎藤, 成也
巻/号	62巻5号
掲載ページ	p. 439-443
発行年月	1987年10月

農林水産省 農林水産技術会議事務局筑波産学連携支援センター
Tsukuba Business-Academia Cooperation Support Center, Agriculture, Forestry and Fisheries Research Council
Secretariat



Patterns of nucleotide substitutions in influenza A virus genes

BY Naruya SAITOU¹⁾

*Center for Demographic and Population Genetics,
The University of Texas Health Science Center at Houston,
P.O. Box 20334, Houston, Texas 77225, USA*

(Received October 9, 1987)

ABSTRACT

The pattern of nucleotide substitution among the four nucleotides (A, T, C, and G) was studied for seven different genes of the influenza A virus. The direction of substitution was inferred by examining phylogenetic trees reconstructed (Saitou and Nei 1986), and the relative frequency of nucleotide substitutions was computed by Gojobori *et al.*'s (1982) method. It is found that the transitional changes (between A and G and between T and C) are much more abundant (about 70%) than expected under random substitution. Among the four types of transitional changes, the G→A change is most frequent. There are negative correlations between the chemical distance of nucleotides and the substitution pattern at the first and second codon positions for all genes examined. This result suggests that the influenza virus genes are subject to the purifying selection even though they show an extraordinary high substitution rate.

The influenza virus is a single-stranded RNA virus and can be divided into three different types (A, B, and C) according to immunological differences (Webster *et al.* 1982). The type A virus is the major cause of influenza epidemic not only in humans but also in other mammals and birds. It is known that this virus has a mutation rate that is approximately one million times that of DNA organisms (Air 1981; Nei 1983; Hayashida *et al.* 1985; Saitou and Nei 1986). In recent years, the nucleotide sequences of many influenza virus genes have been determined for various strains, including those that have been kept in refrigerators for many years. These data provide a unique opportunity for studying the evolutionary history of virus genes. Saitou and Nei (1986) reconstructed phylogenetic trees of various influenza A viral genes by using the maximum parsimony method (Fitch 1977). This analysis made it possible to infer the direction of mutations (substitutions) in the influenza virus genes. In this paper, we use these data to study the pattern of nucleotide substitution among the four nucleotides. The nucleotide sequences used in this study are given in Saitou and Nei (1986).

By examining phylogenetic trees, the ancestral sequences at each branching point of trees were inferred. Ancestral sites which could not be deter-

1) Present address: Department of Anthropology, Faculty of Science, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo 113, Japan.

Table 1. *Relative substitution frequencies in influenza A virus genes*

(A) The H3 subtype of the hemagglutinin 1 gene									
1. First and second codon positions					2. Third codon position				
A	T	C	G	[61.5]	A	T	C	G	[85.5]
A	—	1.8	0.0	19.3	A	—	0.0	2.2	12.9
T	12.1	—	3.2	2.8	T	1.9	—	27.8	0.0
C	11.4	10.9	—	0.0	C	4.9	14.6	—	0.0
G	28.1	7.5	3.1	—	G	30.2	2.8	2.7	—

(B) The N2 subtype of the neuraminidase gene									
1. First and second codon positions					2. Third codon position				
A	T	C	G	[64.7]	A	T	C	G	[84.6]
A	—	5.1	3.9	18.0	A	—	1.4	2.8	12.8
T	1.3	—	8.3	0.0	T	1.2	—	18.1	0.0
C	12.3	16.2	—	1.8	C	8.2	24.0	—	0.0
G	22.2	3.0	8.0	—	G	29.7	0.0	1.8	—

(C) Non-structural protein 1 gene									
1. First and second codon positions					2. Third codon position				
A	T	C	G	[77.0]	A	T	C	G	[76.7]
A	—	2.2	4.8	25.9	A	—	4.5	2.1	16.8
T	0.0	—	6.8	0.0	T	2.2	—	13.3	4.6
C	9.6	19.2	—	6.4	C	9.9	12.5	—	0.0
G	25.2	0.0	0.0	—	G	34.1	0.0	0.0	—

Averaged matrices											
1. First and second codon positions ^a					2. Third codon position ^b						
A	T	C	G	[65.1]	A	T	C	G	[77.7]		
A	—	4.1	2.9	20.2	27.2	A	—	3.5	1.4	16.1	21.0
T	4.7	—	7.1	2.0	13.8	T	3.2	—	21.1	3.0	27.3
C	11.2	14.5	—	2.1	27.8	C	5.8	16.1	—	1.4	23.3
G	23.3	3.0	4.9	—	31.2	G	25.4	1.7	1.3	—	28.4
	39.2	21.6	14.9	24.3	100.0		34.4	21.3	23.8	20.5	100.0

Notes. An element f_{ij} of a matrix is the relative frequency in % of change from nucleotide $i \rightarrow j$. At the upper right corner of each matrix, the proportion of transitional changes is given in brackets.

^a (A), (B), (C) and the N1 subtype of the neuraminidase gene are averaged.

^b (A), (B), (C) the H2 subtype of the hemagglutinin 1 gene, the H3 subtype of the hemagglutinin 2 gene, the N1 subtype of the neuraminidase gene, and the matrix protein 1 gene are averaged.

mined unambiguously by the maximum parsimony method were excluded from the analysis. Nucleotide substitutions were counted for each branch of a tree. The relative frequency of nucleotide substitutions was computed by using Gojobori *et al.*'s (1982) method. The nucleotide substitution matrices obtained from different branches were averaged by weighting each matrix by the number of inferred substitutions for the branch. Further averaging of matrices for different genes was also done by the same weighting method. For the comparison, the cDNA strand was used instead of the viral strand.

Table 1 shows the relative frequencies of the twelve different nucleotide substitutions for three influenza A virus genes, i.e. the H3 subtype of the hemagglutinin 1 gene, the N2 subtype of the neuraminidase gene, and the nonstructural protein 1 gene. It is clear that the transitional changes predominate at all codon positions, particularly at the third position. This tendency has also been observed for pseudogenes of DNA organisms (Gojobori *et al.* 1982; Li *et al.* 1984). Among the four different types of transitional changes, the G→A change is most frequent at all positions. In the third position, the changes between T and C are as frequent as the changes between A and G. Among the transversional changes, the C→A change is always more frequent than the A→C change, and it is often most frequent among all the transversional changes. There are no C→G changes at the third position. This type of change is also low in the other positions.

Because the substitution frequencies for various genes are similar to each other, I computed the weighted average of substitution frequencies (see Table 1). For the first and second codon positions, the data for the N1 subtype of the neuraminidase gene are included, whereas those for the H2 subtype of the hemagglutinin 1 gene, the H3 subtype of the hemagglutinin 2 gene, and the matrix protein 1 gene are included for the third codon position. (Because the total number of nucleotide substitutions observed is small, the substitution matrices are not given separately for these genes.) The numbers of mutations identified for the first and second positions and for the third position were 167 and 268, respectively. It is clear that the T→C change is much higher in the third position than the first and second positions, whereas the C→A change is about two times more frequent in the first and second positions than in the third position.

Gojobori *et al.* (1982) computed a chemical distance between each pair of the four nucleotides using Grantham's (1974) amino acid distances, and showed that there is a negative correlation between the chemical distance (d) and the substitution frequency (f) for the first and second codon positions in functional genes but not in pseudogenes. Table 2 shows the correlation coefficients between d and f for the influenza virus genes. These correlations are similar to those of Gojobori *et al.*'s study for functional genes. Correlation coefficients for the third position is -0.315 and for the other positions is -0.559 for the

Table 2. *Correlations between the frequency of nucleotide substitution and the chemical distance of nucleotides for the first and second positions and for the third position of codons^a*

Gene (subtype)	First and second		Third	
	N	r	N	r
Hemagglutinin 2 (H3)	5	-0.64	19	-0.04
Hemagglutinin 1 (H2)	10	-0.01	53	-0.17
Hemagglutinin 1 (H3)	40	-0.38	43	-0.35
Neuraminidase (N1)	26	-0.37	21	+0.15
Neuraminidase (N2)	65	-0.65	66	-0.39
Non-structural protein 1	36	-0.62	43	-0.45
Matrix protein 1	7	-0.32	23	-0.45
Total	189	-0.56	268	-0.32

^a N: the number of mutations extracted. r: correlation coefficient.

averaged matrices. Graur (1985) studied the amino acid and nucleotide substitution patterns for several genes of retroviruses, and found the same tendency. Therefore, amino acid replacements seem to occur more frequently between similar amino acids than between dissimilar ones irrespective of DNA or RNA organisms. This suggests the existence of purifying selection in RNA virus genes, even if the rate of nucleotide substitution is extraordinarily high.

I would like to thank Dr. Masatoshi Nei for encouragements for this study and detailed comments on this paper. This work was supported by grants from National Science Foundation (U.S.A.) and National Institute of Health (U.S.A.) to Dr. M. Nei.

REFERENCES

- AIR, G. M. (1981) Sequence relationships among the hemagglutinin genes of 12 subtypes of influenza A virus. *Proc. Natl. Acad. Sci. USA* **78**, 7639-7643.
- FITCH, W. M. (1977) On the problem of discovering the most parsimonious tree. *Amer. Natur.* **111**, 223-257.
- GOJOBORI, T., LI, W.-H. and GRAUR, D. (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. *J. Mol. Evol.* **18**, 360-369.
- GRANTHAM, R. (1974) Amino acid difference formula to help explain protein evolution. *Science* **185**, 862-864.
- GRAUR, D. (1985) Pattern of nucleotide substitution and the extent of purifying selection in retroviruses. *J. Mol. Evol.* **21**, 221-231.
- HAYASHIDA, H., TOH, H., KIKUNO, R. and MIYATA, T. (1985) Evolution of influenza virus genes. *Mol. Biol. Evol.* **2**, 289-303.
- LI, W.-H., WU, C.-I. and LUO, C.-C. (1984) Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.* **21**, 58-71.
- NEI, M. (1983) Genetic polymorphism and the role of mutation in evolution. Pp. 165-190 in:

- Evolution of Genes and Proteins* (eds. NEI, M. and KOEHEN, R. K.), Sinaur Associates, Sunderland, Mass.
- SAITOU, N. and NEI, M. (1986) Polymorphism and evolution of influenza A virus genes. *Mol. Biol. Evol.* 3, 57-74.
- WEBSTER, R. G., LAVER, W. G., AIR, G. M. and SCHILD, G. C. (1982) Molecular mechanisms of variation in influenza viruses. *Nature* 296, 115-121.