Evolutionary pathways of N2 neuraminidases of swine and human influenza A viruses: origin of the neuraminidase genes of two reassortants (H1N2) isolated from pigs

Kuniaki Nerome,1* Yumi Kanegae,1 Yasuyuki Yoshioka,1 Shigeyuki Itamura,1 Masatoshi Ishida,1 Takashi Gojobori2 and Akira Oya1

1Department of Virology and Rickettsiology, National Institute of Health, Kamiosaki, Shinagawa-ku, Tokyo 141 and 2Department of Evolutionary Genetics, National Institute of Genetics, Mishima 411, Japan

The complete nucleotide sequences of the neuraminidase (NA) genes of two reassortant (H1N2) and two H3N2 influenza A viruses isolated from pigs were determined and phylogenetic relationships between these and previously reported N2 NA genes were investigated. On the basis of pairwise nucleotide sequence identity, the NA genes of two reassortants, A/sw/Kanagawa/2/78 and A/sw/Ehime/1/80, were most closely related to those of human influenza A virus strains isolated in 1972 and the earliest available swine H3N2 influenza A viruses, respectively. Phylogenetic trees showed that the NA genes can be segregated into three groups, including lineages for (i) swine strains, (ii) the earliest human strain and (iii) recent human strains. The evolutionary tree for the 11 nucleotide and amino acid sequences suggested that the NAs of A/sw/HK/4/76 and A/sw/Kanagawa/2/78 belong to the lineage for recent human viruses. In contrast, the NA genes of the A/sw/HK/3/76 and H1N2 reassortant A/sw/Ehime/1/80 viruses were found to be of a swine lineage. The swine virus NA genes were further characterized by the cocirculation of two distinct lineages. Although the rates of synonymous (silent) substitutions for the swine and human viruses were nearly identical (0.00946 to 0.00884 per site per year), the rate of non-synonymous (amino acid changing) substitutions for swine virus NA genes was about 60% of that for the human virus.

Introduction

Genomic analysis based on competitive RNA–RNA hybridization suggested that the haemagglutinin (HA) gene of the human H3N2 influenza A virus was derived from an avian influenza virus whereas the other seven genes were derived from the H2N2 influenza A virus that had been circulating in man between 1957 and 1967 (Scholtissek et al., 1978). Sequencing studies have shown that it is likely that the PB1 gene of the 1957 human pandemic virus was introduced from avian species (Kawaoka et al., 1989). In addition, nucleotide sequence analysis of the H3 HAs from human and avian influenza viruses strongly suggests that the HA gene was introduced from avian species into humans (Fang et al., 1981; Kida et al., 1987). Furthermore, a close relationship between the HAs of swine and human H3N2 viruses has been established by nucleotide and deduced amino acid sequence analyses (Kida et al., 1988). The above studies strongly support the presence of potential animal reservoirs for future pandemic strains.

The nucleotide sequences of N2 neuraminidase (NA) genes for seven influenza A viruses have been determined (Rompuy et al., 1982; Markoff & Lai, 1982; Bently & Brownlee, 1982; Martinez et al., 1983; Lentz et al., 1984; Air et al., 1985a, b); however, the evolutionary relationships between the NA genes of human and animal influenza viruses are not well understood.

Two H1N2 reassortant viruses were isolated from pigs in Japan between 1978 and 1980; the NA genes from these two reassortants were found to be related antigenically to the early and recent human H3N2 viruses, respectively (Nerome et al., 1983, 1985).

In this study we have sequenced the N2 NA genes of four swine influenza viruses and analysed the phylogenetic relationships between these N2 NA genes as well as N2 NA genes of influenza viruses isolated from humans. This analysis contributes to a better understanding of the origin and evolutionary pathways of the NA of reassor-
tant H1N2 and H3N2 viruses isolated from pigs or humans.

**Methods**

**Virus strains.** For determination of the nucleotide sequences of the NA gene, two H3N2 and two reasortant (H1N2) viruses were selected: A/sw/Hong Kong/3/76 (H3N2) (swHK376), A/sw/Hong Kong/4/76 (H3N2) (swHK476), A/sw/Kanagawa/2/78 (H1N2) (swKA278) and A/sw/Ehime/1/80 (H1N2) (swEH80). They were grown in the allantoic cavity of 11 day old embryonated hens' eggs and purified as described previously (Nerome et al., 1983).

**Nucleotide sequences.** The purified RNA of the above viruses was transcribed into single-stranded cDNA using reverse transcriptase and 12 base synthetic primers complementary to the 3’ terminus of the influenza virus RNA (Davis et al., 1981). Second-strand DNA synthesis was done as described previously (Efstratiadis et al., 1976). The cDNA was cloned into plasmid vector pBR327 and the synthesis was done as described previously (Efstratiadis et al., 1976). The NA cDNAs were recloned into the plasmid vector pUC118 and sequenced by the following methods. First, the sequence was determined from deletion mutants prepared as described previously (Henikoff et al., 1984) and second, a series of synthetic oligonucleotide primers were used for sequencing by the dideoxynucleotide chain termination method (Sanger et al., 1977).

**Calculation of evolutionary rate and construction of phylogenetic trees.** For the analysis of synonymous (silent) and non-synonymous (amino acid-changing) substitutions, we examined four NA sequences and seven previously published sequences (Air et al., 1985a; Bently & Brownlee, 1982; Elleman et al., 1982; Lentz et al., 1984; Markoff & Lai, 1982; Martinez et al., 1983; Rompuy et al., 1982). The number of synonymous and non-synonymous substitutions was calculated by the method described by Nei & Gojobori (1986). Phylogenetic trees were constructed by the neighbour-joining (N-J) method using the synonymous and non-synonymous substitutions (Saitou & Nei, 1987).

**Results**

**Comparison of the nucleotide and deduced amino acid sequence identity of the NAs from reasortant and other viruses**

The N2 NA genes of H3N2 (swHK376 and swHK476) and reasortant (swKA278 and swEH80) viruses were 1468 nucleotides long with a single long open reading frame of 1409 nucleotides, identical in size to those of the previously reported NA genes (Air et al., 1985a; Bently & Brownlee, 1982; Lentz et al., 1984; Markoff & Lai, 1982; Rompuy et al., 1982).

In a previous antigenic analysis using a panel of monoclonal antibodies, the NAs of swEH80 and swKA278 were suggested to have been derived from an A/Aichi/2/68 (H3N2)-like strain and an A/Victoria/3/75 (H3N2) (VIC375)-like strain, respectively (Nerome et al., 1983, 1985). In order to determine their relationship in more detail, we compared the number of nucleotide and amino acid sequence differences between human and swine virus N2 NA genes (Table 1). It was evident that the number of nucleotide and amino acid changes in N2 NAs of human H2N2 and H3N2 viruses appeared to increase with the time (years) between virus isolations. The numbers of nucleotide differences during the periods 1957 to 1967, 1957 to 1972 and 1957 to 1979 were 58, 87 and 104, respectively.

In contrast to this, the nucleotide sequence of the swHK376 strain differed by 94 nucleotides from that of the swHK476 virus, despite their isolation in the same year. Furthermore, there were only 66 nucleotide differences between the NA genes of swHK376 and a reasortant (swEH80) isolated 4 years apart; the NA genes of these two viruses exhibited the highest nucleotide sequence identity (95-4 and 94-8%, respectively) with that of the earliest human H3N2 virus (NT6068).

In addition, the NA gene and encoded protein sequences of swHK476 were most closely related to those of the human strain VIC375, showing 99-9% nucleotide and 98-3% amino acid sequence identity; the amino acid sequence identity between the swHK476 and the A/Udorn/72 (H3N2) (Udorn72) virus NAs was identical to the latter value.

Similarly, the NA gene and amino acid sequences of swKA278 and Udorn72 were most closely related, showing 96-5% and 96-4% identity, respectively, whereas the NA gene of the swEH80 strain was shown to be most closely related to that of the swHK376 strain, and had the greatest nucleotide (94-8%) and amino acid (94-2%) sequence identities.

**Phylogenetic analyses of the NA genes from human and swine viruses**

A comparative analysis of nucleotide and amino acid sequences led us to analyse in more detail the evolutionary relationships between the NA genes and proteins from human and swine influenza viruses. To address this question, we constructed evolutionary trees by the N-J method on the basis of synonymous and non-synonymous substitutions. The values given for each branch in the two trees represent the expected branch lengths (the expected number of nucleotide substitutions per site). The phylogenetic trees based on synonymous (Fig. 1a) and non-synonymous (Fig. 1b) substitutions showed a similar branching pattern. The existence of a putative common origin and branching features indicated that the NA genes could be segregated into three lineages. The first main branch contained NA genes of human strains NT6068 and Tokyo67. At least three human and two swine strains were included in the second main branch, suggesting the possible introduction of sequences from the NA gene of human epidemic influenza virus strains into swine isolates (swHK476 and swKA278).
Table 1. Comparison of nucleotide and amino acid sequence identity between the N2 NAs of two reassortant viruses from swine and other viruses derived from man and swine*

<table>
<thead>
<tr>
<th>NAs from</th>
<th>R15-57</th>
<th>R15+57</th>
<th>Tokyo67</th>
<th>NT6068 N2 (H3)</th>
<th>Udorn72 N2 (H3)</th>
<th>VIC375 N2 (H3)</th>
<th>BK179 N2 (H3)</th>
<th>swHK376 N2 (H1)</th>
<th>swHK476 N2 (H1)</th>
<th>swKA278 N2 (H1)</th>
<th>swEH80 N2 (H1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R15-57</td>
<td>2 (99-6)</td>
<td>58 (95-9)</td>
<td>59 (95-8)</td>
<td>87 (93-8)</td>
<td>100 (92-9)</td>
<td>104 (92-6)</td>
<td>89 (93-7)</td>
<td>96 (93-2)</td>
<td>123 (91-3)</td>
<td>100 (92-9)</td>
<td></td>
</tr>
<tr>
<td>R15+57</td>
<td>2 (99-6)</td>
<td>58 (95-9)</td>
<td>59 (95-8)</td>
<td>87 (93-8)</td>
<td>100 (92-9)</td>
<td>104 (92-6)</td>
<td>89 (93-7)</td>
<td>96 (93-2)</td>
<td>123 (91-3)</td>
<td>100 (92-9)</td>
<td></td>
</tr>
<tr>
<td>Tokyo67</td>
<td>28 (94-0)</td>
<td>28 (94-0)</td>
<td>23 (98-4)</td>
<td>59 (95-8)</td>
<td>71 (95-0)</td>
<td>80 (94-3)</td>
<td>69 (95-1)</td>
<td>71 (95-0)</td>
<td>92 (93-5)</td>
<td>75 (94-7)</td>
<td></td>
</tr>
<tr>
<td>NT6068</td>
<td>44 (90-6)</td>
<td>44 (90-6)</td>
<td>29 (93-8)</td>
<td>58 (95-9)</td>
<td>72 (94-9)</td>
<td>76 (94-6)</td>
<td>65 (94-4)</td>
<td>71 (95-0)</td>
<td>93 (93-4)</td>
<td>73 (94-8)</td>
<td></td>
</tr>
<tr>
<td>Udorn72</td>
<td>48 (89-8)</td>
<td>48 (89-8)</td>
<td>34 (92-8)</td>
<td>9 (98-1)</td>
<td>41 (97-1)</td>
<td>94 (93-3)</td>
<td>69 (95-1)</td>
<td>16 (98-9)</td>
<td>61 (95-7)</td>
<td>104 (92-6)</td>
<td></td>
</tr>
<tr>
<td>BK179</td>
<td>50 (89-3)</td>
<td>50 (89-3)</td>
<td>38 (91-9)</td>
<td>14 (97-0)</td>
<td>15 (96-8)</td>
<td>110 (92-2)</td>
<td>51 (96-4)</td>
<td>73 (94-8)</td>
<td>109 (92-3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>swHK376</td>
<td>45 (90-4)</td>
<td>45 (90-4)</td>
<td>32 (93-2)</td>
<td>8 (98-3)</td>
<td>8 (98-3)</td>
<td>14 (97-0)</td>
<td>34 (92-8)</td>
<td>94 (93-3)</td>
<td>121 (91-4)</td>
<td>66 (95-3)</td>
<td></td>
</tr>
<tr>
<td>swHK476</td>
<td>45 (90-4)</td>
<td>45 (90-4)</td>
<td>33 (93-0)</td>
<td>8 (98-3)</td>
<td>8 (98-3)</td>
<td>14 (97-0)</td>
<td>24 (92-8)</td>
<td>94 (93-3)</td>
<td>121 (91-4)</td>
<td>66 (95-3)</td>
<td></td>
</tr>
<tr>
<td>swKA278</td>
<td>52 (88-9)</td>
<td>52 (88-9)</td>
<td>38 (91-9)</td>
<td>17 (96-4)</td>
<td>21 (95-5)</td>
<td>24 (94-9)</td>
<td>30 (91-7)</td>
<td>21 (95-5)</td>
<td>125 (91-1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>swEH80</td>
<td>36 (92-3)</td>
<td>36 (92-3)</td>
<td>30 (93-6)</td>
<td>27 (94-2)</td>
<td>40 (91-5)</td>
<td>44 (90-6)</td>
<td>48 (90-0)</td>
<td>48 (94-9)</td>
<td>48 (91-3)</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

* Number of nucleotide differences (nucleotides 20 to 1429) between NA genes (right upper half) and amino acid differences between N2 NA proteins (left lower half) are indicated.
† Codes in parentheses represent the haemagglutinin subtype of the virus.
Fig. 1. Phylogenetic trees of 11 N2 NA genes of influenza A viruses constructed by the N-J method (Nei & Gojobori, 1986; Saitou & Nei, 1987) on the basis of synonymous (a) and non-synonymous (b) substitutions. The length of horizontal lines in the phylogenetic trees represents genetic distance and was estimated by the principle of minimum evolution (Kanegae et al., 1990). This is based on the trees being constructed to minimize the total number of nucleotide substitutions for a given tree. The values given for each branch of the tree are the estimated branch lengths (estimated number of nucleotide substitutions per site). Vertical lines joined with horizontal lines are used to separate progeny virus lineages at the position where they diverge from the common ancestor gene. P is the point of common origin.

The following point is of particular interest: the phylogenetic position of the NA gene of swHK476, isolated from a pig, was closest to that of the human VIC375 virus. Our analysis revealed that the NA genes of swHK376 and swEH80 are undoubtedly on a separate, third major evolutionary branch which is distinct from those of the above human viruses.

To estimate the rate of synonymous substitution and the year of divergence, the estimated number of synonymous substitutions per site per year from the common origin was plotted against the year of virus isolation and a linear regression line was calculated (Fig. 2). The numbers of sequence differences in the NA genes are roughly proportional to the period between their isolation, behaving like a molecular clock. On the basis of synonymous substitution, the slope of the regression line of the human and swine influenza virus NA genes is almost the same. The result suggests that the NA genes of the earliest swine strain and earliest human H3N2 virus diverged from a common ancestral virus that was prevalent in the mid-1960s.

Table 2. Comparative analysis of the rate of evolution of synonymous and non-synonymous substitutions* and substitutions at different codon positions between N2 NA genes derived from swine and human viruses

<table>
<thead>
<tr>
<th>Nucleotide substitutions at codon position</th>
<th>Rate of evolution ± S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Swine</td>
</tr>
<tr>
<td>First</td>
<td>0.00121 ± 0.00049</td>
</tr>
<tr>
<td>Second</td>
<td>0.00118 ± 0.00011</td>
</tr>
<tr>
<td>Third</td>
<td>0.00667 ± 0.00068</td>
</tr>
<tr>
<td>All</td>
<td>0.00274 ± 0.00029</td>
</tr>
<tr>
<td>Synonymous</td>
<td>0.00946 ± 0.00154</td>
</tr>
<tr>
<td>Non-synonymous</td>
<td>0.00092 ± 0.00011</td>
</tr>
</tbody>
</table>

* Numbers of synonymous and non-synonymous substitutions per site per year were calculated on the basis of the phylogenetic trees shown in Fig. 1(a and b). Accumulation of synonymous substitutions in swine and human virus NA genes is shown in Fig. 2.
positions for the swine NA genes was estimated to be 0.00121 and 0.00118 per site per year, approximately 33% to 50% of that of the human virus gene at these positions.

The estimated rate of nucleotide substitutions at the third codon position is high and almost the same as that of synonymous substitutions. The rate of non-synonymous substitutions in swine virus NA genes per site per year (0.00092) appeared to be lower than that (0.00156) in human virus NA genes. The rate of non-synonymous substitutions per site per year was estimated to be, on average, 0.00086 for NA genes belonging to the swine virus lineage (swHK376 and swEH80).

To confirm the slower rate of non-synonymous substitution in the swine influenza virus NA genes, we applied another type of regression analysis which was based on a pairwise comparison of the relationship between the number of nucleotide differences per non-synonymous site and the number of nucleotide differences per synonymous site (N/S) in the human and swine NA genes. As seen in Fig. 3, the slopes of the two regression lines were apparently different, showing a tendency for the N/S ratio of swine virus NA genes (0.14720 ± 0.00574) to be less than that of the human virus gene (0.24481 ± 0.00212).

Discussion

In the past few years, comparative analyses of the evolution of human and animal influenza viruses have been made (Kida et al., 1987, 1988) which have shown that duck and swine influenza viruses seem to evolve more slowly than human influenza virus. However, our data have revealed that the evolutionary rate of synonymous substitutions in swine virus NA genes is approximately the same as that in human virus NA genes.

Although the NA antigens of the recombinant (swKA278) and A/Kumamoto/22/76 (Vic375-like strain) viruses gave a similar pattern of reactivity with a panel of monoclonal antibodies, a pairwise comparison of nucleotide sequences suggested that the NA gene of the recombinant is derived from a Udorn72-like virus. However, antigenic analysis of a second recombinant (swEH80) using monoclonal antibodies (Nerome et al., 1985) gave the same result as the pairwise sequence analysis, suggesting that the NA gene of swEH80 may have been introduced from a virus closely related to the swHK376 virus.

Ample evidence has documented the transmission of human H3N2 virus to swine populations (Nerome et al., 1981; Shortridge et al., 1979) and the branching patterns of the evolutionary trees supported the above reports by showing that the NA genes of swHK476 and swKA278 are of a human lineage. Recently, Air et al. (1990) reported a phylogenetic tree consisting of two main branches including the earliest and most recent human H3N2 virus NA genes. Although this observation is essentially similar to the results obtained in the present study, there are the differences. (i) The earliest swine H3N2 virus NA gene has evolved independently from that of the earliest human H3N2 virus and (ii) the remaining swine NA gene isolates were involved in a lineage common to recent human virus strains. These results are compatible with a previous report based on oligonucleotide mapping (Nakajima et al., 1982). In the present study we have demonstrated the cocirculation of two distinct evolutionary lineages of the N2 NA gene in a swine population.

Recently, Air et al. (1990) estimated the rate of nucleotide substitution for the N2 subtype of the NA gene to be 0.00311 (0.311%) per year. These estimates were obtained from an analysis of all codon positions. Table 2 shows our estimates (0.00344) based on all
positions which indicate that the rate of nucleotide substitutions at all codon positions, as well as the first and the second codon positions, tends to give a lower estimate compared with those for the third codon position or synonymous substitution. Our analysis of the rate of non-synonymous substitutions and N/S values obtained in pairwise comparison of the sequence indicated that the NA gene of human virus evolves at a higher rate than that of swine virus. One explanation for this phenomenon is that functional constraints on the NA gene cause the lower rate of protein evolution in the swine population. Another possible explanation is that the higher rate of non-synonymous substitution in human NA genes may reflect the positive selection pressure on the antigenic sites resulting from host immunity; it has been shown that monoclonal antibodies to the NA can neutralize virus growth in embryonated eggs and can permit selection of antigenic variants (Webster et al., 1984).

The authors are grateful to Dr Nancy J. Cox, Centers for Disease Control, Atlanta, Ga., U.S.A. for her comments on the manuscript. This work was supported by a Pharmaceutical and Supply Bureau grant from the Ministry of Health and Welfare.

References


NEROME, K., YOSHIOKA, Y., SAKAMOTO, S., YASUHARA, H. & OYA, A. (1985). Characterization of a 1980 swine recombinant influenza virus possessing H1 hemagglutinin and N2 neuraminidase similar to that of the earliest Hong Kong (H3N2) virus. *Archives of Virology* 96, 197–211.


(Received 20 August 1990; Accepted 13 November 1990)