

Independence of Evolutionary and Mutational Rates after Transmission of Avian Influenza Viruses to Swine

J. STECH,¹ X. XIONG,² C. SCHOLTISSEK,^{1†} AND R. G. WEBSTER^{1*}

*Department of Virology and Molecular Biology¹ and Department of Epidemiology and Biostatistics,²
St. Jude Children's Research Hospital, Memphis, Tennessee 38105-2794*

Received 9 March 1998/Accepted 11 November 1998

In 1979, an H1N1 avian influenza virus crossed the species barrier, establishing a new lineage in European swine. Because there is no direct or serologic evidence of previous H1N1 strains in these pigs, these isolates provide a model for studying early evolution of influenza viruses. The evolutionary rates of both the coding and noncoding changes of the H1N1 swine strains are higher than those of human and classic swine influenza A viruses. In addition, early H1N1 swine isolates show a marked plaque heterogeneity that consistently appears after a few passages. The presence of a mutator mutation was postulated (C. Scholtissek, S. Ludwig, and W. M. Fitch, *Arch. Virol.* 131:237–250, 1993) to account for these observations and the successful establishment of an avian H1N1 strain in swine. To address this question, we calculated the mutation rates of A/Mallard/New York/6750/78 (H2N2) and A/Swine/Germany/2/81 (H1N1) by using the frequency of amantadine-resistant mutants. To account for the inherent variability of estimated mutation rates, we used a probabilistic model for the statistical analysis. The resulting estimated mutation rates of the two strains were not significantly different. Therefore, an increased mutation rate due to the presence of a mutator mutation is unlikely to have led to the successful introduction of avian H1N1 viruses in European swine.

Influenza viruses undergo rapid variation in nature, thereby limiting prevention of epidemics and pandemics (20). Therefore, questions regarding their evolution remain important, and the answers may yield information useful for predicting further antigenic changes and for explaining the occurrence of new pandemic strains. In 1979, an H1N1 influenza virus of avian origin was transmitted to pigs in northern Europe, thereby introducing a new stable lineage (22, 26). Because there is no direct or serologic evidence of previous H1N1 strains in these pigs (35), we do not need to extrapolate the point of introduction from evolutionary changes over time. Therefore, the European swine influenza H1N1 viruses are a model for studying the early evolution of influenza virus strains in a new host after their introduction in the absence of reassortment.

Several characteristics of these H1N1 swine viruses make them particularly interesting for studying the evolution of influenza viruses. Analysis of phylogenetic data on their HA, NP, M, and NS genes revealed that the evolutionary rates of both the coding and noncoding changes of these strains are up to 54% higher than those of human and classic swine viruses (18, 33). Such higher evolutionary rates might be due to positive Darwinian selection in avian-like swine influenza viruses, reflecting adaptation to a new host; to differences in sampling frequencies (10) of the viral lineages compared; or to a higher mutation rate. The elevated rate of noncoding changes might be due to fixation of linked coding changes. In addition, early H1N1 swine viruses show a marked plaque heterogeneity and an unusually high escape rate in the presence of various monoclonal antibodies (18). These observations suggest that at least one other factor, independent from positive selection of advantageous variants, must have influenced the evolution of the

swine H1N1 isolates since 1979. One candidate is the presence of a mutator mutation (24). An initial variant could have a mutation in its polymerase complex that caused it to be more error prone and to generate a broader spectrum of variants. Such a polymerase might be detrimental for an established strain but advantageous under stress conditions such as adaptation to a new environment. After that, variants with a less error-prone polymerase might again become predominant in the population (18).

We wanted to determine whether increased mutation rates due to the presence of a mutator mutation might have contributed to the establishment of the European swine H1N1 viruses. To this end, we calculated the mutation rates of A/Swine/Germany/2/81, a well-characterized early H1N1 swine isolate, and A/Mallard/New York/6750/78 (H2N2), a well-established avian isolate; the evolution of avian influenza viruses is considerably slower than that of classic swine and human isolates (16). An avian H1N1 precursor virus was not available for study. To accommodate the inherent high variability of estimated mutation rates, we developed a new probabilistic model (see Appendix) to calculate the mutation rate and its standard deviation (SD) and screened several parallel clones. To facilitate rapid screening of parallel cultures, we evaluated the frequency with which amantadine-resistant mutants developed. Our results suggest that the mutation rates of A/Swine/Germany/2/81 and A/Mallard/New York/6750/78 are not significantly different and fail to reflect their very different evolutionary backgrounds.

MATERIALS AND METHODS

Viruses. The influenza viruses A/Mallard/New York/6750/78 (H2N2) and A/Swine/Germany/2/81 (H1N1) were plaque purified twice by using monolayers of Madin-Darby canine kidney (MDCK) cells. Culture was with minimal essential medium (MEM) (Sigma, St. Louis, Mo.), 0.9% agar (Difco Laboratories, Detroit, Mich.) overlays containing 0.5 g of TPCK (tolylsulfonil phenylalanyl chloromethyl ketone)-trypsin (Worthington Biochemical Corporation, Freehold, N.J.) per ml, and a 3-day incubation at 37°C. Virus populations were then propagated individually from single well-isolated plaques in the allantoic cavities of 11-day embryonated chicken eggs by incubating for 2 days at 37°C; the eggs

* Corresponding author. Mailing address: Department of Virology and Molecular Biology, St. Jude Children's Research Hospital, 332 N. Lauderdale St., Memphis, TN 38105-2794. Phone: (901) 495-3400. Fax: (901) 495-2622. E-mail: Robert.Webster@stjude.org.

† Present address: Waldstrasse 53, Linden-Leihgestern 35440, Germany.

TABLE 1. Mutation rate of A/Mallard/New York/6750/78

Clone no.	Plaque inhibition assay result (PFU/ml)		Allantoic fluid (ml)	Population size (PFU) ^a	Mutant frequency ^b	Mutation rate (bp/site/replication) ^c	Quotient (control/amantadine)	Quotient × 4	Quotient × 4/genome length
	Control	Amantadine							
1	1.6 × 10 ⁷	1.2 × 10 ⁴	7.0	1.1 × 10 ⁸	7.5 × 10 ⁻⁴	1.4 × 10 ⁻⁵	1.3 × 10 ³	5.3 × 10 ³	0.4
2	4.0 × 10 ⁷	3.2 × 10 ⁴	9.1	3.6 × 10 ⁸	8.0 × 10 ⁻⁴	1.4 × 10 ⁻⁵	1.3 × 10 ³	5.0 × 10 ³	0.4
3	1.0 × 10 ⁷	1.2 × 10 ⁴	7.1	7.1 × 10 ⁷	1.2 × 10 ⁻³	2.3 × 10 ⁻⁵	8.3 × 10 ²	3.3 × 10 ³	0.2
4	2.9 × 10 ⁷	1.2 × 10 ⁴	7.8	2.3 × 10 ⁸	4.1 × 10 ⁻⁴	7.5 × 10 ⁻⁶	2.4 × 10 ³	9.7 × 10 ³	0.7
5	2.0 × 10 ⁷	4.4 × 10 ⁴	8.5	1.7 × 10 ⁸	2.2 × 10 ⁻³	4.0 × 10 ⁻⁵	4.5 × 10 ²	1.8 × 10 ³	0.1
6	9.6 × 10 ⁶	1.6 × 10 ⁴	13.3	1.3 × 10 ⁸	1.7 × 10 ⁻³	3.1 × 10 ⁻⁵	6.0 × 10 ²	2.4 × 10 ³	0.2
7	6.8 × 10 ⁷	8.0 × 10 ³	9.3	6.3 × 10 ⁸	1.2 × 10 ⁻⁴	2.0 × 10 ⁻⁶	8.5 × 10 ³	3.4 × 10 ⁴	2.5
8	8.4 × 10 ⁷	4.0 × 10 ⁴	10.4	8.7 × 10 ⁸	4.8 × 10 ⁻⁴	8.0 × 10 ⁻⁶	2.1 × 10 ³	8.4 × 10 ³	0.6
9	3.2 × 10 ⁷	3.2 × 10 ⁵	7.9	2.5 × 10 ⁸	1.0 × 10 ⁻²	1.8 × 10 ⁻⁴	1.0 × 10 ²	4.0 × 10 ²	0.0
10	1.1 × 10 ⁸	4.0 × 10 ³	13.6	1.5 × 10 ⁹	3.6 × 10 ⁻⁵	6.0 × 10 ⁻⁷	2.8 × 10 ⁴	1.1 × 10 ⁵	8.1
Average	4.2 × 10 ⁷	5.0 × 10 ⁴		4.3 × 10 ⁸	1.8 × 10 ⁻³	3.2 × 10 ⁻⁵	4.5 × 10 ³	1.8 × 10 ⁴	1.3

^a Product of titer from control and amount of allantoic fluid.

^b Quotient of titers from plaque inhibition assays with amantadine and control.

^c Rate for each clone calculated from mutant frequency and population size as described in Materials and Methods.

were stored overnight at 4°C before the allantoic fluid of each egg was harvested individually and used in the plaque inhibition assay.

Plaque inhibition assay with amantadine. The plaque assay was performed on MDCK cell monolayers, which were grown in MEM and stained with crystal violet after 3 days as described by Ho et al. (14). Briefly, cells were inoculated with phosphate-buffered saline-diluted virus, left for 30 min at 37°C to allow adsorption, and then washed twice with phosphate-buffered saline, overlaid with MEM (Sigma) containing 0.9% agar (Difco Laboratories), and incubated at 37°C. After 3 days, the plaques were visualized by crystal violet (0.1% in 10% formaldehyde) staining.

For calculation of the frequency of amantadine-resistant mutants, the numbers of PFU per milliliter were determined for each clone in the absence and presence of 1 µg of amantadine hydrochloride (DuPont, Wilmington, Del.) per ml in the agar overlay.

Calculation of mutation rate and SD and statistical analysis. For calculating the estimated rate of mutation, we have developed a probabilistic model (see Appendix). The estimate of the mutation rate m is $\hat{m} = 2M/(N \log_2 N)$, where M is the number of mutants in the population (i.e., the number of PFU of amantadine-resistant mutants) and N is the population size from one egg-grown clone. Both numbers are the products of the respective PFU per milliliter from the plaque inhibition assay and the total amount of allantoic fluid (milliliters). The SD of that estimate is $2m/\log_2 N$.

For amantadine at a concentration of 1 µg/ml, substitution at one of four amino acids (positions 27, 30, 31, and 34 in the M2 ion channel) leads to resistance (13); the sequences of A/Mallard/New York/6750/78 and A/Swine/Germany/2/81 are identical at these positions. From the comparison with the codons of the amino acid substitutions that lead to resistance (13), it can be concluded that one nucleotide substitution in each triplet, i.e., one of four possible nucleotides, causes the resistant phenotype in both strains. Therefore, m has to be divided by 4, and the formula for the estimated mutation rate (the number of nucleotide substitutions per site per replication) for the experimental approach described here is $\hat{m} = 0.5 M/(N \log_2 N)$. We made no attempt to calculate a maximum mutation rate by considering more than one nucleotide substitution at each site, because transitions are more frequent than transversions (32). Therefore, the assumption of an equal probability of all changes is unlikely.

Because the generation at which a mutation occurs markedly influences the frequency of mutants (6), we grew multiple parallel cultures (C). Therefore, the estimated mutation rate for each viral strain is the arithmetic mean of the mutation rate from every clone,

$$\hat{m} = \frac{0.5}{C} \sum_{i=1}^C \frac{M_i}{N_i \log_2 N_i}$$

and the SD is

$$SD = \sqrt{\frac{m}{2C^2} \sum_{i=1}^C \frac{1}{(\log_2 N_i)^2}}$$

For comparison, the ad hoc SD was calculated as

$$SD_{ad hoc} = \sqrt{\frac{C \left(\sum_{i=1}^C \hat{m}_i^2 \right) - \left(\sum_{i=1}^C \hat{m}_i \right)^2}{C(C-1)}}$$

where \hat{m}_i is the determined mutation rate of the i th clone and C is the whole number of observed clones from one isolate.

For the statistical analysis, the null hypothesis (H_0), $m_{mall} = m_{sw}$, versus the alternative hypothesis (H_a), $m_{mall} \neq m_{sw}$, was tested by using the normal approximation $Z = (\hat{m}_{mall} - \hat{m}_{sw})/\sqrt{SD_{mall}^2 + SD_{sw}^2}$, and the test claims that H_a is true if $|Z| > 1.96$ (97.5% quantile of the normal distribution). (m_{mall} and m_{sw} are the mutation rates for strains A/Mallard/New York/6750/78 and A/Swine/Germany/2/81, respectively.)

The basic assumptions of our mathematical model are as follows.

(i) The mutation rate is the same for all observed sites and at all replications. For the present study, this means that loci related to amantadine resistance mutate at a rate which is representative of those for all other genome loci, including those involved in species transmission. Positions other than 27, 30, 31, and 34 in M2 (e.g., flanking sequences) could affect the generation of amantadine resistance, thus obscuring intrinsic differences in mutation rate. That is assumed to be negligible, because no such data supporting this view are available for influenza virus.

(ii) The mutation rate m is very small compared to 1 (10^{-4} to 10^{-6}); therefore, higher orders of m such as m^2 , m^3 , and m^4 (as would apply to two-, three-, or fourfold substitutions) can be ignored.

(iii) From the second assumption, it follows that reverse mutations are also negligible.

(iv) The model applies only to neutral mutants in an ideal way. Bean et al. showed that amantadine-resistant influenza virus in chickens is capable of competing with wild-type virus and does not revert back in the absence of amantadine over a period of >20 days, thus having no detectable biologic impairment compared to wild-type virus (1). Moreover, negative selection (elimination of deleterious mutation as soon as they occur) is assumed to be identical for the strains compared.

(v) Possible variations in polymerase (or other fidelity-determining proteins) upon virus growth in cell culture need not be considered, because the estimated mutation rate \hat{m} is defined as base substitutions per site per replication.

RESULTS

Evaluation of several independent clones. We measured the frequencies of amantadine-resistant mutants of A/Mallard/New York/6750/78 (H2N2) (Table 1) and A/Swine/Germany/2/81 (H1N1) (Table 2) to indirectly assess the mutation rates of the two parent strains. Because the variability of mutant frequency is inherently high, we evaluated multiple parallel, independently grown viral stocks. Extensive plaque purification ensured that all amantadine-resistant variants occurred during growth in the plaque or allantoic fluid and did not exist prior to culture. A primary resistant plaque would grow to become a

TABLE 2. Mutation rate of A/Swine/Germany/2/81

Clone no.	Plaque inhibition assay result (PFU/ml)		Allantoic fluid (ml)	Population size (PFU) ^a	Mutant frequency ^b	Mutation rate (bp/site/replication) ^c	Quotient (control/amantadine)	Quotient × 4	Quotient × 4/genome length
	Control	Amantadine							
1	9.2×10^8	1.2×10^5	13.0	1.2×10^{10}	1.3×10^{-4}	1.9×10^{-6}	7.7×10^3	3.1×10^4	2.3
2	6.0×10^8	2.6×10^5	9.0	5.4×10^9	4.3×10^{-4}	6.7×10^{-6}	2.3×10^3	9.2×10^3	0.7
3	2.3×10^8	1.8×10^5	13.0	3.0×10^9	7.8×10^{-4}	1.2×10^{-5}	1.3×10^3	5.1×10^3	0.4
4	6.0×10^8	2.6×10^5	11.0	6.6×10^9	4.3×10^{-4}	6.6×10^{-6}	2.3×10^3	9.2×10^3	0.7
5	1.0×10^9	2.0×10^5	12.0	1.2×10^{10}	2.0×10^{-4}	3.0×10^{-6}	5.0×10^3	2.0×10^4	1.5
6	2.2×10^8	8.0×10^4	10.7	2.4×10^9	3.6×10^{-4}	5.8×10^{-6}	2.8×10^3	1.1×10^4	0.8
7	4.2×10^8	1.4×10^5	9.3	3.9×10^9	3.3×10^{-4}	5.2×10^{-6}	3.0×10^3	1.2×10^4	0.9
8	4.0×10^7	4.0×10^4	8.6	3.4×10^8	1.0×10^{-3}	1.8×10^{-5}	1.0×10^3	4.0×10^3	0.3
9	6.0×10^7	1.8×10^4	9.6	5.8×10^8	3.0×10^{-4}	5.2×10^{-6}	3.3×10^3	1.3×10^4	1.0
10	3.2×10^7	2.0×10^4	10.5	3.4×10^8	6.3×10^{-4}	1.1×10^{-5}	1.6×10^3	6.4×10^3	0.5
11	1.0×10^8	1.2×10^4	6.0	6.0×10^8	1.2×10^4	2.1×10^{-6}	8.3×10^3	3.3×10^4	2.5
12	4.0×10^8	4.0×10^4	13.5	5.4×10^9	1.0×10^{-4}	1.5×10^{-6}	1.0×10^4	4.0×10^4	2.9
13	2.8×10^8	2.4×10^5	10.8	3.0×10^9	8.6×10^{-4}	1.4×10^{-5}	1.2×10^3	4.7×10^3	0.3
14	4.0×10^7	8.0×10^3	8.8	3.5×10^8	2.0×10^{-4}	3.5×10^{-6}	5.0×10^3	2.0×10^4	1.5
15	1.2×10^8	4.8×10^4	9.0	1.1×10^9	4.0×10^{-4}	6.7×10^{-6}	2.5×10^3	1.0×10^4	0.7
16	1.3×10^8	1.6×10^4	11.9	1.5×10^9	1.2×10^{-4}	2.0×10^{-6}	8.1×10^3	3.3×10^4	2.4
17	1.4×10^8	9.6×10^4	9.8	1.4×10^9	6.9×10^{-4}	1.1×10^{-5}	1.5×10^3	5.8×10^3	0.4
Average	3.1×10^8	1.0×10^5		3.5×10^9	4.2×10^{-4}	6.8×10^{-6}	3.9×10^3	1.6×10^4	1.2

^a Product of titer from control and amount of allantoic fluid.

^b Quotient of titers from plaque inhibition assays with amantadine and control.

^c Rate for each clone calculated from mutant frequency and population size as described in Materials and Methods.

viral stock that was comprised almost entirely of resistant viruses and therefore would be recognized immediately.

Mutant frequencies and mutation rates. As described in Materials and Methods, we performed plaque inhibition assays to determine the population size and the frequency of amantadine-resistant mutants, and then we calculated the average mutation rate for each strain and its SD (Tables 1 and 2). For A/Mallard/New York/6750/78, the mutation rate (\hat{m}_{mall}) is 3.2×10^{-5} base substitutions per site per replication and its SD (SD_{mall}) is 4.5×10^{-5} ; for A/Swine/Germany/2/81, $\hat{m}_{\text{sw}} = 6.8 \times 10^{-6}$ base substitutions per site per replication and $\text{SD}_{\text{sw}} = 1.5 \times 10^{-5}$. Table 3 shows that our data correspond well with two values previously published for the mutation rate of influenza virus.

Dispersion of data. The mutation rates of clones 9 and 10 of A/Mallard/New York/6750/78 (Table 1) demonstrate the problem of variability in estimates of mutation rate. The estimate for clone 9 was almost 6 times higher than the average mutation rate, whereas that for clone 10 was 50 times lower. Both values were included in our calculations, although one may consider them outliers. Interestingly, the calculation of the ad hoc SD showed a result for A/Mallard/New York/6750/78 similar to that obtained with our formula for SD when the mutation rates from clones 9 and 10 were included ($\text{SD}_{\text{ad hoc}} = 5.3 \times 10^{-5}$ and $\text{SD} = 4.5 \times 10^{-5}$), whereas for A/Swine/Germany/2/81, the values are $\text{SD}_{\text{ad hoc}} = 5.2 \times 10^{-6}$ and $\text{SD} = 1.5 \times 10^{-5}$. This finding indicates that our theoretical formula for SD is applicable to our data and describes the dispersion well. The mutation rates from clones 9 and 10 are very diver-

gent from the mean, indicating the influence of the time point during growth of the viral stock at which mutational events leading to amantadine resistance occurred on the estimated mutation rate. These observations emphasize the randomness of mutational events in nature and demonstrate that measuring the mutation rate of a single clone can be misleading.

Minimum number of replication cycles. In light of the small ratio of the number of sites at which changes lead to amantadine resistance (4 positions) to the genome length (almost 14,000 positions), mutation to resistance is highly improbable compared with a mutation at any other site in the entire viral genome. Therefore, such a mutational event within the viral population is improbable when the population size is considerably smaller than one-fourth of the genome length (for only one position, it would be improbable when the population size is considerably smaller than the entire genome). If all mutants are assumed to have developed from a single mutational event, then the ratio of the PFU per milliliter from plaque inhibition assays without amantadine (control) to that from assays with amantadine in the plaque overlay reveals a theoretical minimum population size in which the first mutant would have arisen. Multiplying the calculated minimum population size of A/Mallard/New York/6750/78 or A/Swine/Germany/2/81 by 4 reveals a value close to the genome length (Tables 1 and 2). The resulting value can be considered to be the minimum number of replication cycles necessary for each additional mutational step (requiring only a single-base substitution) during further growth. Therefore, in light of our results regarding amantadine, a population of approximately 5.0×10^3 PFU

TABLE 3. Comparison of estimated mutation rates for various influenza virus strains

Virus	Reference	Assay	Gene	Mutation rate (bp/site/replication)
A/Victoria/3/75 (H3N2)	28	Monoclonal antibody escape	HA	1.0×10^{-5}
A/WSN/33 (H1N1)	21	Direct sequencing of RNA	NS	0.84×10^{-5}
A/Mallard/New York/6750/78 (H2N2)	This study	Amantadine escape	M	3.2×10^{-5}
A/Swine/Germany/2/81 (H1N1)	This study	Amantadine escape	M	0.68×10^{-5}

probably carries at least one mutated virion; for a mutated phenotype caused by a base substitution at only one possible position, this value would be approximately 2.0×10^4 PFU, and for two mutational steps in one viral genome, it would be approximately 4.0×10^8 PFU. These numbers indicate the importance of the size of a finite viral population in a host and the turnover of virions during infection for the probability of occurrence a particular virus mutant requiring several point mutations. That relationship has been discussed by Coffin (2) and has been described in a mathematical model, derived from a quasispecies equation, by Ribeiro et al. (23) in regard to human immunodeficiency virus population dynamics in view of the frequency of resistant mutant virus before antiviral therapy.

Statistical analysis. The statistical analysis was performed as described in Materials and Methods. For the two-sided test with a significance level of 0.05, because $(\hat{m}_{\text{mall}} - \hat{m}_{\text{sw}}) / \sqrt{SD_{\text{mall}}^2 + SD_{\text{sw}}^2} = 0.5$ (a value smaller than 1.96, the 97.5% quantile of the normal distribution), the null hypothesis ($m_{\text{mall}} = m_{\text{sw}}$) cannot be rejected. Therefore, we conclude that the mutation rates of A/Mallard/New York/6750/78 and A/Swine/Germany/2/81 do not differ significantly.

DISCUSSION

The influenza pandemic of 1918 was caused by an H1N1 influenza virus (31). This isolate was probably closely related to later classic swine viruses and is thought to have been transmitted previously from birds into pigs without reassortment (24, 33). In 1979, an influenza virus of avian origin was introduced as a whole into swine of northern Europe (26), leading to the development of a new virus lineage. Before that time, H1N1 influenza viruses did not exist in pigs in northern Europe (22, 35). Therefore, we can hypothesize that the 1979 event is analogous to that with the 1918 isolate. Use of the northern European isolates as a model for studying the early evolution of influenza viruses is likely to improve our understanding of the conditions and circumstances that surround the emergence of an entirely new pandemic strain.

Evolution of influenza viruses occurs, in various degrees, by Darwinian selection and random drift, including mutation. Elevated rates of mutation, caused by a mutator mutation, may have accelerated adaptation of the H1N1 swine isolates (24). To test this theory, we determined the mutation rates of A/Swine/Germany/2/81 (H1N1), an early representative of the northern European swine influenza lineage, and A/Mallard/New York/6750/78 (H2N2). The latter strain, a well-established avian isolate, likely does not have an elevated mutation rate, because the evolution rates of other avian isolates were very slow (16) or even static (11, 12), and no available information suggests that A/Mallard/New York/6750/78 underwent several host changes in multiple avian species. Our results show that the mutation rates of A/Swine/Germany/2/81 and A/Mallard/New York/6750/78 do not differ significantly. This finding indicates that Darwinian selection rather than an increased mutation rate contributed to the early adaptation and high evolutionary rates of European swine influenza H1N1 viruses.

Our results further demonstrate that directly inferring the rates of mutation of avian-like swine influenza viruses from phylogenetic data is difficult. This is because the evolutionary rate (i.e., number of base substitutions per year per site) is also influenced by fluctuations of the number of replication cycles (i.e., the actual population size over time) and random sampling effects (5) in the field during the observed time span, in contrast to the mutation rate (i.e., number of base substitutions

per site per replication), which is determined under clear lab conditions. Therefore, the evolution rate reflects only cumulated changes of viral sequences year by year in the field. Even if noncoding changes were counted as silent mutations and not approximately as changes of the third position of each codon independent of whether an amino acid replacement occurred there (as most software does), so that fixation of noncoding changes by linked coding changes and selection would be excluded, such a resulting rate would not be a clear correlate to the mutation rate. In light of our results and these considerations, we suggest that, in addition to Darwinian selection, random sampling effects (5) and an increased number of replication cycles (reflecting the effects of factors such as the number of infected animals, the yields of various strains in swine, a greater number of infected target cells in swine than in birds, and the burst size per infected cell) contributed to the high evolutionary rates of European swine influenza H1N1 viruses. Witte et al. discussed that the rapid exhaustive expansion of the avian-like swine influenza viruses from 1979 to 1981 in northern Europe is possibly due to large-scale animal husbandry, in contrast to the few infections observed in Germany in 1957 (35).

The estimated mutation rate has considerable inherent variability because the measured frequency of mutants reflects progeny from one or more mutation events. Further, the specific time point at which a specific mutational event occurs is due to change, and because of the exponential characteristics of virus growth, this contributes greatly to the observed variability among estimated mutation rates (6, 19). To accommodate these factors, we applied a probabilistic model (see Appendix) that enabled us to determine the mutation rate at a given accuracy and to perform a statistical analysis. Our formula for the SD supports the predicted high variability, which was confirmed by the comparison with the ad hoc SD for A/Mallard/New York/6750/78. In addition, we needed to screen several parallel cultures, and we chose amantadine for the indirect determination of the mutation rate because the sites for resistance-inducing mutations are known, the inhibitory effect is complete for both isolates in plaque inhibition assay, and this compound facilitates rapid screening.

In contrast the techniques that we used, Suarez et al. (27) determined mutational rates by using the fluctuation test (19) in the presence of a monoclonal antibody against the HA1 portion of the hemagglutinin. The fluctuation test requires many parallel cultures and therefore is very labor-intensive. Parvin et al. (21) sequenced the NS genes of 108 clones that were derived from a single well-isolated plaque of A/WSN/33. This labor-intensive approach led to the most exact determination of the mutant spectrum, but, as those authors stated, because the mutation rate was evaluated from only one plaque, no discussion of the variability of the estimate of the mutation rate was possible. Table 3 shows that our data are in good accordance with the values for the mutation rate of influenza virus from both groups despite the use of three different virus genes from four different isolates and three different methods. This comparison indicates that differences in mutation rate cannot account for the evolutionary rates of various virus genes. The similarity of the mutation rates for the highly variable HA1 region and the highly conserved ion channel of the M2 gene again emphasizes that a high mutation rate cannot directly account for a high rate of evolution. Therefore, another explanation remains possible, regarding the potential role of a mutator mutation in the early evolution of the northern European swine influenza virus. For a mutator mutation to have contributed to this process, it must have emerged prior to 1981, possibly in 1979 immediately after transmission of the

virus to the new host or even in the avian host before transmission to swine. Simulation studies for finite clonal asexual populations under directional selection (30) have suggested that mutator mutants increase in frequency if they are abundant in the population and associated with a favorable mutation which prevents their extinction (hitchhiking effect) but that they are kept at low frequency when their primary number is low. A mutator mutant could become dominant during a bottleneck event that dramatically reduces the population size; such bottlenecks are probably frequent in the life cycles of RNA viruses because of their airborne or droplet transmission (5). In such situations, an otherwise rare mutator mutant might accelerate adaptation for a brief transitional period until reverting back. To revert easily, the mutator mutation should comprise only a few (at most) point mutations if located in only one segment or should be disrupted by reassortment. For established strains at an equilibrium of mutation and selection, the mutation rate is close to Eigens error threshold (7) (Table 3).

An alternative mechanism for crossing the species barrier may exist in the absence of a mutator mutation. Perhaps the avian virus undergoes several mutations (at the normal mutation rate) in the primary host that are necessary for successful establishment in the new host. The product of the theoretical minimal number of replication cycles (see e.g., Tables 2 and 3) of all necessary mutations would have to be less than the possible finite population size in the avian host. This scenario is likely to apply to mutations that are detrimental or even deadly in avian hosts, whereas changes that also accommodate functional constraints of the new host often occur in multiple avian hosts before the transmission of the virus. From the purpose of quasispecies, to suppress a newly arisen, even superior mutant progeny (4), the new host should be free of another already well-established strain.

An example of the alternate mechanism might be the recent emergence of an H5N1 influenza virus in humans (3). The antecedent strain is a highly pathogenic H5N1 chicken isolate (29). Virulent avian H5 isolates (15) and the new human H5N1 isolate (29) are pantropic in chickens. Such viruses have greater finite population sizes in each chicken than do apathogenic isolates. Therefore, the mutations required for establishment in humans might occur in chickens without the need for an intermediate host (e.g., swine). Both mechanisms (mutator mutation and the alternate mechanism) could coexist in nature, and the one that is used would depend on the number of required adaptive mutations, i.e., on the ratio of the frequency of a mutator mutant to that of an already adapted mutant within the viral population in the avian host. If only a few adaptive mutations were necessary, then a large population size would be sufficient; if multiple changes were required, then a mutator phenotype might be necessary.

Suarez et al. (27) isolated variants of A/Victoria/3/75 whose mutation rates were increased three- to fourfold and estimated that 13% of the viral population had similarly elevated mutation rates. Notwithstanding the ever-present problem of obtaining cultures that have very high mutant frequencies because the mutational events occurred early during the growth of population (thereby mimicking mutator mutants), even those reported values seem to indicate minimal acceleration on adaptation according to simulation studies. According to both Coffin (2) and Taddei et al. (30), the mutation rate should have been at least 10-fold higher. However, Suarez et al. (27) elegantly showed that the mutation rate is a weighted average over the population, and their finding does not exclude the possibility that mutator mutants (with their associated high mutation rates) might have occurred at very low frequencies.

Whether such variants actually contributed to the stable host change seen in 1979 is unknown; however, in light of the data that are available (from 1979 to 1985), changes in mutation rate cannot account for the early evolution of northern European swine influenza viruses. Whether the mutation rates fluctuated during growth of the initial avian influenza H1N1 virus in the earliest-infected pigs or in the avian host and whether those fluctuations in mutation rates were a prerequisite for successful establishment remain to be investigated further.

Ludwig et al. (18) demonstrated that all northern European swine isolates prior to 1985 had a highly heterogeneous plaque morphology, which persisted even after several plaque passages and which was taken as evidence of relatively high genetic instability. Because an increased mutational rate can now be excluded as the basis of this observation, we have to find another explanation, and the one we propose may also be a possible mechanism by which an avian influenza virus can pass the species barrier into swine. Relatively stable partial heterozygotes can be formed after double infection with two different influenza A viruses (25) or by introduction of an additional segment by reverse genetics (8) if bidirectional selection maintains both allelic segments so that they can complement each other's weaknesses. Suppose an influenza virus gains a mutation that enables it to multiply better in swine but that is detrimental for replication to high titers. Such a virus might be able to cross the species barrier if it is partially heterozygous with respect to this gene (RNA segment). Once in the new species, selection pressure for this partial heterozygote persists until suppressor mutations overcome the otherwise detrimental mutation, and one segment gets lost; this process might take several years. Such partial heterozygotes might also be responsible for the observed plaque heterogeneity and for higher frequencies of monoclonal antibody escape mutants, too, if the required mutations are located in the same segment, e.g., the hemagglutinin gene. This alternate mechanism would allow an influenza virus to cross the species barrier to form a stable lineage and would account for the observation described by Ludwig et al. (18).

ACKNOWLEDGMENTS

We thank Krisna Wells and Scott Krauss for excellent advice and expert technical assistance. We thank Amy L. B. Frazier for critical comments and editing the manuscript.

This work was supported in part by Public Health Service research grants AI08831 and AI29680 and by the American Lebanese Syrian Associated Charities (ALSAC).

APPENDIX

The number of virions in culture during maturation is known to grow exponentially (9, 34) and can be described as $N = e^{ct}$. From that exponential function, one can derive an equation, $N = e^{ct} = 2^g$, reflecting an ideal population grown as a binary tree, where each virion gives rise to exactly two progenies of the next generation and all viruses at a given time point t belong to the same generation g . Such a binary tree allows us to consider the number of unmutated viruses in the virus multiplication as a branching process. Using the branching process (17), we have derived estimates of the mutation rate and its variance by counts of mutated and unmutated viruses at a given generation which is \log_2 of the size of such an ideal population. The resulting formula for the mutation rate is similar to the formula described by Drake (6), but the possibility of calculating a variance allows statistical testing. The model agrees with our understanding that during the first generations, mutation to amantadine resistance is very improbable because the number of loci involved (4 loci) is very small in relation to the genome length (~14,000 bases), so that in our approach the mutation frequency is determined many generations after clonal infection during the maturation phase, e.g., the above-mentioned exponential growth,

and not during the eclipse phase. Therefore, the contribution of mutants arising during the eclipse phase to the mutant frequency is negligible.

Notation

m = mutation rate (probability of mutation per nucleotide per copy).

m_{vir} = mutation rate of virus, where the mutants can be defined according to a given criterion. For example, a mutant is defined as a virus that has at least one mutated nucleotide in given sites of the nucleotide sequence.

M = number of mutants in the culture.

N = population size of the culture.

f = frequency of mutants in the population, defined as M/N .

k = number of sites in the sequence of nucleotides where mutation to amantadine resistance is possible.

Estimation of mutation rate. We are interested in m , the mutation rate (probability of mutation per nucleotide per replication), but it cannot be directly estimated. Therefore, we first estimate m_{vir} from M and N and then derive the estimate of m from that of m_{vir} by using the relationship between m and m_{vir} .

(i) **One culture.** Let N be the number of viruses in a culture of virus, and let M be the number of mutants. Then the estimate of m_{vir} is

$$\hat{m}_{vir} = \frac{2f}{\log_2 N} = \frac{2M}{N \log_2 N} \tag{A1}$$

which was derived by using the theory of branching processes (17). The variance of the estimate is

$$V_{\hat{m}_{vir}} = \frac{2m_{vir}}{(\log_2 N)^2} \tag{A2}$$

In the experiment using amantadine, N is the number of viruses in the culture in the absence of amantadine and M is the number of survivors in the presence of amantadine in the culture. In the presence of amantadine, the viruses survive or die according to whether there is any mutation at the k specific sites in the nucleotide sequence of the virus.

When m_{vir} and m are close to 0 and k is not large (e.g., when $m = 10^{-5}$ and $k \leq 100$), then $m \approx m_{vir}/k$. From this relation and equations A1 and A2, the estimate of m is

$$\hat{m} = \frac{\hat{m}_{vir}}{k} = \frac{2M}{kN \log_2 N} \tag{A3}$$

and the variance of the estimate is

$$V_m = \frac{2m}{k(\log_2 N)^2} \tag{A4}$$

In the discussion above, \hat{m}_{vir} , \hat{m} , and M are random variables, while m_{vir} , m , and N are constants. From equation A4, the SD for the estimate of m is

$$SD_m = \frac{\sqrt{2m}}{\log_2 N} \tag{A5}$$

which is much larger than m when m is close to 0. For example, let $m = 10^{-5}$ and $N = 10^8$, which are usual in virus cultures. Then, from equation A5, we have $SD = 8.4 \times 10^{-5}$, which implies that the SD of the estimate is 8.4-fold the rate m . An estimate with such a large variability cannot be used for hypothesis testing.

(ii) **Multiple cultures.** To compare the mutation rates of two types of viruses, we must use multiple cultures of viruses to reduce the errors of estimates. Denote C = total number of parallel cultures, M_i = number of mutants in the i th culture, and N_i = population size in the i th culture, where $i = 1, \dots, C$. We have two types of estimates; one is a simple average of estimates from each culture, and the other is a weighted average of those estimates.

(a) **Simple average.** By taking the average of estimates from equation A3 for each culture, we have the estimate of m :

$$\hat{m}_a = \frac{2}{kC} \sum_{i=1}^C \frac{M_i}{N_i \log_2 N_i} \tag{A6}$$

By equations A3, A4, and A6, the variance of m_a is

$$V_{\hat{m}_a} = \frac{2m}{kC^2} \sum_{i=1}^C \frac{1}{(\log_2 N_i)^2} \tag{A7}$$

Since m in equation A7 is unknown, the variance $V_{\hat{m}_a}$ must be estimated by using equation A7 and by replacing m by \hat{m}_a from equation A6.

(b) **Weighted average.** If there are large differences among the N_i s, simple averaging is not a good way to combine estimates from each culture. Obviously from equation A4, the estimate in equation A3 from a culture with a larger population has smaller variance than the one from a culture with a smaller population. Thus, the former should be weighted more than the latter. With optimal weighting, the estimate is derived as

$$\hat{m}_w = \frac{2}{kT} \sum_{i=1}^C \frac{M_i \log_2 N_i}{N_i} \tag{A8}$$

where

$$T = \sum_{i=1}^C (\log_2 N_i)^2.$$

The variance of \hat{m}_w is

$$V_{\hat{m}_w} = \frac{2m}{kT} \tag{A9}$$

which was minimized by allocating weights according to the inverse of the individual variance. $V_{\hat{m}_w}$ in equation A9 contains m , which is unknown. Replacing m in equation A9 by its estimate in equation A8, the variance in equation A9 is estimated by

$$\hat{V}_{\hat{m}_w} = \frac{4}{(kT)^2} \sum_{i=1}^C \frac{M_i \log_2 N_i}{N_i} \tag{A10}$$

Comparison of mutation rates. Assuming that there are two types of viruses with mutation rates m_{mall} and m_{sw} , we are interested in testing the hypothesis $H_0(m_{mall} = m_{sw})$ versus $H_a(m_{mall} \neq m_{sw})$ for a fixed α level (e.g., 0.001, 0.05, or 0.1). Let \hat{m}_{mall} and \hat{m}_{sw} be the estimates of m_{mall} and m_{sw} , respectively. The estimates of mutation rates, \hat{m}_{mall} and \hat{m}_{sw} can be computed either by the simple average in equation A6 or by the weighted average in equation A8. Accordingly, the estimated variances \hat{V}_{mall} and \hat{V}_{sw} of the estimates are computed either by equation A7 or by equation A10. Using the normal approximation, we formulate the test statistic as

$$Z = \frac{\hat{m}_{mall} - \hat{m}_{sw}}{\sqrt{\hat{V}_{mall} + \hat{V}_{sw}}}$$

H_0 is rejected if $|Z| > Z_{\alpha/2}$, where Z_{α} is the upper α quantile of the normal distribution (e.g., $Z_{0.025} = 1.96$).

REFERENCES

1. **Bean, W. J., St. C. Threlkeld, and R. G. Webster.** 1989. Biologic potential of amantadine-resistant influenza A virus in an avian model. *J. Infect. Dis.* **159**: 1050–1056.
2. **Coffin, J. M.** 1995. HIV population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy. *Science* **267**:483–489.
3. **de Jong, J. C., A. D. M. E. Osterhaus, R. G. Webster, and W. L. Lim.** 1997. A pandemic warning? *Nature* **389**:554.
4. **de la Torre, J. C., and J. J. Holland.** 1990. RNA virus quasispecies population suppresses vastly superior mutant progeny. *J. Virol.* **64**:6278–6281.
5. **Domingo, E., C. Escarmis, N. Sevilla, A. Moya, S. F. Elena, J. Quer, I. S. Novella, and J. J. Holland.** 1996. Basic concepts in RNA virus evolution. *FASEB J.* **10**:859–864.

6. Drake, J. W. 1970. The molecular basis of mutation. Holden-Day, San Francisco, Calif.
7. Eigen, M., and P. Schuster. 1977. The hypercycle A principle of self organization part A: emergence of hypercycle. *Naturwissenschaften* **64**:541–565.
8. Enami, M., G. Sharma, C. Benham, and P. Palese. 1991. An influenza virus containing nine different RNA segments. *Virology* **185**:291–298.
9. Fields, B. N., and D. M. Knipe. 1990. *Virology*, 2nd ed. Raven Press, New York, N.Y.
10. Fitch, W. M., R. M. Bush, C. A. Bender, and N. J. Cox. 1997. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. USA* **94**:7712–7718.
11. Gammel, M., A. Altmüller, U. Reinhardt, J. Mandler, V. R. Harley, P. J. Hudson, W. M. Fitch, and C. Scholtissek. 1990. Phylogenetic analysis of nucleoproteins suggests that human influenza A viruses emerged from a 19th-century avian ancestor. *Mol. Biol. Evol.* **7**:194–200.
12. Gorman, O. T., W. J. Bean, and R. G. Webster. 1992. Evolutionary processes in influenza viruses: divergence, rapid evolution, and stasis. *Curr. Top. Microbiol. Immunol.* **176**:75–97.
13. Hay, A. J., A. J. Wolstenholme, J. J. Skehel, and M. H. Smith. 1985. The molecular basis of the specific anti-influenza action of amantadine. *EMBO J.* **4**:3021–3024.
14. Ho, P. P. K., A. L. Young, and M. Truehaft. 1976. Plaque formation with influenza viruses in dog kidney cells. *J. Gen. Virol.* **33**:143–145.
15. Horimoto, T., and Y. Kawaoka. 1994. Reverse genetics provides direct evidence for a correlation of hemagglutinin cleavability and virulence of an avian influenza A virus. *J. Virol.* **68**:3120–3128.
16. Ito, T., O. T. Gorman, Y. Kawaoka, W. J. Bean, and R. G. Webster. 1991. Evolutionary analysis of the influenza virus M gene with comparison of the M1 and M2 proteins. *J. Virol.* **65**:5491–5498.
17. Karlin, S., and H. Taylor. 1975. *A first course in stochastic processes*, 2nd ed. Academic Press, New York, N.Y.
18. Ludwig, S., L. Stütz, O. Planz, H. Van, W. M. Fitch, and C. Scholtissek. 1995. European swine virus as a possible source for the next influenza pandemic? *Virology* **212**:555–561.
19. Luria, S. E., and M. Delbrück. 1943. Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* **28**:491–511.
20. Palese, P., and J. F. Young. 1982. Variation of influenza A, B, and C viruses. *Science* **215**:1468–1474.
21. Parvin, J. D., A. Moscona, W. T. Pan, J. M. Leider, and P. Palese. 1986. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. *J. Virol.* **59**:377–383.
22. Pensaert, M., K. Ottis, J. Vanderputte, M. M. Kaplan, and P. A. Bachmann. 1981. Evidence for the natural transmission of influenza A virus from wild duck to swine and its potential importance for man. *Bull. W. H. O.* **59**:75–78.
23. Ribeiro, R. M., S. Bonhoeffer, and M. A. Nowak. 1998. The frequency of resistant mutant virus before antiviral therapy. *AIDS* **12**:461–465.
24. Scholtissek, C., S. Ludwig, and W. M. Fitch. 1993. Analysis of influenza A virus nucleoproteins for the assessment of molecular genetic mechanisms leading to new phylogenetic virus lineages. *Arch. Virol.* **131**:237–250.
25. Scholtissek, C., W. Rohde, E. Harms, R. Rott, M. Orlich, and C. B. Boschek. 1978. A possible partial heterozygote of an influenza A virus. *Virology* **89**:506–516.
26. Schultz, U., W. M. Fitch, S. Ludwig, J. Mandler, and C. Scholtissek. 1991. Evolution of pig influenza viruses. *Virology* **183**:61–73.
27. Suarez, P., J. Valcarcel, and J. Ortin. 1992. Heterogeneity of the mutation rates of influenza A viruses: isolation of mutator mutants. *J. Virol.* **66**:2491–2494.
28. Suarez-Lopez, P., and J. Ortin. 1994. An estimation of the nucleotide substitution rate at defined positions in the influenza haemagglutinin gene. *J. Gen. Virol.* **75**:389–393.
29. Subbarao, K., A. Klimov, J. Katz, H. Regener, W. Lim, H. Hall, M. Perdue, D. Swayne, C. Bender, J. Huang, M. Hamphill, T. Rowe, M. Shaw, X. Xu, K. Fukuda, and N. Cox. 1998. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science* **279**:393–396.
30. Taddei, F., M. Radman, J. M. Maynard-Smith, B. Toupance, P. H. Gouyon, and B. Godelle. 1997. Role of mutator alleles in adaptive evolution. *Nature* **387**:700–702.
31. Taubenberger, J. K., A. H. Heid, A. E. Krafft, K. E. Bijwaard, and T. G. Fanning. 1997. Initial genetic characterization of the 1918 Spanish influenza virus. *Science* **275**:1793–1796.
32. Topal, M. D., and J. R. Fresco. 1976. Complementary base pairing and the origin of substitution mutations. *Nature* **263**:285–289.
33. Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka. 1992. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**:152–179.
34. White, D. O. 1960. Some aspects of influenza virus multiplication in the surviving allantois-on-shell system. *Virology* **10**:21–28.
35. Witte, K. H., E. Niendorf, H. Ernst, U. Schmidt, and D. Prager. 1981. Erstmaliges Auftreten einer durch das Schweineinfluenzavirus verursachten Epizootie in Schweinebeständen der Bundesrepublik Deutschland. *Tierärztl. Umschau* **36**:591–606.