Proc. Natl. Acad. Sci. USA
Vol. 96, pp. 1164–1166, February 1999

Commentary

1918 Spanish influenza: The secrets remain elusive

Robert G. Webster

Department of Virology and Molecular Biology, St. Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105

The worst pandemic of influenza ever recorded was caused by the 1918 Spanish influenza virus. Emotional reports of fit and healthy soldiers falling down on parade and dying the same or the next day are recorded (1). An initial mild wave of infection occurred in the spring of 1918 and the lethal wave appeared throughout the world, in the fall. At least 20 million persons died worldwide and probably 10 times that number were affected. The origin of this virus and the basis of high pathogenicity have remained elusive for the virus is not available for study.

Interest in the causative agent of the 1918 pandemic is not new. Because no viruses from the period are available, serological epidemiology has been used to trace the virus. The first childhood infection with an influenza virus leaves an indelible immunological imprint, aptly termed "original antigen sin" (2). Subsequent exposure to an antigenically related variant induces an anamnestic response to the original virus and a primary response to the variant. Sero-archeological studies of samples from humans born in the time period from 1918 to 1920 have shown that the original sinner was a virus closely related to A/Swine/Iowa/30 (H1N1) (3–5). With the advent of reverse transcriptase–PCR technology, archivally stored tissues now can be examined for the footprints of the 1918 influenza virus. Archived material comes in two forms: formalin-fixed, paraffin-embedded tissue samples from infected humans, or tissues from infected persons who were buried in the permafrost. In this issue of the *Proceedings,* Reid *et al.* (6) report the complete sequence of the hemagglutinin (HA) gene of the 1918 influenza virus isolated from samples taken from three people who died of acute influenza in September and November 1918. Two samples were from paraffin blocks containing lung tissue from young soldiers, which had been stored at the Armed Forces Institute of Pathology, and the third was from formalin-fixed frozen lung tissue obtained by an *in situ* biopsy from an Inuit woman buried in the permafrost at Brevig Mission, AK. This report confirms the HA findings in the earlier report from this group (7) on the partial sequence of the HA from a single sample of preserved lung tissue taken from a soldier who died of acute influenza in 1918.

So what new information does this second report provide? The complete nucleotide sequence of one HA gene and of the HA1 portion of two more genes provides new information about the receptor binding site (RBS), the antigenic sites, and the extent of glycosylation, and also confirms the fact that the cleavage site of the HA does not contain polybasic amino acids typically found in highly pathogenic avian H5 and H7 influenza viruses (8). The fact that only two nucleotide differences (one coding for an amino acid difference in the RBS) were found between these viruses originating from different geographical areas suggests the circulation of genetically homogenous viruses, at least in different parts of North America. The results of the second study also reconfirm those of the earlier report (7), which demonstrated a close genetic relationship between the 1918 virus and A/Swine/Iowa/30 (H1N1). However, analysis of the HA sequence information raises more questions than it answers about the origin of this pandemic strain. Phylogenetic analysis indicates that the 1918 HAs are mammalian-like and are located near the root of the human and swine clades, whereas analysis of the antigenic and receptor binding sites indicates that the virus resembles the avian consensus. The pattern of glycosylation is similar to that found in both avian and swine influenza viruses. Additional complexity is added by the postulation that the H1N1 viruses were introduced into humans sometime between 1900 and 1915.

Let us begin by considering the possible dates of introduction into humans. Analysis of the HA places this date at about 1905 (9, 10), and analysis of the nucleoprotein (NP) genes from human and swine H1N1 viruses places it at 1900 (11). Phylogenetic analysis of the HA sequences of the human 1918 isolates, performed by plotting the total number of amino acid changes from node to node against the year of isolation, suggests that the ancestor of the 1918 virus entered humans around 1900 (6), whereas tracing the total number of nucleotide changes from node to node places this date at around 1915 (6). Support for the later date is provided by the analysis of NP nucleotide sequences by Gorman *et al.* (12), which places the date at around 1912 to 1913. Taken together, these findings suggest that the 1918 pandemic strain entered humans a number of years before 1918.

When influenza virus is introduced into a new host, it evolves rapidly, and the proportion of nucleotide changes that result in amino acid changes is usually highest in the HA gene. The emergence of highly pathogenic H5N2 influenza viruses in chickens in Mexico was accompanied by a large proportion of amino acid changes in the HA1 protein (57.3%) and a lower proportion of changes in the products of the internal genes (13). This rapid evolution also was seen in 1979 when the H1N1 avian-like viruses spread into European swine (11). The NP genes of these new viruses evolved at a higher rate than the NP genes in human and classic swine viruses over the period between 1930 and 1988. When the H2N2 virus was introduced into humans in 1957, causing the Asian pandemic, the HA molecule changed rapidly (14). The earliest stages in the evolution of the human lineage appear to have been under greater selective pressures than the later branches, as judged by the ratio of coding to noncoding changes. Initially, only 1.6 nucleotide changes were required to cause a change in an amino acid; later, 3.7 nucleotide changes were required per amino acid change. Rapid evolutionary changes were not detected in the HA of H3N2 viruses in 1968, perhaps because we do not know the date at which the avian H3 virus was transmitted to humans. Serological studies suggest that the H3 virus was introduced into humans several years before the virus was isolated in humans in 1968 (15).

These observations have two consequences for interpretation of the work of Reid *et al.* (6) that is under discussion. First, a higher evolutionary rate would influence the proposed date of introduction of the precursor 1918 virus into humans; Gorman *et al.* (12) estimate this date as 1918. Second, if the precursor viruses had been present in humans since 1900 or even since 1915, antigenic drift in the antigenic domains and in the RBS would be expected. Additional carbohydrate

PNAS is available online at www.pnas.org.

The companion to this Commentary begins on page 1651.

residues on the HA also might be expected. It is difficult to correlate the presence of avian-like antigenic sites and RBSs with an early introduction. Regardless, epidemiological records support the notion of a mild first wave of influenza in humans in the spring and summer of 1918, which provided protection from the highly pathogenic strain that appeared in August 1918 (16).

The H5N1 influenza viruses that infected 18 humans and killed six in Hong Kong in 1997 (17–19) had been isolated 2 months earlier from chickens with lethal influenza (20). This fact provides direct evidence that avian viruses can be transmitted directly to humans. There was no convincing evidence of human-to-human transmission; instead, the available evidence indicated that each case had been caused by independent transmission to humans of virus from birds in the poultry markets. The slaughter of 1.6 million domestic birds in Hong Kong in December 1997 removed the source of virus and interrupted adaptation of the virus to humans.

The receptor specificity of avian influenza viruses is unlike that of human strains; avian viruses bind preferentially to terminal SAα2,3 galactose determinants, whereas humans strains preferentially bind to SAα2,6 galactose terminal sequences (21, 22). The direct transmission of H5N1 viruses from chickens to humans in Hong Kong indicates that receptor specificity is not a definitive host restriction factor and that an intermediate host (23) is not necessarily required for the first stage of transmission to humans. Thus, the RBS data from the 1918 HAs are compatible with the recent transmission to humans of an avian influenza virus.

The suggestion that the 1918 precursor virus was transmitted directly from humans to pigs in 1918 has been supported by events occurring in later human pandemics. The causative virus of the H3N2 human pandemic of 1968 was transmitted to pigs soon after it appeared in humans (24), and descendants of that virus continue to circulate in European pigs. Epidemiologic evidence supports a severe outbreak of swine influenza in the fall of 1918 (25); however, the presence of mild disease before 1918 probably would have gone undetected, as is currently the case with H1N1 in U.S. pigs. Thus, these observations do not rule out the possibility that a mild form of H1N1 may have circulated in pigs before the fall of 1918.

Let us return to an examination of the origin of the 1918 pandemic strain on the basis of the evidence from the HA. Reid *et al.* (6) conclude from the phylogenetic analysis that the "hemagglutinin gene, although more closely related to avian strains than any other mammalian sequence, is mammalian, and may have already been adapting in humans before 1918." One difficulty in supporting this conclusion, as the authors acknowledge, is that no pre-1975 avian H1 influenza viruses are available for study. It can be argued that avian influenza viruses in their natural aquatic bird reservoirs are in evolutionary stasis (26), but this argument applies only to amino acid changes, not to nucleotide conservation. Within each of the avian subtypes, there is also considerable genetic diversity. This reviewer prefers the option (offered by Reid *et al*.) that an avian H1N1 influenza virus of 1918 vintage entered the human population (as the H5N1 strains did in Hong Kong) in early 1918 and developed its pathogenic potential in humans. This conclusion would better fit with Reid *et al*.'s observations of avian-like antigenic domains and RBSs. Further information about this possibility should become available when Reid *et al*. obtain the sequence of the gene(s) encoding the NP and other avian "internal genes," because earlier fowl plague viruses (FPV) are available, including FPV/Brescia–Ascoli/02, FPV/Dutch/27, and FPV/Rostock/34.

The available information suggests that the natural reservoirs of influenza A viruses are the aquatic birds of the world (27), in which each of the 15 known HA subtypes and nine known neuraminidase subtypes of influenza A viruses are perpetuated. Of these, only H1, H2, and H3 are known to have caused pandemics in humans. Both the H2 and H3 HA genes, which caused the 1957 Asian and 1968 Hong Kong pandemics, were closely related to avian counterparts (14, 28). The evidence presented by Reid *et al*. (6) does not rule out the possibility that the precursor of the 1918 H1N1 virus also originally emerged from this avian reservoir.

As implied in the title of this commentary, many of the biological properties of the 1918 pandemic virus remain to be revealed. Disease resulting from influenza virus infection is a complex condition that involves both viral and host gene products. The viral genes associated with pathogenicity vary from virus to virus, leading to the concept that pathogenicity is polygenic and depends on an optimal gene constellation (29, 30). The absence of polybasic amino acids at the cleavage site of the HA of the three 1918 samples examined by Reid *et al.* (6) precludes the contribution of this molecular region to the pathogenicity of the viruses, but what other regions of the HA are involved? Will the neuraminidase have the sequence required for binding plasminogen and sequestering it for plasmin activation of the HA, an activity that Goto and Kawaoka (31) recently showed to be responsible for the pathogenicity of A/WSN/33(H1N1)? Will specific domains on the NP, PB2, and M2 proteins be so unique that they can be associated with pathogenicity? Because we do not know the domains of these proteins that are associated with the host range and pathogenicity of the influenza viruses available for study, we must consider the possibility that biological properties determined by multiple gene segments may not be resolved. Thus, the entire gene sequence is unlikely to reveal the secrets of the high pathogenicity of the 1918 Spanish virus.

Regardless, the search must go on. It has been a tour de force to achieve the entire genome sequence of the HA of one virus and the HA1 sequence of two other H1N1 viruses from 1918. The entire genome sequence will allow us to design prospective vaccines based on the HA, NA, and NP sequences, and to determine whether the currently available anti-influenza drugs that target the M2 and NA would be efficacious if such viruses re-emerge. To answer some of these remaining questions, we need to find both the mild precursor virus that was circulating in the spring of 1918 and also avian and swine influenza viruses from the same time period. It is encouraging that influenza RNA from bodies buried in permafrost is providing so much sequence information. It is hoped that future studies of avian feces and animal tissues from the permafrost will help unlock the secrets of the Spanish influenza virus.

I thank Enid Bedford for preparation of this manuscript and Flo Witte for editorial assistance. The writing of this article was supported by Public Health Research Grant AI29680 and Cancer Center Support (CORE) Grant CA-21765 from the National Institutes of Health and by the American Lebanese Syrian Associated Charities.

1. Crosby, A. (1989) *America's Forgotten Pandemic* (Cambridge Univ. Press, Cambridge, U.K.).
2. Francis, T., Jr. (1953) *Ann. Int. Med.* **39**, 203–221.
3. Shope, R. E. (1936) *J. Exp. Med.* **63,** 669–684.
4. Davenport, F. M., Hennessy, A. V. & Francis, T., Jr. (1953) *J. Exp. Med.* **98,** 641–656.
5. Masurel, N. (1976) *Lancet* **2,** 244–247.
6. Reid, A. H., Fanning, T. G., Hultin, J. V. & Taubenberger, J. K. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 1651–1656.
7. Taubenberger, J. K., Reid, A. H., Krafft, A. E., Bijwaard, K. E. & Fanning, T. G. (1997) *Science* **275,** 1793–1796.
8. Klenk, H. D. & Rott, R. (1988) *Adv. Virus Res.* **34,** 247–281.
9. Kanegae, Y., Sugita, S., Shortridge, K. F., Yoshioka, Y. & Nerome, K. (1994) *Arch. Virol.* **134,** 17–28.
10. Sugita, S., Yoshioka, K., Itamura, S., Kanegae, Y., Oguchi, K., Gojobori, T., Nerome, K. & Oya, A. (1991) *J. Mol. Evol.* **32,** 16–23.

11. Scholtissek, C., Lugwig, S. & Fitch, W. M. (1993) *Arch. Virol.* **131,** 237–250.
12. Gorman, O. T., Bean, W. J., Kawaoka, Y., Donatelli, I., Guo, Y. J. & Webster, R. G. (1991) *J. Gen. Virol.* **65,** 3704–3714.
13. Garcia, M., Crawford, J. M., Latimer, J. W., Rivera-Cruz, E. & Perdue, M. L. (1996) *J. Gen. Virol.* **77,** 1493–1504.
14. Schafer, J. R., Kawaoka, Y., Bean, W. J., Suss, J., Senne, D. & Webster, R. G. (1993) *Virology* **194,** 781–788.
15. Monto, A. S. & Maassab, H. F. (1981) *Am. J. Epidemiol* **113,** 236–244.
16. Shope, R. (1958) *Public Health Rep.* **73,** 165–178.
17. De Jong, J., C., Claas, E. C., Osterhaus, A. D., Webster, R. G. & Lim, W. L. (1997) *Nature (London)* **389,** 554.
18. Subbarao, K., Klimov, A., Katz, J., Regnery, H., Lim, W., Hall, H., Perdue, M., Swayne, D., Bender, C., Huang, J., *et al*. (1998) *Science* **279,** 393–396.
19. Yuen, K. Y., Chan, P. K. S., Peiris, M., Tsang, D. N. C., Que, T. L., Shortridge, K. F., Cheung, P. T., To, W. K., Ho, E. T. F., Sung, R., *et al*. (1998) *Lancet* **351,** 467–471.
20. Claas, E. C. J., Osterhaus, A. D., Van Beek, R., De Jong, J. C., Rimmelzwaan, G. F., Senne, D. A., Krauss, S., Shortridge, K. F. & Webster, R. G. (1998) *Lancet* **351,** 472–477.
21. Matrosovich, M. N., Gambaryan, A. S., Teneberg, S., Piskarev, V. E., Yaminikova, S. S., Lvov, D. K., Robertson, J. S. & Karlsson, K. A. (1997) *Virology* **233,** 224–234.
22. Paulson, J. C. (1985) in *The Receptors,* ed. Conn, P. M. (Academic, Orlando, FL), Vol. 2, pp. 131–219.
23. Scholtissek, C., Schultz, U., Ludwig, S. & Fitch, W. M. (1993) in *Options for the Control of Influenza II: Proceedings of the International Conference on Options of the Control of Influenza*, eds. Hannoun, C., Kendal, A. P., Klenk, H. D. & Ruben, F. L. (Excerpta Medica, Amsterdam), pp. 193–201.
24. Kundin, W. D. (1970) *Nature (London)* **228,** 857.
25. Koen, J. S. (1919) *Am. J. Vet. Med.* **14,** 468–470.
26. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. (1992) *Microbiol. Rev.* **56,** 152–179.
27. Hinshaw, V. S. & Webster, R. G. (1982) in *Basic and Applied Influenza Research*, ed. Beare, A. S. (CRC, Boca Raton, FL), pp. 79–104.
28. Bean, W. J., Schell, M., Katz, J., Kawaoka, Y., Naeve, C., Gorman, O. & Webster, R. G. (1992) *J. Virol.* **66,** 1129–1138.
29. Rott, R., Orlich, M. & Scholtissek, C. (1976) *J. Virol.* **19,** 54–60.
30. Rott, R., Orlich, M. & Scholtissek, C. (1979) *J. Gen. Virol.* **44,** 471–477.
31. Goto, H. & Kawaoka, Y. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 10224–10228.