

- line CDR1 or CDR2 loops (or both) interact with this peptide residue.
35. P. A. Reay, R. M. Kantor, M. M. Davis, *J. Immunol.* **152**, 3946 (1994).
 36. G. J. Kersh and P. M. Allen, *J. Exp. Med.* **184**, 1259 (1996).
 37. K. Matsui, J. J. Boniface, P. Steffner, P. A. Reay, M. M. Davis, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12862 (1994); D. S. Lyons *et al.*, *Immunity* **5**, 53 (1996); G. J. Kersh, E. N. Kersh, D. H. Fremont, P. M. Allen, *Immunity* **9**, 817 (1998).
 38. P. M. Allen *et al.*, *Nature* **327**, 713 (1987).
 39. B. D. Evavold, J. Sloan-Lancaster, K. J. Wilson, J. B. Rothbard, P. M. Allen, *Immunity* **2**, 655 (1995).
 40. K. J. Smith, J. Pyrdol, L. Gauthier, D. C. Wiley, K. W. Wucherpfennig, *J. Exp. Med.* **188**, 1511 (1998).
 41. K. W. Wucherpfennig and J. L. Strominger, *Cell* **80**, 695 (1995).
 42. L. A. Sherman and S. Chattopadhyay, *Annu. Rev. Immunol.* **11**, 385 (1993).
 43. J. A. Speir *et al.*, *Immunity* **8**, 553 (1998).
 44. I. F. McKenzie *et al.*, *J. Exp. Med.* **150**, 1323 (1979); J. Bill *et al.*, *J. Exp. Med.* **169**, 115 (1989).
 45. Reviewed in J. A. Frelinger and M. McMillan, *Immunol. Rev.* **154**, 45 (1996).
 46. C. Daniel, S. Horvath, P. M. Allen, *Immunity* **8**, 543 (1998).
 47. M. T. Scherer, L. Ignatowicz, G. M. Winslow, J. W. Kappler, P. Marrack, *Annu. Rev. Immunol.* **9**, 101 (1993); S. R. Webb, N. R. J. Gascoigne, *Curr. Opin. Immunol.* **6**, 467 (1994).
 48. T. S. Jardetzky *et al.*, *Nature* **368**, 711 (1994).
 49. H. Li *et al.*, *Immunity* **9**, 807 (1998).
 50. D. L. Woodland and M. A. Blackman, *Immunol. Today* **14**, 208 (1993); K. Daly *et al.*, *J. Immunol.* **155**, 27 (1995); M. A. Blackman and D. L. Woodland, *Immunol. Res.* **15**, 98 (1996); D. Donson *et al.*, *J. Immunol.* **158**, 5229 (1997); A. M. Deckhut, Y. Chien, M. A. Blackman, D. L. Woodland, *J. Exp. Med.* **180**, 1931 (1994); N. Labrecque, J. Thibodeau, W. Mourad, R. P. Sekaly, *J. Exp. Med.* **180**, 1921 (1994).
 51. E. L. Reinherz, P. C. Kung, G. Goldstein, R. H. Levey, S. F. Schlossman, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1588 (1980); B. J. Fowlkes and D. M. Pardoll, *Adv. Immunol.* **44**, 207 (1989); A. Kruisbeek, *Curr. Opin. Immunol.* **5**, 227 (1993).
 52. S. H. Chan, D. Cosgrove, C. Waltzinger, C. Benoist, D. Mathis, *Cell* **73**, 225 (1993); S. H. Chan, C. Benoist, D. Mathis, *Immunol. Rev.* **135**, 119 (1993); C. B. Davis, N. Kileen, M. E. Casey Crooks, D. Raulet, D. R. Littman, *Cell* **73**, 237 (1993); H. S. Teh *et al.*, *Nature* **335**, 229 (1988); J. P. M. van Meerwijk and R. N. Germain, *Science* **261**, 911 (1993); J. P. M. van Meerwijk, E. M. O'Connell, R. N. Germain, *Semin. Immunol.* **6**, 231 (1994).
 53. H. Kozono, J. White, J. Clements, P. Marrack, J. W. Kappler, *Nature* **369**, 151 (1994).
 54. H.-C. Chang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11408 (1994); J. Liu *et al.*, *J. Biol. Chem.* **271**, 33639 (1996).
 55. T. A. Jones, J.-Y. Zou, S. W. Cowtan, M. Kjeldgaard, *Acta Crystallogr.* **A47**, 110 (1991).
 56. A. Nicholls, K. A. Sharp, B. Honig, *Proteins Struct. Funct. Genet.* **11**, 281 (1991).
 57. Production and purification of scD10 TCR: The scD10 TCR, constructed by polymerase chain reaction, consists of 237 residues and is organized from NH₂- to COOH-terminus as follows: V β 8.2 (residues 3 to 110)-linker (GSADDAKKDAKKDGG)-V α 2 (AV225) (residues 1 to 112) with a Cys²³⁵→Ser mutation. This linker was modified from that previously used for NMR studies (23) because the longer linker failed to give rise to I-A^k cocrystals of diffraction quality. For bacterial expression, we used the T7 promoter expression vector pET-11a. An overnight preculture of transformant (20 ml) was inoculated into 1 liter of fresh Luria broth supplemented with carbenicillin (50 μ g/ml) at 37°C and grown until the optical density at 600 nm reached 0.6. Isopropyl- β -D-thiogalactopyranoside was added to a final concentration of 1 mM, and the culture was incubated for a further 4 hours. Cells were harvested by centrifugation and lysed by sonication. The inclusion bodies were washed and dissolved in 100 mM tris-HCl (pH 8.0) containing 6 M guanidine chloride, 10 mM EDTA, and 10 mM dithiothreitol. An efficient refolding was achieved by diluting rapidly into a refolding buffer [50 mM tris-HCl (pH 8.0), 400 mM arginine, 2 M urea, 2 mM EDTA, 4 mM reduced glutathione, and 0.4 mM oxidized glutathione]. The refolded material was then applied to a 3D3 affinity column [5 mg/ml mAb coupled to Gamma Bind Plus Sepharose (Pharmacia, Uppsala, Sweden)] followed by gel filtration on Superdex 75 and exchanged to crystallization buffer [20 mM sodium acetate (pH 5.0) with 0.025% sodium azide].
 58. Production and purification of MHC class II I-A^k: A 13-residue hen egg CA peptide (residues 134 to 146) that is recognized by D10 TCR was fused (53) to the NH₂-terminus of the mature I-A^k β chain through a flexible linker. The 37-residue leucine zipper (LZ) sequences (54) were attached to both the α and β chains, with ACID-p1 to the β chain and BASE-p1 to the α chain through flexible thrombin-cleavable linkers. The cDNA constructions were subcloned into the pEE14 vector and expressed in Lec3.2.8.1 Chinese hamster ovary (CHO) cells. The screening of secreted recombinant protein in the culture supernatant was performed by both Sandwich enzyme-linked immunosorbent assay and BiAcCore with antibodies specific for I-A^k (10.2.16) and the LZ epitope (2H11 or 13A12). The yield was ~0.7 mg of I-A^k/l supernatant. The production supernatant was applied to 2H11 affinity column [5 mg/ml mAb coupled to Gamma Bind Plus Sepharose (Pharmacia, Uppsala, Sweden)] and I-A^k protein was eluted by 50 mM citrate, 20 mM tris, 0.5 M NaCl, and 10% glycerol (pH 4.0). The eluted protein was then exchanged to 50 mM tris-HCl (pH 8.0) and cleaved by thrombin (2 U per 50 μ g of I-A^k) at 4°C for 4 hours. Thrombin was then removed by passage through benzamide Sepharose 6B beads and gel filtration on Superdex 75. Subsequently, the purified I-A^k was exchanged to 10 mM Hepes (pH 7.0), 0.025% sodium azide for crystallization.
 59. Z. Ortwinowski and W. Minor, *Methods Enzymol.* **276**, 307 (1997).
 60. J. Navaza, *Acta Crystallogr.* **A50**, 157 (1994).
 61. CA/I-A^k was crystallized by the hanging-droplet vapor diffusion method by equilibrating protein solution (7.2 mg/ml) against a buffer of 16% isopropanol, 18% polyethylene glycol (PEG) 4000, 0.1 M sodium citrate (pH 5.6) in reservoir. The crystals belong to space group P2₁2₁2₁ with unit cell parameters $a = 99.1$ Å, $b = 122.4$ Å, and $c = 68.4$ Å. A single crystal was transferred to the cryoprotectant solution of 30% glycerol, 18% PEG 4000, 0.1 M sodium citrate for half an hour before dipping into liquid nitrogen for freezing. A 95% complete data set was collected at Brookhaven National Synchrotron Light Source X12C beamline by using wavelength 1.072 Å with Brandeis charge-coupled device (CCD) detector. The data were integrated and scaled by DENZO and SCALPACK (55). The R_{merge} of the data was 9.2% to 3 Å resolution for 19,854 unique reflections. The structure was solved by molecular replacement method with AMoRe (56), by using a structure of I-A^k complexed with a peptide from hen egg lysozyme residues 50 to 62 [Protein Data Bank (PDB) code 1iak] as a search model. The structure was refined with X-PLOR. The final R_{free} is 29.5%, and R_{work} is 23.1%. Details will be published elsewhere.
 62. A. T. Brünger, Yale Univ. Press, New Haven, CT (1992).
 63. W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
 64. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
 65. Supported by NIH grant AI19807 (E.L.R.), GM56008 (J.-H.W.), AI/CA37581 (G.W.), and U.S. Department of Energy, Office of Health and Environmental Research, contract W-31-109-Eng-38 (A.J. and R.Z.). We thank B. Benacerraf and M. Eck for careful review of the manuscript. We thank C. Janeway for providing the D10 TCR α and β cDNA clones and the 3D3 hybridoma. The coordinates of the D10 TCR-pMHCII complex have been deposited in the PDB (accession number 1D9K).

21 September 1999; accepted 26 October 1999

Predicting the Evolution of Human Influenza A

Robin M. Bush,^{1*} Catherine A. Bender,² Kanta Subbarao,² Nancy J. Cox,² Walter M. Fitch¹

Eighteen codons in the HA1 domain of the hemagglutinin genes of human influenza A subtype H3 appear to be under positive selection to change the amino acid they encode. Retrospective tests show that viral lineages undergoing the greatest number of mutations in the positively selected codons were the progenitors of future H3 lineages in 9 of 11 recent influenza seasons. Codons under positive selection were associated with antibody combining site A or B or the sialic acid receptor binding site. However, not all codons in these sites had predictive value. Monitoring new H3 isolates for additional changes in positively selected codons might help identify the most fit extant viral strains that arise during antigenic drift.

Antigenic drift due to mutations in the hemagglutinin gene necessitates frequent replacement of influenza A strains in the human vaccine. Antibodies against hemagglutinin are a primary determinant of susceptibility to

infection. However, the effects on antigenicity of specific mutations in the hemagglutinin gene are not well understood because multiple amino acid changes often occur together in important antigenic variants. Changes in antigenicity depend not only on the nature and position of the amino acid replacements but also on the amino acids currently encoded at other key positions in the HA1. We have developed a method for predicting the evolution of the virus that makes use of genetic data in the absence of specific knowledge

¹Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA. ²Influenza Branch, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.

*To whom correspondence should be addressed. E-mail: rmbush@uci.edu

about the antigenic properties of the viruses in question.

The fixation rate in the HA1 domain of hemagglutinin (H3 subtype) is high; about 5.7×10^{-3} nucleotide substitution per site per year (1). At least 18 of the 329 H3 HA1 codons have been under positive selection to change in the past (2, 3) (Table 1). Positive selection is defined as a significant excess of nucleotide substitutions that result in amino acid replacements. If the selective pressure were to evade the host immune response, then viruses sustain-

ing mutations at these codons in the past should have been more fit than other coexisting viruses. If this hypothesis is correct, then screening extant strains for additional mutations in these codons might help to identify the most fit viral strains in circulation.

We tested our prediction method retrospectively by determining whether influenza isolates on phylogenetic lineages that underwent the greatest number of changes in the positively selected codons in past influenza seasons were more fit than other isolates. We

define fitness as follows: if one viral strain is more closely related to future lineages than another strain, it is more fit, regardless of the amount or severity of disease either strain causes. Thus, our goal of predicting the course of evolution is not equivalent to predicting the epidemic strain for the next year, nor can it be used to predict new pandemic strains that may emerge from avian or swine reservoirs.

Retrospective Tests

Our prediction method takes advantage of the fact that phylogenetic trees constructed using hemagglutinin genes from human influenza type A viruses show, over time, a single successful lineage, which we call the "trunk lineage." The trunk lineage is the only lineage from which strains in all subsequent years arise (Fig. 1). All trees in this study were constructed from sequences 987 nucleotides long (4) using the tree bisection-reconnection branch-swapping option of the heuristic search option of the maximum parsimony routine of PAUP (5). Each retrospective test data set contains sequences from isolates collected in 1983 through the end of 1 of 11 consecutive influenza seasons (Table 2), defined here as 1 October through 30 September. We illustrate our method with the 1993–1994 test, which uses isolates collected in 1983 through 30 September 1994. The tree representing the evolutionary history of these isolates is shown in Fig. 1A. All trees used in this study were rooted on the ancestor of A/Oita/3/83, the oldest isolate in the data set.

Considerable genetic heterogeneity may occur during a given epidemic season or calendar year. The most recently collected se-

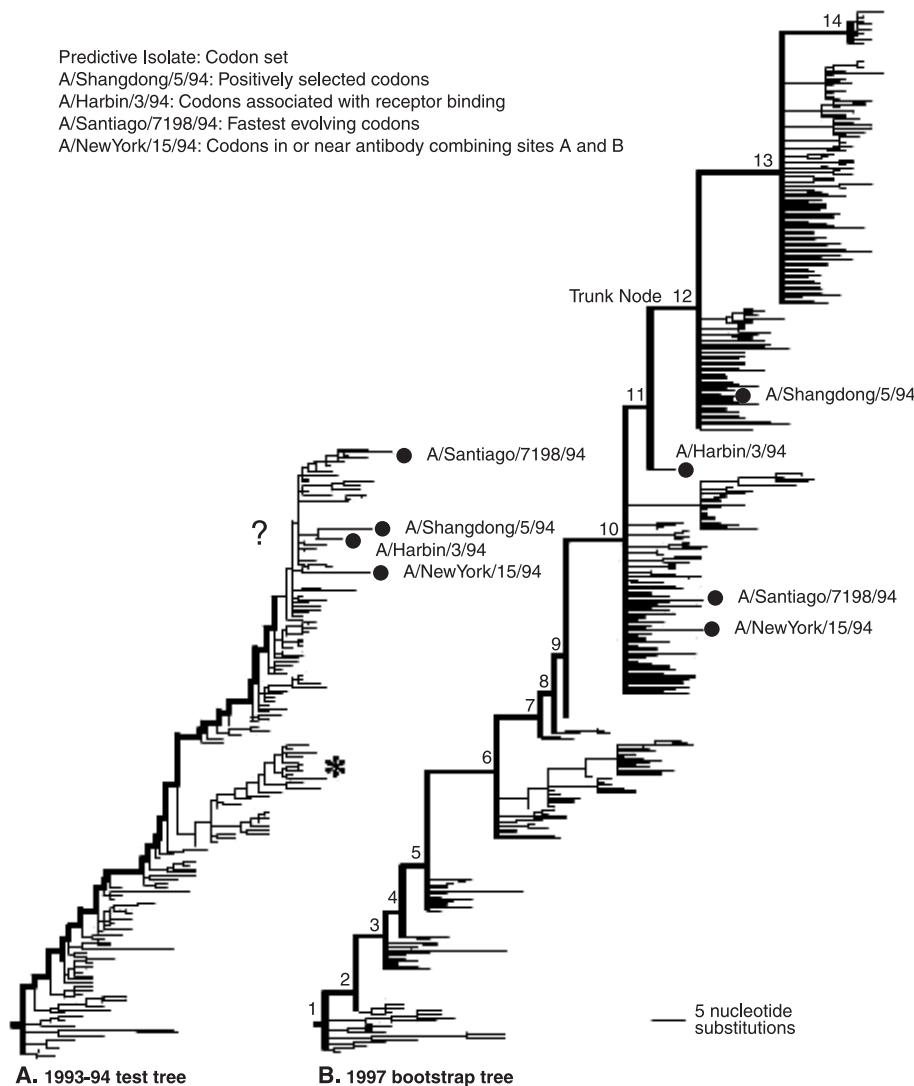


Fig. 1. Predicting the evolution of influenza A hemagglutinin. (A) The 1993–1994 test tree shows the evolution of the HA1 domain of the hemagglutinin gene of human influenza A subtype H3 from 1983 through the 1993–1994 influenza season. The single evolutionarily successful lineage, or trunk lineage, is shown as a bold line. Question mark indicates the point where we lose the ability to visually discern the trunk near the top of the 1993–1994 tree. Asterisk marks a section of the 1993–1994 test tree where two lineages cocirculated for 5 years. (B) The 1997 bootstrap tree, shows the evolution of hemagglutinin from 1983 through the 1996–1997 influenza season. The trunk has been extended through the top of the tree based on subsequently obtained data. Taken together, these trees portray a successful prediction test because the predictive isolate (A/Shangdong/5/94) found by counting replacements in the positively selected codons along all lineages on the 1993–1994 test tree is located on a lineage descending from node 12, the uppermost trunk node from which isolates from the 1993–1994 data set descend. By our definition of fitness A/Shangdong/5/94 is relatively more fit than the other isolates because it is more closely related to the path of the trunk lineage in future influenza seasons.

Table 1. The 18 positively selected codons in the HA1 hemagglutinin gene and their membership in alternative codon sets. R, codons associated with the sialic acid receptor binding site; A or B, codons in or near antibody combining site A or B, respectively; F, codons with rapid rates of amino acid replacement.

Codon	Codon set		
121	—	—	F
124	—	A	—
133	—	A	F
135	R	A	F
138	R	A	F
142	—	A	—
145	—	A	F
156	—	B	F
158	—	B	—
186	—	B	F
190	R	B	F
193	—	B	F
194	R	B	F
197	—	B	—
201	—	—	—
226	R	—	F
262	—	—	—
275	—	—	F

quences typically appear in a region of relatively undifferentiated clusters in the top of the tree (question mark in Fig. 1A). Because of this lack of differentiation, and because the average life-span of a nontrunk lineage is 1.5 years (2), the course of the trunk lineage within the upper portions of a phylogenetic tree constructed with the hemagglutinin gene is identifiable only in retrospect. Usually it becomes discernible within a few years, although one recent nontrunk lineage persisted for 5 years (asterisk in Fig. 1A).

Our prediction method posits that the lineage undergoing the most amino acid replacements in the 18 positively selected codons (Table 1) along the path of branches joining the root of the tree to each of the 173 terminal nodes should identify the section of the 1993–1994 tree from which the future trunk lineage will emerge. That isolate, A/Shangdong/5/94, is called the “predictive isolate” for the 1993–1994 test (Fig. 1A).

To discover whether our prediction was correct, we must determine whether A/Shangdong/5/94 is located in the part of the 1993–1994 test tree from which the trunk eventually emerged. We know the eventual course of the trunk lineage from subsequently collected data used to construct the “1997 bootstrap tree” (6) (Fig. 1B). All nodes on this tree with less than 50% bootstrap support were collapsed to ensure that we evaluated the success of a prediction test (the relative distance of isolates up the trunk lineage) as conservatively as possible (7). It should not be inferred that the 14 trunk nodes divide the nontrunk lineages into classes based on antigenic characteristics, isolation date, or geographic point of sampling.

For the 1993–1994 prediction test to be successful by our criterion, A/Shangdong/5/94 must be at least as fit as any other isolate in the 1993–1994 data set. By our definition of fitness, this means that A/Shangdong/5/94 must be located as far up the trunk lineage of the 1997 bootstrap tree as any other isolate in the 1993–1994 data set. The 173 isolates from the 1993–1994 test tree are located in lineages descending from trunk nodes 1 through 12 on the 1997 bootstrap tree. The predictive isolate, A/Shangdong/5/94, descends from trunk node 12. Thus, according to our criterion, the prediction test for 1993–1994 is successful. We again note that our method predicts only that future lineages will be more closely related to A/Shangdong/5/94 than to other isolates from the 1993–1994 test tree and not that this strain has the antigenic characteristics required to cause the next epidemic.

To ensure that this result is robust with respect to uncertainty in the structure of the test tree, we replicated the 1993–1994 prediction test by using 10 different but equally parsimonious 1993–1994 test trees generated by randomizing the order in which sequences were

entered into PAUP. A/Shangdong/5/94 was the predictive isolate for all 10 of the 1993–1994 replicate trees. Table 3 describes how tests were scored in years when replicate test trees produced different predictive isolates.

We performed retrospective tests for 10 other influenza seasons in the same manner as for the 1993–1994 test. The prediction tests were successful according to our criterion in 9 of 11 influenza seasons (Table 3). The prediction method was not successful in the 1987–1988 and 1988–1989 tests, when the method chose the same predictive isolate as for the 1986–1987 test. Analysis of these failures suggests that we might improve the prediction method by developing criteria with which to decide when lineages have become extinct and rejecting potential predictive isolates that occur on those lineages.

Comparison with Random Expectations

A comparison of our results with random expectations supports the validity of our predictive results. First, we determined the probability of success by using randomly chosen predictive isolates. The 173 isolates in the 1993–1994 test data set are distributed on the lineages descending from trunk nodes 1 through 12 on the 1997 bootstrap tree (Fig. 1B). Only one of the 173 isolates, the predictive isolate A/Shangdong/5/94, is located on the lineages descending from the uppermost possible trunk node, trunk node 12. The probability of obtaining a prediction test that was at least as successful as ours for the 1993–1994 influenza season, assuming that our method actually contained no predictive in-

formation, was thus only 1 of 173, or 0.6%. In other years this probability is much higher, largely because greater proportions of the isolates from the test data set happened to descend from the uppermost trunk lineage. Across influenza seasons, the probability of obtaining results at least as successful as ours by randomly choosing predictive isolates ranged from 0.6% to 42.6%, with an average of 13.9% (Table 2).

Then, we tested the hypothesis that change itself is adaptive. In this case counting mutations using randomly chosen codon sets should produce a similar rate of successful tests. We repeated each of the 11 prediction tests with 1000 sets of 18 randomly sampled codons drawn from the set of 177 codons that encoded amino acid replacements between 1983 and 1997 but were not under positive selection. In the 1993–1994 example, counting replacements at randomly chosen codons resulted in a successful prediction test only twice (0.2% of the time).

Across 11 seasons, the probability of obtaining results at least as successful as ours, had we used randomly chosen codons, averaged 10.6% (Table 2). This probability varied widely among seasons, again in part because of variation in the distribution of isolates from the test data sets on the 1997 bootstrap tree. However, some of the differences must be attributed to the fact that we were counting mutations in a different set of codons. For the 1994–1995 test, none of the random codon sets resulted in a successful prediction test even though there were four isolates from this data set in lineages descending from the uppermost possible trunk node. This is because

Table 2. Sample sizes for retrospective test data sets, and comparison of prediction results with random expectations.

	Influenza season											
	86–87	87–88	88–89	89–90	90–91	91–92	92–93	93–94	94–95	95–96	96–97	Mean
Number of isolates per test tree*	39	50	61	65	76	97	141	173	222	288	357	
Probability of success using randomly chosen isolates†	15.4	30.0	42.6	1.5	1.3	8.3	31.9	0.6	1.8	16.0	3.9	13.9
Probability of success using randomly chosen codons‡	4.1	20.6	41.8	0.8	0.4	14.4	31.0	0.2	0	1.4	1.8	10.6

*Cumulative number of isolates used to construct retrospective test trees representing the evolution of the hemagglutinin gene from 1983 through 11 consecutive influenza seasons. †Percent of randomly chosen predictive isolates that descend from trunk nodes as far or farther up the trunk of the 1997 bootstrap tree (Fig. 1B) as the predictive isolate found by counting replacements at the 18 positively selected codons. ‡Percent of 1000 randomly chosen sets of 18 codons that produced predictive isolates that descend from trunk nodes as far or farther up the trunk of the 1997 bootstrap tree (Fig. 1B) as the predictive isolate found by counting changes at the 18 positively selected codons.

all eight mutations on the lineages joining these isolates to the trunk lineage were in the positively selected codons.

Search for a Causal Explanation

We sought a causal explanation for our results by first determining whether they reflected selection to alter the antibody combining sites in the hemagglutinin. Of the 329 amino acids in HA1 of hemagglutinin, 131 are thought to lie in or near the five overlapping antibody combining sites, labeled A through E (8–10). Changes in 41 codons in or near antibody combining sites A and B have been associated with antigenic drift (11). We repeated the prediction tests, counting changes in just those 41 codons, and obtained seven successful tests across 11 influenza seasons compared with nine successes when the positively selected codons were used (Table 3). This suggests that the selective pressure on the positively selected codons may have been to alter antibody binding.

To determine whether association with antibody combining sites A and B was sufficient to explain the successful prediction tests obtained using the positively selected codons, we repeated the tests with the 28 codons associated with antibody combining sites A and B that were not under positive selection

(9). Only 1 of 11 tests was successful (Table 3). Thus, although selection for antibody binding may be the force causing change in the positively selected codons, association with antibody combining sites A and B is not itself sufficient for a codon to have predictive value using our method. We achieved a successful prediction test in only three seasons (Table 3) when we counted changes in the 90 codons collectively associated with antibody combining sites C, D, and E (9). Only 5 of the 90 codons in this set were also under positive selection, so we did not repeat the test with these codons excluded.

We next examined whether the selective pressure on the positively selected codons reflected selection to alter the sialic acid receptor binding site of the hemagglutinin. Six highly conserved residues comprising the surface of the pocket (Tyr⁹⁸, Trp¹⁵³, Glu¹⁹⁰, Leu¹⁹⁴, His¹⁸³, and Thr¹⁵⁵), along with residues 134 to 138 and 224 to 228, which form the right and left sides of the pocket, respectively, are involved in receptor binding. Five of the positively selected codons and some of the codons associated with antibody combining sites are among this set (9, 12) (Table 1). Selective conservation of the receptor binding site might be expected to preserve function and receptor specificity; however, it is

possible that mutations might be adaptive if antibody binding were reduced or if better binding to respiratory epithelium resulted. Prediction tests counting changes in the receptor binding site codons were successful in 5 of 11 seasons (Table 3). Prediction tests using only the 11 receptor binding site codons that were not also under positive selection (9) failed to produce a single successful test because they underwent very few changes (Table 3). Thus, change appears to be adaptive in the codons within the receptor binding site that are also under positive selection but not for the codon set as a whole.

Among the set of 20 codons having relatively high rates of amino acid replacements relative to other HA1 codons, 12 were also in the set of positively selected codons (9) (Table 1). This association led us to test whether a rapid rate of amino acid replacement is selectively advantageous regardless of the position of the codon in the hemagglutinin gene. Predictive isolates obtained from tests with this alternative codon set produced successful tests in 6 of 11 seasons (Table 3). We repeated the prediction tests with the set of 18 codons with the highest rates of amino acid replacement excluding those under positive selection (9). This resulted in a successful prediction test in only two seasons (Table 3).

Table 3. Results of retrospective prediction tests for 11 recent influenza seasons using codons under positive selection and seven alternative codon sets. Cells indicate the trunk nodes on the 1997 bootstrap tree (Fig. 1B) from which the predictive isolates resulting from each test descended. Successful tests (bold underlined numbers) are those in which the predictive isolate descended from the uppermost possible node on the 1997 tree (bottom row). Right-hand column shows the total number of seasons in which each codon

set produced a successful test. Cell entries with a decimal place are mean trunk nodes (across the 10 replicate test trees) when these tests produced predictive isolates descending from different trunk nodes. Positively selected, the set of 18 codons under positive selection; AB, codons in or near antibody combining sites A and B; CDE, codons in or near antibody combining sites C, D, and E; RBS, codons associated with the sialic acid receptor binding site; Fast, codons undergoing the largest number of amino acid replacements.

	Codons (n)	Influenza season											Success (n)
		86–87	87–88	88–89	89–90	90–91	91–92	92–93	93–94	94–95	95–96	96–97	
Positively selected	18	<u>5</u>	5	5	<u>8</u>	<u>8</u>	<u>10</u>	<u>10</u>	<u>12</u>	<u>13</u>	<u>13</u>	<u>14</u>	9
AB	41	<u>5</u>	5	<u>6</u>	<u>8</u>	7.0	<u>10</u>	<u>10</u>	10	12.0*	<u>13</u>	<u>14</u>	7
AB but not under positive selection	28	<u>5</u>	5	5.5	5.5	5.5	8.7	9.6	10	12	12	12	1
CDE	90	1	1	<u>6</u>	6	6	8	<u>10</u>	10	10	<u>13</u>	13	3
RBS	16	3.7	5	<u>6</u>	6	6	<u>10</u>	<u>10</u>	11	12.5	<u>13</u>	13.1*	4
RBS but not under positive selection	11	4.1	5	5	5	5	5	5	5	5	5	5	0
Fast	20	3	3	<u>6</u>	6.7*	7.0	<u>10</u>	<u>10</u>	10	<u>13</u>	<u>13</u>	<u>14</u>	6
Fast but not under positive selection	18	3	3	<u>6</u>	6	6	6	<u>10</u>	10	10	12	13	2
Top possible trunk node		5	6	6	8	8	10	10	12	13	13	14	

*Prediction tests that failed to identify the same predictive isolate in 7 or more of the 10 replicate tests. For these tests the cell contains the mean trunk node taken across results from all 10 replicate tests.

Thus, a rapid rate of change alone does not explain why change at the set of positively selected codons appears to be adaptive.

Conclusion

We have identified a small set of rapidly evolving codons in the HA1 domain of the hemagglutinin gene of human influenza A subtype H3 in which replacement substitutions in the past appear to have been selectively advantageous. Strains with more mutations in these codons were more likely to be the progenitors of successful new lineages in 9 of 11 influenza seasons. The probability of obtaining a successful test by chance varied across years. However, our method produced successful tests in many years, such as 1993–1994, when it was highly unlikely to do so. A causal explanation for our results is suggested by the significant overlap between the positively selected codons and the codons in or near antibody combining sites A or B and, to a lesser extent, codons associated with the sialic acid receptor binding site. However, codons associated with these sites of known function that are not under positive selection perform poorly in the prediction tests. The positively selected codons are among the most quickly evolving in the HA1 domain of hemagglutinin, but this characteristic also is not a sufficient explanation for our results. It appears that we have identified a subset of rapidly

evolving codons in known functional sites for which change in the past has been associated with a clear selective advantage. Whether additional changes in these codons will confer a selective advantage in the future remains to be seen. Our retrospective tests indicate that, had these data been available in the past, monitoring change in the positively selected codons might have provided potentially useful information about the course of influenza evolution. These methods can be used to examine the evolution of the hemagglutinins of influenza B and influenza A H1 viruses circulating in humans to identify sets of positively selected codons that may have predictive value for these important pathogens.

References and Notes

1. W. M. Fitch, R. M. Bush, C. A. Bender, N. J. Cox, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 7712 (1997).
2. R. M. Bush, W. M. Fitch, C. A. Bender, N. J. Cox, *Mol. Biol. Evol.* **16**, 1457 (1999).
3. The technique for identifying those codons under positive selection excludes amino acid replacements that may have occurred during culture in eggs in the laboratory by ignoring replacements that occur on the branches joining isolate sequences to the tree.
4. Sequences were generated at the Centers for Disease Control and Prevention by the protocols described in (7). They have been deposited in GenBank (accession nos. AF008656 to AF008909 and AF180564 to AF180666).
5. D. L. Swofford, PAUP* (Phylogenetic Analysis Using Parsimony and other methods) version 4.0.0d60.
6. We used trees constructed from the 1983–1997 data

both to derive the set of positively selected codons (2) and to score the prediction tests. This does not bias our scoring because the location of the trunk is not a factor in determining which codons are under positive selection.

7. Trunk nodes present in at least 50% of the bootstrap replicates were consistently present in nucleotide parsimony and neighbor-joining trees and also in trees constructed from amino acids using either parsimony, neighbor-joining (5), or the protML maximum likelihood routine of Molphy 2.2 (made available by J. Adachi and M. Hasegawa). The method we used for evaluating the success of a prediction test is thus robust with regard to the algorithm and type of data used. Computational restraints prevented applying maximum-likelihood algorithms to our entire data set at once, so we used multiple overlapping subsets of sequences to perform these analyses. The data set was also too large for standard bootstrap analysis, so we estimated bootstrap values by using 10,000 PAUP FastStep bootstrap replicates. This number was determined by increasing the number of replicates until we obtained a consistent set of bootstrap values.
8. I. A. Wilson and N. J. Cox, *Annu. Rev. Immunol.* **8**, 737 (1990); D. C. Wiley, I. A. Wilson, J. J. Skehel, *Nature* **289**, 373 (1981).
9. Supplementary tables showing the codons used in alternative prediction tests are available at www.sciencemag.org/feature/data/1044188.shl
10. C. A. Bender and N. J. Cox, unpublished data.
11. N. J. Cox and C. A. Bender, *Semin. Virol.* **6**, 359 (1995).
12. W. Weiss et al., *Nature* **333**, 426 (1988).
13. Supported in part by funds provided by the University of California for the conduct of discretionary research by Los Alamos National Laboratory, conducted under the auspices of the US Department of Energy. We gratefully acknowledge the technical expertise of H. Jing and the critical comments of S. Frank.

3 August 1999; accepted 3 November 1999

REPORTS

Complex Shear Wave Velocity Structure Imaged Beneath Africa and Iceland

Jeroen Ritsema,^{1*} Hendrik Jan van Heijst,¹ John H. Woodhouse²

A model of three-dimensional shear wave velocity variations in the mantle reveals a tilted low velocity anomaly extending from the core-mantle boundary (CMB) region beneath the southeastern Atlantic Ocean into the upper mantle beneath eastern Africa. This anomaly suggests that Cenozoic flood basalt volcanism in the Afar region and active rifting beneath the East African Rift is linked to an extensive thermal anomaly at the CMB more than 45 degrees away. In contrast, a low velocity anomaly beneath Iceland is confined to the upper mantle.

Tomographic models of three-dimensional seismic velocity variations continue to improve our understanding of the structure of flow in Earth's mantle (1–6). It has long been recognized that higher than average seismic velocities are present in regions of the mantle with long-term

subduction (such as the circum-Pacific region and Asia) and that lower than average velocity structures are present in the deep mantle beneath Africa and the Pacific Ocean where subduction has not occurred since the Jurassic. The interpretation of high velocity anomalies as subducting slabs of relatively cold oceanic lithosphere has gained support from recent tomographic models, which show tabular structures of faster than average velocity beneath several subduction zones (7, 8) (Web figure 1). The complex shape and broadening of these slab-like structures may be a manifestation of

the complex subduction zone process, large-scale plate motions, and the resistance to flow caused by the spinel-to-perovskite phase transition at 670 km depth (9). Although laboratory experiments (10, 11) and computer models (12, 13) demonstrate that hot upwellings may ascend into the mantle in a similarly complex manner, it is difficult to image narrow conduits of rapidly rising mantle plumes (14) in the deep mantle. In addition, tomographic models have not fully constrained the dimensions and shape of the large-scale upwellings beneath Africa and the Pacific Ocean (Fig. 1).

We studied the structure of upwellings in the mantle using seismic tomographic model S20RTS (15). We focused on the large-scale low velocity anomaly in the deep mantle beneath Africa, which was imaged in earlier tomographic models (1, 2) (Fig. 1) and has been linked to surface manifestations of hot spots around Africa (16), the high elevation of southern Africa (17), and the ultra-low velocity anomalies at the base of the mantle beneath Africa (18). We compared this anomaly with the low velocity structure beneath Iceland. S20RTS is a degree-20 shear wave velocity model that incorporates surface wave phase velocities, body wave travel times, and free-oscillation splitting measurements (Table 1). These data cover the entire

¹Seismological Laboratory, California Institute of Technology, Pasadena, CA 91125, USA. ²Department of Earth Sciences, Oxford University, Oxford OX1 3PR, UK.

*To whom correspondence should be addressed. E-mail: jeroen@gps.caltech.edu