

Comment on "Large-Scale Sequence Analysis of Avian Influenza Isolates"

Edward C. Holmes,¹ David J. Lipman,^{2*} Dmitriy Zamarin,³ Jonathan W. Yewdell⁴

Obenauer *et al.* (Research Articles, 17 March 2006, p. 1576) reported that the influenza A virus PB1-F2 gene is evolving under strong positive selection, as documented by an extremely high ratio of the number of nonsynonymous nucleotide substitutions to the number of synonymous substitutions (dN/dS). However, we show that this observation is likely to be an artifact related to the location of PB1-F2 in the +1 reading frame of the PB1 gene.

Basic polymerase 1 frame 2 (PB1-F2) is a small influenza A virus (IAV) protein encoded by an overlapping reading frame of a viral polymerase gene (PB1). PB1-F2 is a proapoptotic mitochondrial protein with membrane-disrupting properties that is dispensable for viral replication *in vitro*. Obenauer *et al.* (1) recently reported that the gene encoding PB1-F2 is under strong positive selection pressure, as documented by a higher rate of nonsynonymous (dN) to synonymous (dS) substitutions per site (dN/dS > 9) in a large number of avian and human IAV isolates. However, it is also possible that this is an artifact related to the location of PB1-F2 in the +1 reading frame of the PB1 gene. As expected of genes from RNA viruses with low fidelity polymerases, PB1 exhibits extensive variability in the third synonymous position of each codon. However, because of the nature of the genetic code, these changes will result in nonsynonymous substitutions in the second codon position of PB1-F2, thereby increasing dN/dS if this gene is weakly constrained (2). Similarly, purifying selection against nonsynonymous mutations in PB1 is expected to reduce the rate of synonymous mutation in PB1-F2, again artificially elevating dN/dS ratios.

To test this hypothesis, we collected from GenBank a representative sample of PB1 sequences derived from 150 isolates of avian IAV, including those from 67 different viral subtypes. The PB1 sequence alignment was divided into four gene regions (see Table 1): PB1, excluding the F2 region (PB1-F1 single coding); the PB1-F2 region alone, but translated in Frame 1 (PB1-F1 double coding); the F2 region alone translated in Frame 2 (PB1-F2); and the longest contiguous region (98 codons) of F2 region outside of PB1-F2, where stop codons were infrequent (PB1-F2 control). For each data set, overall and site-specific dN/dS ratios were estimated using the single likelihood ancestor counting (SLAC) method available at the Datamonkey facility (3), incorporating the HKY85 model of nucleotide substitution with phylogenetic trees inferred using the neighbor-joining method.

The entire PB1 gene in frame 1, regardless of whether it is also translated in frame 2, is subject to strong and similar selective constraints manifest as very low dN/dS ratios (Table 1). As expected given such constraints, there was no evidence for positive selection acting on any individual codon site. However, a very different picture was seen in the PB1-F2 region, where a signal for extensive positive selection was re-

vealed, both in the high average dN/dS ratio (6.988) across the gene as a whole, and in that 38 individual codons were found to have statistically significant evidence for positive selection. Notably, an even stronger force of positive selection was found in the PB1-F2 control region, which had an overall dN/dS of 13.296 and 55 positively selected codons.

The simplest explanation for these results is that a combination of strong purifying selection against nonsynonymous mutations in PB1 and a high capacity of PB1-F2 to accept nonsynonymous substitutions resulting from synonymous changes in PB1 have artificially elevated the dN/dS in PB1-F2. Consistent with this conclusion, strong positive selection of PB1-F2 would be expected to alter selection pressures in PB1-F1, most obviously manifest as an increase in the rate of synonymous substitutions. However, there is no clear difference for such a change in selection pressures in either the dN/dS ratio or the overall genetic distance (depicted as the total length of the tree in substitutions per site). Further, if positive selection were acting on PB1-F2, then the observation that 38 out of 90 (42%) codons are accumulating advantageous mutations would perhaps represent the strongest force of adaptive evolution documented to date, although avian IAV is usually considered to be subject to strong selective constraints (4). Finally, that the PB1-F2 control region exhibits an even higher dN/dS ratio confirms that there are residual selective constraints acting on PB1-F2. We therefore conclude that there is no compelling evidence for positive selection acting on PB1-F2.

References

1. J. C. Obenauer *et al.*, *Science* **311**, 1576 (2006).
2. A. M. Pedersen, J. L. Jensen, *Mol. Biol. Evol.* **18**, 763 (2001).
3. S. L. Kosakovsky Pond, S. D. W. Frost, S. V. Muse, *Bioinformatics* **21**, 676 (2005).
4. R. G. Webster *et al.*, *Microbiol. Rev.* **56**, 152 (1992).

26 June 2006; accepted 17 August 2006
10.1126/science.1131729

¹Center for Infectious Disease Dynamics, Department of Biology, The Pennsylvania State University, Mueller Laboratory, University Park, PA 16802, USA. ²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. ³Department of Microbiology, Mount Sinai School of Medicine, New York, NY 10029, USA. ⁴Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.

*To whom correspondence should be addressed. E-mail: lipman@ncbi.nlm.nih.gov

Table 1. Summary of selection pressures acting on the PB1 gene of avian IAV.

Data set	Length (codons)	Overall dN/dS	No. selected sites	Tree length*
PB1-F1 single coding	691	0.021	0	2.659
PB1-F1 double coding	89	0.033	0	2.426
PB1-F2	90	6.988	38	2.418
PB1-F2 control	98	13.296	55	2.271

*Total length of the tree in terms of the numbers of substitutions, per site