Letter to the Editor

Reduced Positive Selection in Vector-Borne RNA Viruses

Christopher H. Woelk[1] and Edward C. Holmes

Department of Zoology, University of Oxford

RNA viruses evolve at greatly elevated rates compared with their DNA counterparts. This is principally due to the high mutation rate associated with RNA polymerase, which introduces approximately one error per genome during each round of viral replication, a rate some two orders of magnitude higher than observed in DNA viruses (Drake et al. 1998). The potential for rapid evolutionary change is thought to provide RNA viruses with unparalleled adaptability (Domingo and Holland 1997), so that a solution can quickly be found to any evolutionary problem, such as those posed by antiviral therapy.

Conversely, studies of long-term rates of nucleotide substitution suggest that certain aspects of viral biology, particularly transmission mode, may impose constraints on RNA virus evolution. In particular, reduced rates of nonsynonymous substitution have been recorded in some RNA viruses transmitted by arthropod vectors (Jenkins et al. 2002), suggesting that the requirement to replicate in disparate hosts, like mammals and invertebrates, imposes greater selective constraints than on viruses that only infect phylogenetically similar species (Scott, Weaver, and Mallampalli 1994; Weaver et al. 1999). However, this theory remains controversial, as reiterated by experimental studies of vesicular stomatitis virus (VSV) in cell culture which showed that alternate replication in insect and mammalian cells did not reduce the rate at which mutations appeared or limit fitness increases (Novella et al. 1999).

To determine whether vector-borne RNA viruses evolve in the same manner as viruses transmitted by other routes (including blood-borne, fecal-oral, respiratory, sexual, and vertical), and to investigate the nature of selection in viruses in general, we analyzed selection pressures in a wide variety of viruses that infect human populations. The sequences of surface structural (envelope glycoprotein or outer capsid) and internal structural (core capsid or nucleoprotein) genes were collected from all human-associated viruses available on GenBank or the Los Alamos National Laboratory in the case of the human immunodeficiency virus type 1 (HIV-1) (http://www.hiv.lanl.gov). In the case of the surface genes, sufficiently large intraspecific data sets were available from 45 viruses (18 vector-borne, 27 nonvector-borne). Of

these, 36 were RNA viruses, two were retroviruses (which have both RNA and DNA stages in their life cycle) and seven were DNA viruses. All the vector-borne viruses had RNA genomes. For the internal genes, 23 viral data sets were collected (12 vector-borne, 11 nonvector-borne), of which 21 were RNA viruses and two were retroviruses. No internal structural genes were available from DNA viruses. Hence, the data studied comprised both slowly evolving DNA viruses which may have cospeciated with their vertebrate hosts over millions of years (for example, herpes simplex viruses, HSV-1 and HSV-2), and rapidly mutating RNA/retroviruses that have only emerged recently (such as HIV-1). In all cases, sequences derived from vaccine strains, as well as sites included in overlapping reading frames, were removed because these factors may introduce artificial signals of positive selection. Similarly, sequences subjected to laboratory passaging were excluded unless the resulting alignment was too small for meaningful analysis. Finally, to limit the possible effects of recombination on our analysis, bootstrapped neighbor-joining trees were inferred for different gene regions in each virus data set and those sequences with conflicting phylogenetic positions were excluded from further analysis (results not shown).

Selection pressures were measured using a maximum likelihood (ML) method that estimates the number of nonsynonymous (d_N) and synonymous (d_S) substitutions among codons using a variety of models of codon evolution that incorporate the phylogenetic relationships of the sequences in question (Yang et al. 2000). We compared two models: the neutral M7 model which assumes that d_N/d_S follows a beta distribution (with 10 categories) but where no category has $d_N/d_S > 1$, and M8 which is equivalent to M7 except that an 11th category of sites is added at which d_N/d_S can exceed 1, thereby accounting for positive selection. M7 and M8 were compared using a likelihood ratio test. Evidence of positive selection in individual viruses existed when M8 was significantly favored over M7 and included a category of codon sites with $d_N/d_S > 1.5$ (arbitrarily chosen to indicate clear-cut positive selection). Mean d_N/d_S ratios for each virus were calculated as the average over all codons under the M8 model and then compiled to compare different groups of viruses (for example, vector- versus nonvector-borne viruses). Because the mean d_N/d_S ratios for each virus are not normally distributed, a nonparametric Mann-Whitney U-test was used to test for a significant difference in overall selection pressure between groups. These analyses were undertaken using the CODEML program from the PAML package (Yang 1997). In all cases, input ML phylogenetic trees were estimated under the HKY85+Γ model of nucleotide substitution using the PAUP* package

[1] Present address: Department of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California.

Key words: virus, positive selection, vector transmission, maximum likelihood, trade-offs.

Address for correspondence and reprints: Edward C. Holmes, Department of Zoology, University of Oxford, South Parks Road, Oxford, OX1 3PS, United Kingdom. E-mail: edward.holmes@zoo.ox.ac.uk.
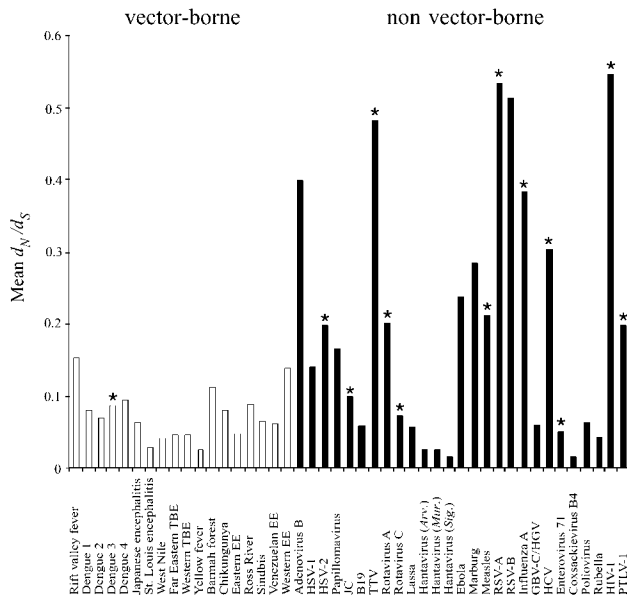
vector-borne   non vector-borne

FIG. 1.—Selection pressures among the surface structural genes of 45 human-associated viruses. Mean d_N/d_S ratios reflect the average value observed across all codons under the M8 model of codon evolution. An asterisk marks those viruses with significant evidence for positive selection. Virus abbreviations are HSV-1 and HSV-2—herpes simplex types 1 and 2; HIV-1—human immunodeficiency virus type 1; PTLV-1—primate T-lymphotropic virus type 1 (comprising human HTLV-1 and simian STLV-1); RSV-A and RSV-B—respiratory syncytial virus A and B; HCV—hepatitis C virus; hantavirus (*Arv.*)—Arvicolinae host, hantavirus (*Mur.*)—Murinae host, Hantavirus (*Sig.*)—Sigmondontinae host; TBE—tick-borne encephalitis; EE—equine encephalitis. Specific genotypes used in some cases were adenovirus B (serotypes 3 and 7; papillomavirus—type 16; TT virus—genotype 1; HIV-1—M group; rotavirus A—genotype 1; influenza A—H3N2; HCV—genotype 1b; Sindbis virus group—Sindbis, Ockelbo, and Karelian fever viruses.

(Swofford 2002). For measles virus, human respiratory syncytial virus A and B, dengue virus, and Venezuelan equine encephalitis virus, the results of selection analysis using CODEML have been published previously and are used here for simplicity (Woelk and Holmes 2001; Woelk et al. 2001; Brault et al. 2002; Twiddy, Woelk, and Holmes 2002). Full results are available as supplementary information at http://www.molbiolevol. org.

For the surface structural genes, the d_N/d_S ratios in the vector-borne viruses were significantly less than in those viruses transmitted by other routes—median d_N/d_S ratios of 0.066 and 0.165, respectively, $P = 0.048$, under a Mann-Whitney U-test. The results of this comparison are presented graphically in figure 1. Furthermore, whereas only one vector-borne virus showed significant evidence for positive selection (Dengue 3), 12 of the nonvector-borne viruses had d_N/d_S ratios significantly >1 at some codons (table 1). Selected viruses included DNA viruses ($n = 3$), RNA viruses ($n = 8$), and retroviruses ($n = 2$), viruses that cause acute ($n = 7$) and persistent ($n = 6$) infections, and those with associated disease syndromes ranging from asymptomatic (TT virus) to severe immunodeficiency (HIV-1). Hence similarities in genome structure, duration of infection, or virulence do not result in common selection pres-

Table 1
Properties of Human-Associated Viruses with Significant Evidence for Positive selection

Virus	Max.[a] d_N/d_S	% Codons Selected	Max.[a] Transmission Mode	Genome	Duration	Human Association	Disease Syndrome
Surface Genes							
Dengue 3	1.985	1.7	Vector	RNA	Acute	Endemic	Febrile, hemorrhagic
HSV-2	9.184	2.1	Sexual, saliva	DNA	Persistent	Endemic	Herpes, neurological
JC virus	1.943	1.8	Respiratory, urine	DNA	Persistent	Endemic	Neuro-denegerative
TT virus.	11.057	2.5	Blood, vertical	DNA	Persistent	Endemic	Asymptomatic, hepatitis?
Rotavirus A	1.973	2.4	Fecal-oral	RNA	Acute	Endemic	Diarrheal
Rotavirus C	1.718	1.5	Fecal-oral	RNA	Acute	Endemic	Diarrheal
Measles	2.113	1.8	Respiratory	RNA	Acute	Endemic	Measles (rare neurological)
RSV-A	2.803	6.8	Respiratory	RNA	Acute	Endemic	Colds, pneumonia
Influenza A	3.146	2.9	Respiratory	RNA	Acute	Endemic	Influenza
Hepatitis C	2.379	7.5	Blood	RNA	Persistent	Endemic	Hepatitis
Enterovirus 71	2.196	0.7	Fecal-oral	RNA	Acute	Endemic	Febrile, aseptic meningitis
HIV-1	2.692	10.4	Blood, sexual, vertical	Retrovirus	Persistent	Endemic	AIDS
PTLV-1	2.073	3.0	Breast milk, sexual	Retrovirus	Persistent	Endemic	T-cell leukaemia
Internal Genes							
Oropouche.	6.946	0.8	Vector	RNA	Acute	Sporadic[b]	Febrile
Measles	2.294	2.0	Respiratory	RNA	Acute	Endemic	Measles (rare neurological)
Hepatitis C	1.580	5.1	Blood	RNA	Persistent	Endemic	Hepatitis

[a] Maximum d_N/d_S values correspond to those observed in the M8 model of codon evolution where there is significant evidence for positive selection (M8 significantly favored over M7 with $d_N/d_S > 1.5$ at a category of codons).
[b] Sporadic viruses are defined here as those that have another animal species as their main host and only occasionally cause outbreaks of disease in humans.

sures. Intriguingly, all 13 of the positively selected viruses show endemic transmission in humans, whereas there was no evidence for adaptive evolution in 20 principally animal viruses that cause only sporadic disease in humans. Although this may be influenced by ascertainment bias, as endemic human viruses have been sampled most intensively, it may also be due in part to the large size and density of human populations, which will counter the stochastic loss of advantageous viral lineages, thereby allowing natural selection to act with more potency. If true, the rapidly increasing size of the human population may allow natural selection to play an ever more important role in viral evolution. This is clearly an area that needs to be investigated in more detail.

To determine whether the difference in d_N/d_S between the vector and nonvector-borne viruses was entirely due to positive selection in the latter group or reflected a more general difference in selective constraints, we compared d_N/d_S ratios between them excluding the 11th (positively selected) category of sites in the M8 model in those viruses where we found evidence for positive selection. Although d_N/d_S in the vector-borne viruses was still less than that in the viruses transmitted by other routes, the difference was not significant (median d_N/d_S of 0.063 and 0.134, respectively; $P = 0.215$). Consequently, the main difference between vector and nonvector-borne viruses is in the extent of positive selection. Furthermore, we found no significant difference in d_N/d_S between DNA and RNA viruses (including retroviruses), either with (median d_N/d_S of 0.165 and 0.134, respectively; $P = 0.501$) or without (median d_N/d_S of 0.140 and 0.098, respectively; $P = 0.740$) positively selected sites. Consequently, although DNA viruses have much lower rates of nucleotide substitution and have often been associated with their hosts for far longer periods of time than RNA viruses, they are not, on average, subject to stronger selective constraints.

An equivalent analysis was undertaken on the internal structural genes. In this case, vector and nonvector-borne viruses did not differ significantly in selection pressure—median d_N/d_S of 0.098 and 0.087, respectively ($P = 0.460$). Furthermore, significant evidence of positive selection was only found in three viruses, one of which was vector-borne (Oropouche) and two nonvector-borne (hepatitis C and measles), with HIV-1 representing a borderline case (maximum d_N/d_S of 1.447, but with 6.4% of codons falling into this class). As expected, the removal of positively selected sites did not result in a major change in overall selection pressure (median d_N/d_S of 0.098 and 0.066 for the vector and nonvector-borne viruses, respectively; $P = 0.218$).

Overall, these results reveal that vector-borne RNA viruses are generally less subject to positive (diversifying) selection than those viruses transmitted by other routes. This is most apparent in the surface structural genes, which frequently contain sites undergoing adaptive evolution in nonvector-borne viruses. The selection pressure in this case is most likely associated with immune evasion because envelope and outer capsid proteins frequently carry epitopes for neutralizing antibody and T-cell responses. As expected, given their critical role in capsid formation, the genes encoding internal structural proteins are relatively well conserved in both types of virus.

We suggest that evolutionary trade-offs inherent in the vector-host association are the most likely explanation for the reduced positive selection in vector-borne RNA viruses. Three such trade-offs can be envisaged. With respect to preferred cell type, vector-borne viruses can be thought of as evolutionary "generalists." Hence, nonsynonymous mutations that enhance infection or replication in mammalian cells might have antagonistic effects in insect cells (and vice versa). Although evidence against this theory was provided by the experimental study of Novella et al. (1999), these authors only considered one virus (VSV) and analyzed rates of mutation in the short-term, which may have included deleterious changes that are removed by purifying selection at transmission. A similar pattern has been documented in rabies virus which infects a wide range of cell types within a single host and exhibits an anomalously low rate of nonsynonymous substitution (Holmes et al. 2002), and an analogous observation has been made in mammals where genes expressed in a range of tissues have lower nonsynonymous substitution rates than tissue-specific genes (Duret and Mouchiroud 2000). A second possible trade-off involves replication strategy, which differs in insect and mammalian hosts because the former may sometimes be noncytolytic to allow persistence at times when vector population sizes are small (Novella et al. 1999). Finally, mammals and insects mount very different immune responses to viral infections. Therefore, mutations facilitating immune tolerance or escape in one host or vector species might have adverse effects on viral biology in another, thereby reducing the extent of positive selection.

Although our study has focused on transmission mode, it is possible that interactions among other aspects of viral biology also result in evolutionary trade-offs. Paradoxically, such trade-offs may be a common by-product of the intrinsically high mutation rates exhibited by RNA viruses. Specifically, although high mutation rates generate a great deal of genetic variability, they also constrain viral genome size by establishing an "error-threshold"; larger genomes than those observed cannot be produced because of the generation of excessive numbers of deleterious mutations (Eigen 1987). Empirical support for the error-threshold comes from the negative correlation between substitution rate and genome size in RNA viruses (Jenkins et al. 2002). In turn, limited genome size means that individual sequence regions will encode multiple functions, such as those governing cell tropism and immune evasion (Baranowski, Ruiz-Jarabo, and Domingo 2001) and hence that single mutations are likely to have multiple and antagonistic affects. Such factors will clearly limit the number of adaptive solutions available and may also explain why convergent evolution is so commonly observed in RNA viruses (Holmes et al. 1992; Bull et al. 1997; Wichman et al. 1999; Fares et al. 2001). Antagonistic effects that limit positive selection may also arise because of RNA

and protein secondary structure. The constraining effect of RNA secondary structure has previously been demonstrated in virus evolution (Simmonds and Smith 1999), and in HIV-1 the fixation of advantageous escape mutants in epitopes recognized by cytotoxic T-lymphocytes is prevented by the deleterious consequences they have on the structure of the capsid protein (Kelleher et al. 2001).

To conclude, our study reveals that viruses differ in selection pressure depending on their mode of transmission. Consequently, rather than assuming that RNA viruses are able to exploit every possible adaptive solution, it is important to consider the potential constraints to phenotypic diversity, such as those that characterize vector-borne viruses. In the long-term, this knowledge may assist in the control of viral infections by vaccines and antiviral agents.

LITERATURE CITED

BARANOWSKI, E., C. M. RUIZ-JARABO, and E. DOMINGO. 2001. Evolution of cell recognition by viruses. Science **292**:1102–1105.

BRAULT, A. C., A. M. POWERS, E. C. HOLMES, C. H. WOELK, and S. C. WEAVER. 2002. Positively charged amino acid substitutions in the E2 envelope glycoprotein are associated with the emergence of Venezuelan equine encephalitis virus. J. Virol. **76**:1718–1730.

BULL, J. J., M. R. BADGETT, H. A. WICHMAN, J. P. HUELSENBECK, D. M. HILLIS, A. GULATI, C. HO, and I. J. MOLINEUX. 1997. Exceptional convergent evolution in a virus. Genetics **147**:1497–1507.

DOMINGO, E., and J. J. HOLLAND. 1997. RNA virus mutations for fitness and survival. Annu. Rev. Microbiol. **51**:151–178.

DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH, and J. F. CROW. 1998. Rates of spontaneous mutation. Genetics **148**:1667–1686.

DURET, L., and D. MOUCHIROUD. 2000. Determinants of substitution rates in mammalian genes: expression pattern affects selection intensity but not mutation rate. Mol. Biol. Evol. **17**:68–70.

EIGEN, M. 1987. New concepts for dealing with the evolution of nucleic acids. Cold Spring Harbor Syn. Quant. Biol. **52**:307–319.

FARES, M. A., A. MOYA, C. ESCARMÍS, E. BARANOWSKI, E. DOMINGO, and E. BARRIO. 2001. Evidence for positive selection in the capsid protein-coding region of the foot-and-mouth disease virus (FMDV) subjected to experimental passage regimens. Mol. Biol. Evol. **18**:10–21.

HOLMES, E. C., C. H. WOELK, R. KASSIS, and H. BOURHY. 2002. Genetic constraints and the adaptive evolution of rabies virus. Virology **292**:247–257.

HOLMES, E. C., L. Q. ZHANG, P. SIMMONDS, C. A. LUDLAM, and A. J. LEIGH BROWN. 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc. Natl. Acad. Sci. USA **89**:4835–4839.

JENKINS, G. M., A. RAMBAUT, O. G. PYBUS, and E. C. HOLMES. 2002. Rates of molecular evolution in RNA viruses: a quantitative phylogenetic analysis. J. Mol. Evol. **54**:152–161.

KELLEHER, A. D., C. LONG, E. C. HOLMES, ET AL. (18 CO-AUTHORS). 2001. Clustered mutations in HIV-1 gag are consistently required for escape from HLA-B27 restricted CTL responses. J. Exp. Med. **193**:375–385.

NOVELLA, I. S., C. L. HERSHEY, C. ESCARMIS, E. DOMINGO, and J. J. HOLLAND. 1999. Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. J. Mol. Biol. **287**:459–465.

SCOTT, T. W., S. C. WEAVER, and V. L. MALLAMPALLI. 1994. Evolution of mosquito-borne viruses. Pp. 293–324 *in* S. S. Morse, ed. Evolutionary biology of viruses. Raven Press, New York.

SIMMONDS, P., and D. B. SMITH. 1999. Structural constraints on RNA virus evolution. J. Virol. **73**:5787–5794.

SWOFFORD, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.

TWIDDY, S. S., C. H. WOELK, and E. C. HOLMES. 2002. Phylogenetic evidence for adaptive evolution of dengue viruses in nature. J. Gen. Virol. **83**:1679–1689.

WEAVER, S. C., A. C. BRAULT, W. KANG, and J. J. HOLLAND. 1999. Genetic and fitness changes accompanying adaptation of an arbovirus to vertebrate and invertebrate cells. J. Virol. **73**:4316–4326.

WICHMAN, H. A., M. R. BADGETT, L. A. SCOTT, C. M. BOULIANNE, and J. J. BULL. 1999. Different trajectories of parallel evolution during viral adaptation. Science **285**:422–424.

WOELK, C. H., and E. C. HOLMES. 2001. Variable immune driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). J. Mol. Evol. **52**:182–192.

WOELK, C. H., L. JIN, E. C. HOLMES, and D. W. G. BROWN. 2001. Immune and artificial selection in the hemagglutin (H) glycoprotein of measles virus. J. Gen. Virol. **82**:2463–2474.

YANG, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS **13**:555–556.

YANG, Z., R. NIELSEN, N. GOLDMAN, and A. M. K. PEDERSEN. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics **155**:431–449.