JOURNAL OF VIROLOGY, Oct. 2003, p. 11296–11298 Vol. 77, No. 20
0022-538X/03/$08.00+0 DOI: 10.1128/JVI.77.20.11296–11298.2003

Patterns of Intra- and Interhost Nonsynonymous Variation Reveal Strong Purifying Selection in Dengue Virus

Edward C. Holmes*

Department of Zoology, University of Oxford, Oxford, United Kingdom

Received 25 April 2003/Accepted 22 July 2003

Considerable uncertainty surrounds the evolutionary rates of and selection pressures acting on arthropod-borne RNA viruses (arboviruses). In particular, it is unclear why arboviruses such as dengue virus show substantial genetic variation within individual humans and mosquitoes yet low long-term rates of amino acid substitution. To address this question, I compared patterns of nonsynonymous variation in populations of dengue virus sampled at different levels of evolutionary divergence. Although nonsynonymous variation was abundant in viral populations within individual humans, there was a marked reduction in the frequency of nonsynonymous mutations in interhost comparisons. Moreover, intrahost genetic variation corresponded to a random pattern of mutation, and most of the sites that exhibited nonsynonymous variation within hosts were invariant at deeper phylogenetic levels. This loss of long-term nonsynonymous variation is the signature of extensive purifying selection such that more than 90% of all nonsynonymous mutations are deleterious. Consequently, although arboviruses are able to successfully adapt to diverse cell types, they are characterized by a high rate of deleterious mutation.

An important concept in the evolution of arthropod-borne RNA viruses (arboviruses) is that due to intrinsic constraints associated with dual replication in mammalian and invertebrate hosts, these viruses evolve more slowly than RNA viruses transmitted by other routes (1, 9, 11, 18, 19). However, although this theory is commonly stated, the evidence claimed to support it is more ambiguous. For example, large-scale phylogenetic studies have revealed that arboviruses have significantly lower rates of nucleotide substitution than other RNA viruses, particularly at nonsynonymous sites (6), and that they are subject to significantly less positive selection pressure (and hence lower levels of amino acid diversity) on their surface structural genes (21). In contrast, analyses of individual human and mosquito hosts infected with dengue virus (DENV) have found levels of nonsynonymous and synonymous genetic variation comparable to those seen in other, highly variable RNA viruses (3, 16, 17, 20). Moreover, while a variety of experimental studies confirm that selection pressures on arboviruses differ when the viruses are grown in insect and mammalian cells, there is conflicting evidence as to whether alternating replication among hosts produces negative fitness trade-offs such that traits favored in one host are selected against in another (2, 8, 18).

In an attempt to resolve the discrepancies concerning the evolutionary rates of and selection pressures acting on arboviruses, I analyzed the patterns of genetic variation in a representative and well-studied arbovirus—DENV—at different levels of evolutionary divergence. In particular, I compared the respective numbers of nonsynonymous and synonymous mutations within and among hosts and used the changing proportions of nonsynonymous mutations as a key indicator of the selection pressures affecting this virus.

Three different types of DENV data sets were compiled, representing increasing levels of evolutionary divergence: (i) one for virus populations within individual human hosts, (ii) one for epidemiologically linked virus populations, and (iii) one for populations representing the full extent of genetic diversity of two of the four DENV serotypes (Table 1). Four intrahost data sets were collected, all corresponding to DENV-3 in humans, and included full-length (20) and partial (17) envelope (E) gene sequences and partial sequences of the capsid and NS2B genes (16). I also collected four data sets representing DENV diversity within epidemiologically linked host populations, that is, viruses collected from the same country over a limited time span. All sequences in these sets came from the E gene, and sequences in three data sets, namely, those corresponding to viral populations collected from Thailand in 1987 (data compiled from GenBank), Venezuela during the period from 1997 to 2000 (14), and Vietnam during the period from 1997 to 1998 (12), were from DENV-2, for which most sequence data are available. I also analyzed a single data set, corresponding to viral populations collected in Venezuela during the period from 2000 to 2001 (15), of DENV-3 sequences. Finally, I examined genetic variation at a more distant phylogenetic level by compiling two data sets encompassing the full range of genetic diversity in DENV-2 and DENV-3. Where possible, five sequences were taken from each of the six defined genotypes of DENV-2 (12) and the five genotypes of DENV-3 (20). For genotypes with fewer than five sequences available, all sequences were included in the analysis. All sequences were aligned by eye and no insertions or deletions needed to be introduced.

To measure the change in selection pressures on DENV at different levels of evolutionary divergence, I compared the numbers of nonsynonymous and synonymous mutations. In the simplest analysis, the proportion of nonsynonymous mutations in each alignment (designated *pN*) was computed. To obtain a more accurate estimation of changing selection pressures and

* Mailing address: Department of Zoology, University of Oxford, South Parks Rd., Oxford OXI 3PS, United Kingdom. Phone: 44 1865 271282. Fax: 44 1865 310447. E-mail: Edward.Holmes@zoo.ox.ac.uk.

TABLE 1. Analysis of patterns of nonsynonymous variation in DENV at different levels of evolutionary divergence

Data set[a]	DENV type (place and date of collection)	No. of sequences	Length (bp)	π	pN	d_N/d_S[b]
Intrahost						
1	3	180	318	0.003	0.757	1.062
2	3	180	366	0.004	0.732	1.183
3	3	69	393	0.003	0.747	1.262
4	3	20	1,479	0.002	0.520	0.562
Average				0.003	0.689	1.017
Interhost (epideniologically linked populations)						
5	2 (Thailand, 1987)	19	1,485	0.008	0.067	0.030
6	2 (Venezuela, 1997–2000)	35	1,485	0.010	0.133	0.066
7	2 (Vietnam, 1997–1998)	16	1,485	0.017	0.164	0.077
8	3 (Venezuela, 2000–2001)	15	1,479	0.002	0.200	0.131
Average				0.009	0.141	0.076
Interhost (among genotypes)						
9	2	25	1,485	0.090	0.120	0.054
10	3	21	1,479	0.081	0.148	0.076
Average				0.086	0.134	0.065

[a] Sequences in data sets 1 and 2 were from capsid and NS2B genes, respectively. Sequences in all other data sets were from E genes. Data for each of sets 1 and 2 were collected from 18 patients; data for set 3 were collected from 6 patients.

[b] For data sets 1 to 3, d_N/d_S was calculated by dividing the numbers of nonsynonymous and synonymous mutations in samples from each patient by the average number of nonsynonymous and synonymous sites calculated across all patients. In individual pairwise comparisons in which there were no synonymous substitutions, d_N/d_S was set at 1.0.

to take account of multiple substitutions at single sites, I next compared the ratio of nonsynonymous (d_N) to synonymous (d_S) substitutions per site (d_N/d_S). This analysis was undertaken by using the CODEML program from the PAML package (22), as this enabled me to obtain a maximum likelihood estimate of d_N/d_S averaged across all sites in the alignment while taking into account the phylogenetic relationships of the sequences in question (as specified in the M0 model of codon evolution [23]). Finally, to analyze how overall levels of genetic diversity changed through time, I also calculated π, the average pairwise distance among the sequences in each data set.

The levels of nonsynonymous variation at different levels of evolutionary divergence are presented in Table 1. The most striking observation was that far more nonsynonymous variation was present within hosts than among them (in epidemiologically linked populations or DENV genotypes). On average, ~69% of mutations at the intrahost level were nonsynonymous (~75% if the small E gene data set from Wittke et al. [20] is excluded). Such a high proportion of nonsynonymous changes is very close to that expected (~70%) if both nonsynonymous and synonymous mutations occurred at random. In contrast, an average of only 14.1% of mutations within epidemiologically linked populations were nonsynonymous, and a similarly low proportion (13.4%) was detected among the different genotypes of DENV-2 and DENV-3. The reduction in frequency of nonsynonymous changes was also revealed in a comparison of d_N/d_S ratios at different levels of evolutionary divergence. In this case, the mean intrahost d_N/d_S ratio was 1.017 (1.169 if the E gene data set from Wittke et al. [20] is excluded), which is again very close to the value of ~1.0 expected if all mutations occurred randomly. Far lower d_N/d_S ratios were observed within local populations (mean $d_N/d_S = 0.076$) and among genotypes of DENV-2 and DENV-3 (mean $d_N/d_S = 0.065$). That both pN and d_N/d_S are similar among epidemiologically linked DENV populations and genotypes of DENV-2 and DENV-3 but differ substantially from the values seen for populations within hosts indicates that very different selection pressures are imposed on viral populations in the short term and in the long term. Moreover, the interhost d_N/d_S ratios are lower than those seen in many other RNA viruses (21) and confirm previous suggestions that DENVs are subject to relatively strong selective constraints such that most amino acid changes are deleterious.

More evidence that the majority of intrahost nonsynonymous mutations may be deleterious was the fact that they generally occurred as singletons, suggesting relatively low frequency in the population, and that the sites with these mutations were usually invariant in comparisons involving more distantly related DENVs. For example, 23 amino acid sites were variable in the virus sample from patient ID17 from the E gene data set of Wang et al. (17), yet only one of these sites exhibited any variation in an alignment of 68 sequences from the five genotypes of DENV-3 sampled on a global scale. That the remaining 22 amino acid sites were invariant in the global DENV-3 data set strongly suggests that these sites are subject to major selective constraints so that purifying selection purges nonsynonymous variation in the long term. Likewise, of the 13 amino acid changes observed in the E gene data set of Wittke et al. (20), only two occurred at sites that are variable in the global DENV-3 alignment. Finally, and as expected, overall genetic diversity within DENV populations increased with evolutionary divergence. The only exception was the DENV-3 population from Venezuela that was sampled during 2000 and 2001, in which no more genetic diversity was observed than that within individual hosts.

Taken together, these results reveal that the intrahost genetic variation in DENV represents the random mutation frequency in the population prior to the action of widespread natural selection, either positive or negative. Indeed, stop codons and deletion mutations, which are presumably deleterious, were observed in all the intrahost data sets (16, 17, 20). That the frequency of nonsynonymous mutations is greatly reduced in comparisons involving more distantly related viruses further indicates that the majority of these mutations are deleterious and eventually purged by purifying selection. To estimate the proportion of all nonsynonymous mutations that are deleterious in DENV, I compared the difference in mean d_N/d_S ratios within and among hosts (Table 1). As the mean d_N/d_S ratio in the interhost comparisons is less than 10% of its mean value within hosts, it can be roughly estimated that more than 90% of the nonsynonymous mutations that occur within hosts are removed by purifying selection in the long term. Moreover, because intrahost genetic variation must have built up over a number of replication cycles, the high numbers of nonsynonymous mutations observed at this time imply that any purifying selection within individual hosts must be relatively weak and that this selective purging most likely occurs when the virus replicates in the alternate host. However, there is currently insufficient data to ascertain whether different sites are selected in human and mosquitoes.

My calculation that over 90% of nonsynonymous mutations in DENV are deleterious is higher than the average figure of ~80% estimated for the human genome (5). Because all my interhost comparisons involved the E gene, which shows greater variability in corresponding amino acids than many other DENV genes because of sporadic positive selection (12, 13), it is also likely that an even greater frequency of deleterious mutations will be found in whole genome comparisons. Moreover, the fact that the DENV-3 population from Venezuela which had a low level of genetic variation also had the highest d_N/d_S ratio further suggests that the full force of purifying selection has yet to be experienced in this population. Finally, it is also possible that a proportion of synonymous changes are deleterious because of such factors as RNA secondary structure (10) and codon usage bias (6). While the frequency of deleterious synonymous mutations can be estimated by comparing changing base compositions through time (4), the available sample size of DENV (or any arbovirus) currently precludes this analysis.

Overall, the reduction in frequency of nonsynonymous mutations through time strongly supports the notion that there are major constraints acting on the long-term evolution of DENV and most likely of arboviruses in general. Although the high level of intrahost genetic variation confirms that arboviruses are not in evolutionary stasis (8), my finding that most of these mutations are deleterious explains why long-term rates of amino acid substitution are lower than those seen in many other RNA viruses (7). As such, measures of intrahost genetic variation should not be used to estimate long-term rates of nucleotide substitution. Moreover, while it is clear that arboviruses are able to adapt to diverse cell types, as revealed in experimental studies, many of the mutations produced will be deleterious in the alternate host. As such, perhaps the most telling measure of the selective constraints acting on arboviruses is the high frequency of deleterious mutations; although these mutations may only cause negative fitness trade-offs in some experimental systems (depending on the extent of epistasis and pleiotropy), arboviruses in general are likely to produce many more deleterious mutations than RNA viruses that replicate only in phylogenetically similar host species.

I thank Oliver Pybus, Michael Worobey, and two referees for constructive comments.

REFERENCES

1. **Clinis, M. J., W. Kang, and S. C. Weaver.** 1996. Genetic conservation of Highlands J viruses. Virology **218:**343–351.
2. **Cooper, L. A., and T. W. Scott.** 2001. Differential evolution of eastern equine encephalitis virus populations in response to host cell type. Genetics **157:** 1403–1412.
3. **Craig, S., H. M. Thu, K. Lowry, X.-F. Wang, E. C. Holmes, and J. Aaskov.** 2003. Diverse dengue type 2 virus populations contain recombinant and both parental viruses in a single mosquito host. J. Virol. **77:**4463–4467.
4. **Eyre-Walker, A. C.** 1999. Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. Genetics **152:**675–683.
5. **Fay, J. C., G. J. Wyckoff, and C.-I. Wu.** 2001. Positive and negative selection on the human genome. Genetics **158:**1227–1234.
6. **Jenkins, G. M., and E. C. Holmes.** 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res. **92:**1–7.
7. **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.
8. **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.
9. **Scott, T. W., S. C. Weaver, and V. L. Mallampalli.** 1994. Evolution of mosquito-borne viruses, p. 293–324. *In* S. S. Morse (ed.), Evolutionary biology of viruses. Raven Press, New York, N.Y.
10. **Simmonds, P., and D. B. Smith.** 1999. Structural constraints on RNA virus evolution. J. Virol. **73:**5787–5794.
11. **Strauss, J. H., and E. G. Strauss.** 1994. The alphaviruses: gene expression, replication, and evolution. Microbiol. Rev. **58:**491–562.
12. **Twiddy, S. S., J. F. Farrar, N. V. Chau, B. Wills, E. A. Gould, T. Gritsun, G. Lloyd, and E. C. Holmes.** 2002. Phylogenetic relationships and differential selection pressures among genotypes of dengue-2 virus. Virology **298:**63–72.
13. **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.
14. **Uzcategui, N. Y., D. Camacho, G. Comach, E. C. Holmes, and E. A. Gould.** 2001. The molecular epidemiology of dengue-2 virus in Venezuela: evidence for *in situ* viral evolution and recombination. J. Gen. Virol. **82:**2945–2953.
15. **Uzcategui, N. Y., G. Comach, D. Camacho, M. Salcedo, M. Cabello deq Uintana, M. Jimenez, G. Sierra, R. Cuello de Uzcategui, W. S. James, S. Turner, E. C. Holmes, and E. A. Gould.** 2003. The molecular epidemiology of dengue-3 virus in Venezuela. J. Gen. Virol. **84:**1569–1575.
16. **Wang, W.-K., T.-L. Sung, C.-N. Lee, T.-Y. Lin, and C.-C. King.** 2002. Sequence diversity of the capsid gene and the nonstructural gene NS2B of dengue-3 virus *in vivo*. Virology **303:**181–191.
17. **Wang, W.-K., S.-R. Lin, C.-M. Lee, C.-C. King, and S.-C. Chang.** 2002. Dengue type 3 virus in plasma is a population of closely related genomes: quasispecies. J. Virol. **76:**4662–4665.
18. **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.
19. **Weaver, S. C., R. Rico-Hesse, and T. W. Scott.** 1992. Genetic diversification and slow rates of evolution in New World alphaviruses. Curr. Top. Microbiol. Immunol. **176:**99–117.
20. **Wittke, V., T. E. Robb, H. M. Thu, S. Nimmannitya, S. Kalayanrooj, D. W. Vaughn, T. P. Endy, E. C. Holmes, and J. G. Aaskov.** 2002. Extinction and rapid emergence of strains of dengue 3 virus during an interepidemic period. Virology **301:**148–156.
21. **Woelk, C. H., and E. C. Holmes.** 2002. Reduced positive selection in vector-borne RNA viruses. Mol. Biol. Evol. **19:**2333–2336.
22. **Yang, Z.** 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:**555–556.
23. **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.