Phylogenetic Evidence for Recombination in Dengue Virus

Edward C. Holmes, Michael Worobey, and Andrew Rambaut

The Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, U.K.

A split decomposition analysis of dengue (DEN) virus gene sequences revealed extensive networked evolution, indicative of recombination, among DEN-1 strains but not within serotypes DEN-2, DEN-3, or DEN-4. Within DEN-1, two viruses sampled from South America in the last 10 years were identified as recombinants. To map the breakpoints and test their statistical support, we developed a novel maximum likelihood method. In both recombinants, the breakpoints were found to be in similar positions, within the fusion peptide of the envelope protein, demonstrating that a single recombination event occurred prior to the divergence of these two strains. This is the first report of recombination in natural populations of dengue virus.

Introduction

Dengue fever (DF) is the most common vector-borne viral disease of humans, with at least 100 million cases recorded each year in Africa, Latin America, Oceania, and most notably, Southeast Asia. Most patients develop an acute febrile illness, which, although debilitating, rarely causes a fatal outcome. Others, however, go on to experience two more serious clinical conditions—dengue hemorrhagic fever (DHF) and dengue shock syndrome (DSS), both of which are major causes of morbidity and mortality if untreated and appear to have increased in prevalence since World War II (Monath 1994). The virus is transmitted between human and monkey hosts by mosquitoes of the genus *Aedes*, and principally *Aedes aegypti*, which often breeds in areas of human inhabitation. The ever-increasing size, urbanization, and mobility of human populations, as well as the lack of an effective vaccine, are major factors in the emergence of dengue (DEN) as one of the most serious health problems facing the developing world (Zanotto et al. 1996; Gubler 1998).

The causative, single-stranded, positive-sense RNA virus (of approximately 11 Kb in length) comprises four serotypes (DEN-1 to DEN-4), which now cocirculate in many localities. The genetic variation within each serotype can be further partitioned into different viral genotypes; some have restricted geographical distributions, while others are more cosmopolitan, reflecting their dispersal across the tropical world (Rico-Hesse 1990).

Despite this genetic diversity, dengue and the other *Flaviviridae* (which include those viruses responsible for yellow fever, Japanese encephalitis, and hepatitis C) have, to date, only been shown to accumulate variation through mutation, with no role for recombination (Blok et al. 1992). Indeed, recombination has been rarely reported in positive-strand RNA viruses (Lai 1992), with most documented examples occurring within the picornaviruses, coronaviruses, and the alphaviruses, the latter of which are also transmitted by mosquito vectors (Hahn et al. 1988). However, the rising prevalence of dengue virus and the greater mixing of host and vector populations have led to suggestions that genetic exchange between strains is an increasing possibility (Kuno 1997).

To test the hypothesis that dengue represents a clonal RNA virus we employed methods which detect conflicting phylogenetic signals in gene sequence data and which have proven useful in the documentation of recombination in other viruses (Robertson et al. 1995; Bollyky et al. 1996). In the first part of this analysis, the evolutionary relationships among gene sequences from the four serotypes of dengue virus were reconstructed using split decomposition, a method which depicts all the shortest pathways linking sequences, including those that produce an interconnected network, as expected under recombination (Bandelt and Dress 1992; Dopazo, Dress, and Von Haeseler 1993). To further characterize putative recombination events, we have developed and employed a new statistical method.

Methods

The main set of dengue sequences we analyzed were from the viral envelope (E) gene. These data were mainly collected from GenBank (one sequence was entered by hand) and aligned manually. The E gene data set comprised 15 sequences from DEN-1, 69 from DEN-2, 26 from DEN-3, and 20 from DEN-4 viruses, with an average length of 1485 bp (base pairs). Supplementary sequence data was also available from the capsid (C) gene (33 sequences representing all serotypes [342 bp]), the pre-membrane/membrane (prM/M) gene (58 sequences from all serotypes [498 bp]), and the NS1 gene (31 sequences from DEN-1 and DEN-2 [1056 bp]). All sequence alignments used here may be obtained from the authors on request.

Because recombination produces networks of sequences rather than strictly bifurcating evolutionary trees (Fitch 1997), the initial search for recombination in dengue virus was conducted using split decomposition, a method which depicts parallel edges between sequences if there are conflicting phylogenetic signals in the data. This analysis was undertaken with the SplitsTree program (versions 1.0.1 and 2.4; Huson 1998) with the input pairwise distances between sequences, on which the splits graph is constructed, estimated by the Kimura 3-ST method (Kimura 1981), although similar results were found with other distance measures.

Key words: dengue virus, phylogeny, maximum likelihood, networks, recombination, split decomposition.

Address for correspondence and reprints: Dr. Edward C. Holmes, The Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS U.K. E-mail: Edward.Holmes@zoo.ox.ac.uk.

Mol. Biol. Evol. 16(3):405–409. 1999

Although split decomposition represents a useful way to visualize the non-treelike evolutionary relationships between sequences, it does not allow individual recombination events to be mapped nor the statistical support for them to be assessed. To do this, we have developed a likelihood method which tests the disparity in phylogenetic signal for different parts of the alignment, an extension of the method of Maynard Smith (1992). The analysis consists of estimating the branch lengths of a tree containing the putative recombinant sequence and two ''parental sequences'' (the nature of which we will return to). With just three sequences, there is only one tree topology, massively reducing the computational complexity of the problem.

The likelihood of the null hypothesis, H_0, that there has been no recombination event is therefore assessed as a simple, unrooted tree of three sequences in the manner described by Felsenstein (1981). This null-hypothesis model has three free parameters (the tree branch lengths). To obtain the likelihood of the alternative hypothesis, H_1, we allow a different tree (more specifically, different branch lengths) either side of a breakpoint, *k*, along the sequences. We now have six free parameters in the model, which will produce a better fit to the data and hence a better likelihood. By moving *k* along the alignment, always optimizing the branch lengths, we obtain the maximum likelihood position of the recombination breakpoint and the likelihood of H_1. We now need to assess whether H_1 is a significantly better fit to the data, i.e., that the likelihood ratio of H_1 to H_0 is greater that we would expect by chance. This can be done with Monte Carlo simulation (for example, Goldman 1993; method described in Rambaut and Grassly 1997), in which sequences are simulated 500 times each under the maximum likelihood H_0 model and then analyzed by the same procedure as for the real data to obtain a null distribution of likelihood ratios between the hypotheses of recombination and no recombination.

The method described is actually a simplification of one in which all sequences in the data set are used and where two different tree topologies (as well as branch lengths) are fitted to either side of the breakpoint. Computational constraints, however, make estimating two trees separately for each possible breakpoint prohibitively slow. Reducing the problem to three representative sequences (the recombinant and two parents) is computationally less intensive but will make the test of phylogenetic disparity more conservative. To minimize this loss of power, it is therefore important to identify suitable parental sequences, namely those in the data set that are closest to each recombinant region of the mosaic sequence. Candidate parental sequences should be readily identifiable from the prior split decomposition analysis.

Although rate heterogeneity among sites is accommodated, our method will not be able to distinguish recombination from cases in which some contiguous regions of the alignment show elevated or reduced rates in only one of the three lineages. However, once the breakpoints have been identified, phylogenies can then be constructed on the different regions and the extent of incongruence between them assessed. A computer program, written in C, which performs this analysis is available at the authors' web site (http://evolve.zoo.ox.ac.uk/Lard/Lard.html).

All maximum likelihood trees used here were constructed with the test version 4.0d63 of PAUP*, kindly provided by D. L. Swofford, incorporating the HKY85 model of DNA substitution with the substitution rates for each codon position and the transition:transversion ratio estimated from the empirical data. No significant difference in the results was obtained by using the gamma distribution of rate heterogeneity among sites.

Results and Discussion

The split decomposition analysis provided strong evidence for networked evolution among DEN-1 viruses (fig. 1). To determine which of these strains were the most likely recombinants, we mapped each nucleotide site successively onto the splits graph. This analysis revealed that Brazil (BR/90), isolated from a DF patient infected during the first wave of an epidemic that swept through Rio de Janeiro in 1990/91, and French Guiana (FGA/89), collected from a DF patient in 1989 (both viruses described in Desprès, Frenkiel, and Deubel 1993), contained two distinct regions in their E genes, i.e., regions that showed closest connections to different viruses in the sample. For both BR/90 and FGA/89, the majority of their E gene sequence closely matched that of a virus isolated from a DHF patient in Singapore in 1990 (Singapore S275/90) (Fu et al. 1992) and which can be assigned to genotype I of DEN-1, while the remainder most strongly resembled a cluster of viruses collected from the Caribbean region during 1977 to 1985 (Chu, Rourke, and Trent 1989) and which represent strains from genotype IV. No good evidence for recombination was found in the other dengue serotypes or genes, with any networking observed more likely due to stochastic processes than genetic exchange (the splits graph for DEN-2, which is illustrative, is shown in figure 2, and the others are available from the authors on request).

To characterize the mosaic history of BR/90 and FGA/89 in more detail, we undertook a likelihood analysis of these and five other DEN-1 strains for which contiguous C, prM/M, and E gene sequences were available (a total length of 2325 bp). These five strains (with simplified names as used in fig. 3) were Singapore S275/90 (''Singapore''), Caribbean CV1636/770 (''Jamaica''), Philippines 836–1 (''Philippines'') , AHF82–80 (''Thailand''), and WestPac (''Nauru'').

For both BR/90 and FGA/89, the split decomposition analysis identified Jamaica and Singapore as the closest parental sequences for the different recombinant segments. By applying our new maximum likelihood method, we were then able to find the breakpoints in the sequence which gave the highest likelihood under an evolutionary model incorporating recombination. These were located in the same region for both recombinants, before nucleotides 1146 and 1152 for BR/90 and FGA/89, respectively (corresponding to nucleotides

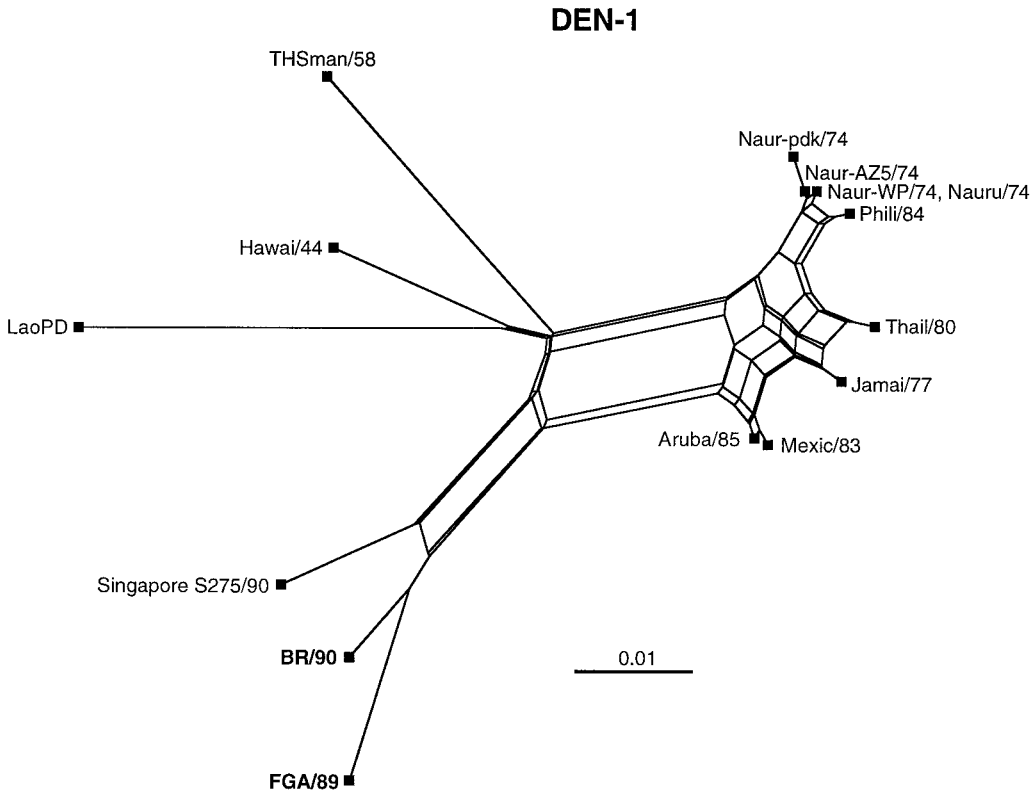

FIG. 1.—Split decomposition analysis of 15 DEN-1 E gene sequences. The observation that the viruses in the sample are linked to each other by multiple pathways, thereby forming an interconnected network rather than a single bifurcating tree, is suggestive of recombination. The two viruses with the strongest signal of recombination, BR/90 and FGA/89, are shown in boldface. All branch lengths are drawn to scale.

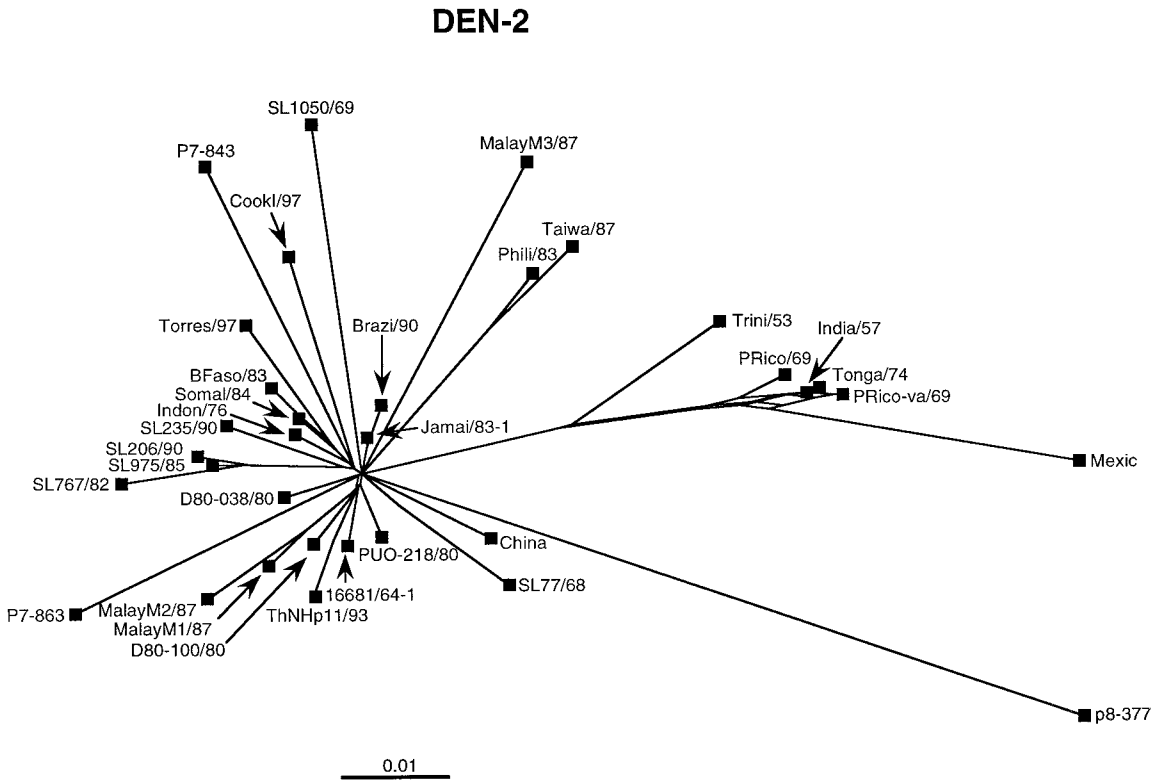

FIG. 2.—Split decomposition analysis of 33 DEN-2 E gene sequences. Because of the large number of DEN-2 E gene sequences available, many of which are very closely related, only a representative sample are presented here for purposes of clarity. Similar findings were obtained in a larger analysis of 69 strains and for the other serotypes (not shown). All branch lengths are drawn to scale.

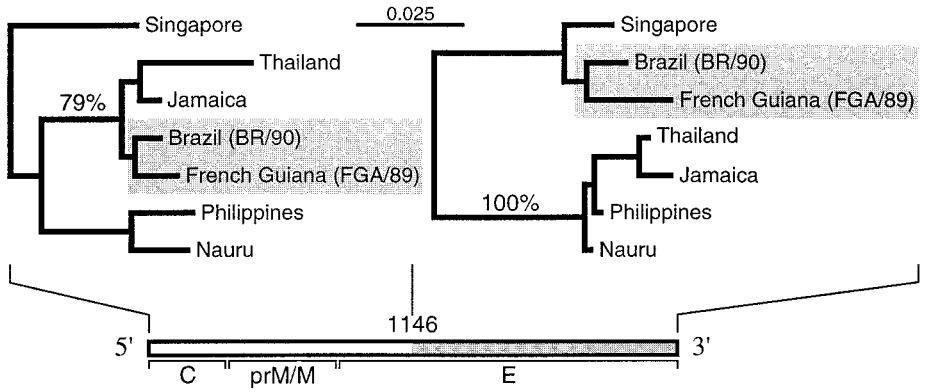

FIG. 3.—Phylogenetic support for recombination in DEN-1 virus. The maximum likelihood trees (with associated bootstrap values) supporting the contrasting phylogenetic positions of strains BR/90 and FGA/89 either side of nucleotide 1146 in a contiguous alignment of the C, prM/M, and E genes are shown. The nucleotide 1146 breakpoint was identified in a prior maximum likelihood-based analysis. All branch lengths drawn to scale. The GenBank accession numbers for each sequence are as follows: Brazil (BR/90)—S64849, French Guiana (FG/89)—not available on GenBank (sequence entered by hand from the original publication), Singapore—M87512, Jamaica—D00501, Philippines—D00503, Thailand—D00502, Nauru—M23027.

306 and 312 in the E gene—part of the conserved fusion peptide), indicating that these strains are descendants of a single chimeric virus (bottom part of fig. 3). Monte Carlo simulation revealed that the log likelihood ratios of the two recombinants (42.2 for BR/90 and 42.7 for FGA/89) were significantly greater than for any of the simulated data (maxima of 14.1 and 12.1, respectively). Comparable results were obtained using the method of Maynard Smith (1992), with all P values < 0.001 under a randomization test. Finally, and to provide another indication of the conflicting phylogenetic signals in these data, maximum likelihood trees were constructed (by an ''exhaustive'' search) for each of the two proposed recombinant regions, arbitrarily choosing the nucleotide 1146 breakpoint, along with a bootstrap analysis involving 1000 replicate maximum likelihood trees (constructed by an ''heuristic'' search strategy; top part of fig. 3).

The significant discrepancy between phylogenetic trees inferred for nucleotide sequences upstream and downstream of a breakpoint in the E gene constitutes powerful evidence for recombination in these DEN-1 viruses. Although BR/90 and FGA/89 were sequenced in the same laboratory, the possibility that contamination resulted in the production of ''false'' recombinants seems remote, given that both viruses were plaque purified prior to PCR amplification and differ by 2.4% in nucleotide sequence. Furthermore, the sequences of their parents, Jamaica and Singapore, were determined in different laboratories and differ from BR/90 and FGA/89 by an average of 2.0% and 3.3% in their respective parental regions. Finally, it is noteworthy that when a split decomposition analysis was undertaken with BR/90 and FGA/89 removed, networked evolution was still observed, suggesting that other DEN-1 viruses may also be recombinants, although with a weaker phylogenetic signal.

It is unclear whether the recombination event that produced the ancestor of BR/90 and FGA/89 took place in a human host coinfected with multiple strains, as has been observed on occasion with different serotypes (Kuno 1997), or within a mosquito vector which had fed on different hosts. However, when two viruses infect a single cell, it is theoretically possible for recombination to occur through copy-choice if the RNA polymerase switches between genomes during viral replication. Although recombination in dengue may be rare, especially between serotypes which differ substantially in sequence, it could play some role in the development of strains with new phenotypic properties, including increased virulence and transmission potential. Furthermore, the observation that the recombinants were collected in different locations in South America shows that these viruses can spread successfully through human populations.

While no dengue vaccine has been approved to date, clinical trials are currently taking place in Thailand with a tetravalent candidate vaccine, which simultaneously introduces live attenuated strains of each of the four serotypes (Bhamarapravati and Yoksan 1997). Our results suggest that the possibility of genetic exchange between vaccine strains, as well as between wild viruses, merits careful consideration.

Acknowledgments

This work was supported by grants from The Royal Society, The Wellcome Trust, and The Rhodes Trust.

LITERATURE CITED

BANDELT, H.-J., and A. W. M. DRESS. 1992. Split decomposition: a new and useful approach for the phylogenetic analysis of distance data. Mol. Phylogenet. Evol. **1**:242–252.

BHAMARAPRAVATI, N., and S. YOKSAN. 1997. Live attenuated tetravalent dengue vaccine. Pp. 367–377 *in* D. J. GUBLER and G. KUNO, eds. Dengue and dengue hemorrhagic fever. CAB International, Wallingford, U.K.

BLOK, J., Y. H. MCWILLIAM, H. C. BUTLER, A. J. GIBBS, G. WEILLER, B. L. HERRING, A. C. HEMSLEY, J. G. AASKOV, S. YOKSAN, and N. BHAMARAPRAVATI. 1992. Comparison of Dengue-2 virus and its candidate vaccine derivative: sequence relationships with the flaviviruses and other viruses. Virology **187**:573–590.

BOLLYKY, P. L., A. RAMBAUT, P. H. HARVEY, and E. C. HOLMES. 1996. Recombination between sequences of hepatitis B virus from different genotypes. J. Mol. Evol. **42**: 97–102.

CHU, M. C., E. J. O'ROURKE, and D. W. TRENT. 1989. Genetic relatedness among structural protein genes of dengue 1 virus strains. J. Gen. Virol. **70**:1701–1712.

DESPRÈS, P., M.-P. FRENKIEL, and V. DEUBEL. 1993. Differences between cell membrane fusion activities of two dengue type-1 isolates reflect modifications of viral structure. Virology **196**:209–219.

DOPAZO, J., A. DRESS, and A. VON HAESELER. 1993. Split decomposition: a technique to analyse viral evolution. Proc. Natl. Acad. Sci. USA **90**:10320–10324.

FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

FITCH, W. M. 1997. Networks and viral evolution. J. Mol. Evol. **44**:S65-S75.

FU, J., B.-H. TAN, E.-H. YAP, Y.-C. CAN, and Y. H. TAN. 1992. Full-length cDNA sequence of dengue type 1 virus (Singapore strain S275/90). Virology **188**:953–958.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. **36**:182–198.

GUBLER, D. J. 1998. Dengue and dengue hemorrhagic fever. Clin. Microbiol. Rev. **11**:480–496.

HAHN, C. S., S. LUSTIG, E. G. STRAUSS, and J. H. STRAUSS. 1988. Western equine encephalitis virus is a recombinant virus. Proc. Natl. Acad. Sci. USA **85**:5997–6001.

HUSON, D. H. 1998. SplitsTree: a program for analyzing and visualizing evolutionary data. Bioinformatics **14**:68–73.

KIMURA, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA **78**:454–458.

KUNO, G. 1997. Factors influencing the transmission of dengue viruses. Pp. 61–88 *in* D. J. GUBLER and G. KUNO, eds. Dengue and dengue hemorrhagic fever. CAB International, Wallingford, U.K.

LAI, M. M. C. 1992. RNA recombination in animal and plant viruses. Microbiol. Rev. **56**:61–79.

MAYNARD SMITH, J. 1992. Analysing the mosaic structure of genes. J. Mol. Evol. **34**:126–129.

MONATH, T. P. 1994. Dengue: the risk to developed and developing countries. Proc. Natl. Acad. Sci. USA **85**:7627–7631.

RAMBAUT, A., and N. C. GRASSLY. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. CABIOS **13**:235–238.

RICO-HESSE, R. 1990. Molecular evolution and distribution of dengue viruses type 1 and type 2 in nature. Virology **174**: 479–493.

ROBERTSON, D. L., P. M. SHARP, F. E. MCCUTHAN, and B. H. HAHN. 1995. Recombination in HIV-1. Nature **374**:124–126.

ZANOTTO, P. M. DE A., E. A. GOULD, G. F. GAO, P. H. HARVEY, and E. C. HOLMES. 1996. Population dynamics of flaviviruses revealed by molecular phylogenies. Proc. Natl. Acad. Sci. USA **93**:548–553.

MANOLO GOUY, reviwing editor

Accepted December 7, 1998