Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 4702–4706, August 1984
Biochemistry

An evolutionary tree relating eight alphaviruses, based on amino-terminal sequences of their glycoproteins

(togaviruses/Sindbis virus/equine encephalitis viruses/mosquito-borne viruses)

JOHN R. BELL*†, RICHARD M. KINNEY‡, DENNIS W. TRENT‡, ELLEN G. STRAUSS*, AND JAMES H. STRAUSS*§

*Division of Biology, California Institute of Technology, Pasadena, CA 91125; and ‡Division of Vector-Borne Viral Diseases, Center for Infectious Diseases, Centers for Disease Control, Fort Collins, CO 80522

Communicated by James Bonner, April 17, 1984

ABSTRACT **The NH_2-terminal amino acid sequences of both structural glycoproteins of each of eight alphaviruses have been obtained. These sequences demonstrate that the alphaviruses are all closely related and have in all probability descended from a common ancestor. Cysteines are conserved as well as several other residues important for secondary structure, suggesting that the three-dimensional conformations of the alphavirus glycoproteins are conserved while considerable variation in the primary sequence has evolved. Secondary structure predictions based upon the amino acid sequences are consistent with this hypothesis. An evolutionary tree for these eight alphaviruses has been constructed from the amino acid sequence data and, at many positions in the sequence, the amino acids present in the ancestral glycoproteins have been deduced.**

The *Alphavirus* genus of the family Togaviridae contains about 25 members, many of which have geographical variants (1). The vertebrate host range of this group of viruses is quite wide, and alphaviruses have been isolated from numerous species of birds and mammals as well as from reptiles and amphibians (2). Many of these viruses are important human or veterinary pathogens, including three equine encephalitis viruses, western, eastern, and Venezuelan, which are epidemic in the United States (3). These viruses are transmitted in nature by mosquitos and must replicate in the arthropod vector as well as in the vertebrate host. Because of this alternation between arthropod and vertebrate hosts and the very wide vertebrate host range, the selective pressures operative during alphavirus evolution are expected to be quite different in scope from those that affect viruses restricted to a few host organisms.

To date, inclusion within the *Alphavirus* genus and the relationships among the various alphaviruses have been defined by serology, using the extent of cross-reaction between members in neutralization, complement fixation, and hemagglutination-inhibition tests and cross-protection experiments in host animals. This has allowed the grouping of these viruses into six subgroups (4). Although this approach has proven extremely useful in the identification and classification of virus isolates, it is clear that much more information on the relationships of these viruses can be obtained from a comparison of the primary amino acid sequences of their structural proteins. Such an approach, the comparison of the primary structures of common proteins, has proven useful in the determination of phylogenetic relationships among eukaryotes (5). Similarly, recent studies of influenza virus (6, 7) have compared the protein sequences of the hemagglutinin (HA) protein from a number of strains.

Efforts have been made to examine relationships between alphaviruses by examining sequence similarities indirectly by nucleic acid hybridization (8) or by RNase digestion patterns (9). However, the extent of divergence at the nucleotide level is so great among the alphaviruses that little or no similarity could be detected except between the most closely related members of the genus. This divergence arises in part from the fact that alphavirus nucleotide sequences encoding either nonstructural proteins (10–12) or structural proteins (13, 14) have diverged markedly and even in regions in which the amino acid sequences are highly conserved, the codon used for any conserved amino acid has been essentially randomized. Thus the evolutionary divergence among the alphaviruses is extensive and the conservation of amino acid sequence results from functional requirements and is of particular relevance.

We have now obtained the NH_2-terminal amino acid sequences of the two virus glycoproteins, E1 and E2, for eight alphaviruses, including representatives of five subgroups, and these data have allowed the construction of an evolutionary tree for these viruses.

METHODS

Virus Strains. The eight alphaviruses used in this study are listed in Table 1 together with abbreviations to be used in this report. EEE and VEE (two strains, the virulent Trinidad Donkey strain and the TC-83 vaccine strain) were grown in Fort Collins. WEE (McMillan strain) was grown in Tokyo (15), SIN (AR339 and HR strains, obtained originally from E. R. Pfefferkorn), RR (strain T48, obtained from R. Shope) (14), MID (obtained from P. Marcus), SF (obtained from L. Dalgarno), and BF (14) were grown in Pasadena.

Preparation of Virus Glycoproteins. EEE and VEE were grown and purified by the methods of Trent *et al.* (16, 17) and the glycoproteins were separated by polyacrylamide gel electrophoresis. SIN, MID, RR, SF, and BF were grown and purified by the methods of Bell *et al.* (18) and the glycoproteins were separated by chromatography on glass wool. This procedure utilizes the fact that alphavirus E2 binds to glass wool under appropriate conditions whereas E1 does not. WEE glycoproteins were also separated by glass wool chromatography, as previously described (15, 19).

Amino Acid Sequencing. Purified glycoproteins were subjected to automated Edman degradation on a noncommercial spinning cup sequencer (20). Reversed-phase high-pressure liquid chromatography was used to separate the mixture of phenylthiohydantoin amino acid derivatives released at each cycle (21). In some cases, to permit the determination of cysteines, the glycoprotein was reduced with dithiothreitol and the cysteines were modified with iodoacetamide prior to sequencing; in other cases cysteine was not determined.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: see Table 1 for abbreviations of viruses.

†Present address: Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080.
§To whom reprint requests should be addressed.

Table 1. Alphaviruses studied

Virus	Abbreviation	Subgroup*	Geographical distribution
Sindbis	SIN	WEE	Old World†
Western equine encephalitis	WEE	WEE	Americas
Venezuelan equine encephalitis	VEE	VEE	Americas
Barmah Forest	BF	?	Australia
Eastern equine encephalitis	EEE	EEE	Americas
Semliki Forest	SF	SF	(Africa)‡
Ross River	RR	SF	Australia
Middelburg	MID	MID	South Africa

*From Calisher *et al.* (4).
†Africa, Eastern Europe, India, Southeast Asia, Philippines, and Australia.
‡Isolated originally in Uganda; exact geographical distribution undefined.

Construction of the Phylogenetic Tree. The sequence data from the E1 glycoproteins of the eight alphaviruses through residue 50 were used to construct a phylogenetic tree by manual application of the algorithm described by Foulds *et al.* (22). In this approach the tree is built by sequentially adding branches. At each iteration, the unplaced sequence most closely related to another sequence is added, and the overall length of the tree is minimized before the next iteration. No attempt was made to weight distances by the minimum number of nucleotide changes required, so all distances reflect amino acid substitutions only. The location of the root of the tree, the ancestral sequence, cannot be fixed by this procedure. We have placed the root so that the total distances on each side of the root are equal.

RESULTS

The NH_2-terminal amino acid sequences of the two glycoproteins of the eight alphaviruses are shown in Fig. 1, using a format that emphasizes homology. For SIN, RR, SF, and MID the sequences have been extended by using nucleotide sequence data as described in the figure legend. The data for each protein are presented as a comparison with a consensus sequence, which lists at each position the most common (or one of the equally most common) amino acids found at that position in the alphavirus sequences. At many positions it has been possible to deduce the ancestral amino acid (see below) for the genes for E1 and E2; these ancestral amino acids are boxed in the two consensus sequences. Certain highly conserved regions or residues have been shaded to emphasize their conservation. Cysteines, aromatic residues, and certain prolines and glycines tend to be conserved, suggesting that the secondary structure of the glycoproteins is conserved.

The sequence data show that all eight alphaviruses are closely related and that all of the E1 glycoproteins are almost certainly derived from a common ancestral gene, as is also the case for the E2 glycoproteins (27). The degree of homology among these viruses in the NH_2-terminal region examined is shown in Table 2 for both E1 and E2. The homology varies

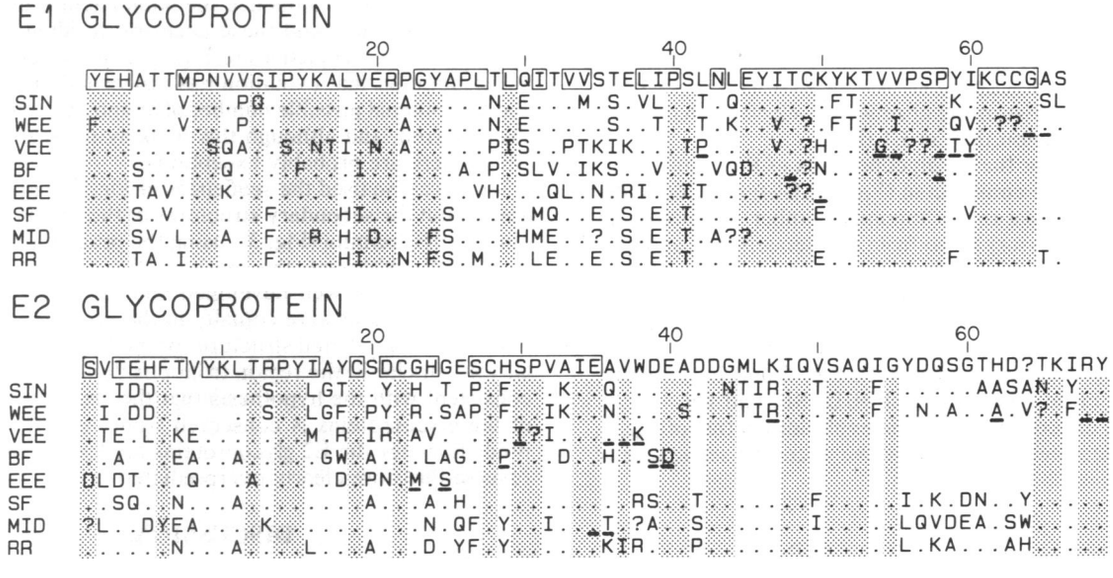

FIG. 1. NH_2-terminal amino acid sequences of E1 and E2 for eight different alphaviruses. The single-letter amino acid code (A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr) is used. A "consensus" sequence is shown in the first line and all virus sequences are shown in relation to this: a dot means the amino acid is the same as in the consensus sequence; if the amino acid at any position differs from this consensus sequence, the changed amino acid is shown. Gaps have been introduced as necessary to maintain the alignment. ? indicates no assignment was made; underlining indicates some uncertainty in the assignment. In the case of SIN (13), SF (23), and RR (14), the sequences have been confirmed and extended by using the amino acid sequences deduced from nucleotide sequence analysis. For MID E2, positions 2–40 were determined by amino acid sequencing and positions 39–69 by nucleotide sequencing. For this single-stranded complementary DNA was synthesized, and *Hae* III restriction fragments were sequenced and aligned by using methods previously described (10). The amino acid sequence data for SIN (24, 25), WEE (15), and BF (26) have been previously reported. The SF data obtained by us include residues 1–33 for E1 (except that residue 26 was not determined) and 4–11 for E2 and agrees with the sequence reported by Garoff *et al.* (23) from nucleotide sequence analysis. For RR we obtained the amino acid sequence of residues 1–11 and 13–15 of E1 and 1–36 of E2, which was in perfect agreement with that found by nucleotide sequence analysis (14). For VEE the sequence was obtained for both the TRD and TC-83 strains; where these sequences differ (see text) the sequence given here is that of the TRD strain. TC-83 E2 residues 1–17 were confirmed by nucleotide sequence analysis, using the methods described above for MID E2. Certain highly conserved residues or areas are shaded. In the case of E1 of WEE, VEE, BF, and EEE the protein was not reduced and alkylated so that cysteines were not recovered. The failure to identify an amino acid at position 49 in these cases suggests that cysteine is present as it is in the other viruses examined. Similarly, residues 62 and 63 of WEE E1 are likely to be cysteine. Residues boxed in the consensus sequences in the regions 1–60 in E1 and 1–35 in E2 are those amino acids deduced to be present in the ancestral genes for these proteins (see text).

Table 2. Homology between alphavirus proteins

		E1/E1							
		SIN	WEE	VEE	BF	EEE	SF	MID	RR
E2/E2	SIN	—	81	47	59	56	55	40	53
	WEE	72	—	49	54	56	57	45	51
	VEE	23	26	—	47	44	46	36	40
	BF	34	29	30	—	48	63	48	56
	EEE	26	35	40	32	—	63	45	56
	SF	40	36	49	60	44	—	74	82
	MID	40	39	43	42	42	60	—	71
	RR	42	44	41	50	40	74	64	—

The homology, in percent, between the alphavirus glycoproteins over the residues identified in Fig. 1. When gaps are required for alignment, each gap, regardless of length, is considered as one comparison and is considered to equal three differences. Numbers above and to the right of the diagonal refer to E1/E1 comparisons and those below and to the left of the diagonal refer to E2/E2 comparisons.

from 32% to 83%, depending upon which glycoproteins are being compared, and E1 is more highly conserved than E2. An excellent example of the latter is the fact that no gaps or insertions are required to align the E1 sequences, but four gaps (one each in SIN, WEE, VEE, and BF) are required to align the E2 sequences.

We have also examined two strains of VEE, the virulent Trinidad donkey strain (TRD) and the avirulent TC-83 vaccine strain. The TC-83 sequence of E1 extended for 36 residues and was identical to that for the TRD strain. The E2 sequence for TC-83 extended for 26 residues, and there were two amino acid substitutions: Lys-7 of TRD E2 becomes Asn in TC-83, and Ile-20 of TRD becomes Pro in TC-83. It is of interest that the only changes found occur in E2, the less highly conserved glycoprotein. A similar situation has been found for SIN in that in the absence of selection changes in E2 are more common than in E1 (25).

From the data for the E1 glycoproteins of Fig. 1 an evolutionary tree can be constructed (Fig. 2). Distances along the tree are proportional to the minimal number of amino acid substitutions necessary to go from one sequence to another. Sequence data for E2 are in many cases more limited than for E1, and the divergence among the E2 glycoproteins is so extensive that there are few positions that supply the information necessary for tree construction. Nonetheless, the E2 data are in general consistent with Fig. 2 and in particular support the close relationship of MID, SF, and RR.

Using the data of Fig. 1 and the evolutionary tree of Fig. 2, we have attempted to determine the ancestral sequences of

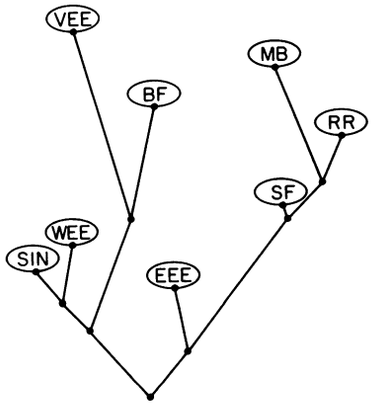

FIG. 2. Evolutionary tree for eight alphaviruses. The distance along any line is proportional to the number of amino acid substitutions along that line.

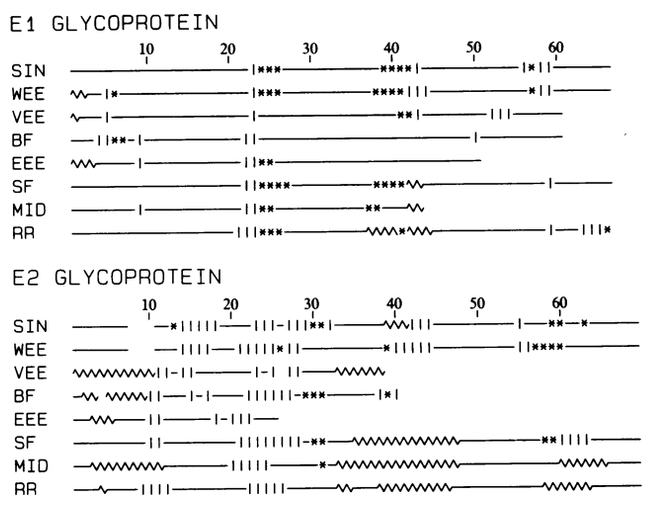

FIG. 3. Structural predictions for the NH_2-terminal regions of the alphavirus proteins E1 and E2. ——, β-sheet; $\mathcal{M}$, α-helix; |, reverse turns; and *, random coil.

the NH_2-terminal regions of the genes for E1 and E2. To accomplish this, we used the algorithm described by Fitch (28) to deduce the possibilities at each amino acid position for all of the intermediate sequences in the tree in such a way that we account for the sequences in Fig. 1 with the minimal number of amino acid substitutions. At some positions in the proteins the ancestral sequences cannot be determined from among two or more possibilities. However, at many positions we were able to deduce the amino acid present in the ancestral sequence, and these ancestral residues are boxed in the consensus sequences of Fig. 1. This exercise requires the assumption that evolution has proceeded in such a way that the minimal possible number of amino acid substitutions has occurred, and although this is likely to be true in most cases, there are probably exceptions.

We have also used the sequences of Fig. 1 to make predictions about the secondary structures of the alphavirus glycoprotein NH_2 termini (Fig. 3), using the procedure of Garnier *et al.* (29). Although the methods for predicting protein structure are not well developed and probably only about 50% of the residues are actually in the predicted conformation (29), the predicted structures of the E1 glycoproteins are all very similar in the region from positions 10 to 50 and are consistent with the hypothesis that the secondary structure of the glycoproteins is conserved from virus to virus. The predictions for the E2 glycoproteins are less satisfactory, as reverse-turns are clearly overpredicted.

DISCUSSION

The two surface glycoproteins of the alphaviruses, E1 and E2, are synthesized as part of a polyprotein precursor from the subgenomic 26S mRNA (reviewed in ref. 30). These glycoproteins are inserted during translation into the endoplasmic reticulum, glycosylated and acylated, and transported to the plasma membrane in the same way as the normal cell surface proteins of the host cell. Virus maturation occurs as preformed nucleocapsids bud through the plasma membrane, acquiring the viral envelope of lipid and E1 and E2. Host proteins are rigorously excluded from budding virions (31).

As has been noted, E1 is more highly conserved than E2 (Fig. 1 and Table 2). Although only the NH_2-terminal regions have been examined here, results with three viruses for which the complete amino acid sequence of the glycoproteins is known from nucleotide sequence analysis, SIN (13), SF (23), and RR (14), reveal that this is also true of the glyco-

proteins in their entirety. This is illustrated in Fig. 4, where the glycoproteins of SIN and RR are compared throughout their length with a homology routine in which the percent homology over consecutive 25 amino acid stretches is plotted as a moving average. These two viruses have been chosen for illustration because from Fig. 2 they are distantly related alphaviruses. Also indicated on the figure are some landmarks in the glycoproteins. Certain regions of the glycoproteins are more highly conserved than others, but E1 shows greater conservation overall than E2. Note also that the homology in the NH_2-terminal regions is not atypical of that found for the complete protein.

In view of the different rates of sequence divergence of E1 and E2, it is of interest to examine the functions of these two glycoproteins in the virus life cycle. These two glycoproteins form a heterodimer (33, 34), which appears to be the functional unit on the surface of the virus and which possesses the activities of hemagglutination and cell attachment (believed to be manifestations of the same activity, that for attachment to susceptible cells to initiate the infection process), fusion (initiation of infection by fusion of the virus membrane with an endosomal or lysosomal membrane to introduce the virus nucleic acid into the cytoplasm, see ref. 35), and neutralization (interaction with antibodies leading to loss of virus infectivity). The physical domain for hemagglutination may reside in E1, as isolated E1 from SF or SIN can hemagglutinate (36, 37). Garoff *et al.* (23) have suggested that the sequences responsible for fusion also reside in E1. In either case, however, the heterodimeric unit appears to be required for virus attachment and fusion (see for example refs. 38 and 39). On the other hand, the site for neutralization appears to be located on E2, since antibodies to E2 are in general neutralizing whereas those to E1 are not (reviewed in ref. 40). Furthermore, antibodies to E1 are cross-reactive, whereas those to E2 tend to be type-specific (36, 38, 41), in agreement with the observation that E1 is more highly conserved than E2. All of this suggests that E2 functions during virus evolution to generate strain diversity and is the primary target of the immune system in that it elicits the production of neutralizing antibodies, whereas E1 is more highly conserved, possibly because it carries domains required for interaction with cellular receptors and for fusion.

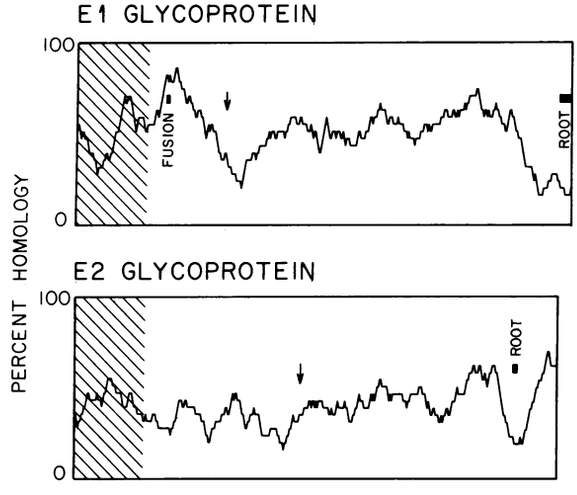

FIG. 4. Homology between glycoproteins E1 and E2 of SIN and RR throughout the entire protein. The glycoproteins of the two viruses are compared 25 amino acids at a time and the percent homology is plotted as a moving average. Certain landmark features of the glycoprotein are noted, including the conserved attachment sites of the carbohydrate chains (arrows) (14), the hydrophobic anchors that span the lipid bilayer (ROOT) (32), and the region believed to be involved in membrane fusion (23). The regions corresponding to the sequenced regions shown in Fig. 1 and used for construction of the evolutionary tree (E1 only) are shaded.

Inspection of Fig. 1 reveals that there is considerable variation in the amino acid sequence with the exception of certain conserved residues or conserved regions. All of the cysteines in these NH_2-terminal regions appear to be conserved, as well as many or most of the prolines, aromatic residues, and glycines, and certain of the charged residues. Since these residues are important in the determination of secondary structure this suggests that the overall conformation of the glycoproteins is conserved while considerable variation in the primary sequence can be accommodated. Secondary structure predictions (Fig. 3), within the limits of accuracy of such predictions, are consistent with this suggestion.

The evolutionary tree derived from the amino acid sequence data illustrates the relationships among these viruses, and it is of interest to compare these relationships with those deduced from serological cross-reactions. Calisher *et al.* (4) recognize six subgroups of alphaviruses: the WEE group (which includes SIN), the VEE group, the EEE group, the SF group (which includes RR), the MID group, and the Ndumu group (for which group we have no data). Porterfield (1) has grouped MID, albeit distantly, with the WEE group, and the SF group with the VEE complex. Our grouping, based on protein sequence data and illustrated in Fig. 2, is closer to that suggested by Calisher *et al.* (4), but the sequence data show that MID, SF, and RR are closely related and these three viruses form a separate branch in the evolutionary tree.

BF is of particular interest. Limited serological data exist for this virus; it was originally classified as a bunyavirus (42) because it cross-reacts with Umbre virus (43), a member of the Turlock group. The amino acid sequences of the glycoproteins and recent serological data demonstrate it is in fact an alphavirus (26) and the cross-reaction with Umbre virus is presumably adventitious. Thus although serological results are usually quite accurate for the classification of viruses, exceptions can occur. Within the alphaviruses BF is most closely related to VEE (Fig. 2) but appears to form a separate subgroup.

A number of interesting questions remain concerning the origin and evolution of the alphaviruses, particularly with regard to the current geographic distribution of these viruses. Alphaviruses isolated in the New World have not been isolated in the Old World, and vice versa, and, in addition, several New World alphaviruses may lead to encephalitis in man, while Old World alphaviruses do not. However, for example, SIN (an Old World virus) and WEE (a New World virus) have clearly diverged only relatively recently (Fig. 2). Clearly, the worldwide geographic distributions, and the tendencies toward neurotropism, of the alphaviruses have been established after the evolutionary divergence of these viruses. It is possible that the mosquito vector is of importance in these events, and it has been suggested that the differing virulence of geographically different strains of WEE may arise from transmission by different species of mosquitos in the various areas, differences in the vector–host relationships in different areas, or both (2).

In this regard it is of interest that in a study of the genomic divergence among Sindbis virus strains, using the extent of RNase resistance of heterologous RNA·RNA hybrids as the criterion of relatedness, Rentier-Delrue and Young (8) found that strains from a single geographic region were more closely related than strains from different regions. These relationships thus appear analogous to studies of clinal distributions within a species. However, our data on different alphaviruses appear more analogous to studies of different species, and in these instances evolutionary divergence is not always correlated with geographic distribution, as illustrated by the SIN–WEE example discussed above.

Our data indicate that all of the alphavirus E1 genes arose from a single ancestral gene, as did all of the E2 genes. Although other models are conceivable, we consider it likely that an ancestral alphavirus arose only once and that all alphaviruses are derived from this single virus ancestor. Perhaps the most interesting question concerning the origin of alphaviruses is whether they arose from organisms such as vertebrates or arthropods or from another virus. The only possible approach toward answering this question lies with a comparison, at the sequence level, of the alphavirus ancestral genes with the genes of other viruses and other organisms. The evolutionary data presented here are a first step in obtaining this information.

We are grateful to E. M. Lenches for technical assistance, to T. Hunkapiller for the computer programs used in construction of the figures, and to L. Hood for use of sequencing and computer facilities. This work was supported in part by Grants GM06965 and AI10793 from the National Institutes of Health and Grant PCM8316856 from the National Science Foundation.

1. Porterfield, J. S. (1980) in *The Togaviruses*, ed. Schlesinger, R. W. (Academic, New York), pp. 13–46.
2. Chamberlain, R. W. (1980) in *The Togaviruses*, ed. Schlesinger, R. W. (Academic, New York), pp. 175–227.
3. Shope, R. E. (1980) in *The Togaviruses*, ed. Schlesinger, R. W. (Academic, New York), pp. 47–82.
4. Calisher, C. H., Shope, R. E., Brandt, W., Casals, J., Karabatsos, N., Murphy, F. A., Tesl, R. B. & Wiebe, M. E. (1980) *Intervirology* **14**, 229–232.
5. Dayhoff, M. O., ed. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, DC), Vol. 5.
6. Krystal, M., Elliott, R. M., Benz, E. W., Jr., Young, J. R. & Palese, P. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4800–4804.
7. Air, G. M. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7639–7643.
8. Rentier-Delrue, F. & Young, N. A. (1980) *Virology* **106**, 59–70.
9. Wengler, G., Wengler, G. & Filipe, A. R. (1977) *Virology* **78**, 124–134.
10. Strauss, E. G., Rice, C. M. & Strauss, J. H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 5271–5275.
11. Ou, J.-H., Rice, C. M., Dalgarno, L., Strauss, E. G. & Strauss, J. H. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 5235–5239.
12. Ou, J.-H., Strauss, E. G. & Strauss, J. H. (1983) *J. Mol. Biol.* **168**, 1–15.
13. Rice, C. M. & Strauss, J. H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2062–2066.
14. Dalgarno, L., Rice, C. M. & Strauss, J. H. (1983) *Virology* **129**, 170–187.
15. Bell, J. R., Bond, M. W., Hunkapiller, M. W., Strauss, E. G., Strauss, J. H., Yamamoto, K. & Simizu, B. (1983) *J. Virol.* **45**, 708–714.
16. Trent, D. W., Clewley, J. P., France, J. K. & Bishop, D. H. L. (1979) *J. Gen. Virol.* **43**, 365–381.
17. Trent, D. W. & Grant, J. A. (1980) *J. Gen. Virol.* **47**, 261–282.
18. Bell, J. R., Strauss, E. G. & Strauss, J. H. (1979) *Virology* **97**, 287–294.
19. Yamamoto, K. & Simizu, B. (1980) *Appl. Environ. Microbiol.* **40**, 240–243.
20. Hunkapiller, M. W. & Hood, L. E. (1980) *Science* **207**, 523–525.
21. Hunkapiller, M. W. & Hood, L. E. (1983) *Methods Enzymol.* **91**, 486–493.
22. Foulds, L. R., Hendy, M. D. & Penny, D. (1979) *J. Mol. Evol.* **13**, 127–149.
23. Garoff, H., Frischauf, A.-M., Simons, K., Lehrach, H. & Delius, H. (1980) *Nature (London)* **288**, 236–241.
24. Bell, J. R., Hunkapiller, M. W., Hood, L. & Strauss, J. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2722–2726.
25. Arias, C., Bell, J. R., Lenches, E. M., Strauss, E. G. & Strauss, J. H. (1983) *J. Mol. Biol.* **168**, 87–102.
26. Dalgarno, L., Short, N. J., Hardy, C. M., Bell, J. R. & Strauss, J. H. (1984) *Virology* **133**, 416–426.
27. Doolittle, R. F. (1981) *Science* **214**, 149–159.
28. Fitch, W. M. (1971) *Syst. Zool.* **20**, 406–416.
29. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120**, 97–120.
30. Strauss, E. G. & Strauss, J. H. (1983) *Curr. Top. Microbiol. Immunol.* **105**, 1–98.
31. Strauss, E. G. (1978) *J. Virol.* **28**, 466–474.
32. Rice, C. M., Bell, J. R., Hunkapiller, M. W., Strauss, E. G. & Strauss, J. H. (1982) *J. Mol. Biol.* **154**, 355–378.
33. Ziemiecki, A. & Garoff, H. (1978) *J. Mol. Biol.* **122**, 259–269.
34. Rice, C. M. & Strauss, J. H. (1982) *J. Mol. Biol.* **154**, 325–348.
35. Helenius, A., Kartenbeck, J., Anderson, R. G. W., Simons, K. & Fries, E. (1980) *J. Cell Biol.* **84**, 404–420.
36. Dalrymple, J. M., Schlesinger, S. & Russell, P. K. (1976) *Virology* **69**, 93–103.
37. Helenius, A., Fries, E., Garoff, H. & Simons, K. (1976) *Biochim. Biophys. Acta* **436**, 319–334.
38. Roehrig, J. T., Day, J. W. & Kinney, R. M. (1982) *Virology* **118**, 269–278.
39. Mann, E., Edwards, J. & Brown, D. T. (1983) *J. Virol.* **45**, 1083–1089.
40. Strauss, J. H. & Strauss, E. G. (1984) in *Immunochemistry of Viruses: The Basis of Serodiagnosis and Vaccines*, eds. Van Regenmortel, M. H. V. & Neurath, A. R. (Elsevier, Amsterdam), in press.
41. Chanas, A. C., Gould, E. A., Clegg, J. C. S. & Varma, M. G. R. (1982) *J. Gen. Virol.* **58**, 37–46.
42. Bishop, D. H. L. & Shope, R. E. (1979) in *Comprehensive Virology*, eds. Fraenkel-Conrat, H. & Wagner, R. R. (Plenum, New York), Vol. 14, pp. 1–156.
43. Marshall, I. D., Woodroofe, G. M. & Hirsh, S. (1982) *Aust. J. Exp. Biol. Med. Sci.* **60**, 457–470.