Phylogeny of North American Powassan virus

Gregory D. Ebel, Andrew Spielman and Sam R. Telford III

Laboratory of Public Health Entomology, Department of Immunology and Infectious Diseases, Harvard School of Public Health, 665 Huntington Ave, Boston, MA 02115, USA

To determine whether Powassan virus (POW) and deer tick virus (DTV) constitute distinct flaviviral populations transmitted by ixodid ticks in North America, we analysed diverse nucleotide sequences from 16 strains of these viruses. Two distinct genetic lineages are evident, which may be defined by geographical and host associations. The nucleotide and amino acid sequences of lineage one (comprising New York and Canadian POW isolates) are highly conserved across time and space, but those of lineage two (comprising isolates from deer ticks and a fox) are more variable. The divergence between lineages is much greater than the variation within either lineage, and lineage two appears to be more diverse genetically than is lineage one. Application of McDonald–Kreitman tests to the sequences of these strains indicates that adaptive evolution of the envelope protein separates lineage one from lineage two. The two POW lineages circulating in North America possess a pattern of genetic diversity suggesting that they comprise distinct subtypes that may perpetuate in separate enzootic cycles.

Introduction

Powassan virus (POW), the sole North American member of the tick-borne encephalitis (TBE) complex of the genus Flavivirus, was isolated from a fatal case of encephalitis in 1958 (McLean et al., 1959). Although cases of human disease due to POW infection have occurred sporadically (Anon., 1995; Artsob, 1989; Deibel et al., 1979; Embil et al., 1983; Gholam et al., 1999; Goldfield et al., 1973; Jackson, 1989; Partington et al., 1980; Rossier et al., 1974; Smith et al., 1974), the threat to human health posed by POW appears to be less than that from certain other tick-borne infections, such as Lyme disease, because the main enzootic vector of POW, Ixodes cookei, feeds mainly on groundhogs (Marmota monax), attacking other hosts only rarely.

In 1996 a virus closely related to POW was discovered in deer ticks (Ixodes dammini or I. scapularis of some authors) in Massachusetts and Connecticut (Telford et al., 1997). A similar isolate was subsequently established from deer ticks infesting distant sites where other deer tick-transmitted agents perpetuate (Ebel et al., 1999). This virus was provisionally designated deer tick virus (DTV) based on apparent differences in vertebrate reservoir (Johnson, 1987; Main et al., 1979; McLean & Donahue, 1959; McLean et al., 1966), arthropod vector (Artsob, 1989; Ebel et al., 1999; Telford et al., 1997), the lack of clinically apparent cases of encephalitis among human residents of transmission foci, and molecular genetic data indicating that DTV strains may form a distinct clade within the TBE complex (Ebel et al., 1999; Telford et al., 1997).

Molecular genetic studies have contributed greatly to our understanding of the epidemiology and evolutionary history of flaviviruses and other zoonotic viral agents (Heinz & Kunz, 1982; Ksiazek et al., 1997; Kuno et al., 1998; Lanciotti et al., 1994, 1997; Monroe et al., 1999; Zanotto et al., 1995, 1996). Our previously published analyses of POW and DTV (Ebel et al., 1999; Telford et al., 1997), however, seem severely limited because gene sequence data from only one POW strain was available (Mandl et al., 1993). To determine whether POW and DTV constitute distinct tick-borne flaviviral populations co-circulating in North America, we analysed nucleotide sequence data from eleven POW strains and five DTV strains. Specifically, we constructed phylogenetic trees based on fragments of the envelope (env) and nonstructural protein 5 (ns5) genes, and an untranslated region. We then determined
whether the pattern of polymorphism observed within POW nucleotide sequences supports the observation that strains isolated from deer ticks seem to perpetuate in a cycle different from that of other POW strains (Ebel et al., 2000).

Methods

- **Virus strains.** To assemble a collection of RNA samples for analysis, we combined DTV strains collected by our group with RNA from 11 POW isolates graciously donated by Nick Karabatsos and Goro Kuno (CDC, Ft Collins, CO, USA). The sources and methods for isolation from 11 POW isolates were described (Ebel et al., 1997; Telford et al., 1999). Strainsspo-B-9901 and cf9901 were isolated from adult deer ticks collected near Spooner and Chippewa Falls, WI, USA during October, 1999. Ticks were assayed for infection (Ebel et al., 2000) and virus was isolated (Telford et al., 1997) as described.

- **RNA amplification and sequencing.** To obtain gene sequence data, RNA samples were reverse transcribed using random hexamers, and nucleotides 1027–1548 (env) and 8672–8966 (nucleotide numbering after Mandl et al., 1993) (ns5) genes were amplified using primers env-A, POW-e, DTVPN5-F and DTVPN5-2 as previously described (Ebel et al., 1999; Telford et al., 1997). Nucleotides 10356–10806 (3’ noncoding region) were amplified using primers utr-f (5’ to 3’ gcactggaagctacgttagagag) and utr-r (5’ to 3’ agcgggtttttcttgactaca) designed to amplify approximately 500 bp from this region. Reaction products from all regions were electrophoretically separated on a 2% agarose gel, excised, purified and sequenced as described (Ebel et al., 1999).

- **Phylogenetic analyses.** To construct phylograms, alignments were generated with the PItheLUP program of the Wisconsin Genetics Computer Group and initially analysed by the distance method using MEGA (Kumar et al., 1993). Evolutionary distances were computed using the Kimura 2-parameter method including both transitions and transversions. Distance trees were constructed using the neighbour-joining method, and their robustness was estimated by performing 500 bootstrap replicates. Maximum parsimony and maximum likelihood analyses were performed using PAUP version 4.0b3a for Windows (Sinauer Associates). Parsimony and likelihood analyses were performed with 500 bootstrap replicates of a heuristic (maximum parsimony) or fast-heuristic (maximum likelihood) search. Trees generated using PAUP were displayed using TreeView version 1.5.2 (Roderick D. M. Page). MEGA was used to analyse predicted amino acid sequences derived from nucleotide sequences aligned as above. Amino acid sequences were computed using the number of differences algorithm and trees were constructed by the neighbour-joining method. The robustness of these trees was estimated as above. McDonald–Kreitman tests (McDonald & Kreitman, 1991) were performed using DnaSP 3.00 (Rozas & Rozas, 1999) and the P-values reported are Fisher’s exact test values. For comparison, we used this same method to compare the TBE subtypes central European encephalitis virus (CEE), Russian spring-summer encephalitis virus (RSSE) and lopping ill virus (LI). GenBank accession numbers for the sequences used are: U27491, U39292, U27495, X60286 (CEE); X07755, M38310, M97356 (RSSE); X69975, M59376, M94956 (LI). The neutrality index reports the degree of amino acid polymorphism departure from the neutral expectation (Kimura, 1983).

Results

We assembled POW and DTV samples from diverse locations, years and hosts (Table 1). Five strains isolated from deer ticks (DTV) and eleven previously characterized POW strains were included. Samples included were isolated over the 41 year period beginning in 1958 and ending in 1999 from five states and one Canadian province. Most samples derive from various tick species, but other sources, such as squirrels, woodchucks, a fox and human brain are included. Our sample of isolates adequately represents the virus population circulating in the northeastern and north central regions of North America during the last five decades.

We carried out sequence-based phylogenetic analyses of a structural gene (the envelope gene), a nonstructural gene (ns5) and an untranslated region (3′ untranslated region) (Fig. 1) of the various POW and DTV strains. Phylogenetic reconstructions based on each region yielded similar overall topologies. Two distinct and well-supported clades are the most striking feature in each tree. These clades are recognized by analyses using each of three optimality criteria, and consist of identical taxa for all regions analysed. These two clades are also apparent in a neighbour-joining tree of distances derived from deduced envelope protein sequences including other members of the TBE complex (Fig. 2). Based on phylogenetic trees constructed using nucleotide and amino acid sequences from three regions of the POW genome, we conclude that two distinct POW lineages are present in North America.

We noted the assemblage of strains within each lineage. Strains originating in New York and Canada (lineage 1) were more closely related to each other than to strains isolated in Wisconsin and on the eastern seaboard (lineage 2). Lineage one includes isolates deriving from ticks, groundhogs, human brain and a squirrel, isolated from 1958 until 1981. Lineage two comprises isolates from the eastern seaboard and Wisconsin, deriving mainly from deer ticks collected from 1994 to 1999, but also including a fox-derived strain isolated in 1977.

**Geographical- and host-associations may define each of the two lineages**

We then evaluated the substructure within each major lineage. Within lineage one, analysis of the nucleotide sequences of all three regions recognized two minor clades separating strains isolated in Canada from New York-derived strains. The neighbour-joining tree constructed from the envelope protein, however, failed to resolve these two sublineages. In this protein analysis, isolates from lineage one are more closely related to each other than to strains isolated in Wisconsin and on the eastern seaboard (lineage 2). Lineage one includes isolates deriving from ticks, groundhogs, human brain and a squirrel, isolated from 1958 until 1981. Lineage two comprises isolates from the eastern seaboard and Wisconsin, deriving mainly from deer ticks collected from 1994 to 1999, but also including a fox-derived strain isolated in 1977.
England and Wisconsin isolates. Although the nucleotide and amino acid sequences of New York and Canadian POW isolates are highly conserved across time and space, those deriving from deer ticks and a fox are more variable.

We then quantified the divergence between lineages by computing the mean pairwise percent nucleotide and amino acid differences between and within each of the two main lineages. For all genes studied, the percent nucleotide differences between lineages were at least 11% and were at least 3-fold greater than the corresponding values within each lineage (Table 2). The percent amino acid sequence difference was also greater between lineages than within either lineage. Between lineages one and two the variation in the envelope protein sequence was 13-fold greater and the variation in the ns5 protein sequence was 8-fold greater than the difference within lineage one. We conclude that the divergence between lineages is much greater than the variation within either lineage, and that lineage two appears to be more genetically diverse than lineage one.

We then characterized this divergence by means of the McDonald–Kreitman (MK) test. When applied to the envelope gene, this test rejects the hypothesis that the differences occurred through random processes (geographical isolation, genetic drift, etc.) in favour of the alternative that natural selection has yielded the two lineages that exist in the dataset (Table 3). MK tests narrowly fail to reject the null when applied to the ns5 gene. As a comparison, MK tests were performed for the same portion of the envelope genes of CEE, RSSE and LI. Only the comparison of RSSE to LI produced a result similar to that observed between the two POW lineages. We conclude that natural selection of the envelope protein may account for

### Table 1. Characteristics of the various POW strains considered in this study

<table>
<thead>
<tr>
<th>Strain designation</th>
<th>Geographical origin</th>
<th>Year of isolation</th>
<th>Source</th>
<th>Accession numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>12542</td>
<td>WV, USA</td>
<td>1977</td>
<td><em>Vulpes</em> spp.</td>
<td>ay004099, ay004077, ay004095</td>
</tr>
<tr>
<td>1247-62</td>
<td>Ontario, Canada</td>
<td>1962</td>
<td><em>Ixodes</em> dammini</td>
<td>ay004109, ay004075, ay004088</td>
</tr>
<tr>
<td>1982-64</td>
<td>Ontario, Canada</td>
<td>1964</td>
<td><em>M. monax</em></td>
<td>ay004104, ay004074, ay004086</td>
</tr>
<tr>
<td>22287-91</td>
<td>Unknown</td>
<td>1964</td>
<td>Unknown</td>
<td>ay004106, ay004071, ay004083</td>
</tr>
<tr>
<td>64-7062</td>
<td>NY, USA</td>
<td>1964</td>
<td>Ticks on <em>Marmota</em> sp.</td>
<td>ay004102, ay004069, ay004081</td>
</tr>
<tr>
<td>64-7483</td>
<td>NY (?)</td>
<td>Unknown</td>
<td>Unknown</td>
<td>ay004103, ay004070, ay004082</td>
</tr>
<tr>
<td>LB</td>
<td>Ontario, Canada</td>
<td>1958</td>
<td>Human brain</td>
<td>l06436</td>
</tr>
<tr>
<td>m11665</td>
<td>Ontario, Canada</td>
<td>1960s</td>
<td>Tick</td>
<td>ay004105, ay004073, ay004087</td>
</tr>
<tr>
<td>m1409</td>
<td>Ontario, Canada</td>
<td>1960s</td>
<td>Tick</td>
<td>ay004107, ay004072, ay004084</td>
</tr>
<tr>
<td>t18-23-81</td>
<td>Ontario, Canada</td>
<td>1981</td>
<td><em>Ixodes</em> cookei on <em>Marmota</em> sp.</td>
<td>ay004108, ay004078, ay004085</td>
</tr>
<tr>
<td>m8898</td>
<td>Ontario, Canada</td>
<td>1960s</td>
<td>Tick</td>
<td>ay004110, ay004076, ay004089</td>
</tr>
<tr>
<td>DTV-IPS</td>
<td>Ipswich, MA</td>
<td>1994</td>
<td><em>I. dammini</em></td>
<td>ay004100, u93288, ay004093</td>
</tr>
<tr>
<td>DTV-CT</td>
<td>Connecticut</td>
<td>1994</td>
<td><em>I. dammini</em></td>
<td>ay004101, u93289, ay004094</td>
</tr>
<tr>
<td>DTV-SPO</td>
<td>Spooner, WI</td>
<td>1997</td>
<td><em>I. dammini</em></td>
<td>ay004097, af135461, ay004091</td>
</tr>
<tr>
<td>spo-B-9901</td>
<td>Spooner, WI</td>
<td>1999</td>
<td><em>I. dammini</em></td>
<td>ay004098, ay004080, ay004090</td>
</tr>
<tr>
<td>wic9901</td>
<td>Chippewa Falls, WI</td>
<td>1999</td>
<td><em>I. dammini</em></td>
<td>ay004096, ay004079, ay004092</td>
</tr>
</tbody>
</table>

### Discussion

Our conclusions are based on a relatively small number of POW strains. We have included, however, all strains that were available to us, and most (perhaps all) that currently exist. POW appears to circulate in South Dakota (Artsob, 1989), the Western United States, Western Canada (Johnson, 1987; McLean et al., 1970) and Siberia (L’vov et al., 1974; Leonova et al., 1981). We were not able to obtain stock virus or RNA samples from these strains, and do not know whether they have been preserved. Their inclusion would have strengthened our analyses and conclusions. Including data from other tick species would have similarly strengthened our study. However, we have obtained no virus isolates from ticks such as *Dermacentor variabilis* during our studies of the epizootiology of tick-borne infections in the Northeastern and upper midwestern USA.

Only a small portion (approximately 12%) of the virus genome was analysed in this study. Although the complete sequences for the ns5 and env genes might provide more information, the homogeneity of our findings across three functionally diverse regions suggests that the availability of extended sequence data would be unlikely to significantly alter our conclusions. Phylogenetic studies of other zoonotic viruses have relied on analyses of single gene sequences or gene sequence fragments, and our tree-building approach is consistent with many of these studies (Childs et al., 1994; Ecker et
Fig. 1. For legend see facing page.
Fig. 1. Phylogenetic relationships among POW strains based on a 522 bp fragment of the envelope gene (a), a 292 bp fragment of the ns5 gene (b) and a 451 bp fragment of the 3’noncoding region (c) of POW strains. Trees are neighbour-joining analyses with branch-lengths proportional to genetic distance. Numbers at each node are bootstrap confidence estimates based on 500 replicates of neighbour-joining/maximum parsimony/maximum likelihood analysis. Asterisks indicate that the node was not recognized by the optimality criteria at that position. GenBank accession numbers are provided in Table 1. For all analyses, louping ill virus was used as the outgroup. Accession numbers for this virus are M59376 (envelope) and Y07863 (ns5 and untranslated region).

al., 1999; Lanciotti et al., 1994, 1997; Plyusnin et al., 1996; Weaver et al., 2000). Analysis of nucleotide sequences from each of three regions of the virus genome produced similar tree topologies showing that two distinct POW lineages circulate in North America. Lineage one comprises solely isolates from New York and Canada, while lineage two includes strains isolated from the Atlantic coast of the United States and from Wisconsin. This result was not contingent upon the function of the sequence studied or on the analytical method employed. These observations establish unambiguously that two main POW lineages perpetuate in cycles involving ixodid ticks in North America.

The mode of perpetuation of each isolate appears to be associated with membership of a lineage. Strains in lineage one appear to have been isolated from I. cookei (subgenus Pholeoixodes) and members of the Sciuridae, and were obtained mainly before the range expansion of deer ticks (subgenus Ixodes) that occurred during the 1970s. Strains in lineage two, on the other hand, appear to be associated with deer ticks. Five of the six strains constituting this lineage were isolated from deer ticks, and the remaining strain was isolated from a region and at a time where deer ticks may have been present. Strains in lineage one seem to be mainly associated with a groundhog–I. cookei cycle, and strains in lineage two appear to be associated with deer ticks and the white-footed mice upon which they most frequently feed (Ebel et al., 2000).

Nucleotide variation within natural POW populations appears to be maximal within the envelope gene, but the protein sequences are much less variable than are ns5 protein sequences, both within and between lineages. The conservation of the amino acid sequences within each lineage is consistent with proposed functions of the envelope protein in host cell binding and membrane fusion. It may be, then, that the observed changes in the env gene between lineage one and lineage two are not due to immune-mediated selection (antigenic drift), but to functional constraints arising from different modes of perpetuation, specifically the reliance of each lineage on a different tick vector. This mode of virus–vector association was proposed when strains of RSSE virus were recovered from Ixodes persulcatus ticks where CEE was the sole virus expected (Ecker et al., 1999). Further field and laboratory observations are required to determine whether specific vector–virus relationships have evolved between TBE complex viruses and their ixodid tick vectors.
Fig. 2. Phylogenetic relationships among tick-borne flaviviruses based on a 174 amino acid fragment of the envelope protein.
Table 2. Variation within and between POW lineages
A mean pairwise percent difference was calculated for each region studied.

<table>
<thead>
<tr>
<th>Percent difference in:</th>
<th>Lineage</th>
<th>Envelope protein</th>
<th>Nonstructural region 5</th>
<th>3’ Untranslated region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Nucleotides</td>
<td>Amino acids</td>
<td>Nucleotides</td>
</tr>
<tr>
<td>1</td>
<td>2:6</td>
<td>0:2</td>
<td>2:6</td>
<td>0:7</td>
</tr>
<tr>
<td>1 vs 2</td>
<td>15:0</td>
<td>2:9</td>
<td>11:1</td>
<td>5:4</td>
</tr>
<tr>
<td>2</td>
<td>5:6</td>
<td>0:4</td>
<td>3:2</td>
<td>1:2</td>
</tr>
</tbody>
</table>

Table 3. McDonald–Kreitman tests for evidence that divergence in POW lineages is driven by natural selection

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Gene</th>
<th>Neutrality index</th>
<th>MK P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>POW(1)–POW(2) env (522 bp)</td>
<td>0:103</td>
<td>0:024</td>
<td></td>
</tr>
<tr>
<td>POW(1)–POW(2) ns5 (292 bp)</td>
<td>0:244</td>
<td>0:065</td>
<td></td>
</tr>
<tr>
<td>CEE–RSSE env (522 bp)</td>
<td>0:892</td>
<td>1:000</td>
<td></td>
</tr>
<tr>
<td>CEE–LI env (522 bp)</td>
<td>1:068</td>
<td>1:000</td>
<td></td>
</tr>
<tr>
<td>RSSE–LI env (522 bp)</td>
<td>0:270</td>
<td>0:007</td>
<td></td>
</tr>
</tbody>
</table>

MK tests on envelope gene- and protein-derived sequence data support the observation that the two POW lineages perpetuate differently by rejecting the hypothesis that isolation and random genetic drift have been the most important factors in producing the extant taxa. The envelope protein sequences studied contained more molecular evidence for differentiating the two POW lineages than for differentiating CEE and RSSE [now considered as subtypes of a single virus species (Ecker et al., 1999)] or CEE and LI, virus pairs known to differ in mode of perpetuation and outcome of human infection. Based on the envelope gene and protein sequences, it seems likely that the two POW lineages perpetuate differently.

The ns5 data, however, do not add support to the adaptive evolution hypothesis. It may be that a longer sequence, or one spanning a different region of the ns5 gene might have produced a different result in the MK test, but the precise effect is difficult to predict in the absence of this data. Alternatively, it may be that due to its role in virus replication, the ns5 gene, which codes a protein with RNA-dependent RNA polymerase activity (Khromykh et al. 2000; Lai et al., 1999; Steffens et al., 1999), is subject to different selective pressures than the env protein, and would not necessarily provide evidence for an evolutionary history that is concordant to one derived from the env sequences.

These results clearly illustrate the division of North American POW strains into two genetic lineages. Within each lineage, the virus nucleotide and amino acid sequences are highly conserved. The biological significance of the genetic differences between and within POW subtypes requires further investigation, but is likely to be great. Only one subtype (lineage one) has been associated with human disease, while only the other (lineage two) appears to have epidemic potential through its apparently unique association with aggressively human-biting deer ticks. The genetic diversity within our POW samples is well within that which has been reported within other flavivirus species (Ecker et al., 1999; Lanciotti et al., 1994, 1997; Lepiniec et al., 1994; Nam et al., 1996), indicating that the two lineages do not constitute distinct species. The two POW lineages circulating in North America constitute distinct POW subtypes that may perpetuate in distinct enzootic cycles.

The authors thank Goro Kuno, Harvey Artsob and Nick Karabatsos for graciously sharing POW RNA and for thoughtful discussions of this work; and Marta Guerra for helping pinpoint where we might find CF9901-infected ticks and for helping collect the ticks. Heidi Goethert also helped with tick collection in Chippewa Falls, WI.

This work was supported in part by NIH grants AI 39002, AI37993 and AI19693. Dr Ebel was supported while conducting this work by NIH AI 07350 ‘Doctoral Training Program in Tropical Public Health’.

Neighbour-joining tree with branch lengths proportional to amino acid distance. Branch numbers are bootstrap confidence estimates based on 500 replicates. GenBank accession numbers are in parentheses. Abbreviations for viruses are as follows: KFD, Kyasanur Forest disease; LGT, Langat; OGF, Omek haemorrhagic fever; LI, Louping Ill; LI-Neg, Louping Ill-Negishi; SSE, Spanish sheep encephalitis; GGE, Greek goat encephalitis; TSE, Turkish sheep encephalitis; RSSE, Russian spring-summer encephalitis; CEE, Central European encephalitis; SRE, Saumarez Reef; TYU, Tyuleniy; YF, yellow fever.
References


Received 26 October 2000; accepted 14 March 2001