Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 47 (2008) 717–728
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Phylogenetic and evolutionary analyses of St. Louis encephalitis virus genomes

Gregory J. Baillie [a,*,1], Sergios-Orestis Kolokotronis [a], Eric Waltari [a], Joseph G. Maffei [b], Laura D. Kramer [b], Susan L. Perkins [a]

[a] *Division of Invertebrate Zoology, and Sackler Institute for Comparative Genomics, American Museum of Natural History, Central Park West at West 79th Street, New York, NY 10024, USA*
[b] *The Arbovirus Laboratories, Wadsworth Center, New York State Department of Health, 5668 State Farm Road, Slingerlands, NY 12159, USA*

Received 15 August 2007; revised 24 January 2008; accepted 14 February 2008
Available online 29 February 2008

Abstract

St. Louis encephalitis virus belongs to the Japanese encephalitis virus serocomplex of the genus *Flavivirus*, family *Flaviviridae*. Since the first known epidemic in 1933, the virus has been isolated from a variety of geographical, temporal, and host origins. We have sequenced 10,236 nucleotides of the open reading frame (93.6% of the full-length genome) of 23 of these strains, and have used the sequences to conduct phylogenetic analyses, in order to investigate the forces shaping the evolution of St. Louis encephalitis virus. Contrary to previous reports, we found little evidence for recombination in these isolates. Most of the amino acid sites in the SLEV polyprotein appeared to be under negative selection, with some sites evolving neutrally, and a small number under positive selection. The strongest signal for positive selection was evident in the *N*-linked glycosylation site of the envelope protein. Intra-strain sequence variability within strains was observed at this site, and analyses suggested that it is under selection *in vitro*. Furthermore, using heterochronous sequence data, we estimated the most recent expansion of St. Louis encephalitis virus in North America to have happened towards the end of the 19th century.

Keywords: St. Louis encephalitis virus; *Flavivirus*; Phylogeny; Positive selection; Recombination; *N*-linked glycosylation; Divergence times

1. Introduction

St. Louis encephalitis virus (SLEV) belongs to the Japanese encephalitis virus serocomplex within the genus *Flavivirus* (Family: *Flaviviridae*) (Lindenbach and Rice, 2001). Its known distribution includes parts of North America, Central America, South America, and islands of the Caribbean, and its primary invertebrate and vertebrate hosts are mosquitoes of the genus *Culex* and various species of wild birds, respectively (Reisen, 2003). Mammals, including humans, are also occasionally infected, but are considered dead-end hosts (Reisen, 2003). The virus is named after an epidemic that occurred in St. Louis, Missouri in 1933, and there have been at least 41 outbreaks in North America since that time (Day, 2001; Reisen, 2003).

For the last 74 years, samples of SLEV from a variety of geographical, temporal, and host origins have been collected and stored. These strains show wide diversity in the degree of viremia, neurovirulence, and severity of symptoms induced in both avian and mammalian hosts, with some correlation between geographic origin and these phenotypes (Bowen et al., 1980; Monath et al., 1980), although isolates made after 1977 were not included in these analyses. Oligonucleotide fingerprinting (Trent et al., 1980) and phylogenetic analysis of the envelope gene (Kramer

* Corresponding author. Fax: +44 1223 764667.
E-mail address: gb354@cam.ac.uk (G.J. Baillie).
[1] Present address: Cambridge Infectious Diseases Consortium, Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK.

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.02.015

and Chandler, 2001) show that, in general, geographically proximal strains are also genetically similar.

The genome of SLEV is, like that of all members of the genus *Flavivirus*, a non-segmented, single-stranded, positive-sense RNA, approximately 11 kb in length. It encodes a single 3429 amino acid polyprotein, which is co- and post-translationally cleaved by viral and cellular proteases into three structural (C, PrM/M, E) and seven non-structural (NS1, NS2A, NS2B, NS3 [protease/helicase], NS4A, NS4B, NS5 [polymerase]) proteins (Lindenbach and Rice, 2001) (Fig. 1).

The E protein is the primary determinant of cell receptor binding and immune recognition (Lindenbach and Rice, 2001). In some strains of SLEV, the E protein is glycosylated at an asparagine at position 154 of the protein, when it is part of a Asn-X-Ser/Thr (NXS/T) tripeptide, whereas in other strains this site is lacking or is variably glycosylated (Vorndam et al., 1993). Glycosylation does not appear to affect formation or release of viral particles, although virions lacking glycosylated E infect SW-13 (human adenocarcinoma) and CRE (hamster) cells significantly less efficiently than those with glycosylated E (Vorndam et al., 1993). No correlation is observed between glycosylation of E and virulence (Vorndam et al., 1993). In other flaviviruses, the glycosylation state of the envelope protein is also heterogeneous (Shirato et al., 2004), can be altered by passage of viruses in cell culture (Chambers et al., 1998; Halevy et al., 1994; Shirato et al., 2004), affects virus replication *in vitro* (Li et al., 2006), and can influence viremia, neuroinvasiveness, and neurovirulence *in vivo* (Beasley et al., 2005; Chambers et al., 1998; Halevy et al., 1994; Hanna et al., 2005; Li et al., 2006; Scherret et al., 2001; Shirato et al., 2004).

Recent studies have highlighted the potential roles of recombination and selection in the ongoing evolution of flaviviruses. Twiddy and Holmes (2003) report evidence of recombination in dengue virus, Japanese encephalitis virus, and SLEV. Their results for SLEV, based on previously published (Kramer and Chandler, 2001) envelope gene sequences from 61 isolates, suggest that the recombi-

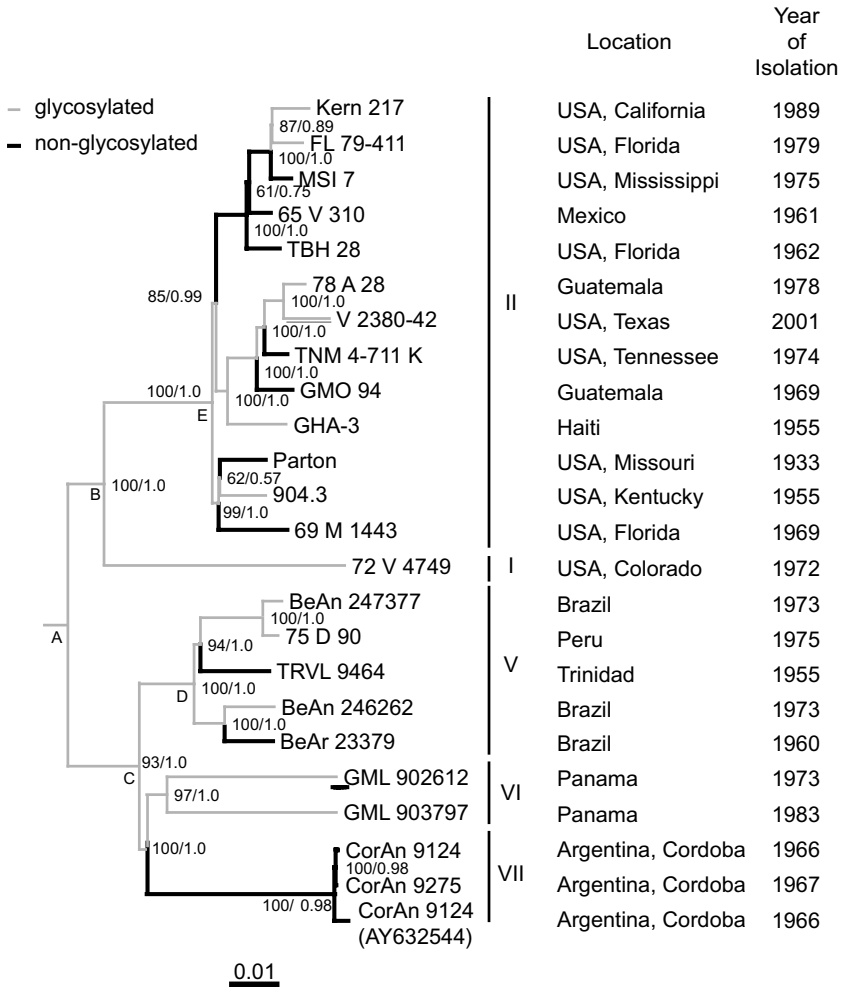

Fig. 1. Maximum likelihood tree based on open reading frame sequences. The tree is mid-point rooted for purposes of illustration. Glycosylation state at position 154–156 of E protein, as estimated by ML analysis, is indicated. Numbers at nodes indicate bootstrap support in percentage of 1000 pseudoreplicates for the ML analysis and posterior probabilities for the Bayesian inference of phylogeny. The scale bar indicates the number of substitutions per site.