Available online at www.sciencedirect.com
ScienceDirect
Gene 399 (2007) 152–161
GENE
www.elsevier.com/locate/gene

Rapid fixation of a distinctive sequence motif in the 3′ noncoding region of the clade of West Nile virus invading North America

Austin L. Hughes [a,*], Helen Piontkivska [b], Ivo Foppa [c]

[a] Department of Biological Sciences, University of South Carolina, Columbia SC 29208, United States
[b] Department of Biological Sciences, Kent State University, Kent OH 44242, United States
[c] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia SC 29208, United States

Received 18 April 2007; received in revised form 14 May 2007; accepted 15 May 2007
Available online 25 May 2007
Received by S. Yokoyama

Abstract

Phylogenetic analysis of complete genomes of West Nile virus (WNV) by a variety of methods supported the hypothesis that North American isolates of WNV constitute a monophyletic group, together with an isolate from Israel and one from Hungary. We used ancestral sequence reconstruction in order to obtain evidence for evolutionary changes that might be correlated with increased virulence in this clade (designated the N.A. clade). There was one amino acid change (I→T at residue 356 of the NS3 protein) that occurred in the ancestor of the N.A. clade and remained conserved in all N.A. clade genomes analyzed. There were four changes in the upstream portion of the 3′ noncoding region (the AT-enriched region) that occurred in the ancestor of the N.A. clade and remained conserved in all N.A. clade genomes analyzed, changes predicted to alter RNA secondary structure. The AT-enriched region showed a higher rate of substitution in the branch ancestral to the N.A. clade, relative to polymorphism, than did the remainder of the noncoding regions, synonymous sites in coding regions, or nonsynonymous sites in coding regions. The high rate of occurrence of fixed nucleotide substitutions in this region suggests that positive Darwinian selection may have acted on this portion of the 3′NCR and that these fixed changes, possibly in concert with the amino acid change in NS3, may underlie phenotypic effects associated with increased virulence in North American WNV.

Keywords: Evolution of virulence; Flavivirus; Natural selection; RNA helicase; West Nile virus

1. Introduction

West Nile virus (WNV), a mosquito-borne single-stranded positive-sense RNA virus classified in the genus *Flavivirus* (family Flaviviridae), was named for its initial discovery in the West Nile Province of Uganda (Smithburn et al., 1940). It was first observed in the Western Hemisphere in 1999, when an outbreak was reported in the New York City area (Lanciotti et al., 2002; Nash et al., 2001). Since that time, WNV has spread rapidly across North America, causing extensive mortality in wild bird species, which are its primary natural hosts (McClean, 2006). The American Crow *Corvus brachyrhynchos* has suffered particularly high mortality from WNV (Yaremych et al., 2004; McClean, 2006). Experimental infections of birds of various species have provided additional evidence of a high vulnerability of the American Crow, as well as in certain other members of the crow family (Corvidae) to North American WNV [6]. High mortality was also observed in some other species such as the House Finch *Carpodacus mexicanus* and the Ring-billed Gull *Larus delawarensis* (Komar et al., 2003). High rates of mortality in avian hosts might be explainable by two not necessarily mutually exclusive factors: (1) an unusual lack of resistance to this virus on the part of certain North American bird species; and (2) unusually high virulence of WNV invading North America.

Abbreviations: 3′NCR, 3′ noncoding region; 5′NCR, 5′ noncoding region; d_N, the number of nonsynonymous nucleotide substitutions per nonsynonymous site; d_S, the number of synonymous nucleotide substitutions per synonymous site; JEV, Japanese encephalitis virus; KUNV, Kunjin virus; WNV, West Nile virus; YFV, yellow fever virus.

* Corresponding author. Tel.: +1 803 777 9186; fax: +1 803 777 4002.
E-mail address: austin@biol.sc.edu (A.L. Hughes).

0378-1119/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2007.05.007

Evidence of differential vulnerability among avian species to infection with WNV provides support for the hypothesis that certain hosts are especially susceptible to WNV infection (Komar et al., 2003). On the other hand, there is also experimental evidence that North American WNV has distinctive biological properties causing greater virulence than seen in certain other WNV isolates. For example, the North American WNV has been shown to be substantially more virulent in American Crows than was a Kenyan isolate (Brault et al., 2004). One factor contributing to this difference may be the enhanced ability of North American WNV to replicate at the high body temperatures (>43 °C) found in infected American Crows (Kinney et al., 2006). There are reports that both naturally occurring (Davis et al., 2003) and experimentally induced (Wicker et al., 2006) mutations attenuate virulence of North American WNV.

Comparing sequences of North American WNV isolates with those from the Old World may help identifying nucleotide substitutions that may account for the distinctive biological characteristics of the former (Brault et al., 2004). Phylogenetic methods provide a particularly powerful tool for such comparisons, because they make it possible to reconstruct changes ancestral to a group of related sequences. If virulence has played a role in the rapid of WNV in the new world, any sequence changes that occurred in the ancestor of North American WNV and have remained fixed as the virus has spread across North America are particularly strong candidates for causing high virulence.

Here we use phylogenetic methods to reconstruct the evolutionary relationships of complete WNV genomes from throughout the world. We use these methods to test the hypothesis that the North American WNV represent a monophyletic group or clade (i.e., the group of sequences descended from a common ancestor) within worldwide WNV and to reconstruct nucleotide sequence changes in both the coding sequence and noncoding regions of the WNV genome. We use statistical methods to test the hypothesis that distinctive features of North American WNV were fixed by positive Darwinian selection and that they thus represent adaptive characteristics of the virus.

2. Methods

Analyses were based on 81 genomes of West Nile virus (WNV); two genomes of Kunjin virus (KUNV); and four genomes of Japanese encephalitis virus (JEV; see Fig. 1 for accession numbers). The WNV genome encodes a single polyprotein which is subsequently cleaved to form the 10 proteins Core, PreM, Env, NS1, NS2a, NS2b, NS3, NS4a, NS4b, and NS5 (Brinton, 2002). Polyprotein sequences were aligned at the amino acid level using CLUSTAL X (Thompson et al., 1997), and the

Fig. 1. NJ tree of WNV polyprotein amino acid sequences (based on 3413 aligned sites). Branches with less than 50% bootstrap support are condensed, and topology only is shown. Accession numbers and geographic origin (including abbreviations for states in the case of isolates from the United States) are given. Lineages are identified following Bakonyi et al. (2006). Selected JEV isolates were used as an outgroup to root the tree.