Evolutionary Genetics of Eastern Equine Encephalitis Virus in the Americas Suggest Madariaga Virus, the South America Lineages as Distinct Species in Genus Alphavirus

Yi Tan[1*], Albert J. Auguste[2], Naomi L. Forrester[2], Mathilde Guerbois[2], Amy Ransier[3], Rebecca A. Halpin[3], Susmita Shrivastava[3], Timothy B. Stockwell[3], Vivien G. Dugan[3], David E. Wentworth[3], Scott C. Weaver[2], Suman R. Das[3]

[1.] J. Craig Venter Institute, [2.] Institute for Human Infections and Immunity, Center for Biodefense and Emerging Infectious Diseases, and Department of Pathology, University of Texas Medical Branch, Galveston, Texas, USA, [3.] J. Craig Venter Institute, Rockville, Maryland USA

Background/Objective

Eastern equine encephalitis virus (EEEV) is a mosquito-borne zoonotic virus in genus Alphavirus, present in the Americas and the Caribbean. Previous studies suggest that EEEV consists of four distinct genetic lineages: one lineage circulated in North America and the Caribbean (NA EEEV) and three lineages circulated in Central and South America (SA EEEV). However, those analyses were only based on the partial or complete sequence of structural polyprotein open reading frames, since the publicly available full genome sequences are very limited.

Method

We sequenced 99 full genomes of historical EEEV strains collected from 1933 to 2010, which expended public available full genome sequences of EEEV by ten folds. Here, using these full genome sequence data, we performed a comprehensive phylogenetic analysis of total 110 full genomes.

Result

Our full genome analysis confirmed the four distinct genetic lineages with single, monophyletic NA EEEV lineage restricted in North American/Caribbean and three divergent SA EEEV lineages restricted in Central/South America. NA EEEV lineage and three SA EEEV lineages share 76.4% similarity in nucleotide level across the genome and 89% similarity in amino acid level in complete CDS. The overall nucleotide substitution rate per year is estimated at 1.4E-4 with 95% HPD interval from 1.2E-4 to 1.6E-4. NA EEEV lineage appears to have diverged from SA EEEV lineages at 2,434-4,858 years ago. The time of most recent common ancestor (tMRCA) of three SA EEEV lineages was estimated at 1,508-2,667 years ago and tMRCA of monophyletic NA EEEV lineage was narrowly estimated at the year of 1920 with 95% HPD interval from 1916 to 1938.

Conclusion

The difference of geographic profile, genetic diversity and evolutionary histories of NA EEEV and SA EEEV lineages suggests that NA EEEV and SA EEEV may actually be distinct viruses. Our findings support the renaming of SA EEEV to Madariaga Virus (MADV).