Infection, Genetics and Evolution 17 (2013) 46–50

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Epidemiological history and phylogeography of West Nile virus lineage 2

Massimo Ciccozzi [a,*,1], Simone Peletto [b,1], Eleonora Cella [a], Marta Giovanetti [a], Alessia Lai [c], Elena Gabanelli [c], Pier Luigi Acutis [b], Paola Modesto [b], Giovanni Rezza [a], Alexander E. Platonov [d], Alessandra Lo Presti [a], Gianguglielmo Zehender [c]

[a] Department of Infectious, Parasitic, and Immunomediated Disease, National Institute of Health, Rome, Italy
[b] Istituto Zooprofilattico Sperimentale del Piemonte, Liguria e Valle d'Aosta, Turin, Italy
[c] Department of Clinical Sciences Luigi Sacco, Section of Infectious Diseases, University of Milan, Milan, Italy
[d] Central Research Institute of Epidemiology, Moscow, Russia

ARTICLE INFO

Article history:
Received 4 December 2012
Received in revised form 14 March 2013
Accepted 20 March 2013
Available online 29 March 2013

Keywords:
West Nile virus
Phylogeography
Epidemiology

ABSTRACT

West Nile virus (WNV) was first isolated in Uganda. In Europe WNV was sporadically detected until 1996, since then the virus has been regularly isolated from birds and mosquitoes and caused several outbreaks in horses and humans. Phylogenetic analysis showed two main different WNV lineages. The lineage 1 is widespread and segregates into different subclades (1a–c). WNV-1a includes numerous strains from Africa, America, and Eurasia. The spatio-temporal history of WNV-1a in Europe was recently described, identifying two main routes of dispersion, one in Eastern and the second in Western Europe. The West Nile lineage 2 (WNV-2) is mainly present in sub-Saharan Africa but has been recently emerged in Eastern and Western European countries. In this study we reconstruct the phylogeny of WNV-2 on a spatio-temporal scale in order to estimate the time of origin and patterns of geographical dispersal of the different isolates, particularly in Europe. Phylogeography findings obtained from E and NS5 gene analyses suggest that there were at least two separate introductions of WNV-2 from the African continent dated back approximately to the year 1999 (Central Europe) and 2000 (Russia), respectively. The epidemiological implications and clinical consequences of lineage 1 and 2 cocirculation deserve further investigations.

1. Introduction

West Nile virus (WNV) is a single-stranded RNA Flavivirus belonging to the Japanese encephalitis virus group. Its genome encodes for a single polyprotein which is then processed by viral and cellular proteases in order to produce three structural (capsid, C; membrane, M; and envelope, E) and seven non-structural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5) (Brinton, 2002). In nature WNV is transmitted among birds by culicine mosquitoes. Recent studies in the United State have found infection in >1300 species of birds and >60 species of mosquitoes. The waterfowl, especially cormorants and storks, are the elements of an enzootic "sylvatic" cycle including ornithophilic mosquitoes, particularly *Culex pipiens*, as a vector. Members of the order Passeriformes (jays, blackbirds, finches, sparrows, crows) and Columbiformes (collared doves) seem to be also important in maintaining the virus in nature because of their high viraemias. Many of these species are synanthropic, that is ecologically associated with humans, and function as the hosts in an "urban cycle". Because of their presumed low and transient viraemias, humans and horses are not considered important in the natural transmission cycle but are accidental dead-end hosts (Hayes et al., 2005).

The virus was first isolated in Uganda in 1937 and, during the early 1950s, it was isolated in sporadic cases and epidemic outbreaks involving humans and/or horses in Africa and the Middle East (Zehender et al., 2011). WNV has been identified in both sub-Saharan and Northern Africa, Europe, Asia, Australia (Zeller and Schuffenecker, 2004) and in the North-America since the summer of 1999 (Hayes, 2001; Lanciotti et al., 1999).

In Europe, it was only sporadically detected until 1996, when a major outbreak of infection characterized by a high (~10%) fatality rate occurred in Romania (Zeller and Schuffenecker, 2004). Since then, the virus has been isolated from horses, humans birds, mosquitoes and other vectors in a number of Eastern and Western European countries (Calistri et al., 2010; Murgue et al., 2001).

From a phylogenetic point of view, several WNV lineages have so far been described (Lanciotti et al., 2002). The lineage 1 is widespread and further segregates into different subclades (1a–c): WNV-1a, which includes strains from all over the world, the Australian Kunjin (1b) and some Indian strains (1c) (Lanciotti et al.,

* Corresponding author. Address: Department of Infectious, Parasitic and Immunomediated Diseases, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Roma, Italy.
E-mail address: massimo.ciccozzi@iss.it (M. Ciccozzi).
[1] Both authors contributed equally to this work.

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.03.034

2002; Kramer et al., 2008). Recently, by using a "phylodynamic" framework we have reconstructed the spatio-temporal history of WNV-1a in Europe, identifying two main routes of dispersion, one in Eastern and the second in Western Europe (Zehender et al., 2011). Whereas lineage 2 is mainly present in sub-Saharan Africa, Madagascar and has only recently also been identified in other geographic areas such as Eastern Europe, where it was detected in 2004 in Hungary (Bakonyi et al., 2006), in 2010 in Rumania and Greece (Sirbu et al., 2011; Chaskopoulou et al., 2011; Valiakos et al., 2011), in 2011 in Italy (Bagnarelli et al., 2011) and in 2004–2011 in Russia (Platonov et al., 2008, 2011). Moreover animal cases of WNV-2 infection have been identified more recently in Austria and Italy (Wodak et al., 2011; Savini et al., 2012). WNV lineage 3 is represented by a single isolate from the Rabensburg region in Czech Republic (Bakonyi et al., 2005), whereas lineage 4 was identified in the Southern Russia (Lvov et al., 2004; Shopenskaya et al., 2008), and lineage 5 was only found in India (Bondre et al., 2007). Other putative lineages have been described from other regions (Mackenzie and Williams, 2009; Vazquez et al., 2010).

The aim of this study was to reconstruct the phylogeny of WNV-2 on a spatio-temporal scale in order to estimate the time of origin and patterns of geographical dispersal of the different isolates, particularly in Europe.

2. Material and methods

2.1. Sequence data-sets

Two different WNV-2 data-sets were built and examined separately in order to estimate individual substitution rates and to better investigate phylogeographic relationships.

The first data-set consisted of 51 WNV-2 envelope protein (E) sequences available in GenBank (<http://www.ncbi.nlm.nih.gov/>). The sampling dates ranged from 1937 to 2011. The sampling locations were Uganda ($n = 2$), Central African Republic ($n = 3$), Madagascar ($n = 4$), South Africa ($n = 28$), Hungary ($n = 1$), Senegal ($n = 2$), Russia ($n = 6$), Mozambique ($n = 1$), Botswana ($n = 1$), Greece ($n = 1$), Italy ($n = 1$), Congo ($n = 1$).

The second data-set included 31 WNV-2 NS5 sequences available in GenBank. The sampling dates ranged from 1937 to 2011. The sampling locations were Uganda ($n = 3$), Italy ($n = 2$), Greece ($n = 2$), Hungary ($n = 2$), South Africa ($n = 10$), Congo ($n = 1$), Central African Republic ($n = 1$), Austria ($n = 6$), Russia ($n = 1$), Madagascar ($n = 1$), Senegal ($n = 1$), Tanzania ($n = 1$). All the sequences downloaded were selected on the basis of the following inclusion criteria: (1) the sequences had to have been already published in peer-reviewed journals; (2) there had to be no uncertainty about the lineage assignment of each sequence; (3) the city/state of origin were known and clearly established in the original publication.

All the sequences were aligned using Clustal X software (Thompson et al., 1994). Manual editing was performed using BioEdit software v. 7.0 (Hall, 1999). ModelTest v. 3.6 (Posada and Buckley, 2004) was used to select the simplest evolutionary model that adequately fitted the sequence data. All the sequences used are listed in Table S1.

2.2. Evolutionary rate estimates, time-scaled phylogeny reconstruction and Bayesian phylogeography

The dated trees and evolutionary rates were co-estimated by using a Bayesian MCMC approach (Beast v. 1.6.1, <http://beast.bio.ed.ac.uk>) (Drummond et al., 2005; Drummond and Rambaut, 2007) implementing for the first dataset a HKY + Invariant + Gamma model and for the second dataset a TN93 + I + G model.

For each dataset we compared both a strict and uncorrelated log-normal relaxed clock models and as coalescent priors, we compared four parametric demographic models of population growth (constant size, exponential, logistic growth, expantional) and a Bayesian skyline plot (BSP, a non-parametric piecewise-constant model).

Chains were conducted for at least 100×10^6 generations, and sampled every 10,000 steps. Convergence was assessed on the basis of the effective sampling size (ESS) and only parameter estimates with ESS's of >200 were accepted. Uncertainty in the estimates was indicated by 95% highest posterior density (95% HPD) intervals and the best fitting models were selected by a Bayes factor (BF, using marginal likelihoods) implemented in Beast (Drummond and Rambaut, 2007). In accordance with Kass and Raftery (1995), the strength of the evidence against H_0 was evaluated as follows: 2lnBF < 2 no evidence; 2–6 weak evidence; 6–10 strong evidence, and >10 very strong evidence. A negative 2lnBF indicates evidence in favour of H_0. Only values of $\geqslant 6$ were considered significant. The trees were summarized in a target tree by the Tree Annotator program included in the Beast package by choosing the tree with the maximum product of posterior probabilities (maximum clade credibility) after a 10% burn-in. The time of the most recent common ancestor (tMRCA) estimates were expressed as mean and 95% HPD years before the most recent sampling dates, corresponding to 2011 in this study.

The continuous-time Markov Chain (CTMC) process over discrete sampling locations implemented in Beast (Drummond and Rambaut, 2007) was used for the geographical analysis for the first and second dataset, implementing the Bayesian Stochastic Search Variable Selection (BSSVS) model which allows the diffusion rates to be zero with a positive prior probability. Comparison of the posterior and prior probabilities of the individual rates being zero provided a formal BF for testing the significance of the linkage between locations.

For NS5 phylogeographic analysis the sequences from Congo, Senegal, Central African Republic and Tanzania were grouped into a single location group called CA (Central Africa).

The final trees were manipulated in FigTree v. 1.3 for display. The most probable location of the main nodes were highlighted by a label.

3. Results

3.1. Estimated rates of WNV-2 E and NS5 gene evolution

The mean evolutionary rate of the E gene sequences of WNV-2, was estimated on the first dataset of 51 sequences. BF analysis showed that the relaxed clock fitted the data significantly better than the strict clock (2lnBF between the strict and relaxed clock was 9.75 in favor of the second). Under the relaxed clock, the BF analysis showed that the BSP was better than the other models (2lnBF > 40). The estimated mean value of the WNV-2 E gene evolutionary rate was 1.2×10^{-3} substitution/site/year (95% HPD: 2.5×10^{-4}–2.1×10^{-3}).

The mean evolutionary rate of the NS5 gene sequences of WNV-2 was estimated on the second dataset of 31 sequences.

BF analysis showed that the relaxed clock fitted the data significantly better than the strict clock (2lnBF between the strict and relaxed clock was 8 in favor of the second). Under the relaxed clock, the BF analysis showed that the BSP was better than the other models (2lnBF > 6).

The estimated mean value of the WNV-2 NS5 gene evolutionary rate was 5.8×10^{-4} substitution/site/year (95% HPD: 2.5×10^{-4}–8.9×10^{-4}).