Infection, Genetics and Evolution 22 (2014) 30–39

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Global phylogeography of Dengue type 1 and 2 viruses reveals the role of India

Atul M. Walimbe [a], Mrunalini Lotankar [a], D. Cecilia [b], Sarah S. Cherian [a,*]

[a] Bioinformatics and Data Management Group, National Institute of Virology, 20-A, Dr. Ambedkar Road, Post Box No. 11, Pune 411001, Maharashtra, India
[b] Dengue Group, National Institute of Virology, 20-A, Dr. Ambedkar Road, Post Box No. 11, Pune 411001, Maharashtra, India

ARTICLE INFO

Article history:
Received 17 May 2013
Received in revised form 2 January 2014
Accepted 2 January 2014
Available online 10 January 2014

Keywords:
DENV-1
DENV-2
Bayesian phylogeny
Phylogeography
Migration patterns
India

ABSTRACT

Patterns in virus dispersal and epidemiology of viral diseases can be revealed by phylogeographic studies. Currently knowledge about phylogeography of Dengue virus (DENV) Types 1 and 2 is limited. We carried out the phylogeographic analyses for DENV-1 and DENV-2, by the Bayesian Markov Chain Monte Carlo (MCMC) approach, with emphasis on Indian isolates in relation to the global evolutionary dynamics of the viruses. More than 250 E-gene sequences of each virus, available in GenBank, were used for the analyses. The study was focused on understanding the most likely geographical origin for the major genotypes and sub-lineages of DENV-1/DENV-2 and also the possible pathways in the dispersal of the virus.

The results showed that for DENV-1, Southeast Asia was the most likely geographical origin and India was determined to be the ancestral location of the Cosmopolitan genotype circulating in India, Sri Lanka, West and East Africa, Caribbean region, East and Southeast Asia. For DENV-2, the ancestral source could not be precisely inferred. Further, in spite of the earliest isolate from Trinidad-1953 of the American genotype, it was depicted that India may have been the probable ancestor of this genotype. India was also determined to be the ancestral location of a subgroup of the Cosmopolitan genotype. It was noted that DENV-1 and DENV-2 were introduced into India during 1940s and 1910s respectively. Subsequently, dispersal of both the viruses between India and different regions including West, East and Central Africa, Southeast and East Asia and Caribbean was inferred. Overall, the current study provides insight into the spatial as well as temporal dynamics of dengue virus serotypes 1 and 2.

1. Introduction

The Dengue virus (DENV) belonging to family *Flaviviridae* is the most prevalent arbovirus, infecting nearly 50–100 million people yearly. Dengue fever which is the most common manifestation of the infection is caused by any of the four antigenically distinct serotypes of Dengue virus (DENV 1–4) (World Health Organization, 2013; Henchal and Putnak, 1990).

The rampage of dengue has been recorded since the late 17th century with epidemics of dengue-like illness being reported in Asia, Africa and North America (Gubler, 2006). The virus attained a global distribution during the 18th century though over the past two decades, regular epidemics of Dengue fever (DF) and dengue hemorrhagic fever (DHF) have been reported in Sri Lanka, India, the Maldives Islands, Bangladesh and Pakistan (Messer et al., 2002, 2003; Rahman et al., 2002).

Several studies have been aimed at understanding the epidemiology of DENV, rates and dates of evolution, and selection pressure in its different genes and genotypes (Rico-Hesse, 1990; de Zanotto et al., 1996; Rico-Hesse et al., 1997; Twiddy et al., 2003; Weaver and Vasilakis, 2009). Further, a recent review (Chen and Vasilakis, 2011) reports large scale phylogenetic studies of all dengue serotypes and discusses how factors such as rates of evolution, selection pressures, population sizes and evolutionary constraints influences DENV transmission, pathogenesis and emergence. New approaches reveal the patterns in the spread and epidemiology of viral diseases by phylogeographic studies. Currently, knowledge about the phylogeography of DENV-1 is mainly limited to studies done for a particular country or region (America by Allicock et al., 2012; Brazil by Drumond et al., 2012 and Asia by Sun and Meng, 2013). Therefore, the knowledge about DENV-1 phylogeography is elaborative for those regions. In addition, a recent work based on data from 45 distinct geographic locations isolated during the period 1944 to 2009 (Villabona-Arenas et al., 2013) analyzed the world-wide spread of DENV-1. Though the study revealed the demographic history of the major DENV-1 genotypes circulating in the Asian and South-Pacific regions, the source of dispersal into the Caribbean could not be ascertained unanimously probably

* Corresponding author. Tel.: +91 020 26127301/26006213; fax: +91 020 26122669.
E-mail addresses: atulmw@gmail.com (A.M. Walimbe), mrunallotankar@gmail.com (M. Lotankar), Cecilia.dayaraj@gmail.com (D. Cecilia), cheriansarah@yahoo.co.in (S.S. Cherian).

1567-1348/$ - see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2014.01.001

because of missing temporal data from certain countries. In case of DENV-2, the only available reports have investigated the phylogeography in a restricted geographic area (for Caribbean basin by Foster et al., 2004; Vietnam by Rabaa et al., 2010; Brazil by Drumond et al., 2013 and Peru by Cruz et al., 2013).

Phylogeography studies have similarly been carried out for DENV-3 and DENV-4 (Araújo et al., 2009; Villabona-Arenas et al., 2011) inferring the evolutionary history of these serotypes in the context of global spread. As per reports, the geographic structure of current dengue diversity is strongly biased towards Asia, supporting the hypotheses of Asian origins for each serotype (Bennett, 2010). Our own previous studies, done for DENV-3 and DENV-4, revealed the important role of India and Sri Lanka in the evolution and dispersal of these serotypes (Patil et al., 2012). On the other hand, our previous studies for DENV-1 (Patil et al., 2011) and DENV-2 (Kumar et al., 2010) had applied molecular clock approaches to determine rates of nucleotide substitution and the time to most recent common ancestor (tMRCA), along with site-specific selection pressure studies. These studies also indicated transmission of DENV-1 and DENV-2 among India, Caribbean Islands and the Americas. However, the dispersal remained to be elucidated. Hence, the current study was carried out to understand the direction of migration of DENV-1 and DENV-2 and the role of India in its geographical spread. The current Bayesian analyses study utilized a larger and/or more temporally and geographically spread data than studies done prior. Hence, the study was expected to complete the picture of molecular epidemiology and phylogeography of both these dengue serotypes.

2. Materials and methods

2.1. DENV-1 and DENV-2 molecular data

Full length E-gene sequences (~1485 bp) of time-stamped isolates of DENV-1 and -2, representing worldwide geographical regions were considered for this study. All such sequences of DENV-1 and DENV-2 were downloaded from GenBank with the accession number, year of isolation, host species, geographical region, and strain name. Attenuated, patented, clones and artificial sequences were discarded. Initially a distance based tree was constructed, from which down sampled datasets were built by selection of representatives based on year of isolation, country and the assignment to a monophyletic group in the tree. Tables S1 and S2 give the details regarding the sequences of DENV-1 (n = 269) and DENV-2 (n = 307) used in this study respectively.

The geographic regions assigned for DENV-1 and DENV-2 phylogeography analyses were East Asia, South-Central Asia, Southeast Asia, India, Middle East, East Africa, Central Africa, West Africa, Caribbean Islands, Central America, North America, South America and Oceania (Table S3). The time span covered for DENV-1 and DENV-2 ranges from 1944–2011 and the details regarding the same are provided in Tables S4 and S5.

2.2. Bayesian phylogenetic analysis

MEGA 5 (Tamura et al., 2011) was used for multiple sequence alignment based on ClustalW. Percent nucleotide identities (PNI) were calculated using p-distances in MEGA 5. The best-fit model of nucleotide substitution was selected on the basis of Akaike Information Criterion (AIC) available in PAUP v.4b10 (Swofford, 1998) and ModelTest 3.5 (Posada and Crandall, 1998). For DENV-1 and DENV-2 datasets the GTR + G + I model (general time-reversible model with gamma-distributed rates of variation among sites and a proportion of invariable sites) was found to be the best-fit model.

The temporal information of the sequence data was used to estimate the evolutionary rate and the ancestral time for different genotypes of DENV-1 and DENV-2, by generating a Maximum Clade Credibility (MCC) tree using the Bayesian Markov Chain Monte Carlo approach as implemented in BEAST 1.6.2 (Drummond and Rambaut, 2007). For this, we employed both strict and relaxed (uncorrelated exponential and uncorrelated lognormal) clock (Drummond et al., 2006) models with the Bayesian Skyline tree prior. Three independent runs of the chain were carried out, each with at least 100 million generations and sampling frequency of 5000. The posterior probability and marginal likelihood of the models were used to choose the most suitable model for the data (Suchard et al., 2001). The convergence of the chain was assessed by Tracer 1.5 and the MCC tree was visualized in FigTree 1.2.3. For both datasets, the Bayes Factor analysis indicated that the uncorrelated exponential clock model fits better than strict clock or uncorrelated lognormal clock model (data not shown). The corresponding output files generated by BEAST were utilized for further analysis.

2.3. Phylogeographic and migration pattern analyses

The phylogeny-trait association tests (AI, Association Index and PS, Parsimony Score) available in BaTS (Parker et al., 2008) were performed to evaluate the association between phylogeny and geographical locations of the sequence data and hence the suitability of phylogeography analysis. The tests indicated strong association between sampling location and phylogeny for both DENV-1 and DENV-2 (Table S6), thereby implying the suitability of phylogeographic analysis.

The spatial information of the isolates was hence used to infer the geographic spread patterns of the virus by fitting a standard continuous-time Markov chain (CTMC) model with the Bayesian stochastic search variable selection (BSSVS) (Lemey et al., 2009) in BEAST. The outputs of the Bayesian phylogeographic analyses, as generated by BEAST, were summarized using maximum clade credibility (MCC) trees. An MCC tree is a point-estimate characterizing the posterior distribution of trees and represents the tree topology yielding the highest product of individual clade probabilities in their posterior sample; branch lengths in these MCC trees are posterior median estimates (Lemey et al., 2009). Further the tree nodes were annotated with their most probable (modal) location states via color labeling. The MCC tree obtained under the CTMC model was also input to the program SPREAD 1.0.3 (Bielejec et al., 2011) to visualize and analyze the dispersion pathways. Using Bayes Factor test available in SPREAD, the well-supported migration links between the different geographical regions were identified.

3. Results and Discussion

3.1. Bayesian phylogeny of DENV-1

The Maximum Clade Credibility (MCC) tree, under the relaxed exponential clock, revealed delineation of this serotype into five genotypes designated as I–V (Goncalvez et al., 2002) or as described based on their geographical distribution (Domingo et al., 2006; Patil et al., 2011) (Fig. 1). The strains from North America, East Asia, South Central Asia, Middle East, India, and East Africa along with some strains from Southeast Asia were included in the Asian genotype (Genotype I). The South Pacific genotype (Genotype IV) comprised of strains from East and Southeast Asia, East Africa, Central and North America and Oceania region. The cosmopolitan genotype (Genotype V) mainly contained strains from India, East Africa, West Africa, and Southeast Asia, South