Infection, Genetics and Evolution 18 (2013) 202–212

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Genomic mosaicism in two strains of dengue virus type 3

Christian Julián Villabona-Arenas, Anderson Fernandes de Brito, Paolo Marinho de Andrade Zanotto *

Laboratory of Molecular Evolution and Bioinformatics, Department of Microbiology, Institute of Biomedical Sciences, University of São Paulo, Av. Prof. Lineu Prestes, 1734, CEP: 05508-000, São Paulo, SP, Brazil

ARTICLE INFO

Article history:
Received 5 March 2013
Received in revised form 14 May 2013
Accepted 15 May 2013
Available online 30 May 2013

Keywords:
Dengue virus type 3
Phylogeny
Recombination
Split decomposition

ABSTRACT

Recombination is a significant factor driving genomic evolution, but it is not well understood in Dengue virus. We used phylogenetic methods to search for recombination in 636 Dengue virus type 3 (DENV-3) genomes and unveiled complex recombination patterns in two strains, which appear to be the outcome of recombination between genotype II and genotype I parental DENV-3 lineages. Our findings of genomic mosaic structures suggest that strand switching during RNA synthesis may be involved in the generation of genetic diversity in dengue viruses.

1. Introduction

The availability of increasing sequence information from different localities and refined phylogenetic tools make it possible to address many of the processes that lead to genetic diversification, adaptation and spread of *Dengue virus* (DENV) (Weaver and Vasilakis, 2009). Populations of RNA viruses usually harbor genetic diversity, associated with high mutation rates, small effective population and large population sizes (Simon-Loriere and Holmes, 2011). The exchange of genetic material is an important mechanism that allows adaptive change among natural populations and evidences its ubiquitous role in evolution (Awadalla, 2003).

Dengue fever in humans is caused by enveloped, positive single stranded RNA ([+] ssRNA) mosquito-borne viruses, which belong to the genus *Flavivirus* (Family: Flaviviridae) and comprise four serotypes (DENV-1 to DENV-4) (Karabatsos, 1985). Their genomes (~11 kbp) encode a single open reading frame, flanked by highly structured 5′ and 3′ untranslated regions (UTRs) (Lindenbach and Rice, 2003; Rice et al., 1985). The N-terminal of the polyprotein encodes the three structural proteins (C-prM-E), followed by seven non-structural (NS) proteins (NS1–NS2A–NS2B–NS3-NS4A–NS4B–NS5) (Chambers et al., 1990; Rice et al., 1985).

Concomitant infection with two different strains (a prerequisite for DENV recombination) was demonstrated both in humans and in mosquitoes (Bharaj et al., 2008; Loroño-Pino et al., 1999; Laille et al., 1991; Montenegro et al., 2006; Wang et al., 2003). Recently, Thai et al. (2012) documented distinct lineages within the same host. Evidence of recombination between diverse dengue strains has arisen (Aaskov et al., 2007; AbuBakar et al., 2002; Carvalho et al., 2010; Chen et al., 2008; Craig et al., 2003; Lavergne et al., 2006; Uzcategui et al., 2001; Worobey et al., 1999) since the first report of recombination in natural populations in 1999 (Holmes et al., 1999). Also, recombination was shown in other [+] ssRNA viruses, (Becher and Tautz, 2011; Carney et al., 2011; Chuang and Chen, 2009; Gould et al., 2004; Kalinina et al., 2002; Reiter et al., 2011), Retroviruses (Nájera et al., 2002) and [−] ssRNA viruses (Sibold et al., 1999). Two main mechanisms have been proposed for viral RNA recombination: (i) non-replicative breakage and rejoining and (ii) replicative template switching (Cooper, 1974; Nagy and Simon, 1997). The first model is a viral replication-independent RNA recombination process in which RNAs spontaneously rearrange their sequences, as found in tissue culture cells infected by polioviruses (Gmyl et al., 1999) and pestiviruses (Becher and Tautz, 2011). The second mechanism predicts that during replication the viral RdRp pauses RNA synthesis and eventually jumps, together with the incomplete nascent RNA, to another template in order to continue RNA synthesis. This mechanism has also been studied for [+] RNA such as the *Brome mosaic virus* and polioviruses, pestiviruses, carmoviruses and tombusviruses (Arnold and Cameron, 1999; Cheng and Nagy, 2003; Dahourou et al., 2002; Duggal et al., 1997; Simon and Nagy, 1996).

Due to modern transportation, wide-range dispersion of viruses can take place within 24–36 h resulting in several epidemic outbreaks, some of which involve the co-circulation of different viral strains, which increases the likelihood of mixed infections (Gould and Gritsun, 2006). Herein, we made use of available data and distinct analytical approaches to identify for the first time recombinant viruses in whole-genome alignments of DENV-3.

* Corresponding author. Tel.: +55 11 3091 8453.
E-mail address: pzanotto@usp.br (P.M. de Andrade Zanotto).

1567-1348/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2013.05.012

2. Materials and methods

2.1. Data set

Genomic DENV-3 strains were retrieved from GenBank (Table S1, in the supplemental material) and their source feature inspected, as a caution required in obtaining data from a public database, to check each entry name and to exclude mutants and non-human/non-mosquitoes isolates. Alignments were built with the Muscle 3.8.31 program (Edgar, 2004a,b), manually inspected, and trimmed to include only coding regions.

2.2. Detecting recombination

We used the automated algorithms contained within the recombination detection program 3.0 (RDP3) (Martin et al., 2010) to screen for putative recombination events in the coding region alignment. We used the default settings for each algorithm with a highest acceptable *p*-value of 0.01 and the standard Bonferroni correction (a *p*-value modifier that decrease the *p*-value cutoff according to the size of the dataset being examined) and set the analysis to report all hits detected by 6 or more algorithms. Once a putative recombination event was detected, a smaller dataset was built from selected sequences that included the likely recombinant, surrogate parental taxa, and reference strains.

2.3. Detection of recombinant breakpoints in the whole-genome alignment

Two Hidden Markov Models procedures (HMM) were used to corroborate the results from the automated tests and accurately locate breakpoints: (i) The Husmeier and Wright approach (Husmeier and McGuire, 2003) with four taxa (a likely recombinant, the surrogate parental taxa and one reference strain) and a time-reversible F84 substitution model (Felsenstein and Churchill, 1995) implemented in the Topali 2.5 software (Milne et al., 2009); and (ii) the Westesson and Holmes approach (86) that follows the structural EM heuristic (Friedman et al., 2002) using the RecHMM software by setting the minimum cutoff length for a recombinant region to 400 bp and using two additional reference strains. Significant tree topology changes allow partition of the alignment into genomic regions with apparent distinct evolutionary histories. For instance, for four taxa, there are three different unrooted tree topologies; in the presence of recombination, the tree topology can change and thus become a random variable that depends on the sites under consideration (the hidden states of the HMM represent the three topologies).

2.4. Phylogenetic network analyses

Alignments were built for each detected recombinant, including the most plausible surrogate parental taxa and three reference sequences from each genotype (free of recombination signal as evidenced by further automated detection tests). A pairwise homoplasy index (PHI) based on the principle of refined incompatibility was calculated for each alignment to test for recombination (Bruen et al., 2006). Afterwards, Split decomposition analyses (Bandelt and Dress, 1992) were done to assess the presence of a 'phylogenetic network' as an indication of reticulate evolution. This approach enhances phylogenetic analyses of distance data by unveiling incompatible groupings of taxa (splits) that arise when recombination and/or horizontal gene transfer have occurred. Analyses were done with the SplitsTree 4.0 software (Huson and Kloepper, 2005), with a bootstrapping of 100 replicates, using a general time reversible substitution model (GTR). The BEAST software package (Drummond et al., 2012) was used to estimate the proportion of invariable sites (I) among site rate variation (γ); nucleotide frequencies; and substitution rates for the model.

2.5. Phylogenetic reconstruction

A Maximum Clade Credibility tree (MCC) and divergence times for each alignment were inferred by using a Bayesian Markov chain Monte Carlo (MCMC) approach as implemented in BEAST. The estimated GTR + I + γ model was used together with a relaxed molecular clock (Drummond et al., 2006) and a Bayesian Skyline population growth model (Drummond et al., 2005). Four MCMC chain lengths were run for 80 million generations sampling every 8000th generation. Four independent runs were done to improve convergence and effective sample sizes were assessed using Tracer 1.4 (http://beast.bio.ed.ac.uk/Tracer).

3. Results

3.1. DENV-3 recombinants

We obtained 636 genomic sequences from GenBank (Table S1, in the supplemental material) and the final alignment included 10,170 coding nucleotide positions. We detected a significant recombination event, with an exceedingly small probability of 5.7E-18 of random occurrence obtained with multiple comparisons correction by all the algorithms implemented in RDP 3.0 (RDP, GENECONV, BootScan, MaxChi, Chimaera, Siscan and 3Seq). For this single event (Fig. 1) DQ401690 (Indonesia 1982) and AY496879 (Philippines 1997) appeared as recombinants, while AY858038 (Indonesia 1988), AY496877 (Bangladesh 2002) and FN429900 (Malaysia 1999) appeared as surrogate parental taxa.

3.2. Phylogenetic network of recombinants

We sought to characterize these events using different criteria and the dataset from Table 1. We used a phylogenetic network approach to provide an implicit picture of evolutionary relationships where the sets of parallel edges (splits) represented incompatible and ambiguous signals in the dataset. In this analysis, we included the complete monophyletic groups to which the surrogate parental taxa belong. These groups comprised the surrogate parental lineage and were retrieved from a maximum likelihood tree inferred with Garli 2.0 (Zwickl, 2006) and the 636 whole-genome sequences (Fig. S1). Our findings provided statistically significant evidence for recombination as indicated by a PHI test (p-value <0.05) and the split decomposition output showed multiple bands of parallel edges associating PH-1997 and ID-1982 with two different genotypes (Fig. 2A). Nevertheless, these split networks did not conclusively support genotype assignment either to genotype II or I, but suggested that there was evidence for both.

We created a chimera to better scrutinize our findings in a computer simulation, since the comparative data did not portray a strong tree-like structure. It was built *in silico* by joining together genomic regions from DENV-1 and DENV-4 serotypes and using it alongside other DENV sequences in a Split Decomposition analysis (Fig. S1, in the supplemental material). Fig. S2A illustrates the resulting phylogenetic network. The interesting result was that the chimera had a region of split decomposition due to conflicting evolutionary histories involving two diverse ancestral monophyletic groups, as in our real data.