Virus Research 192 (2014) 46–51

Contents lists available at ScienceDirect

Virus Research

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

Short communication

Evolution of the vesicular stomatitis viruses: Divergence and codon usage bias

Yang Liang, Mei He, Chun-Bo Teng*

College of Life Science, Northeast Forestry University, Harbin, China

ARTICLE INFO

Article history:
Received 1 July 2014
Received in revised form 5 August 2014
Accepted 18 August 2014
Available online 27 August 2014

Keywords:
Vesicular stomatitis viruses
Divergence
Codon usage bias

ABSTRACT

Four *Vesiculovirus* species causing vesicular stomatitis in the Americas, together with two closely related insect isolates, can be phyletically classified into two major groups: New Jersey (NJ) and Indiana (IN). Here, Bayesian coalescent analyses were conducted to the time-stamped entire coding sequences of the *G* gene of these vesiculoviruses, with emphasis on their divergence scenario. The primary bifurcation was a much ancient event that might have taken place around 1.8 million years ago between NJ and IN, which shared a similar high mean rate. Interestingly, the overall codon usage bias pattern of these viruses resembled that of the insect vectors rather than the livestock hosts.

Vesicular stomatitis (VS) is a zoonotic disease endemic in the Western Hemisphere. Infecting a wide range of domestic and wild mammals, it is rarely fatal but highly painful. Horses, cattle and swine are the most afflicted and often suffer from lameness, weight loss and productivity reduction. In the latter two hosts, VS is of particular economic significance for clinical resemblance to the notorious exotic foot-and-mouth disease that spells animal quarantine and trade barrier. This urges rapid diagnosis for differentiation between the two diseases (Letchworth et al., 1999; Rodriguez and Pauszek, 2012).

Among the ten species of the *Vesiculovirus* genus that have been recognized by the 9th ICTV report (King et al., 2011), four are the etiologic agents of VS: *Vesicular stomatitis New Jersey virus* (VSNJV), *Vesicular stomatitis Indiana virus* (VSIV, formerly classified as subserotype IN1), *Cocal virus* (except the prototype COCV, formerly IN2), and *Vesicular stomatitis Alagoas virus* (VSAV, formerly IN3). The latter three, together with two singular insect isolates, *Maraba virus* (MARV) and CoAr_171638 (CoAr), phylogenetically compose a large IN group (Pauszek et al., 2011) (Fig. 1). Here, we continue to use the serological classification of IN for convenience.

Both NJ and IN1 are enzootic in the Americas and can be further divided into four major genotypes designated I–IV (Bilsel and Nichol, 1990), whereas IN2 and IN3 appear occasionally in South America (Rodriguez et al., 2002). Moreover, IN1 is not only the type species of the *Vesiculovirus* genus and the *Rhabdoviridae* family (Van Regenmortel et al., 2000), but a laboratory model for scientific research (Letchworth et al., 1999). Its enveloped virion is in typical bullet shape and encapsidates a non-segmented, negative-sense, single-stranded RNA genome of ~11 kb, in which the five common genes are organized as 3′-N-P-M-G-L-5′, denoting nucleo-, phospho-, matrix-, glyco-, and RNA polymerase protein, respectively (Rodriguez et al., 2002; Rose and Schubert, 1987).

Until now, the divergence of the six closely related vesiculoviruses, VSV-NJ, IN1, IN2, IN3, MARV, and CoAr, has not yet been dated, which will open the possibility to infer their evolutionary history and origin. For this purpose, as described previously (He et al., 2013), complete *G* coding sequences of 101 field isolates (Fig. 2) collected between 1949 and 2005 (Nichol, 1987; Pauszek et al., 2011; Pauszek and Rodriguez, 2012; Rodriguez et al., 2002) were retrieved from GenBank and aligned with CLUSTAL W (Thompson et al., 1997). First, intragenic recombination event was searched by multiple approaches as applied before (He et al., 2012) and no significant signal was detected.

Then, to estimate the time to the most recent common ancestor (TMRCA) of the six vesiculoviruses, the whole dataset was subjected to BEAST v1.7.4 (Drummond et al., 2012) under the best-fit nucleotide substitution model GTR + G + I (General Time Reversible model with a gamma distribution of rates and a proportion of invariable sites) determined by MODELTEST in HyPhy (Pond et al., 2005). When the exponential growth model was employed for better performance under the best-fit uncorrelated exponential clock recommended by the Bayes factor tests (Baele et al., 2012; Suchard et al., 2001), the mean TMRCA was calculated to be 3259 years up to

* Corresponding author at: College of Life Science, Northeast Forestry University, Harbin 150040, China. Tel.: +86 451 8219 1784; fax: +86 451 8219 1784.
E-mail address: chunboteng@nefu.edu.cn (C.-B. Teng).

http://dx.doi.org/10.1016/j.virusres.2014.08.013
0168-1702/

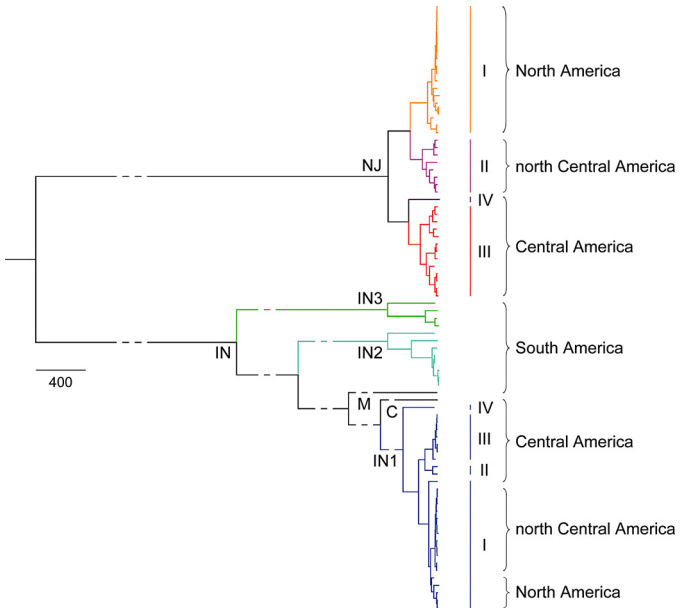

Fig. 1. Maximum clade credibility phylogeny of the six closely related vesiculoviruses based on the *G* gene. The tree is scaled to time generated under the relaxed uncorrelated exponential clock. Except *Vesicular stomatitis New Jersey virus* (NJ), the other five species, *Vesicular stomatitis Indiana virus* (IN1), *Cocal virus* (IN2), *Vesicular stomatitis Alagoas virus* (IN3), *Maraba virus* (M), and CoAr_171638 (C), compose a large Indiana (IN) group, as indicated about their nodes/branches. Further, both NJ and IN1 can be divided into four genotypes (I–IV). The lineages are in fine correlation with their geographical origins as shown on the right. The dashed branches are saturated due to purifying selection and cannot be reliably dated at present.

2005, with the 95% highest probability density (HPD) values varying from 1185 to 6128. As was clear in the maximum clade credibility (MCC) tree (Fig. 1), such ancient bifurcation took place between NJ and IN.

Seen from the NJ line (Fig. 2A), the first divergence episode was that the progenitor of I and II separated from that of III and IV in the early 17th century. However, the four genotypes did not form until more than a century later. When it came to the IN trunk (Fig. 2B), IN3 was the first one to branch off ~1700 years ago, followed by IN2 at an interval of ~500 years. Sequentially, MARV diverged out ~750 years ago. Having been separated from CoAr in the early 1500s, IN1 commenced to diversify around the 1720s, as marked by the emergence of an IV-type virus, the precursor of 59-PN-L. Among the other three genotypes emerging in the last century, II and III were sister taxa, whose ancestor diverged from that of I ~200 years ago.

However, as suggested by the much higher frequency of synonymous substitution ($d_N/d_S < 0.2$, Table 1) detected by Hyphy (Pond et al., 2005), the vesiculoviral *G* gene was under strong purifying selection which could lead to underestimation of the lengths of the long, deep branches (Wertheim et al., 2013). Thus, we re-estimated branch lengths in HyPhy using the branch site random effects likelihood (BS-REL) model accounting for variation in selection pressures across sites and lineages (Wertheim et al., 2013). The new estimates were then compared to those produced under the GTR + G + I model. It turned out that the long branches associated with the emergence of the six vesiculoviruses were all subjected to length expansion in varying degrees (Table S1).

Currently, the lengths of these saturated branches could only be calculated in a simple yet inaccurate way. Taking the total tree length as an example, it was estimated to be 1404 substitutions/site under the BS-REL model. Compared to that at 2.52 substitutions/site under the GTR + G + I model, there was a 557-fold expansion, that is, the divergence time between NJ and IN might be advanced by a big margin from 3259 years to 1.8 million years, suggesting far more ancient existence of these vesiculoviruses.

Notably, the shorter branches were little affected by substitution saturation, as pointed out by Wertheim et al. (2013). Therefore, age estimates of the four vesicular stomatitis viruses as well as the subtypes of NJ were more reliable (Table 1). Although NJ-type virus might have existed for millions of years, the 59 NJ isolates collected in modern times were descended from a most recent common ancestor (MRCA) at just 382 (95% HPD: 160–688) years old up to 2005, not to mention the much younger MRCAs of its three analyzed genotypes (I–III). Coincidentally, the MRCAs of IN2 and IN3 were of a similar age to that of NJ, whereas the MRCA of IN1 was ~100 years younger.

In spite of the inherent error-prone replication, NJ and IN1 can remain highly stable for many years in specific enzootic areas (Letchworth et al., 1999). Owing to the fine correlation between genetic relationship and geographical region (Bilsel and Nichol, 1990; Rodriguez, 2002), viral origin could be inferred from the MCC tree despite the currently uncertain time scale (Fig. 1). We supported the hypothesis that these viruses originated in the broadly defined equatorial America (Bilsel and Nichol, 1990) from Columbia to northern Brazil.

Hence, although not only the observation of the disease but the isolation of the pathogen took place firstly in the USA (Cotton, 1927; Hanson, 1952), both NJ and IN1 were not indigenous but exotic as the outcomes of two independent northward colonizing incidents. To be the dominant epidemic cause in waves at ~10-year intervals (Rodriguez, 2002; Rodriguez and Pauszek, 2012), NJ reached the USA in the late 18th century (Fig. 2A), before 1801 when outbreak of 'sore tongue', mimic symptom of VS, was observed in livestock (Letchworth et al., 1999). Notably, IN1 arrived there ~100 years later (Fig. 2B). Therefore, the agent that disabled ~4000 army horses required for the Civil War in 1862 (Hanson, 1952) was likely to be an ancestral NJ-type virus.

Moreover, high nucleotide substitution rates of the species/subtypes were presented by the accordingly partitioned *G* subsets. From the overall perspective, NJ and IN shared a similar speed at nearly 6×10^{-4} subs/site/year, whilst their subtypes exhibited rate differences varying in a wide range (Table 1). Among the three analyzed genotypes of NJ, II was estimated with the highest rate at 2.75×10^{-3} subs/site/year, about 3-fold higher than that of I, while among the three sub-serotypes of IN, IN3 was calculated to evolve fastest at 5.62×10^{-3} subs/site/year, nearly 9 times faster than that of IN1. However, as the panels of II and IN3 were represented by small numbers of isolates (Table 1), more entire sequence data are required to confirm the rate differences.

Furthermore, as a better knowledge of codon usage bias is essential to understanding the processes that govern the evolution of RNA viruses (Jenkins and Holmes, 2003), in terms of the effective number of codons used by a gene (*Nc*) and the frequency of (G + C) at the synonymous 3rd codon position (GC_{3S}), codon usage bias in each of the five common genes was measured by using CodonW 1.4.4 (http://codonw.sourceforge.net). *Nc* can take values from 20 for top bias when one codon is exclusively used for each amino acid, to 61 for no bias when all synonymous codons are equally used (Wright, 1990). Therefore, all *Nc* values here being higher than 40 (Table 2), each viral gene just had slight bias in codon usage.

Then, *Nc* was plotted against GC_{3S}, indicator of the extent of mutational bias (Fig. 2). If codon choice was constrained only by uneven G + C composition, the point would lie on or just below the curve of the expected values under the assumption of no selection (He et al., 2014b). Thus, most points being away from the curve suggested that usage variation of the five genes should have influences aside from mutational pressure. Actually, when the most dispersed *P* points were not included, *Nc* was in significantly positive correlation with not only GC_{3S} ($R = 0.54$, $P < 0.01$) but protein aromaticity