Arch Virol (2012) 157:2247–2251
DOI 10.1007/s00705-012-1420-x

ANNOTATED SEQUENCE RECORD

Full-length genome analysis of vesicular stomatitis New Jersey virus strains representing the phylogenetic and geographic diversity of the virus

Steven J. Pauszek · Luis L. Rodriguez

Received: 3 May 2012 / Accepted: 10 June 2012 / Published online: 24 July 2012

We describe the complete genomic sequence of nine isolates of vesicular stomatitis New Jersey virus (VSNJV) representing six distinct phylogenetic groups and spanning the known geographic range of the virus. The total genomic length (11,119-11,123 nt) and structure of these isolates were very similar to those of other vesicular stomatitis viruses. There were no differences in the number of amino acids in any of their viral proteins. As expected, the genomic termini, possessing initiation signals for viral transcription and replication, were highly conserved. These complete sequences, the first for natural isolates of VSNJV, should prove useful in the design of molecular detection tests and provide the basis for future functional genomics and pathogenesis studies.

Vesicular stomatitis (VS) virus causes disease in cattle, horses and pigs, and economic losses throughout North, Central and South America [16]. The clinical presentation of VS in cattle and pigs is markedly similar to foot-and-mouth disease, a devastating reportable disease, making a rapid differential diagnosis crucial. The etiologic agents of VS are viruses in the family *Rhabdoviridae*, genus *Vesiculovirus*, which have been classified into two serotypes, New Jersey (VSNJV) and Indiana (VSIV), both of which are endemic in northwest South America (Bolivia, Colombia, Ecuador, Peru, Venezuela), throughout Central America and in southern Mexico [8]. Throughout this work, virus nomenclature follows the Eight Report of the International Committee on Taxonomy of Viruses [4]. VSNJV accounts for the majority of clinical cases of vesicular disease in livestock throughout the Americas [5] [16]. In the United States, outbreaks of VS were reported in the southeastern states (Georgia, Florida, North Carolina, South Carolina) until the late 1970s. In the southwestern United States (Texas, New Mexico, Colorado, Wyoming, Utah), VS outbreaks are caused by cyclical incursions of VSV from endemic areas in Mexico every 8-10 years, causing significant economic losses to the cattle and horse industries [14, 16]. Although previous reports have documented overwintering of specific genetic lineages of VSNJV in the northwestern US for up to three years, disease outbreaks are usually associated with newly introduced viral lineages from endemic regions [13].

Multiple reports describing phylogenetic analysis of VSV field strains have shown that specific genetic lineages exist in different geographical areas [16]. Furthermore, studies suggest that ecological factors play an important role on the evolution of VSV [15]. In previous studies, we have described the complete genome sequences of VSIV strains from South, Central and North America [17] and the prototype viruses for the Indiana 2 and Indiana 3 subtypes [12]. However, despite the fact that VSNJV accounts for the vast majority of VS cases in the western hemisphere and that VSNJV was shown to be more pathogenic than VSIV in a natural host [10], very little work has been done on the full-length genome of this economically important virus. Current full-length genomic sequences deposited in GenBank are assembled from partial sequences of multiple VSNJV isolates of mixed and/or indeterminate passage history. Here, we report for the first time the characterization of the full-length genomic sequences of nine naturally occurring isolates of VSNJV, representing six

S. J. Pauszek · L. L. Rodriguez
Agricultural Research Service, US Department of Agriculture, Plum Island Animal Disease Center, Greenport, NY 11957, USA

S. J. Pauszek · L. L. Rodriguez (✉)
Agricultural Research Service, US Department of Agriculture, Plum Island Animal Disease Center, Orient, NY 11957, USA
e-mail: luis.rodriguez@ars.usda.gov

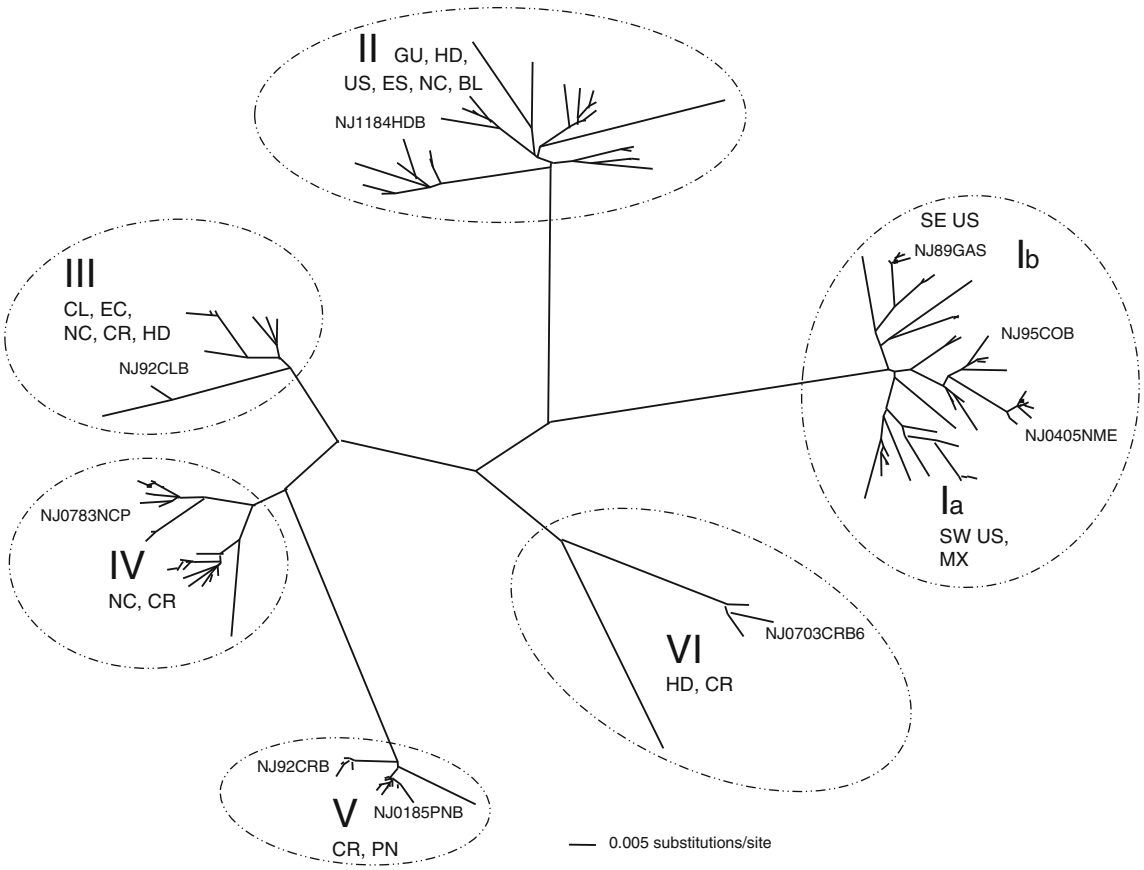

Fig. 1 Neighbor-joining tree of the phosphoprotein hypervariable region of the nine viruses analyzed in this study and all other available VSNJV sequences. Isolates were selected for sequencing from each of the six major clades, encircled by a broken line and labeled I-VI. For each clade, the country of origin of a viral isolate within the clade is indicated by a two-letter code as follows: BL, Belize; CL, Colombia; CR, Costa Rica; EC, Ecuador; ES, El Salvador; GU, Guatemala; HD, Honduras; MX, Mexico; NC, Nicaragua; PN, Panama; SE US, southeastern US; SW US, southwestern US. Additional isolate names were removed for clarity

phylogenetic groups and comprising the known geographic breadth of the virus.

To select the viruses for full-length genomic analysis, we aligned all available VSNJV phosphoprotein hypervariable region sequences using ClustalX [22] and reconstructed the phylogeny with PAUP* 4.0b10 [21]. The viruses formed six distinct genetic groups or clades (Fig. 1). Group I contains the viruses from the United States and Mexico. This group can be further divided into two subgroups: group Ia, containing viruses from the southeastern US, and group Ib, containing viruses from Mexico and the southwestern US. The greatest genetic diversity is observed in Central America, where viruses in group II principally originate from northern Central America, and groups IV, V and VI are mostly comprised of viruses from the middle and southern countries in Central America. Interestingly, group III contains viruses from the middle countries of Central America (Honduras, Nicaragua, and Costa Rica) and northern South America (Colombia, Ecuador) but not Panama.

At least one virus from each of the six groups in Fig. 1 was selected for full-length genomic sequencing to represent the known phylogeographic spectrum of the virus. Detailed isolate information is shown in Table 1. Virus stocks for full-length sequencing were made by infecting cells at a multiplicity of infection of 0.1 or less. Genomic sequences, including termini, were obtained and analyzed as described previously [12]. Primers for PCR and sequencing reactions were designed based on the published sequences of VSNJV or based on newly obtained sequences (primer sequences are available from the corresponding author upon request). Bioinformatic analysis was conducted using PAUP* 4.0b10 or MEGALIGN (DNASTAR).

The basic genomic structure of the VSNJV isolates was similar to that of other VS viruses, with a 47-nt 3' leader (Le) sequence followed by the five structural genes (Nucleoprotein [N], Phosphoprotein [P], Matrix [M], Glycoprotein [G], Polymerase [L]) and a 57-nt 5' trailer (Tr) sequence. Total genomic lengths were 11,119-11,123 nt which is shorter than previously seen in field isolates of

Table 1 Description of the viruses sequenced in this study

Virus	Year of Isolation	Area of isolation	Host Species	Pass History	Clade*
NJ89GAS	1989	Georgia, US	Sandflies	BHK p3	Ia
NJ95COB	1995	Colorado, US	Bovine	BHK p1	Ib
NJ0405NME	2005	New Mexico, US	Equine	Vero p1	Ib
NJ1184HDB	1984	Honduras	Bovine	BHK p3	II
NJ92CLB	1992	Colombia	Bovine	BHK p2	III
NJ0783NCP	1983	Nicaragua	Porcine	BHK p2	IV
NJ0185PNB	1985	Panama	Bovine	BHK p3	V
NJ92CRB	1992	Costa Rica	Bovine	BHK p3	V
NJ0703CRB6	2003	Costa Rica	Bovine	BHK p1	VI

Table 2 Bioinformatics analysis of the coding sequences of the sequenced viruses

Gene	CDS length	%GC	Ti/Tv ratio	Average, Minimum nt% similarity
N	1269	43 %	3.92	86.3, 82.0
P	825	41 %	3.13	85.0, 77.3
M	690	41 %	3.40	84.7, 79.3
G	1554	43 %	3.16	84.4, 78.6
L	6330	40 %	3.18	85.4, 80.6
Protein	No. of invariant aa (%)	Positions under one alternate aa replacement (%)	Positions under multiple aa replacement (%)	Average, Minimum aa% similarity
N	399 (95 %)	33 (8 %)	0 (0 %)	98.1, 96.7
P	201 (73 %)	53 (19 %)	20 (7 %)	89.2, 82.8
M	205 (90 %)	22 (9 %)	2 (1 %)	96.1, 91.7
G	441 (85 %)	67 (13 %)	9 (2 %)	95.0, 91.5
L	1831 (87 %)	219 (10 %)	59 (3 %)	94.9, 92.5

VSIV. Interestingly, there is significantly less variability in the overall lengths of VSNJV isolates compared to VSIV isolates (11,155-11,336 nt) [17]. There were no differences in length of the predicted aa sequence in any of the structural proteins of the VSNJV isolates (Table 2).

Elements governing viral replication have been mapped to the genomic termini of vesiculoviruses [11, 25]. For VSIV, the 15 terminal-most nucleotides of the 3' Le and the 5' Tr have been shown to be essential for replication of the positive- and negative-sense viral genome, respectively [25]. Despite the broad geographic and phylogenetic span of the isolates, the terminal 15 nucleotides were completely conserved in both the Le and Tr in all of the isolates (data not shown). Previous work has shown that increasing the complementarity between the 3' and 5' termini enhances the levels of positive- and negative-strand viral replication [24]. All of the isolates analyzed in this study showed 100 % complementarity through the 18 terminal-most nucleotides. While all vesiculoviruses possess a certain degree of terminal complementarity, the conservation of complementarity observed in the VSNJV isolates is interestingly high.

At each intergenic region, vesiculoviruses exhibit a conserved sequence that contains cis-acting signals for transcription termination, mRNA polyadenylation and downstream transcription reinitiation [18] [19]. This intergenic sequence is characteristically 23 nt in length (3'-AUAC$(U)_7$NNUUGUCNNNAG-5') with the $(U)_7$ polyadenylation signal followed by two non-transcribed nucleotides. These intergenic dinucleotides have a significant role in the cis-acting signals involved in the staggered termination and reinitiation of transcription in VSV [2, 20]. These dinucleotides were found to be quite heterogeneous in VSIV field isolates [17], with one virus even having an intergenic trinucleotide. However, for all of the viruses sequenced in this study, all intergenic dinucleotides were the same, (3'-GA-5'), even at the G-L junction, where variability was observed in VSIV. Interestingly, studies

have shown that the intergenic dinucleotide 3'-GA-5' is the most efficient sequence for transcription termination but not reinitiation [2].

Prior to this report, the solitary VSNJV G-L sequence available in GenBank was the Ogden strain, in which a unique variation of the conserved VS intergenic sequence was observed [9]. The Ogden sequence has a 19-nt insertion (with respect to VSIV) between the non-transcribed dinucleotides and the 3'-UUGUCNNAG-5' mRNA transcription start signal in L. Inserts of variable length seen in this region have been described previously for one specific Central American clade of VSIV [3]. Interestingly, this 19-nt sequence at the G-L junction of VSNJ isolates was not unique to one specific clade, but rather, it was present in all sequenced viruses. The length of the insert was quite conserved, as eight of the viruses also possessed a 19-nt insert, while NJ0703CRB6 had an 18-nt insert.

There was no difference in the predicted lengths of the five structural proteins, and the N and P proteins were the most and least conserved, respectively, at both the nucleotide and amino acid level (Table 2). The C'/C proteins, whose function remains elusive, were present in a second ORF in the P mRNA in all of the isolates analyzed here. There were no changes to predicted glycosylation sites in the glycoprotein.

Previous studies have shown that the multifunctional M protein has conserved methionines in the NH_2-terminal portion (positions 33 and 51 for VSIV, positions 33, 48 and 51 for VSNJV). Mutational analysis showed that these methionines are involved in cell rounding (position 33) [6] and inhibition of host-cell gene expression (position 51 and, to a lesser extent, position 48) [7]. We found that all three of these methionine positions (33, 48 and 51) were completely conserved in the viruses sequenced in this study.

The purpose of this work was to characterize the complete full-length sequence of natural isolates of VSNJV, the virus serotype responsible for the vast majority of outbreaks in the Americas. Isolates from every major geographic and evolutionary cluster, collected over a twenty-year span and from four different host species, were selected to give a comprehensive representation of the virus. Detailed knowledge of the primary structure of the VSNJV genome will aid in functional genomics studies using reverse genetics, allowing the identification of genetic determinants of viral virulence. This information should prove useful for the understanding of the VSNJV life cycle that could be applied to disease mitigation or prevention of disease in livestock. In addition, as VSV isolates find applications in the medical field as vaccine vectors or as therapeutic agents for cancer, more detailed knowledge of their primary structure and function should support the development of better vectors for medical use [1, 7, 23, 26].

Acknowledgments The authors would like to acknowledge the assistance of J. Zamparo, T. Bunch and Z. Lu for valuable assistance in sequencing. This project was supported by USDA-ARS Project 1940-32000-058-00D.

References

1. Ahmed M, McKenzie MO, Puckett S, Hojnacki M, Poliquin L, Lyles DS (2003) Ability of the matrix protein of vesicular stomatitis virus to suppress beta interferon gene expression is genetically correlated with the inhibition of host RNA and protein synthesis. J Virol 77:4646–4657
2. Barr JN, Whelan SP, Wertz GW (1997) Role of the intergenic dinucleotide in vesicular stomatitis virus RNA transcription. J Virol 71:1794–1801
3. Bilsel PA, Nichol ST (1990) Polymerase errors accumulating during natural evolution of the glycoprotein gene of vesicular stomatitis virus Indiana serotype isolates. J Virol 64:4873–4883
4. Fauquet CM, Fargette D (2005) International Committee on Taxonomy of Viruses and the 3,142 unassigned species. Virol J 2:64
5. Hanson RP, Estupinan J, Castaneda J (1968) Vesicular stomatitis in the Americas. Bull Off Int Epizoot 70:37–47
6. Jayakar HR, Whitt MA (2002) Identification of two additional translation products from the matrix (M) gene that contribute to vesicular stomatitis virus cytopathology. J Virol 76:8011–8018
7. Kim GN, Kang CY (2007) Matrix protein of VSV New Jersey serotype containing methionine to arginine substitutions at positions 48 and 51 allows near-normal host cell gene expression. Virology 357:41–53
8. Letchworth GJ, Rodriguez LL, Del cbarrera J (1999) Vesicular stomatitis. Vet J 157:239–260
9. Luk D, Masters PS, Gill DS, Banerjee AK (1987) Intergenic sequences of the vesicular stomatitis virus genome (New Jersey serotype): evidence for two transcription initiation sites within the L gene. Virology 160:88–94
10. Martinez I, Rodriguez LL, Jimenez C, Pauszek SJ, Wertz GW (2003) Vesicular stomatitis virus glycoprotein is a determinant of pathogenesis in swine, a natural host. J Virol 77:8039–8047
11. Pattnaik AK, Ball LA, LeGrone A, Wertz GW (1995) The termini of VSV DI particle RNAs are sufficient to signal RNA encapsidation, replication, and budding to generate infectious particles. Virology 206:760–764
12. Pauszek SJ, Allende R, Rodriguez LL (2008) Characterization of the full-length genomic sequences of vesicular stomatitis Cocal and Alagoas viruses. Arch Virol 153:1353–1357
13. Perez AM, Pauszek SJ, Jimenez D, Kelley WN, Whedbee Z, Rodriguez LL (2010) Spatial and phylogenetic analysis of vesicular stomatitis virus over-wintering in the United States. Prev Vet Med 93:258–264
14. Rainwater-Lovett K, Pauszek SJ, Kelley WN, Rodriguez LL (2007) Molecular epidemiology of vesicular stomatitis New Jersey virus from the 2004–2005 US outbreak indicates a common origin with Mexican strains. J Gen Virol 88:2042–2051
15. Rodriguez LL, Fitch WM, Nichol ST (1996) Ecological factors rather than temporal factors dominate the evolution of vesicular stomatitis virus. Proc Natl Acad Sci USA 93:13030–13035
16. Rodriguez LL (2002) Emergence and re-emergence of vesicular stomatitis in the United States. Virus Res 85:211–219
17. Rodriguez LL, Pauszek SJ, Bunch TA, Schumann KR (2002) Full-length genome analysis of natural isolates of vesicular stomatitis virus (Indiana 1 serotype) from North, Central and South America. J Gen Virol 83:2475–2483
18. Rose JK (1980) Complete intergenic and flanking gene sequences from the genome of vesicular stomatitis virus. Cell 19:415–421

19. Schnell MJ, Buonocore L, Whitt MA, Rose JK (1996) The minimal conserved transcription stop-start signal promotes stable expression of a foreign gene in vesicular stomatitis virus. J Virol 70:2318–2323
20. Stillman EA, Whitt MA (1997) Mutational analyses of the intergenic dinucleotide and the transcriptional start sequence of vesicular stomatitis virus (VSV) define sequences required for efficient termination and initiation of VSV transcripts. J Virol 71:2127–2137
21. Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts
22. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882
23. von Kobbe C, van Deursen JM, Rodrigues JP, Sitterlin D, Bachi A, Wu X, Wilm M, Carmo-Fonseca M, Izaurralde E (2000) Vesicular stomatitis virus matrix protein inhibits host cell gene expression by targeting the nucleoporin Nup98. Mol Cell 6:1243–1252
24. Wertz GW, Whelan S, LeGrone A, Ball LA (1994) Extent of terminal complementarity modulates the balance between transcription and replication of vesicular stomatitis virus RNA. Proc Natl Acad Sci USA 91:8587–8591
25. Whelan SP, Wertz GW (1999) Regulation of RNA synthesis by the genomic termini of vesicular stomatitis virus: identification of distinct sequences essential for transcription but not replication. J Virol 73:297–306
26. Wollmann G, Rogulin V, Simon I, Rose JK, van den Pol AN (2010) Some attenuated variants of vesicular stomatitis virus show enhanced oncolytic activity against human glioblastoma cells relative to normal brain cells. J Virol 84:1563–1573