Virus Genes (2010) 40:293–297
DOI 10.1007/s11262-009-0435-1

5047

# Complete genome sequence of an Argentinean isolate of *Solenopsis invicta* virus 3

**Steven M. Valles · Clare Allen · Laura Varone · Juan Briano**

Received: 30 October 2009 / Accepted: 11 December 2009 / Published online: 24 December 2009
© Springer Science+Business Media, LLC 2009

**Abstract** *Solenopsis invicta* virus 3 (SINV-3) is a recently described positive-strand RNA virus that infects the red imported fire ant, *S. invicta*. The genome of an Argentinean isolate of *Solenopsis invicta* virus 3 (SINV-$3^{ArgSF}$) obtained from the Santa Fe region of Argentina was sequenced in entirety. Assembly of nine overlapping fragments yielded a consensus genome sequence 10,386 nucleotides long, excluding the poly(A) tail present on the 3′ end (Genbank accession number GU017972). With the exception of the poly(A) tail, the genome length of SINV-$3^{ArgSF}$ was identical to the North American isolate (SINV-$3^{USDM}$). The SINV-$3^{ArgSF}$ genome possessed three major open reading frames (ORFs) (comprised of $\geq$100 codons) in the sense orientation; SINV-$3^{USDM}$ possessed only two. ORFs 1 and 2 had identical start and stop genome positions for both isolates. Blastp analysis of the translated ORF 1 of SINV-$3^{ArgSF}$ recognized conserved domains for helicase, protease, and RNA-dependent RNA polymerase. These domains and their corresponding positions were identical to those reported for SINV-$3^{USDM}$. ORF 2a, unique to the SINV-$3^{ArgSF}$ genome, was also found in frame 2 and had a canonical start codon located at nucleotide position 8,351 and a stop codon ending at position 8,827. Blastp analysis of the translated amino acid sequence of ORF 2a revealed no significant similarity in the Genbank database. The two SINV-3 isolates exhibited 96.2% nucleotide sequence identity across the entire genome. The amino acid sequences of ORFs 1 and 2 exhibited higher identities (99.0 and 98.2%, respectively) than the corresponding nucleotide regions within the genome. These data indicated that the nucleotide differences between the SINV-3 isolates were largely synonymous. This observation was corroborated by codon substitution rate analysis. Thus, the majority of the SINV-3 codon changes were silent in the two polyproteins, indicating purifying selection pressure on the viral genome.

**Keywords** *Solenopsis invicta* · RNA virus · Genome sequence · SINV-3

S. M. Valles (✉) · C. Allen
Center for Medical, Agricultural and Veterinary Entomology, USDA-ARS, 1600 SW 23rd Drive, Gainesville, FL 32608, USA
e-mail: Steven.Valles@ars.usda.gov

L. Varone · J. Briano
South American Biological Control Laboratory, USDA-ARS, Bolivar 1559, Hurlingham, Buenos Aires Province, Argentina

## Introduction

*Solenopsis invicta* virus 3 (SINV-3) is a recently described positive-strand RNA virus that infects the red imported fire ant, *S. invicta* Buren [1]. SINV-3 is associated with significant mortality of adult and larval *S. invicta* [1] often resulting in colony collapse reminiscent of colony collapse disorder described in honey bees [2]. The SINV-3 genome has been sequenced in entirety (Genbank accession number: FJ528584) from a *S. invicta* colony in the United States (DM isolate) where this ant has been introduced [3]. However, SINV-3 was reported recently infecting *S. invicta* in Argentina [4], its native range [5]. The genome of the SINV-3 DM isolate is monopartite (10,386 nucleotides) encoding two large polyproteins [1]. The 5′-proximal open reading frame (ORF) contained domains with strong identity to an RNA-dependent RNA polymerase (RdRp), helicase, and 3C-like protease, whereas the 3′-proximal ORF ostensibly encodes for viral structural proteins. Phylogenetic analysis of the RdRp sequence placed SINV-3 within a unique clade (with Kelp Fly Virus) separated from

all described positive-strand RNA viruses, including *Solenopsis invicta* viruses 1 and 2 [1].

Viral genes can be quite useful in revealing population structure, phylogenetic relationships, and demographic histories of their hosts [6]. The objective of this study was to sequence the SINV-3 genome from SINV-3-infected *S. invicta* collected in northern Argentina. *S. invicta* was introduced into the United States in the 1930s [3] which provides a unique opportunity to examine changes in the SINV-3 genome since the bottleneck event. Furthermore, the sequence of a SINV-3 isolate from the native range, Argentina [5], provides additional baseline data for this virus.

## Materials and methods

Several entire colonies of *S. invicta* were collected by excavation from San Javier, Santa Fe Province, Argentina, on November 13, 2008. One colony, colony SF, was determined to be infected with SINV-3 by reverse-transcription polymerase chain reaction (RT–PCR) with SINV-3-specific oligonucleotide primers [4].

To acquire the SINV-3 genome sequence from the Argentinean colony of *S. invicta*, RNA was prepared from a pooled group of ten worker ants from the SF colony using Trizol reagent according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). cDNA was synthesized from 150 ng of total RNA using the appended GeneRacer Oligo dT oligonucleotide primer (Table 1, Invitrogen). The cDNA synthesis reaction was allowed to proceed for 1 h at 55°C and terminated by heating the sample to 70°C for 15 min. RNase H was added to the tube and incubated for 20 min at 37°C.

A series of nine PCR reactions were conducted with the cDNA as template yielding overlapping fragments that covered the entire genome. Oligonucleotide primer pairs and corresponding positions on the genome included: p791 and p767 (1–1,167), p764 and p756 (1,039–2,624), p769 and p752 (2,003–3,696), p770 and p748 (2,938–4,683), p771 and p738 (4,042–5,746), p772 and p734 (5,038–6,442), p858 and p855 (6,186–8,144), p775 and p787 (8,038–9,730), p789 and GeneRacer 3′ primer (8,743–polyA tail). Amplification was completed in a thermal cycler under the following temperature regime: 1 cycle at 94°C for 2 min, 35 cycles of 94°C for 15 s, 60°C for 15 s, 68°C for 2 min, followed by a final elongation step of 68°C for 5 min. Gel-purified amplicons were ligated into the pCR4-TOPO vector, transformed into TOP10 competent cells (Invitrogen), and sequenced by the Interdisciplinary Center for Biotechnology Research (University of Florida). At least three clones of each genome region were sequenced.

Assembly of the genome sequences was conducted with the Vector NTI suite (Invitrogen, version 9.1) software and ContigExpress program. Contig pairwise assembly parameters were set at a minimum 20 nucleotide overlap and 95% sequence identity.

**Table 1** Oligonucleotide primers used to acquire the SINV-3 (SF isolate) genome

| Oligonucleotide primer designation | Genome position | Orientation | Sequence |
|---|---|---|---|
| GeneRacer oligo dT | | Reverse | 5′GCTGTCAACGATACGCTACGTAACGGCATGACAGTG$(T)_{24}$ |
| 3′ primer | Poly(A) tail | Reverse | 5′GCTGTCAACGATACGCTACGTAACG |
| p791 | 1–37 | Forward | 5′TTTTAAAATAGGAAATTAAAGTCCAGTAAGGTTACTG |
| p764 | 1039–1075 | Forward | 5′TGATAATTTCCTTGGTATTCCTAATTATAAATTTGCT |
| p767 | 1135–1167 | Reverse | 5′ATTTCACTTTGTTTATCTTTAGGACAATCACGA |
| p769 | 2003–2036 | Forward | 5′GCAAATGCTACTAATCTTAAAATTGGATCAGAGG |
| p756 | 2589–2624 | Reverse | 5′CCAATTCAAAAGGATTATTGAAAGAAACTCTATGAA |
| p770 | 2938–2975 | Forward | 5′TGATGGTGATTATGTTTATATTTCTGAGCATAAAATTC |
| p752 | 3660–3696 | Reverse | 5′AATCCATTTACATCGGGTAAAGTAAGAACTTCCTGCT |
| p771 | 4042–4079 | Forward | 5′AGAAGAGCTTTATGATGCTGAAAATTGTAATACTGTTC |
| p748 | 4648–4683 | Reverse | 5′GCAGTTGTAAGTTTCCAACCATTAGTAACCGTCAAT |
| p772 | 5038–5076 | Forward | 5′AGTTGAAAATGAAACAGTTGAAATTTTAGGTATAACTCA |
| p738 | 5715–5746 | Reverse | 5′ACGAGCTGTGAACTCACCAAGAATCCAACGTT |
| p858 | 6186–6221 | Forward | 5′TAGCAGTTGATGATTCAATTGCTGACATCTTTAATT |
| p734 | 6411–6442 | Reverse | 5′ATGTTCAGGAGTATCATCAGTTGCATAACAAA |
| p775 | 8038–8072 | Forward | 5′CATAATATTACTGATGTTGTGGTTTCTTCAAAACC |
| p855 | 8111–8144 | Reverse | 5′TGAGTTATATCACCTCTCCAAACACTAGCAGAAT |
| p789 | 8743–8782 | Forward | 5′ATGGTTAGACAGGGTTCTTCAACAACCCAATCTACTGATA |
| p787 | 9689–9730 | Reverse | 5′CATGATTTTGTTGTTCAATAGGTTCATAAATATGTTCCTCAT |

The nonsynonymous and synonymous substitution rates of the genome from each SINV-3 isolate (DM from North America and SF from South America) were estimated using the Phylogenetic Analysis by Maximum Likelihood (PAML) program [7, 8]. Nucleotide sequences of the SINV-3 DM isolate from North America (Genbank accession number FJ528584) and the corresponding sequences for the SF isolate from Argentina (Genbank accession number GU017972) were downloaded into the Vector NTI database. Two regions of the SINV-3 genome were compared, the polyprotein encoded by ORF 1 (nucleotide sequence positions 92 to 7,834; non-structural proteins), and the polyprotein encoded by ORF 2 (nucleotide positions 8,308 to 10,263; structural proteins).

The North American and Argentinean SINV-3 sequences were aligned by ClustalW through the MEGA 4 program, version 4.0.2 [9]. Nonsynonymous ($d_N$) and synonymous ($d_S$) substitution rates were estimated using the methods of Nei and Gojobori [10] and Yang and Nielson [11] as integrated into the yn00 program of the PAML collection [7, 8]. To process the sequences through the yn00 program, the *stop* codon of each sequence was removed from the 3′ end (ORF 1—TGA, ORF 2—TAA). Thus, the final fragment sizes for ORFs 1 and 2 were 7,740 and 1,953, respectively.

The North American and Argentinean SINV-3 sequences were also examined for regions with homology or identity with internal ribosomal entry sites (IRES). Sequences were queried against a database of known, well-characterized IRES elements, http://iresite.org/IRESite_web.php [12, 13].

## Results and discussion

Assembly of the nine overlapping fragments amplified with cDNA reverse transcribed from RNA purified from SINV-3-infected fire ants collected in the Santa Fe region of Argentina (hereafter referred to as SINV-3$^{ArgSF}$) yielded a consensus genome sequence 10,386 nucleotides long, excluding the poly(A) tail present on the 3′ end (Genbank accession number GU017972). Excluding the poly(A) tail, the genome length of SINV-3$^{ArgSF}$ was identical to the North American isolate (hereafter referred to as SINV-3$^{USDM}$, [1]). As with SINV-3$^{USDM}$, the SINV-3$^{ArgSF}$ genome sequence was highly A/U rich (34.7% A, 35.9% U, 14.0% C, 15.4% G). The SINV-3$^{ArgSF}$ genome possessed three major ORFs (comprised of ≥100 codons) in the sense orientation; SINV-3$^{USDM}$ possessed only two ORFs. ORFs 1 and 2 had identical start and stop genome positions for both isolates; ORF 1 was in frame two of the genome sequence and ORF 2 was in frame one. Blastp analysis [14] of the translated ORF 1 of SINV-3$^{ArgSF}$ recognized conserved domains for helicase (genome region 396), protease (genome region 1,379), and RNA-dependent RNA polymerase (genome region 2,027). These domains and their corresponding positions were identical to those reported for SINV-3$^{USDM}$ [1]. ORF 2a, unique to the SINV-3$^{ArgSF}$ genome, was also found in frame two and had a canonical start codon (AUG) located at nucleotide position 8,351 and a stop codon ending at position 8,827. ORF 2a yielded a predicted peptide comprised of 158 amino acids with a molecular mass of 18.8 kDa. Blastp analysis of the translated amino acid sequence of ORF 2a revealed no significant similarity in the Genbank database. Untranslated regions (UTRs) were present at the 5′ end (nts 1–91), 3′ end (nts 10,264–10,386), and between ORFs 1 and 2 (nts 7,835–8,307) of the SINV-3$^{ArgSF}$ genome.

Two regions were identified in ORF 2 of both isolates that contained a high proportion of asparagine (N) and glutamine (Q) residues. An asparagine-rich region (41% N) was identified between amino acid positions 383 and 416 and a glutamine-rich region (32% Q) between amino acid positions 299 and 353. ORF 2 ostensibly encodes for the capsid proteins, but this conclusion has not been determined empirically. The functions of these Q/N-rich areas are not known. However, Q/N-rich protein domains are associated with the formation of self-propagating protein aggregations often associated with prion-based diseases [15] or improper protein folding [16]. Perhaps, these highly polar regions are involved in capsid aggregation and assembly or even participate in the disease process in the ants (e.g., target receptor binding). These are interesting questions that should be investigated.

The two SINV-3 isolates exhibited 96.2% nucleotide sequence identity across the entire genome (Table 2). The 5′ and 3′ UTRs of the genomes exhibited very high identities (100 and 99.2%, respectively). Conversely, the nucleotide identity of the intergenic UTR (between ORFs 1 and 2) was lower (92.6%). The nucleotide identity of ORF 1 (96.8%) was higher than ORF 2 (94.4%); this result was somewhat anticipated because ORF 1 encodes for the nonstructural proteins that are more conserved than the structural proteins of ORF 2 (94.4% identity) which is typical among the positive-strand RNA viruses [17].

The amino acid sequences of ORFs 1 and 2 exhibited higher identities (99.0 and 96.6%, respectively) than the corresponding nucleotide regions of the genome. These data indicate that the nucleotide differences between the SINV-3 isolates are largely synonymous. Indeed, the proportion of amino acid residues that were similar in ORFs 1 and 2 were even higher (99.6 and 98.2%, respectively). Strong sequence identity/similarity indicates that the SINV-3 polyproteins are experiencing purifying (or neutral) selection pressure. This observation is corroborated by codon substitution rate analyses [10, 11]. The nonsynonymous-to-synonymous ratios were less than 1 for ORF 1 (0.02 and

**Table 2** Sequence comparisons of different genomic and polyprotein regions of SINV-3 between the Argentinean (SF) and the United States (DM) isolates

| SINV-3 genomic region | Identity (%) | | Similarity (%) | $d_N/d_S^d$ by the method of | |
|---|---|---|---|---|---|
| | nt[a] | AA[b] | AA[c] | Nei and Gojobori [10] | Yang and Nielsen [11] |
| Entire genome | 96.2 | | | | |
| 5′ UTR | 100 | | | | |
| Intergenic UTR | 92.6 | | | | |
| 3′ UTR | 99.2 | | | | |
| ORF 1 | 96.8 | 99.0 | 99.6 | 0.0339 | 0.0156 |
| RdRp domain[e] | 97.9 | 99.7 | 100 | | |
| Protease domain[f] | 95.8 | 99.4 | 99.4 | | |
| Helicase domain[g] | 95.7 | 100 | 100 | | |
| ORF 2 | 94.4 | 96.6 | 98.2 | 0.0923 | 0.0215 |

[a] Nucleotide identity

[b] Amino acid identity

[c] Amino acid similarity

[d] Ratio of nonsynonymous to synonymous changes in the corresponding region

[e] Amino acid region 1,659–2,309

[f] Amino acid region 1,238–1,403

[g] Amino acid region 385–500

0.03) and ORF 2 (0.09 and 0.02) using two different analytical methods [10, 11], respectively. This ratio (also identified as $\omega$) may be interpreted as $\omega = 0$ (neutral), $\omega < 1$ (negative), and $\omega > 1$ (positive) type selection pressures. Therefore, these values suggest that the polyproteins are being adaptively preserved. However, it is worth noting that these statistical analyses provide an estimation of the $\omega$ value applied over the entire sequence (polyproteins 1 and 2); it does not provide detailed analysis of individual regions of the polyproteins that may indicate local directional selection pressure. Analyzed as single sequences, the majority of codon changes were silent in the two polyproteins. Although neither polyprotein (ORFs 1 and 2) was under directional selection, the ratio of nonsynonymous to synonymous changes was slightly greater in ORF 2 (1.4- to 2.7-fold depending on the model, Table 2). These results were again somewhat expected because ORF 2 ostensibly encodes the structural, or capsid proteins, and ORF 1 encodes the nonstructural proteins. The nonstructural proteins are quite well conserved among the positive-strand RNA viruses [17], whereas the structural proteins of viruses generally exhibit a higher mutation rate [18]; a conclusion that appears to be supported by the data. Indeed, this notion is further supported by high amino acid sequence identity values for the RdRp (99.7%), protease (99.4%), and helicase (100%) domains between the two isolates (Table 2).

SINV-3 which infects the red imported fire ant, *S. invicta*, is found in both South America (Argentina) and North America (United States) [4]. *S. invicta* was introduced into the United States sometime during the 1930s or 1940s [19]. Assuming SINV-3 was present in the founding colonies, SINV-3 has been separated geographically (Argentina/United States) and temporally (approximately 60 years). Despite this separation, directional selection has not apparently occurred in the genome of SINV-3$^{USDM}$. However, the possibility that SINV-3 was absent in the founding population and was introduced later during secondary host invasions must be considered.

Blast [14] analysis of the SINV-3$^{USDM}$ and SINV-3$^{ArgSF}$ genome sequences against an IRES sequence database yielded no significant expectation scores [12]. Further analysis of the intergenic regions of the SINV-3 genomes and the 5′ UTRs against sequences in the IRES database also yielded no significant similarity. Although virus IRES elements can be predicted based on primary sequence [13], SINV-3 is a unique virus that has not been taxonomically classified [1]. Indeed, little is currently known of the biology of SINV-3.

*S. invicta* has been discovered recently in Australia [20], mainland China [21], and Taiwan [22], and possesses the potential to invade additional areas globally [23]. To gain an understanding of the dispersal and to determine the source populations in newly introduced areas, arduous comparative genetic analyses of *S. invicta* are required (e.g., [5]). The high mutation rate of positive-strand RNA viruses [24] provides a unique opportunity to assist the construction of phylogenetic relationships of the host organism [6]. Thus, the genome sequences of the United States and Argentinean SINV-3 isolates may find utility as

phylogenetic tools to facilitate identification of source populations of *S. invicta* in newly infested areas.

**Acknowledgments** We thank Drs. Man-Yeon Choi and Rod Nagoshi (USDA-ARS, Gainesville, FL), for critical reviews of the manuscript. The use of trade, firm, or corporation names in this publication are for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the United States Department of Agriculture or the Agricultural Research Service of any product or service to the exclusion of others that may be suitable.

## References

1. S.M. Valles, Y. Hashimoto, Virology **388**, 354–361 (2009)
2. D.L. Cox-Foster, S. Conlan, E.C. Holmes, G. Palacios, J.D. Evans, N.A. Moran, P. Quan, T. Briese, M. Hornig, M. Geiser, V. Martinson, D. VanEngelsdorp, A.L. Kalkstein, A. Drysdale, J. Hui, J. Ahai, L. Cui, S.K. Hutchinson, J.F. Simons, M. Egholm, J. Pettis, W.I. Lipkin, Science **318**, 283–287 (2007)
3. D.F. Williams, H.L. Collins, D.H. Oi, Am. Entomol. **47**, 146–159 (2001)
4. S.M. Valles, L. Varone, L. Ramirez, J. Briano, J. Virol. Methods **162**, 276–279 (2009)
5. E.J. Caldera, K.G. Ross, C.J. DeHeer, D. Shoemaker, Biol. Invasions **10**, 1457–1479 (2008)
6. R. Biek, A.J. Drummond, M. Poss, Science **311**, 538–541 (2006)
7. Z. Yang, Comput. Appl. Biosci. **13**, 555–556 (1997)
8. Z. Yang, Mol. Biol. Evol. **24**, 1586–1591 (2007)
9. K. Tamura, J. Dudley, M. Nei, S. Kumar, Mol. Biol. Evol. **24**, 1596–1599 (2007)
10. M. Nei, T. Gojobori, Mol. Biol. Evol. **3**, 418–426 (1986)
11. Z. Yang, R. Nielsen, Mol. Biol. Evol. **17**, 32–43 (2000)
12. M. Mokrejs, T. Masek, V. Vopalensky, P. Hlubucek, P. Delbos, and M. Pospisek, Nucleic Acids Res. (in press)
13. M. Mokrejs, V. Vopalensky, O. Kolenaty, T. Masek, Z. Feketova, P. Sekyrova, B. Skaloudova, V. Kriz, M. Pospisek, Nucleic Acids Res. **34**, D125–D130 (2006)
14. S.F. Altschul, T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Nucleic Acids Res. **25**, 3389–3402 (1997)
15. M.D. Michelitsch, J.S. Weissman, Proc. Natl Acad. Sci. USA **97**, 11910–11915 (2000)
16. L. Cruzeiro, J. Phys.: Condens. Matter **17**, 7833–7844 (2005)
17. E.V. Koonin, V.V. Dolja, Crit. Rev. Biochem. Mol. Biol. **28**, 375–430 (1993)
18. L.A. Shackelton, C.R. Parrish, U. Truyen, E.C. Holmes, Proc. Natl Acad. Sci. USA **102**, 379–384 (2005)
19. A.M. Callcott, H.L. Collins, Fla. Entomol. **79**, 240–251 (1996)
20. K.I. McCubbin, J.M. Weiner, Med. J. Aust. **176**, 518–519 (2002)
21. R. Zhang, Y. Li, N. Liu, S.D. Porter, Fla. Entomol. **90**, 723–731 (2007)
22. T.C. Huang, Y.C. Chou, and H.C. Chou, C.J. Shih, and W.J. Wu (eds.), in *Proceedings of the Symposium on the Control of the Red Imported Fire Ant. Bureau of Animal and Plant Health Inspection and Quarantine*, Council of Agriculture, Taipei, 2004, pp. 1–13
23. L.W. Morrison, S.D. Porter, E. Daniels, M.D. Korzukhin, Biol. Invasions **6**, 183–191 (2004)
24. J.J. Holland, K. Spindler, F. Horodyski, E. Grabau, S. Nichol, S. VandePol, Science **215**, 1577–1585 (1982)