Journal of Invertebrate Pathology 105 (2010) 105–111

Contents lists available at ScienceDirect

Journal of Invertebrate Pathology

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

Exploitation of a high genomic mutation rate in *Solenopsis invicta* virus 1 to infer demographic information about its host, *Solenopsis invicta*

Clare Allen [a], Juan A. Briano [b], Laura Varone [b], David H. Oi [a], Steven M. Valles [a,*]

[a] USDA-ARS, Imported Fire Ant and Household Insects Research Unit, 1600 SW 23rd Drive, Gainesville, FL 32608, USA
[b] USDA-ARS, South American Biological Control Laboratory, Bolivar 1559, Hurlingham, Buenos Aires Province, Argentina

ARTICLE INFO

Article history:
Received 24 March 2010
Accepted 26 May 2010
Available online 1 June 2010

Keywords:
RNA-dependent RNA polymerase
Dicistroviridae
Phylogenetic analysis

ABSTRACT

The RNA-dependent RNA polymerase (RdRp) region of *Solenopsis invicta* virus 1 (SINV-1) was sequenced from 47 infected colonies of *S. invicta*, *S. richteri*, *S. geminata*, and *S. invicta*/*richteri* hybrids collected from across the USA, northern Argentina, and northern Taiwan in an attempt to infer demographic information about the recent *S. invicta* introduction into Taiwan by phylogenetic analysis. Nucleotide sequences were calculated to exhibit an overall identity of >90% between geographically-separated samples. A total of 171 nucleotide variable sites (representing 22.4% of the region amplified) were mapped across the SINV-1 RdRp alignment and no insertions or deletions were detected. Phylogenetic analysis at the nucleotide level revealed clustering of Argentinean sequences, distinct from the USA sequences. Moreover, the SINV-1 RdRp sequences derived from recently introduced populations of *S. invicta* from northern Taiwan resided within the multiple USA groupings implicating the USA as the source for the recent introduction of *S. invicta* into Taiwan. Examination of the amino acid alignment for the RdRp revealed sequence identity >98% with only nine amino acid changes observed. Seven of these changes occurred in less than 4.3% of samples, while 2 (at positions 1266 and 1285) were featured prominently. Changes at positions 1266 and 1285 accounted for 36.2% and 34.0% of the samples, respectively. Two distinct groups were observed based on the amino acid residue at position 1266, Threonine or Serine. In cases where this amino acid was a Threonine, 90% of these sequences possessed a corresponding Valine at position 1285; only 10% of the Threonine^{1266}-containing sequences possessed an Isoleucine at the 1285 position. Among the Serine^{1266} group, 76% possessed an Isoleucine at position 1285, while only 24% possessed a Valine. Thus, it appears that the Threonine^{1266}/Valine^{1285} and Serine^{1266}/Isoleucine^{1285} combinations are predominant phenotypes.

Published by Elsevier Inc.

1. Introduction

Solenopsis invicta virus 1 (SINV-1) was the first virus to be isolated and characterized in the host *S. invicta*, or, more commonly, the red imported fire ant (Valles et al., 2004). This RNA virus is known to infect *S. invicta* in Argentina and other *Solenopsis* species such as *S. richteri*, *S. geminata*, and the *S. invicta*/*richteri* hybrid throughout the southern region of the USA (Valles et al., 2007). The prevalence of this virus appears seasonal and temperature-dependent with levels highest between May and August (Valles et al., 2007). The virus is composed of a single, positive-strand RNA genome, with two large open reading frames (ORF) residing in the sense direction (Valles et al., 2004) that specifically place it in the *Dicistroviridae* family (Mayo, 2002). Currently, 15 viruses comprise this family and all infect arthropods (Bonning, 2009).

S. invicta was introduced into the USA in the 1930s through the likely location of Mobile, Alabama (Callcott and Collins, 1996). Since that time, the invasion has spread throughout the south-Atlantic states north to Virginia and west to California. Within the last decade, the red imported fire ant has expanded its range globally. Recent introductions have been reported in Australia (McCubbin and Weiner, 2002), Taiwan (Yang et al., in press) and China (Wong and Yuen, 2005).

The use of pathogen population genomics is a useful and indirect method capable of providing information about their hosts (Wirth et al., 2005). Indeed, such methods have been successfully employed to construct demographic histories of a host population (Biek et al., 2006). Similarly, human host history and demography have been studied by observing genomic changes in viruses, such as the JC Virus, and bacteria, like *Helicobacter pylori* (Kitchen et al., 2008; Wirth et al., 2005).

SINV-1 and its host, *S. invicta*, are ideally suited for examination of potential host-pathogen geographic linkages. *S. invicta* is an introduced species in the USA which has been separated geographically from the native range population (in Argentina (Caldera et al., 2008)) for nearly eight decades. SINV-1 is widely distributed,

* Corresponding author. Fax: +1 352 374 5818.
E-mail address: steven.valles@ars.usda.gov (S.M. Valles).

0022-2011/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jip.2010.05.018

prevalent, and confined to ants in the *Solenopsis* genus (Valles et al., 2007). Positive-strand RNA viruses, like SINV-1, are known to exhibit high mutation rates (Holland et al., 1982). Therefore, examination of SINV-1 genomic changes in different populations of *S. invicta* may afford the unique opportunity to infer demographic information about new or incipient infestations of *S. invicta*. We hypothesized that phylogenetic analysis of a conserved region of the SINV-1 genome from virus-infected fire ant colonies in Argentina and the USA would exhibit sufficient divergence to permit demographic origin assignment to colonies from a recent Taiwanese introduction. Furthermore, examination of the RdRp region of SINV-1 samples from the USA, Argentina, and Taiwan revealed a large number of variable nucleotide sites and distinct phenotypic groupings based on amino acid changes at two positions in the RdRp.

2. Materials and methods

2.1. Total RNA extraction and cDNA production from infected ants

Fire ant species were identified using taxonomic keys of Trager (1991) and Wojcik et al. (1976). Ant samples from Alabama were further evaluated by venom alkaloid and cuticular hydrocarbon analysis to discern *S. richteri*, *S. invicta*, and the *S. invicta*/*richteri* hybrids (Vander Meer et al., 1985; Ross et al., 1987; Vander Meer and Lofgren, 1990). The distribution of *S. richteri* in the USA is limited to areas of northern Alabama and western Tennessee. *S. invicta* and *S. richteri* are sympatric in these areas and require chemotaxonomic methods to definitively discern each species and the hybrids produced (Vander Meer and Lofgren, 1990). Ant collections were accomplished by plunging a scintillation vial (20 ml) into a nest for several minutes allowing fire ants to fall into the vial. Ant samples were taken to the laboratory and evaluated for SINV-1 infection by the multiplex RT-PCR method described previously (Valles et al., 2009). Total RNA was extracted from 10 to 30 worker ants by the Trizol method according to the manufacturer's instructions (Invitrogen, CA). Sample collection dates were as follows: *S. invicta*, Alabama (July 2005), California (April 2005), Florida (October 2003–August 2009), Georgia (June 2006), Louisiana (April–May 2005), Oklahoma (May–June 2005), South Carolina (June 2005), Texas (December 2004), Argentina (February 2005–December 2005) Taiwan (August 2006–November 2007); *S. richteri*, Alabama (July 2005); *S. invicta*/*richteri* hybrid, Alabama (August 2005); *S. geminata*, Florida (August 2005), Darwin, Australia (June 2009), Veracruz, Mexico (October 2005 and August 2009).

Two-step RT-PCR was employed to amplify the SINV-1 RdRp region. First, 0.5 μl (50–100 ng) of the total extracted RNA was mixed with 10 mM dNTPs and 1 μM reverse oligonucleotide primer A02 (5′CGCTCCTCATTGAAGATAAATCCTCTC) and heated to 65 °C for 5 min and then placed on ice or at −20 °C for at least 1 min. Afterwards, 5× first strand buffer and Superscript Reverse Transcriptase (RT, Invitrogen) were added and the reaction incubated at 55 °C for 1 h before inactivating the RT reaction at 70 °C for 15 min.

2.2. Amplification of the SINV-1 RdRp region and sequencing

To amplify the SINV-1 RdRp region, the corresponding cDNA was added to a Platinum *Taq* (Invitrogen) PCR reaction with the primer combination: A01 (5′AGAATTGAAAACGCAAAACAAGGAA TACGG) and A04 (5′GAAGATAAATCCTCTCTTGAGAAACGAAAC) under the following cycling conditions: 94 °C for 2 min, 35 cycles of: 94 °C for 15 s, 54 °C for 15 s and 68 °C for 1 min, 1 cycle of 68 °C for 5 min. Amplicons were separated on 1% Agarose gels and visualized with ethidium bromide staining. The amplified SINV-1 RdRp product (762 bp) was ligated into the pCR4-TOPO vector (Invitrogen) and transformed into TOP10 *Escherichia coli* competent cells (Invitrogen). Colonies were picked and inoculated in *Luria* broth with ampicillin (75 μg/ml), incubated at 37 °C plus shaking at 225 rpm, and positively determined for the presence of the insert by PCR with the original oligonucleotide primers.

Insert-positive plasmid DNA was purified using the QIAprep Miniprep Spin kit (Qiagen) and the final clones (three per sample) were sequenced by the University of Florida, Interdisciplinary Center for Biotechnology Research (Gainesville, FL). Consensus sequences for each clone were used in subsequent analyses.

2.3. Phylogenetic analysis

All nucleotide sequences were originally edited to remove the oligonucleotide primer sequences in Vector NTI, version 11 (Invitrogen). The edited nucleotide sequences were multiple aligned using ClustalW in the MEGA 4 suite (Tamura et al., 2007) and the aligned sequences were exported as a PHYLIP file. Using the Méthodes et Algorithms pour la Bio-informatique LIRMM (MABL) suite with 'Advanced Mode' under the following conditions to estimate the phylogenetic relationship (Castresana, 2000; Guindon and Gascuel, 2003; Edgar, 2004; Anisimova and Gascuel, 2006 and Dereeper et al., 2008): Alignment: Muscle in full model, Curation: G Blocks with 'do not allow many contiguous non-conserved positions', Phylogeny: PhyML using statistical test for branch support with the Approximate Likelihood Ratio Test (aLRT) SH-like and, separately, with bootstrapping using 100 replicates, Substitution model HKY85 and a transition/transversion ratio of 8.1. Amino acid sequences were translated directly from the edited nucleotide sequences and exported as a FASTA file before the phylogeny was calculated as previously stated with the substitution model WAG (protein). The phylogenetic tree was visualized in TreeDyn and the results saved as a Newick tree, before being placed into the treeview program (Page, 1996) where the phylogenetic tree was observed.

2.4. Secondary structure protein prediction program

Secondary structure of the RdRp region encompassing amino acids at positions 1266 and 1285 of ORF 1 was calculated using the Self Optimized Protein Method from Alignments (SOPMA) program (Geourjon and Deléage, 1995; Combert et al., 2000).

2.5. Geographical maps

Maps to illustrate the approximate locations of SINV-1-positive *S. invicta*, *S. richteri*, *S. geminata* and *S. invicta*/*richteri* hybrids were obtained from various sources: USA (Open Clip Art Library, no date), Taiwan (Central Intelligence Agency, no date) World (Open Clip Art Library, no date), Argentina (Alpern, 2007) and South America (Open Clip Art Library, no date).

3. Results and discussion

A total of 323 colonies of *Solenopsis* species collected from the USA ($n = 39$), Argentina ($n = 43$), Mexico ($n = 200$), Australia ($n = 15$), and Taiwan ($n = 26$) between the years 2003 and 2009 were examined for the presence of SINV-1. Among these colonies, 46 were infected with SINV-1 as determined by amplification by RT-PCR of the RdRp region of the genome. These sequences have been deposited into the Genbank database (Table 1). SINV-1-positive colonies were detected across the USA ($n = 28$) [to clarify, we also included in subsequent analyses the originally-described SINV-1 sequence collected from Florida providing a total of 29 samples from the USA (Valles et al., 2004)], Taiwan ($n = 6$), and

Table 1
Collection information and Genbank accession numbers of host sources for SINV-1 samples from USA, Argentina and Taiwan.

Sample designation	Collection location	Collection date	Host	Genbank accession number
UFL1Si	Gainesville, FL, USA	September-2008	*Solenopsis invicta*	GU932981
ULA2Si	Clinton, LA, USA	May-2005	*Solenopsis invicta*	GU932982
USC1Si	Orangebury County, SC, USA	June-2005	*Solenopsis invicta*	GU932983
USC2Si	Orangebury County, SC, USA	June-2005	*Solenopsis invicta*	GU932984
UOK1Si	Ardmore, Carter, OK, USA	May-2005	*Solenopsis invicta*	GU932985
UOK2Si	Ardmore, Carter, OK, USA	May-2005	*Solenopsis invicta*	GU932986
UGA1Si	Savannah, GA, USA	June-2006	*Solenopsis invicta*	GU932987
UTX1Si	Laredo, TX, USA	December-2004	*Solenopsis invicta*	GU932988
UTX2Si	Laredo, TX, USA	December-2004	*Solenopsis invicta*	GU932989
ACO1Si	Yahape, Corrientes, Argentina	May-2005	*Solenopsis invicta*	GU932990
ACO2Si	Route 12 km 998, Corrientes, Argentina	December-2005	*Solenopsis invicta*	GU932991
ACO3Si	Route 12 km 1000, Corrientes, Argentina	December-2005	*Solenopsis invicta*	GU932992
ACO4Si	Route 12 km 1034, Corrientes, Argentina	December-2005	*Solenopsis invicta*	GU932993
UCA1Si	Lake Elsinore, CA, USA	April-2005	*Solenopsis invicta*	GU932994
UCA2Si	Lake Elsinore, CA, USA	April-2005	*Solenopsis invicta*	GU932995
ACO5Si	Machuca Camping, Corrientes, Argentina	December-2005	*Solenopsis invicta*	GU932996
ACO6Si	Road to CENADAC, Corrientes, Argentina	December-2005	*Solenopsis invicta*	GU932997
UFL2Si	Gainesville, FL, USA	December-2008	*Solenopsis invicta*	GU932998
TTA1Si	Taoyuan, Taiwan	August-2006	*Solenopsis invicta*	GU932999
TTA2Si	Taoyuan, Taiwan	August-2006	*Solenopsis invicta*	GU933000
UAL1Si	Frayette, AL, USA	July-2005	*Solenopsis invicta*	GU933001
UVA1Si	Virginia Beach, VA, USA	July-2008	*Solenopsis invicta*	GU933002
UVA2Si	Chesapeake, VA, USA	October-2007	*Solenopsis invicta*	GU933003
UFL3Si	Gainesville, FL, USA	December-2008	*Solenopsis invicta*	GU933004
UFL4Si	Gainesville, FL, USA	April-2009	*Solenopsis invicta*	GU933005
UFL5Si	Gainesville, FL, USA	April-2009	*Solenopsis invicta*	GU933006
UAL1Sh	Auburn, AL, USA	August-2005	*S. invicta/richteri hybrid*	GU933007
UAL2Sh	Auburn, AL, USA	August-2005	*S. invicta/richteri hybrid*	GU933008
UAL3Sh	Auburn, AL, USA	August-2005	*S. invicta/richteri hybrid*	GU933009
UFL6Si	Gainesville, FL, USA	August-2009	*Solenopsis invicta*	GU933010
THS1Si	Guansi Town, Hsinchu County, Taiwan	November-2007	*Solenopsis invicta*	GU933011
THS2Si	Guansi Town, Hsinchu County, Taiwan	November-2007	*Solenopsis invicta*	GU933012
THS3Si	Guansi Town, Hsinchu County, Taiwan	November-2007	*Solenopsis invicta*	GU933013
THS4Si	Guansi Town, Hsinchu County, Taiwan	November-2007	*Solenopsis invicta*	GU933014
UMS1Sr	Tishomingo County, MS, USA	July-2005	*Solenopsis richteri*	GU933015
UMS2Sr	Tishomingo County, MS, USA	July-2005	*Solenopsis richteri*	GU933016
UAL1Sr	Lawrence County, AL,USA	July-2005	*Solenopsis richteri*	GU933017
UFL1Sg	Gainesville, FL, USA	August-2005	*Solenopsis geminata*	GU933018
UFL7Si	Gainesville, FL, USA	August-2009	*Solenopsis invicta*	GU933019
ASJ1Si	San Javier, Santa Fe, Argentina	February-2005	*Solenopsis invicta*	GU933020
ASJ2Si	San Javier, Santa Fe, Argentina	February-2005	*Solenopsis invicta*	GU933021
ASJ3Si	San Javier, Santa Fe, Argentina	February-2005	*Solenopsis invicta*	GU933022
ASJ4Si	San Javier, Santa Fe, Argentina	February-2005	*Solenopsis invicta*	GU933023
ASJ5Si	San Javier, Santa Fe, Argentina	February-2005	*Solenopsis invicta*	GU933024
ASJ6Si	San Javier, Santa Fe, Argentina	February-2005	*Solenopsis invicta*	GU933025
ULA1Si	Clinton, LA, USA	May-2005	*Solenopsis invicta*	GU933026
UFL8Si	Gainesville, FL, USA		*Solenopsis invicta*	NC_006559

from the region within Argentina (n = 12) from which the USA population of *S. invicta* has been reported to have originated (Caldera et al., 2008). Specifically, six of the Argentinean samples were from Corrientes and six samples were from San Javier which is approximately 450 km southwest of Corrientes. These results confirm the global distribution of SINV-1 in the *Solenopsis* genus of ants (Valles et al., 2007). Colonies sampled from southern Mexico and Darwin, Australia, all of which were *S. geminata*, were negative for the presence of SINV-1.

The amplified region of the RdRp encompassed 762 nucleotides (SINV-1 genomic coordinates 3157–3918) and spanned motifs I to VII, which are highly conserved among the positive-strand RNA viruses (Koonin, 1991) (Fig. 1). SINV-1 nucleotide sequence identities within this region ranged from 92% to 100% among the 47 samples examined. There were no consistent patterns or divisions between host species or geographical location when nucleotide sequence identity values were compared. Thus, nucleotide sequence identity alone was incapable of discerning geographical or species-based differences. For example, despite separation for 70–80 years, nucleotide identities from Argentinean SINV-1 samples exhibited 92–97% identity with samples collected from the USA.

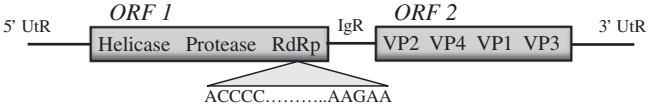

Fig. 1. Schematic representation of the SINV-1 genome and its component proteins encoded by each open reading frame (VP = viral protein). The conserved RNA-dependent RNA polymerase (RdRp) region of ORF 1 (nucleotide position 3157–3918) was amplified and sequenced for comparative and phylogenetic analyses. UtR, untranslated region; IgR, intergenic region.

A total of 171 nucleotide variable sites were mapped across the SINV-1 RdRp alignment and no insertions or deletions were detected. Variable sites (n = 171) within the RdRp sequence amplified represented 22.4% of the nucleotide positions. A high number of variable sites were similarly reported for a single-stranded RNA virus infecting the honeybee (Sacbrood virus) sampled from across Europe and Africa (Grabensteiner et al., 2001). Grabensteiner et al. (2001) examined five regions covering 38% of the Sacbrood virus genome; the number of variable sites ranged from 15.3% to 30.5% depending on the location within the genome. The vast majority of the changes observed in the SINV-1 RdRp were transitional

(T ↔ C, A ↔ G) accounting for 162 (94.7%) of all changes with 15 (9.26%), 4 (2.47%), and 143 (88.27%) mapped to the first, second and third codon positions, respectively. In contrast, only 20 transversions (T ↔ A, T ↔ G, C ↔ A, C ↔ G) were observed over the same region; nine (45%) were at unique sites and 11 (55%) were at sites in which transitions also occurred (i.e., shared sites). The majority (80%, $n = 16$) of the transversional changes were mapped to the third codon position. Thus, the observed changes across the SINV-1 RdRp alignment gave a calculated transition/transversion (*ti*/*tv*) ratio of 8.1:1. Bias toward nucleotide transitions over transversions has long been known in eukaryotic nuclear and mitochondrial genes (Fitch, 1967). However, among RNA viruses, reports of *ti*/*tv* ratios are limited. Indeed, no *ti*/*tv* values have been reported for the *Dicistroviridae*.

Despite the large number of nucleotide variable sites in the SINV-1 RdRp region, only nine amino acid changes resulted in the translated RdRp portion of the polyprotein (Table 2). Amino acid identities among the sequences of the translated RdRp from all 47 samples of SINV-1 ranged from 98% to 100%. Thus, the vast majority of these nucleotide changes were silent (or synonymous). Furthermore, of the nine amino acid changes observed, seven were similar. As reported previously with the structural polyprotein (Valles and Strong, 2005), the majority of the nucleotide changes in the RdRp region were synonymous and did not alter the amino acid sequence. The amino acid changes that were observed resulted from nucleotide changes at two codon positions, four at position 1 and five at position 2 (Table 2). Transitional ($n = 6$, amino acid positions 1105, 1114, 1218, 1269, 1285, and 1291) and transversional ($n = 3$, amino acid positions 1219, 1261, and 1266) mutations were observed. Seven (78%) of the amino acid changes were found in a small proportion of the samples ($\leqslant$4.3%) while two changes (amino acid positions 1266 and 1285) were found in a large proportion ($\geqslant$34%) of the samples.

As the nucleotide sequences exhibited the most diversity, the phylogenetic relationship was conducted with this data set to examine possible host-specific or geographic-based distinctions among the samples. Fig. 2 illustrates the phylogenetic relationship among the 47 SINV-1 RdRp sequences taken from the USA, Argentina, and Taiwan. A clear distinction between Argentinean and USA/Taiwan sequences was observed (Fig. 2). Phylogenetic trees obtained by Neighbor-joining, maximum likelihood, and maximum parsimony methods yielded similar topologies and relationships with varying bootstrap support values at the nodes (data not shown). Branch support with the Approximate Likelihood Ratio Test was 93% at the node separating the Argentinean and North American clades. Bootstrap values for this node position ranged from 50% to 75% depending on the analysis which is arguably an acceptable level of support (Soltis and Soltis, 2003). Lower bootstrap values are often observed among species that have not diverged extensively (Soltis and Soltis, 2003). Indeed, our comparisons are among samples of the same virus separated in "shallow time" (approximately 70–80 years) (Biek et al., 2006) which likely contribute to lower bootstrap support. However, two distinct and highly differentiated clades were clearly observed. These major clades were labeled North American and Argentinean. The North American clade was comprised of SINV-1 samples taken from ant species in the *Solenopsis* genus from across the USA and Taiwan.

Further subdivisions among the North American sequences did not show a strong assortment by host (*S. geminata*, *S. invicta*, *S. richteri*, and *S. invicta*/*richteri* hybrid) or geographic location. However, there were 4 North American clades. Grabensteiner et al. (2001) reported phylogenetic separation of Sacbrood virus genotypes based on host preference rather than geography from samples examined in Europe and Africa. The formation of distinct Argentinean and North American clades supports our hypothesis that SINV-1 sequence comparisons by phylogenetic analysis can be used to infer source introduction information of the host, *S. invicta*. Specifically, the Taiwanese samples (collected at different times and locations) assort strongly within the North American clade (Fig. 2). Thus, these data suggest that the recent introduction of *S. invicta* into Taiwan (Yang et al., in press) likely originated from the United States and not Argentina. Indeed, this conclusion has been reported previously based on phylogenetic analysis of microsatellites of *S. invicta* genes (Yang, 2006). While SINV-1 RdRp sequence presents a geographic-based separation, no distinctions among sequences were observed in the different *Solenopsis* species or hybrids. These data suggest that SINV-1 may readily move from species to species within the *Solenopsis* genus. Interestingly, although SINV-1 infects *S. geminata* in Florida, it was not detected among *S. geminata* collected from Mexico, nor was SINV-1 found in *S. geminata* from Hawaii (Valles et al., 2007) or Darwin, Australia. To date, SINV-1 has not been found on the Australian continent in *S. invicta* (Yang et al., in press). Perhaps *S. geminata* is not the natural host for SINV-1 and became infected from interactions in introduced regions with *S. invicta* or *S. richteri*.

In addition to inferring demographic information about the ant host from SINV-1 RdRp nucleotide sequences, multiple alignments of the amino acid sequences for the RdRp revealed distinct phenotypes (Table 3, Fig. 3). A high proportion of the samples exhibited amino acid changes at positions 1266 and 1285 (all amino acid position references correspond to the polyprotein of ORF 1) accounting for 36.2% and 34.0% of the samples, respectively. On closer comparison of these two sites, it became clear that two residue combinations and two potential intermediate states were present (Table 3). Two groups were observed based on the amino acid

Table 2
Amino acid changes observed in the RNA-dependent RNA polymerase coding region of the nonstructural polyprotein (ORF 1) among 47 SINV-1 isolates collected from the United States, Argentina, and Taiwan.

Amino acid		Proportion of samples exhibiting change (%)	Type of mutation[b]	Codon position[b]	Type of change[c]
Position[a]	Change				
1105	T → I	1/47 (2.1)	Transition	2nd	Dissimilar
1114	K → R	2/47 (4.3)	Transition	2nd	Similar
1218	K → R	2/47 (4.3)	Transition	2nd	Similar
1219	A → G	1/47 (2.1)	Transversion	2nd	Similar
1261	L → I	1/47 (2.1)	Transversion	1st	Similar
1266	T → S	17/47 (36.2)	Transversion	1st	Similar
1269	K → R	1/47 (2.1)	Transition	2nd	Similar
1285	V → I	16/47 (34.0)	Transition	1st	Similar
1291	D → N	1/47 (2.1)	Transition	1st	Dissimilar

[a] Amino acid position in the nonstructural polyprotein (ORF 1).
[b] The type of nucleic acid mutation required to result in the observed amino acid change and the corresponding codon position of the mutation.
[c] Refers to the type of amino acid change resulting from the mutation.

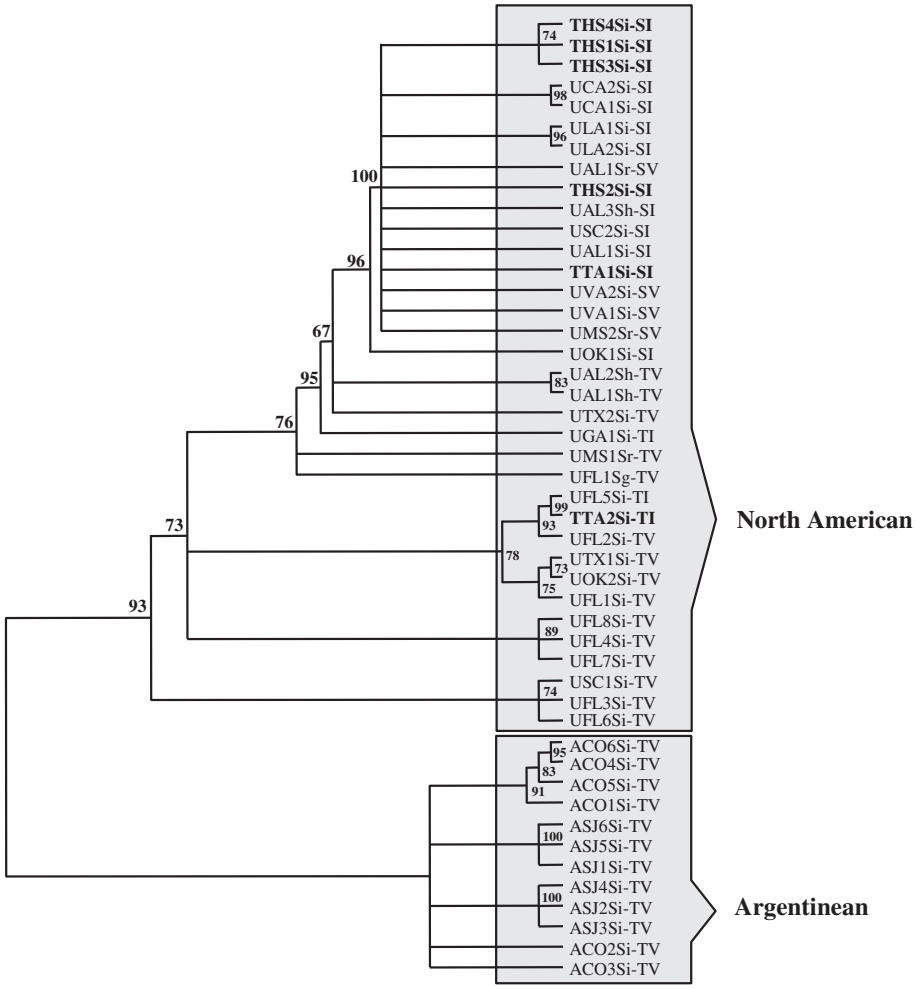

Fig. 2. Rectangular cladogram of proposed lineages of SINV-1 among nucleotide sequences of the RdRp acquired from SINV-1-infected *Solenopsis* ants from the USA, Argentina and Taiwan. The phylogenetic relationship was constructed from the RdRp nucleotide sequences using MABL and Treeview. The internal values represent the statistical test for branch support as calculated by aLRT expressed as a percentage. Branch supports below 50% have been collapsed for clarity. Samples are defined in detail in Table 1. Taiwanese samples are indicated by bold type. Briefly, the first letter indicates the country of origin (A = Argentina, U = USA, and T = Taiwan), followed by state or province abbreviation, followed by the sample number from that location, followed by the host *Solenopsis* species (Si, *S. invicta*; Sr, *S. richteri*; Sg, *S. geminata*; Sh, *S. invicta*/*richteri* hybrid), followed by the amino acid residues found at positions 1266 and 1285 for that respective sequence.

Table 3
Comparison among RNA-dependent RNA polymerase amino acid residues at positions 1266 (T or S) and 1285 (V or I) among 47 SINV-1 samples collected from the United States, Argentina, and Taiwan.

Amino acid residues at positions 1266/1285	Proportion of SINV-1 isolates exhibiting residue combinations (%)			
	Overall	US	Argentina	Taiwan
Serine/Isoleucine (S/I)	13/47 (27.7)	8/29 (27.6)	0/12 (0)	5/6(83.3)
Serine/Valine (S/V)	4/47 (8.5)	4/29 (13.8)	0/12 (0)	0/6 (0)
Threonine/Valine (T/V)	27/47 (57.4)	15/29 (51.7)	12/12 (100)	0/6 (0)
Threonine/Isoleucine (T/I)	3/47 (6.4)	2/29 (6.9)	0/12 (0)	1/6 (16.7)

residue at position 1266, Threonine or Serine (Fig. 3), labeled Threonine^{1266} and Serine^{1266}. In cases where the amino acid was a Threonine, 90% of these sequences possessed a corresponding Valine at position 1285; only 10% of the Threonine^{1266}-containing sequences possessed an Isoleucine at the 1285 position. Among the Serine^{1266} group, 76% possessed an Isoleucine at position 1285, while only 24% possessed a Valine. Thus, it appears that the Threonine^{1266}/Valine^{1285} and Serine^{1266}/Isoleucine^{1285} combinations are predominant phenotypes. Despite being similar changes (T → S and V → I), perhaps conformational constraints in the RdRp molecule make it necessary for corresponding changes at these positions. The amino acid region encompassing the different residue combinations is found primarily between motifs VI and VII (Koonin, 1991) or motifs C, D, and E in the highly conserved 'palm' region of the RdRp (O'Reilly and Kao, 1998). Using the secondary structure prediction model of the Self Optimized Protein Method from Alignments (SOPMA) program (Geourjon and Deléage, 1995; Combert et al., 2000) the consequence of such alterations was estimated (Fig. 4). Throughout all four residue combinations, a basic pattern of two extended strands, then α-helix followed by an extended strand and finally an α-helix was consistent. However, more revealing changes were apparent towards the C-terminus of the sequence with the omission of an extended strand (from T/V to S/V) to appearance of a β-turn (S/I and S/V) just after position 1285.

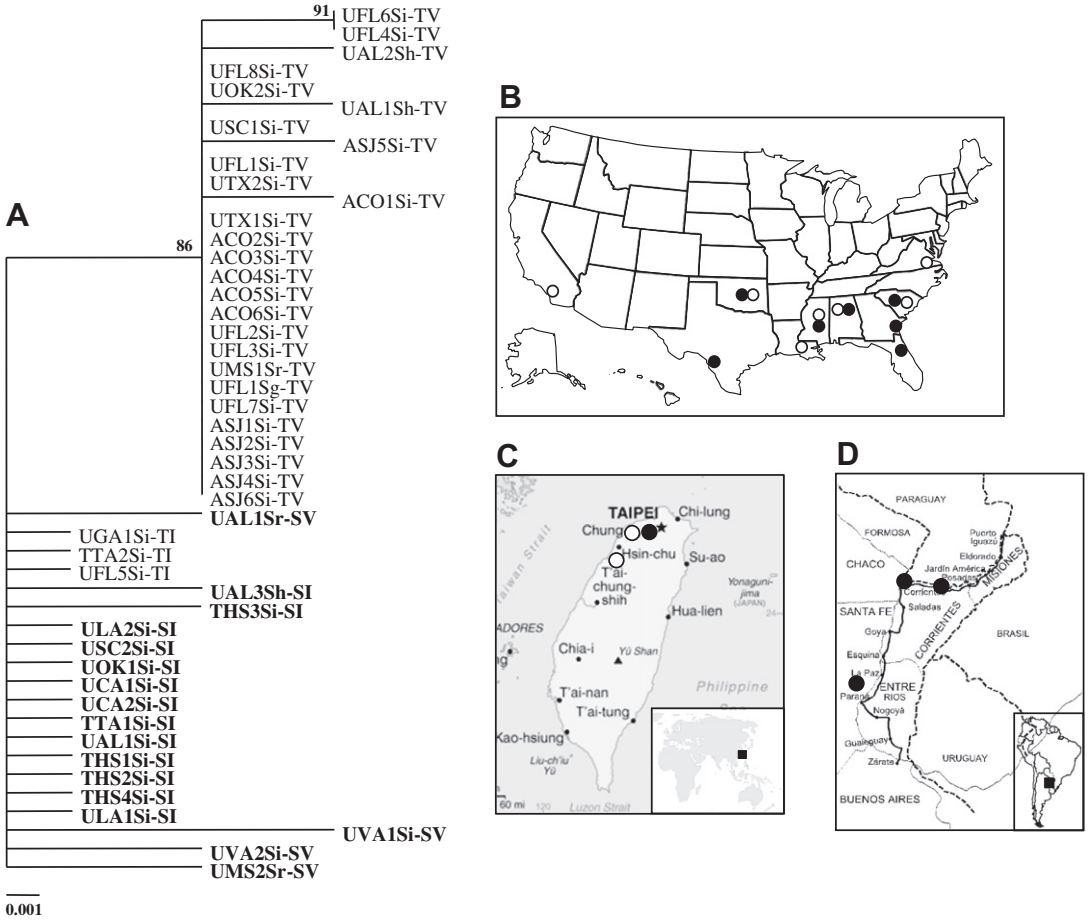

Fig. 3. Panel A represents the proposed lineages of SINV-1 based on amino acid sequences of the RdRp acquired from SINV-1-infected *Solenopsis* ants from the USA, Argentina and Taiwan. The phylogenetic relationship was constructed from the RdRp amino acid sequences using MABL and Treeview with the tree represented as a phylogram. The internal values represent the statistical test for branch support as calculated by aLRT, expressed as a percentage. The linear scale represents the number of amino acid changes per site. Samples are defined in detail in Table 1. Briefly, the first letter indicates the country of origin (A = Argentina, U = USA, and T = Taiwan), followed by state or province abbreviation, followed by the sample number from that location, followed by the *Solenopsis* species from which the SINV-1 infection was taken (Si, *S. invicta*; Sr, *S. richteri*; Sg, *S. geminata*; Sh, *S. invicta*/*richteri* hybrid), followed by the amino acid residues found at positions 1266 and 1285 for that respective sequence. (A) Panel A further illustrates the grouping of SINV-1 samples based on phenotypic variants at amino acid position 1266 (Threonine [regular font] or Serine [bold font]). Panels B, C, and D represent the approximate locations of samples collected from the USA, Taiwan, and Argentina, respectively, and their corresponding phenotype as based on the amino acid residue at position 1266 (Threonine = ●; Serine = ○).

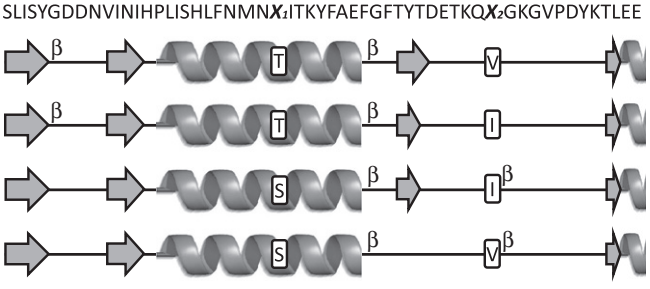

Fig. 4. Predictions of secondary protein structural topology as a consequence of amino acid changes at positions 1266 (X_1) and 1285 (X_2) within the SINV-1 RdRp region. Block arrows represent β-strands, coils represent α-helices, unaltered lines represent random coils, and β symbols represent β-turns. Rectangles within the structure indicate the amino acid residues and position for each phenotype observed.

How significant these changes are to the tertiary structure and the successful functioning of the RdRp remains unknown. However, it is quite interesting to observe that the two combinations (T/V and S/I) differ by a single methyl group. Threonine has one additional methyl group than Serine, while Isoleucine has one additional methyl group than Valine. The predominant combinations balance this difference. Specifically, Threonine^{1266} is predominantly paired with the 'methyl-short' Valine^{1285}, while Serine^{1266} is predominately paired with 'methyl-long' Isoleucine^{1285}.

Interestingly, all of the Argentinean samples possessed the residue combination of Threonine^{1266}/Valine^{1285}, compared with 51.7% of USA samples and none of the Taiwanese samples (Table 3). The Serine^{1266}/Isoleucine^{1285} combination was present in 27.6% of the USA and 83.3% of the Taiwan samples. The intermediate states (Serine^{1266}/Valine^{1285} and Threonine^{1266}/Isoleucine^{1285}) were only observed in a small proportion of the USA and Taiwan sequences. Both predominant phenotypes were found throughout the USA and in the samples collected from Taiwan (Fig. 3) and they did not appear to exhibit an association for a particular species.

Biek et al. (2006) reported the utility of using a viral infection to ascertain demographic information about a host in "shallow time." SINV-1 and its host, *S. invicta*, are ideally suited for utilization of this method. *S. invicta* was introduced into the USA in the 1930s, ostensibly from the Formosa region of Argentina (Caldera et al., 2008). SINV-1 has been discovered recently in both native (Argentina) and introduced (USA) populations of *S. invicta*. This geographic separation for nearly 80 years and the high mutation rate characteristic of positive-strand RNA viruses has apparently resulted in sufficient divergence among the SINV-1 populations to permit discrimination by phylogenetic analysis of the RdRp

nucleotide sequence (Fig. 2). Indeed, 2 distinct clades were observed – Argentinean and North American. Taiwan samples assorted within the North American clade suggesting an introduction of *S. invicta* into Taiwan from a source in the USA. These conclusions corroborate those of the Yang (2006) who concluded that the Taiwan *S. invicta* infestation likely had origins in the USA and not Argentina.

Acknowledgments

We thank Drs. M.Y. Choi and R. Nagoshi USDA-ARS, Gainesville, FL, for critical reviews of the manuscript and C. Strong for technical assistance. We also thank Chin-Cheng Yang for the use of two *S. invicta* samples from Taiwan, C.P. Russell for graphical assistance, and Dr. R.K. Vander Meer for conducting chemotaxonomic analyses of ants. 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

Alpern, D., 2007. Map of National Route 12. <http://en.wikipedia.org/wiki/National_Route_12_(Argentina)> (accessed January–February 2010).
Anisimova, M., Gascuel, O., 2006. Approximate likelihood ratio test for branches: a fast, accurate and powerful alternative. Syst. Biol. 55, 539–552.
Biek, R., Drummond, A.J., Poss, M., 2006. A virus reveals population structure and recent demographic history of its carnivore host. Science 311, 538–541.
Bonning, B.C., 2009. The *Dicistroviridae*: an emerging family of invertebrate viruses. Virol. Sinica 24, 415–427.
Caldera, E.J., Ross, K.G., Deheer, C.J., Shoemaker, D.D., 2008. Putative native source of the invasive fire ant *Solenopsis invicta* in the USA. Biol. Invas. 10, 1457–1479.
Callcott, A.-M. A., Collins, H.L., 1996. Invasion and range expansion of imported fire ants (Hymenoptera: Formicidae) in North America From 1918–1995. Fla. Entomol. 79, 240–251.
Castresana, J., 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540–552.
Central Intelligence Agency (no date). The World Fact book. <http://www.cia.gov/library/publications/the-world-factbook/geos/tw.html> (accessed January–February 2010).
Combert, C., Blanchet, C., Geourjon, C., Deléage, G., 2000. NPS@: network protein sequence analysis. TIBS 25, 147–150.
Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.-M., Gascuel, O., 2008. Phylogeny fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36 (Web Server Issue), W465–W469 (July 1, Epub 2008 April 19).
Edgar, R.C., 2004. Multiple sequences alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Fitch, W.M., 1967. Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutation. J. Mol. Biol. 26, 499–507.
Geourjon, G., Deléage, G., 1995. Sopma: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. Comput. Appl. Biosci. 11, 681–684. <http://pbil.ibcp.fr/htm/index.php?page=pbil_ibcp_Methods.html>.
Grabensteiner, E., Ritter, W., Carter, M.J., Davison, S., Pechhacker, H., Kolodziejek, J., Boecking, O., Derakhshifar, I., Moosbeckhofer, R., Licek, E., Nowotny, N., 2001. Sacbrood virus of the honeybee (*Apis mellifera*): rapid identification and phylogenetic analysis using reverse transcription-PCR. Clin. Diag. Lab. Immunol. 8, 93–104.
Guindon, S., Gascuel, O., 2003. A simple, fast and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696–704.
Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., VandePol, S., 1982. Rapid evolution of RNA genomes. Science 215, 1577–1585.
Kitchen, A., Miyamoto, M.M., Mulligan, C.J., 2008. Utility of DNA viruses for studying human host history: case study of JC virus. Mol. Phylogenet. Evol. 46, 673–682.
Koonin, E.V., 1991. The phylogeny of RNA-dependent RNA polymerase of positive-strand RNA viruses. J. Gen. Virol. 72, 2197–2206.
Mayo, M.A., 2002. A summary of taxonomic changes recently approved by ICTV. 2002. Arch. Virol. 147, 1655–1656.
McCubbin, K.I., Weiner, J.M., 2002. Fire ants in Australia: a new medical and ecological hazard. Med. J. Aust. 176, 518–519.
Open Clip Art Library (no date). <http://www.openclipart.org> (accessed January–February 2010).
O'Reilly, E.K., Kao, C.C., 1998. Analysis of RNS-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. Virology 252, 287–303.
Page, R.D.M., 1996. TREEVIEW: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
Ross, K.G., Vander Meer, R.K., Fletcher, D.J.C., Vargo, E.L., 1987. Biochemical phenotypic and genetic studies of two introduced fire ants and their hybrid (Hymenoptera: Formicidae). Evolution 41, 280–293.
Soltis, P.S., Soltis, D.E., 2003. Applying the bootstrap in phylogeny reconstruction. Stat. Sci. 18, 256–267.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599. <http://www.kumarlab.net/publications>.
Trager, J.C., 1991. A revision of the fire ants, *Solenopsis geminata* group (Hymenoptera: Formicidae: Myrmicinae). J. NY Entomol. Soc. 99, 141–198.
Valles, S.M., Strong, C.S., 2005. *Solenopsis invicta* virus-1A (SINV-1A): distinct species or genotype of SINV-1? J. Invertebr. Pathol. 88, 232–237.
Valles, S.M., Strong, C.A., Dang, P.D., Hunter, W.B., Pereira, R.M., Oi, D.H., Shapiro, A.M., Williams, D.F., 2004. A picorna-like virus from the red imported fire ant, *Solenopsis invicta*: initial discovery, genome sequence, and characterization. Virology 328, 151–157.
Valles, S.M., Strong, C.A., Oi, D.H., Porter, S.D., Pereira, R.M., Vander Meer, R.K., Hashimoto, Y., Hooper-Bùi, L.M., Sánchez-Arroyo, H., Davis, T., Karpakakunjaram, V., Vail, K.M., Graham, L.C. Fudd, Briano, J.A., Calcaterra, L.A., Gilbert, L.E., Ward, R., Ward, K., Oliver, J.B., Taniguchi, G., Thompson, D.C., 2007. Phenology, distribution, and host specificity of *Solenopsis invicta* virus-1. J. Invertebr. Pathol. 96, 18–27.
Valles, S.M., Varone, L., Ramírez, Briano, J., 2009. Multiplex detection of *Solenopsis invicta* viruses-1, -2, and -3. J. Virol. Methods 162, 276–279.
Vander Meer, R.K., Lofgren, C.S., 1990. Chemotaxonomy applied to fire ant systematics in the United States and South America. In: Vander Meer, R.K., Jaffe, K., Cedeno, A. (Eds.), Applied Myrmecology, a World Perspective. Westview Press, Boulder, pp. 75–84.
Vander Meer, R.K., Lofgren, C.S., Alvarez, F.M., 1985. Biochemical evidence for hybridization in fire ants. Fla. Entomol. 68, 501–506.
Wirth, T., Meyer, A., Achtman, M., 2005. Deciphering host migrations and origins by means of their microbes. Mol. Ecol. 14, 3289–3306.
Wojcik, D.P., Buren, W.F., Grissell, E.E., Carlysle, T., 1976. The fire ants (Solenopsis) of Florida (Hymenoptera: Formicidae). Florida Dept. Agricul. Cons. Serv. Entomology Circular No. 173.
Wong, S.S.Y., Yuen, K.Y., 2005. Red imported fire ants in Hong Kong. Hong Kong Med J. 11, 131–132.
Yang, C.-C. 2006. Population genetics and the origin of invasion of red imported fire ant *Solenopsis invicta* in Taiwan: evidence from microsatellite loci. Master's Thesis, National Taiwan University, Taiwan, 66 pp.
Yang, C.-C., Yu, Y.-C., Valles, S.M., Oi, D.H., Chen, Y.-C., Shoemaker, D., Wu, W.-J., Shih, C.-J., in press. Loss of microbial (pathogen) infections associated with recent invasions of the red imported fire ant *Solenopsis invicta*. Biol. Invas. doi:10.1007/s10530-010-9724-9.