Molecular epidemiology of vesicular stomatitis New Jersey virus from the 2004–2005 US outbreak indicates a common origin with Mexican strains

Kaitlin Rainwater-Lovett,1 Steven J. Pauszek,1 William N. Kelley2 and Luis L. Rodriguez1

1Plum Island Animal Disease Center, Agricultural Research Service, United States Department of Agriculture, PO Box 848, Greenport, NY 11944, USA
2Veterinary Services, Animal Plant Health Inspection Service, United States Department of Agriculture, 2150 Centre Avenue, Building B, Fort Collins, CO 80526, USA

Correspondence
Luis L. Rodriguez
luis.rodriguez@ars.usda.gov

Received 14 October 2006
Accepted 6 March 2007

Vesicular stomatitis (VS) outbreaks of unknown origin occur at 8–10-year intervals in the south-western USA with the most recent outbreak beginning in 2004. A previous study has suggested that strains causing US outbreaks are closely related to strains causing outbreaks in Mexico [Rodriguez (2002) Virus Res 85, 211–219]. This study determined the phylogenetic relationships among 116 vesicular stomatitis New Jersey virus (VSNJV) strains obtained from the 2004 outbreak and from endemic areas in Mexico. All 69 US viruses showed little sequence divergence (<1.3%), regardless of their location or time of collection, and clustered with 11 Mexican viruses into a genetic lineage not previously present in the USA. Furthermore, viruses with identical phosphoprotein hypervariable region sequences to those causing the US outbreaks in 1995–1997 and 2004–2005 were found circulating in Mexico between 2002 and 2004. Molecular adaptation analysis provided evidence for positive selection in the phosphoprotein and glycoprotein genes during a south-to-north migration among 69 US viruses collected between the spring and autumn of 2004 and 2005. Phylogenetic data, temporal–spatial distribution and the finding of viral strains identical to those causing major outbreaks in the USA circulating in Mexico demonstrated that VS outbreaks in the south-western USA are the result of the introduction of viral strains from endemic areas in Mexico.

INTRODUCTION

Vesicular stomatitis (VS) is an infectious disease of cattle, swine and horses occurring throughout North and South America caused by vesicular stomatitis virus (VSV). A member of the family Rhabdoviridae and the genus Vesiculovirus, VSV has an 11 kb RNA genome of negative polarity that encodes five structural proteins: the nucleocapsid (N), phosphoprotein (P), matrix (M), glycoprotein (G) and large polymerase (L) (Wagner & Rose, 1996). The P gene is used most extensively for phylogenetic analyses due to a 450 nt long hypervariable region that accumulates a high number of nucleotide substitutions that may reflect genetic changes elsewhere in the genome (Bilsel et al., 1990; Nichol et al., 1993).

Two serotypes of VSV, New Jersey (VSNJV) and Indiana (VSIV), cause epidemics approximately every decade in the south-western USA. The New Jersey serotype is responsible for the majority of US cases, and outbreaks caused by VSIV have been reported in the USA on only two occasions in the past 40 years, 1966 and 1997–1998. The clinical presentation of VS resembles that of foot-and-mouth disease with vesicular lesions appearing on the mouth, tongue, teats and hooves, making the rapid identification of VSV outbreaks increasingly important (Rodriguez & Nichol, 1999). Furthermore, when VS occurs, it causes significant economic and production losses of livestock due not only to veterinary costs, but also to trade and animal movement restrictions (Jenney et al., 1984; Thurmond et al., 1987).

The factors responsible for the cyclic epidemics of VS in the USA remain unclear, despite previous investigations into the genetic, environmental and host influences on the occurrence of the virus (Mccluskey et al., 2003; Mead et al., 2000; Rodriguez, 2002; Sellers & Maarouf, 1990). Currently, two hypotheses exist regarding the natural cycle of VSV in the south-western USA. The first proposes that VS has an endemic transmission cycle in reservoir species from which the virus periodically infects domesticated animals (Webb et al., 1987). The second hypothesis maintains that each VSV outbreak is an introduction of the
Molecular epidemiology of 2004–2005 US VSV epidemic

viruses into the USA from endemic areas elsewhere (Rodriguez et al., 2000). Phylogenetic analyses have shown that viruses causing each US outbreak group into distinct genetic lineages. These viruses are distantly related to strains causing previous US outbreaks, but are closely related to strains from endemic areas of Mexico (Rodriguez, 2002; Rodriguez et al., 2000).

In 2004, VSNJV re-emerged in the south-western USA for the first time since 1997, with subsequent episodes of occurrence in 2005 and to a lesser extent in 2006. In this study, we report the geographical and spatial characteristics as well as the phylogenetic relationships of VSV strains responsible for the 2004–2005 US outbreak and those circulating in endemic regions of Mexico. The implications of these relationships are discussed in terms of the evolution of the virus and the epidemiology of disease occurrence in the south-western USA.

METHODS

Viral strains and RNA extraction. A total of 116 VSNJV samples from the USA and Mexico were sequenced in this study. US equine and donkey samples (n=44) were obtained from the USDA National Veterinary Services Laboratory in Ames, IA, USA, and US bovine and alpaca samples (n=25) from the USDA’s Foreign Animal Diseases Diagnostic Laboratory at Plum Island Animal Disease Center in Greenport, NY, USA. The US strains were collected from eight of the nine states affected by the 2004–2005 VSNJV outbreak (samples from Idaho did not yield viable nucleic acid). Bovine strains (n=47) from Mexico collected between the years 2000 and 2004 were obtained from the Exotic Animal Disease Commission Laboratory in Palo Alto, Mexico. Samples were either epithelia taken from clinically infected animals and kept frozen in transport media or tissue culture supernatant from the first passage in Vero cells when epithelia was not available. Viral RNA was extracted using a Qiagen RNeasy kit. RNA pellets were resuspended in sterile water and kept at ~70 °C until tested.

RT-PCR. Reverse transcription was performed using random hexamers (Invitrogen) and SuperScript II RNase H reverse transcriptase (Invitrogen) following the manufacturer’s instructions. PCR was performed using Pfu DNA polymerase (Stratagene) according to the manufacturer’s instructions. Alternatively, RT-PCR was carried out using the one-tube GeneAmp EZ rTth RNA PCR kit (Perkin-Elmer) as described previously (Rodriguez et al., 2000). The previously described primers NJP-102F and NJP-831R were used to amplify the hypervariable region used for phylogeny (Rodriguez et al., 1993, 2000). Primers based on the VSNJV Ogden strain were used to sequence parts of the M and G genes (Rodriguez et al., 1993, 2000). Products were analysed on agarose gels and visualized by ethidium bromide staining.

Sequencing and phylogenetic analysis. Products were purified directly from the RT-PCR using QIAquick PCR Purification kits (Qiagen). PCR products were sequenced by dideoxy sequencing using a BigDye Terminator Sequencing kit on a 3730A automated sequencer (Applied Biosystems). SEQUENCER software v4.1 (GeneCodes) was used to analyse the chromatograms. Alignments were performed using CLUSTAL_X (Thompson et al., 1997). Additional hypervariable regions of previously determined VSNJV P sequences (n=47) from GenBank were also used in the phylogenetic analysis, which was performed by maximum likelihood using PAUP* version 10 on a Macintosh G5 (Swofford, 1998). We used the maximum-likelihood optimality criterion (HKY85 model) to reconstruct the most likely tree. Settings included a 2 : 1 transition/transversion ratio and a three-bisection-reconnection branch-swapping algorithm, and the initial starting tree was constructed by the neighbour-joining method. A 450 nt fragment of the hypervariable region within the P gene was used for sequence analysis in all samples and a 1437 nt fragment of the G gene was used for sequence analysis in selected samples. This section began at nt 117 within the G gene open reading frame. Combined P, M and G gene sequences of ten US viruses and one Mexican virus and maximum-likelihood, maximum-parsimony and neighbour-joining algorithms used for phylogenetic analyses provided similar tree topologies to those obtained using the P hypervariable region. Additionally, a pairwise nucleotide sequence divergence (NSD) matrix was produced from the P hypervariable region sequence data using PAUP* version 10 (Swofford, 1998).

Selection patterns. Evidence of selective pressure among the 2004–2005 US strains was examined by determining the synonymous/non-synonymous substitution rate ratio (ω = dS/dN) within the amino acid sites of the P and G genes, whereby ω > 1 indicates positive selection (Yang & Bielawski, 2000). Pairwise analysis was computed and the mean ω was determined using the SNAP package (http://hcv.lanl.gov/content/hcv-db/SNAP/SNAP.html), which is based on the Nei–Gojobori method (Korber, 2002; Nei & Gojobori, 1986). A Wilcoxon signed-rank test was performed to assess the significance of the differences between the non-synonymous and synonymous rate ratios (Pagano & Gauvreau, 2000).

Geographical analysis. The location of each premises was measured as a latitude/longitude coordinate pair (WGS84 datum). Coordinates were captured primarily with global positioning system (GPS) receivers or through geocoding the address of the premises. Geocoding was performed using the TeleAtlas EZ- Locate geocoding service (http://www.geocode.com). This service uses the address of the premises in conjunction with TeleAtlas Multinet street data to generate a coordinate for the location of the address on the street. The premises coordinates were captured at the point where the premises driveway intersects with the public road to ensure continuity between the GPS and geocoding methods. All coordinates were managed and maintained as an Environmental Systems Research Institute personal GeoDatabase. Available premises data for the 2004 and 2005 VSV outbreaks were combined and sorted based on the month and week (ignoring year) that the quarantine was started. As the exact infection date is unknown, the closest recorded date is when the premise quarantine began. From these combined data, we calculated the mean latitude for each week and plotted the results using Microsoft EXCEL.

Correlation between geographical and genetic distance. In order to determine whether geographical movement influenced genetic changes, the latitude and longitude coordinates of the locations of all viruses from 2004 from the US (n=48) and Mexico (n=17) were collected. Pairwise matrices measuring the geographical and genetic distances between each of these viruses within the US and Mexico were created using Google Earth (http://www.earth.google.com) and PAUP* version 10, respectively. The geographical and genetic distances were plotted using Microsoft EXCEL.

RESULTS

Geographical and temporal distribution of cases

The first case of VSNJV in 2004 was reported in south-west Texas during May. This was the first appearance of VSNJV
in the USA since November 1997 (Rodriguez et al., 2000). Although no virus was collected at this time, the diagnosis was based on clinical signs and confirmed by serology. The first viral strain sequenced in this study was from a case in New Mexico in June 2004. Subsequently, there was a rapid appearance of cases in New Mexico and Colorado, with the largest number of new cases reported during the third week of July. The last case of 2004 was reported in Colorado in December 2004. The virus reappeared in the spring of 2005 in Texas, New Mexico and Colorado and progressed northward along river-valley systems into Arizona, Utah, Wyoming, Idaho, Montana and Nebraska (Fig. 1). The weekly mean latitude was calculated from cases occurring in both years and indicated a northward migration that peaked in September 2004 and in December 2005, respectively (Fig. 2). Although the latitudinal trend was similar in both years, cases occurred further north throughout every week of 2005 (Fig. 2).

**Phylogenetic analysis**

We reconstructed the phylogeny of 116 VSNJ viruses obtained from the 2004–2005 outbreak in the USA and from cases in Mexico between 2000 and 2004. This phylogenetic analysis was based on the hypervariable region of the P gene. Additional VSNJV strains deposited in GenBank from earlier US outbreaks (n=24) and from endemic areas of Mexico (n=23) were also included. As has been described previously, viruses from the southeastern USA that had been collected between 1943 and 1989 grouped into a separate lineage with no recent common ancestry with strains from the south-western USA or Mexico, indicating that these viruses are maintained in independent, natural cycles (Fig. 3) (Rodriguez, 2002). Each US outbreak (1982–1983, 1984–1985, 1995–1997 and 2004–2005) was represented by a different genetic lineage (see shaded areas in Fig. 3) with little intra-lineage genetic variation.

![Map of the western USA indicating the premises from where individual strains were collected during the 2004–2005 VSNJV outbreak. Four main genotypes and several unique genotypes were identified throughout eight of the nine states affected by the outbreak (see Fig. 5 for specific genotype differences).](image-url)
variation (<1.5 % NSD), and contained identical or near-identical sequences of viruses from endemic areas in Mexico. Many other viruses collected in endemic areas of Mexico showed a high level of genetic variation (up to 6.9 % NSD) with up to 16 different genetic lineages co-existing in Mexico between 2002 and 2004 (Fig. 3).

A detailed phylogenetic analysis of the 2004–2005 US outbreak and viruses occurring in Mexico between 2000 and 2004 demonstrated that all 69 US viruses and 11 Mexico viruses clustered into a single viral lineage with very low NSD (0.0–1.3 %) (Fig. 4). A unique sequence represented by four Mexican viruses and 26 US viruses indicated a common origin. This is consistent with the introduction of this strain into the USA from endemic areas of Mexico (Fig. 4). Similar topologies were observed when phylogenetic trees were reconstructed from a combination of the P, M and G gene sequences (data not shown).

One Mexican virus from the state of Tabasco (03TB in Fig. 3) had 100 % sequence identity with the virus causing the 1995–1997 US outbreak (Fig. 3). This sequence showed a relatively high level of nucleotide sequence divergence (4.9 %) from the 2004–2005 US epidemic, demonstrating a second instance of a Mexico strain causing an outbreak in the USA (Fig. 3). Whilst the Tabasco virus was collected 6 years after the 1995 US outbreak, previous studies have shown foci of genetically distinct VSNJV strains in endemic areas that are maintained over many years (Nichol et al., 1993; Rodriguez et al., 1996).

### Microevolution

Whilst the NSD among all 2004–2005 US viruses was less than 1.3 %, four main genotypes were identified by phylogenetic analysis, each occurring in at least two US states (Fig. 5). Two non-synonymous substitutions in the hypervariable region of the P gene at nt 204 and 284 differentiated three genotypes termed G1, G2 and G3. One synonymous substitution at nt 303 differentiated G4 from G2. Genotype 1 (n=26 viruses) was first identified in southern Mexico in 2002 (state of Veracruz) and then detected in south-central Mexico in 2004 (Veracruz, Colima and Jalisco). This same viral strain appeared in New Mexico during June 2004 and was later identified in Colorado and Texas throughout November of the same year. Genotype 3 (n=4) appeared only in Texas and Colorado in 2004. Sixteen additional genotypes differing from G1 by less than three nucleotides were identified only once in either 2004 or 2005 (Fig. 1).

Genotype 2 (n=17) was identified in New Mexico and Colorado during 2004. This was the only genotype to overwinter and give rise to a sublineage, G4 (n=6), which emerged in the late summer of 2005 in Montana, Wyoming and Nebraska. Of the 16 viruses from 2005 that were sequenced, 15 (94 %) were included in G2 or G4, whereas only 12/42 viruses (29 %) from 2004 grouped within these genotypes (Fig. 4).

### Evolutionary rate and molecular adaptation

The mean number of nucleotide changes among all 2004–2005 US viruses and their most recent common ancestor remained at approximately $3.33 \times 10^{-3}$ substitutions per site per month throughout 2004. However, towards the end of 2005, the number of nucleotide substitutions more than doubled to $7.78 \times 10^{-3}$ substitutions per site per month. Most of this increase was associated with the appearance of genetic sublineage G4 (Fig. 4). In order to determine the evolutionary pattern during the epidemic, we sequenced the envelope G gene, which contains all previously described neutralizing epitopes (Wagner & Rose, 1996). Fifteen variable sites were identified among the 1437 nt sequenced in the G gene. Ten of these 15 (67 %) mutations were transitions (Table 1). Of the 450 nt region sequenced in the P gene, 20 variable sites were detected, 15 (75 %) of which were transitions. The transition : transversion ratios for the P and G genes were estimated to be 1.267 and 1.770, respectively. These results should be taken with caution, however, as the higher rate of transitional compared with transversional substitutions detected among the sequences could be a reflection of the low level of nucleotide diversity, as well as the relatively short length of the genes sequenced (450 and 1437 nt for P and G, respectively) (Wakeley, 1994). No changes appeared in any of the glycosylation sites in the G gene and only one substitution was observed 3 aa upstream from the previously described neutralizing epitope VII (Nichol et al.,...
No insertions or deletions (indels) were detected in either gene. An analysis for detecting molecular adaptation among strains from the USA provided statistically significant evidence (P<0.001) for positive selection in the P and G genes (ω=2.24 and 1.53, respectively) (Korber, 2002; Nei & Gojobori, 1986; Yang & Bielawski, 2000). Among the pairwise comparisons between the US viruses, ω ranged from 0.00 to 13.21 and from 0.91 to 3.98 for the P and G genes, respectively.

**Genetic and geographical distance**

Geographical distances were measured between the 2004 US viruses (n=48) and between the 2004 Mexican viruses (n=17) and compared in combination with a pairwise nucleotide distance matrix to determine whether the geographical distance between any two given viruses had an effect on their genetic relationship (Fig. 6). The geographical distance between the US viruses ranged from 0 to 1026 miles and the genetic divergence was no greater than 1.1%. The geographical distance between the Mexican viruses ranged from 0 to 1213 miles and the maximum genetic divergence was 4.9%. The US strains showed a relatively small amount of genetic variation in comparison with the Mexican strains, despite having a similar geographical range.

**DISCUSSION**

**Epidemiology**

Outbreaks of VSNJV occur in the south-western USA at approximately 10-year intervals and typically begin during the spring in states that border Mexico such as Arizona, New Mexico and Texas. Phylogenetic analyses of previous US outbreaks have suggested that the causative viruses shared recent common ancestry with certain viral strains from Mexico (Rodriguez, 2002; Rodriguez et al., 2000). Here, four of 47 (8.5%) Mexican viruses analysed showed identical sequences to 26 of the 48 (54%) US viruses from the 2004 outbreak that were sequenced. Viruses from this
genetic lineage were circulating in the state of Veracruz in southern Mexico as early as 2002 (02VCB), reappeared in Veracruz during 2004, moved north-west into the states of Colima and Jalisco and then emerged in the south-western USA during the same year (0604NME). After appearing in the USA, these strains travelled northward along river-valley systems during the summer and autumn of 2004, reappeared in the south-west during the spring of 2005 and moved northwards again over the next several months. The northward latitudinal progression of VSNJV in 2005 mimicked the 2004 movement except for an approximately 2° northward shift throughout that year. This south-to-north pattern of disease movement along with the stepwise phylogenetic relationship between the strains is consistent with the introduction of the virus from Mexico in 2004 followed by overwintering in the south-western USA and re-emergence in 2005. Overwintering of VSV in the south-western USA was reported previously in 1995 when the same VSNJ viral lineage reappeared in 1997 (Rodriguez et al., 2000).

The tree topology observed in the 2004–2005 US outbreak is consistent with that of previous VSV epidemics during which viruses from a single genetic lineage spread throughout the western USA (Fig. 3) (Rodriguez et al., 2000). The low level of genetic variation in the USA during 2004 (1.1% NSD) is in contrast to that of endemic areas such as Mexico where, during the same year, high levels of genetic diversity (up to 4.9%) were observed among strains collected at sites near to one another at approximately the same time.

Despite the relatively low level of genetic variation, our microevolution analysis showed four distinct genotypes (G1, G2, G3 and G4) that could be tracked as they moved from state to state. Genotype 1 was detected circulating in Mexico in 2002 and 2004, providing direct evidence that the genetic lineage causing the US outbreak in 2004 originated in endemic areas of Mexico. Genotype 2 overwintered and re-emerged in 2005, giving rise to a small sublineage (G4) containing eight viruses. These two genotypes dominated the 2005 US epidemic. Interestingly, G2 did not include any Mexican strains. Whilst it is possible that this virus may have been reintroduced during the spring of 2005 from Mexico, it is highly unlikely given the large genetic diversity of VSNJV circulating in Mexico during this time.
Viral evolution

The hypervariable region of the P gene was used to reconstruct the phylogeny and identify evolutionary patterns occurring during the 2004–2005 US outbreak. Pairwise analyses of the synonymous/non-synonymous substitution rate ratio in the P gene indicated a relatively high level of positive selection ($\omega = 13.21$) between strains from the beginning of the outbreak (0704TXE and 1104NME) and those from the latter stages of the outbreak (1005WYB), providing evidence for selective pressures on VSJV as it moved northward during the outbreak. Our data are consistent with a previous study by Nichol et al. (1993), which suggested positive selection in the P gene. Strains from that study had up to 25% nucleotide sequence divergence and were collected from various endemic and epidemic areas throughout the Americas spanning a 50-year time frame. Despite the fact that our dataset represented a much smaller time frame (2 years) and narrower geographical distribution (USA and Mexico), evidence of positive selection was still obtained.

A portion of the G gene of 11 viruses was sequenced in order to determine whether the nucleotide substitutions in the P gene were an accurate depiction of the evolution occurring in other parts of the genome. Consistent with previous investigations, no indels were detected and the level of divergence observed in the G gene was lower than that observed in the P gene ($\omega = 2.24$). This lower level of positive selection coupled with the fact that no amino acid substitutions were observed at or near neutralizing epitopes in G suggest that immunological pressures probably do not play a predominant role in the evolution of this virus during US epidemics.

Other arboviruses have adapted as they have spread throughout the USA from their point of introduction.
For example, phylogenetic analyses of West Nile virus (WNV; family Flaviviridae, genus Flavivirus) have revealed that a single genetic lineage closely related to strains from Israel was first introduced into the USA during 1999 (Lanciotti et al., 1999). As with our data, this single introduction of viruses from a distinct genetic lineage subsequently gave rise to dominant variants as the virus spread across the USA (Beasley et al., 2003; Davis et al., 2003, 2005; Deardorff et al., 2006). However, in contrast to our analysis, the authors did not perform analyses to detect selective pressures acting on this arbovirus, but presumed that these variants were the result of gradual sequence drifts introduced by the error-prone RNA polymerase (Davis et al., 2005; Steinhauer & Holland, 1986).

Unlike VSV, WNV has become established in the USA, whilst VSV has continually been unable to do so, despite multiple documented introductions over the past 50 years. This may be explained by the fact that WNV has adapted to native US mosquitoes and found amplifying reservoirs in local bird populations whereas VSV has not (McLean et al., 2001; Turell et al., 2001). Presumably, the vectors and reservoirs for long-term maintenance of VSV are lacking in the south-western USA. Therefore, future VSV studies should focus on endemic areas to determine the ecological factors associated with viral maintenance.

Whilst little is known about the determinants of VSV occurrence, patterns are emerging following examination of outbreaks in the western USA and the analysis presented here. This is the third well-documented cycle of VSV outbreaks in the south-western USA since 1982–1983 (Rodriguez, 2002; Rodriguez et al., 2000). In all of these outbreaks, it has been consistently observed that epidemics begin during the spring in states bordering Mexico and are caused by strains from a single genetic lineage, distinct from viruses causing previous US outbreaks but closely related to viruses from endemic areas of Mexico. We have now reported the finding of strains circulating in endemic areas of Mexico with identical sequences to those causing outbreaks in 1995–1997 and 2004–2005. These findings strongly suggest that each VSNJ outbreak in the south-western USA is caused by a novel introduction of VSV into the USA from endemic areas of Mexico (Nichol et al., 1993;
Rodriguez et al., 1996, 2000). Understanding the mechanisms mediating the emergence of VSV in the USA will allow the development of better quarantine policies and vector control procedures to reduce the impact of future epidemics.

ACKNOWLEDGEMENTS

We acknowledge the technical support provided by George Smoliga from USDA Agricultural Research Service in Plum Island, NY. We thank Sabrina Swenson and Leo Koster from the USDA National Veterinary Services Laboratory in Ames, IA, for providing equine strains; Samia Metwally and Bill Cerino from USDA Foreign Animal Disease Diagnostic Laboratory in Plum Island, NY, for providing bovine strains; Roberto Navarro and Pedro Paz in Mexico for providing bovine strains; and Lynn Creekmore from USDA Veterinary Services in Fort Collins, CO, for providing epidemiologic data. K. R.-L. was the recipient of a Plum Island Animal Disease Center Research Participation Program fellowship, administered by the Oak Ridge Institute for Science and Education (ORISE) through an interagency agreement between the US Department of Energy (DOE) and the US Department of Agriculture (USDA). ORISE is managed by the Oak Ridge Associated Universities (ORAU) under DOE contract number DE-AC05-06OR23100. All opinions expressed in this paper are those of the authors and do not necessarily reflect the policies and views of the USDA, DOE or ORAU/ORISE. This research was made possible by ARS-CRIS project 914-32000-040-00D.

REFERENCES


