RESEARCH LETTER—Virology

Molecular variability and genetic structure of white spot syndrome virus strains from northwest Mexico based on the analysis of genomes

Delia Patricia Parrilla-Taylor1, Norberto Vibanco-Pérez2, Maria de Jesús Durán-Avelar2, Bruno Gomez-Gil3, Raúl Llera-Herrera4 and Ricardo Vázquez-Juárez1,*


*Corresponding author: Vázquez-Juárez, R. Centro de Investigaciones Biológicas del Noroeste, S.C. Instituto Politécnico Nacional 195, Col. Playa Palo de Santa Rita Sur La Paz, Baja California Sur, México C.P. 23096. Tel: +52 6121238500; E-mail: rvazquez04@cibnor.mx

One sentence summary: This paper uses genome sequencing to show diversity of WSSV in Northwest Mexico, suggesting the virus has been introduced via multiple events and that specific genome regions can accommodate substantial deletions without compromising viability.

Editor: Rich Boden

ABSTRACT

White spot syndrome virus (WSSV) has a ~300 kb double-stranded DNA genome. It originated in China, spread rapidly through shrimp farms in Asia, and subsequently to America. This study determined complete genome sequences for nine historic WSSV strains isolated from Pacific white shrimp (Litopenaeus vannamei) captured in farm ponds in northwest Mexico (Sinaloa and Nayarit). Genomic DNA was captured by an amplification method using overlapping long-range PCR and sequencing by Ion Torrent-PGM. Complete genome sequences were assembled (length range 255–290 kb) and comparative genome analysis with WSSV strains revealed substantial deletions (3 and 10 kb in two regions) in seven strains, with two strains differing from the rest. Phylogenetic analysis identified that the WSSV strains from the northern area of the state of Sinaloa clustered with strains from China (LC1, LC10, DVI) and Korea (ACF2, ACF4), while those from the southern region of the state of Nayarit (AC1 and JP) differed from both of those and from strains found in Taiwan and Thailand. Our data offer insights into the diversity of the WSSV genome in one country and their divergent origin, suggest that it entered Mexico via multiple routes and that specific genome regions can accommodate substantial deletions without compromising viability.

Keywords: white spot syndrome virus; DNA capture; northwest Mexico; genome sequence; Whiteleg shrimp (Litopenaeus vannamei)
INTRODUCTION

Viral diseases are a serious threat to shrimp farming worldwide. Intensive aquaculture and the international mobility of organisms have altered natural equilibriums, allowing viruses to spread around the planet (Reno 1998). White spot syndrome virus (WSSV) is the most prevalent virus in penaeid shrimp cultures, and one that causes mass mortalities in closed systems within 3–10 days of infection. The first reported appearances of WSSV in penaeid shrimp occurred in China and Taiwan in 1992–1993 (Leu et al. 2009), but it quickly spread to Southeast Asia and America, with severe socioeconomic consequences for the industry (Lightner 1996; Marks et al. 2004; Leu et al. 2009).

WSSV contains a large, double-stranded DNA genome of about 300 kb that covers approximately 184 Open Reading Frames (ORFs; Van Hulten et al. 2001; Yang et al. 2001). It has one of the largest genomes among viruses that infect animals, and is the only member of the genus Whispovirus (Van Hulten et al. 2001). Several lineages of WSSV viral genomes described on the basis of whole genome sequences have shown that while isolates share 99% identity, their size differences are due mostly to repeated number variations, several small insertions, and deletions (Marks et al. 2004; Lightner 2011).

Initial attempts to genotype WSSV patterns –based on restriction fragment length polymorphism and compositional analysis of dinucleotides– suggested that aside from small genetic differences, geographically distant strains were similar and closely-related. Previous analyses of strain variability have identified genomic differences and shown differential virulence among distinct geographic WSSV isolates (Marks et al. 2005; Pradeep, Karunasagar and Karunasagar 2009; Li et al. 2017), thus generating a context in which the molecular basis for differences in virulence remains unclear (Marks et al. 2005; Verbruggen et al. 2016). Reports have mentioned that competitive fitness depends on the size of the genome (Zwart et al. 2010), and it has been suggested that the small size of the WSSV-TH genome (WSSV Thailand strain) could have a replication advantage in the infection success of WSSV (Marks et al. 2005).

Durán-Avelar et al. (2015) reported a significant diversity of haplotypes in northwest Mexico (Nayarit, Sinaloa, Sonora) after analysing ORFs 75, 94 and 125. They suggested a high rate of mutation of the WSSV genome in isolates from Sonora, Sinaloa and Nayarit that could be the result of frequent introductions from other countries, and the mobility of viral genotypes through the study area.

Whole genome sequencing is one of the most useful tools for identifying the range of genetic variations in viral populations caused by evolutionary challenges that include selective pressure. The process of capturing specific viral DNA (by PCR amplification) as a preliminary step to deep-sequencing using Next-Generation sequencing has improved enrichment efficiency and allowed the sequencing of complete viral genomes with more precise alignments (Depledge et al. 2011; Beerenwinkel et al. 2012; Horn 2012).

Despite the enormous economic importance of WSSV infection in shrimp-farming in northwest Mexico, there is insufficient information on possible differences among WSSV strains in the area. One isolate (WSSV-MX08) from the Mexican state of Sonora was recently sequenced (Rodríguez-Anaya et al. 2016), but no genome-wide comparisons have yet been reported. Therefore, our study analysed the whole genomes of WSSV isolates from Sinaloa and Nayarit states (Mexico). The information generated by identifying these isolates from different geographic origins in northwest Mexico is important for both shrimp production and disease control. Isolating genomic differences among isolates may also help identify regional variability and conserved sites (Marks et al. 2004) that could be related to variations in pathogenicity. To ascertain the relationship among several WSSV strains and their genomic diversity, we adopted a purpose-designed bioinformatics strategy to analyse the next-generation sequencing data of nine Mexican WSSV strains that had previously shown high haplotypic diversity (Durán-Avelar et al. 2015), through a robust and accurate, full-length sequencing based on long-range PCR (Kvisgaard et al. 2013; Uribe-Convers et al. 2014). This work enabled us to compare the whole genomes of WSSV strains from this high-production zone in order to establish a possible relation among differences in the ancestral Asiatic sequenced strains and determine genetic relatedness with those isolates for future research on pathogenicity.

MATERIALS AND METHODS

To obtain WSSV inocula, ~100 mg of gill tissue were sampled from Litopenaeus vannamei collected in ponds at farms in the Guasave (GVE05, ACF2, ACF4 strains), Angostura (LC1, LC10 strains) and El Dorado (DVI strain) regions of Sinaloa, and from northern (AC1), central (JP) and southern (LC) areas of Nayarit (Fig. 1). WSSV inocula were prepared as described previously (Escobedo-Bonilla et al. 2005) and stored at ~86°C.

White spot syndrome virus infection

Juvenile L. vannamei (5–10 g mean body weight) were obtained from a commercial farm in Sinaloa and acclimated in 200-L tanks (synthetic sea water at 33 ± 1 ppt salinity, 26 ± 1°C and continuous aeration) for five days. Absence of the virus was confirmed by WSSV nested PCR (Kimura et al. 1996). Groups of six WSSV-free shrimp were then injected intramuscularly with 50 μL of each inoculum. When lethargic/moribund shrimp were observed at 48 h post-injection, they were collected at 6-h intervals and DNA was extracted as described below.

DNA extraction and white spot syndrome virus quantification

DNA was extracted from pleopod tissue using a DNA Blood and Tissue Kit (Qiagen, Hilden, Germany). The WSSV DNA was quantified by real-time qPCR using the IQ Real WSSV quantitative system (GeneReach Biotechnology Corp) and samples containing the highest amounts of WSSV DNA were chosen. To obtain WSSV DNA, samples were incubated with DNase I (2 U/μL) to digest shrimp genomic DNA, and then with Proteinase K (600 μAU/mL solution) to digest DNase and the WSSV capsid envelope. Viral DNA was recovered using a DNeasy Blood and Tissue Kit (Qiagen), following the manufacturer’s protocol.

Amplification of viral sequences

The WSSV genomic DNA was amplified with overlapping primers designed using the WSSV-China strain sequence (WSSV-CN; GenBank: AF332093) because the ICTV Whispovirus study group committee chose the China isolate, WSSV-CN, as the type strain (Leu et al. 2009). The WSSV-CN genome (305,107 bp) was fragmented into 31 contiguous sections (30 of 10 kb, 1 of 5107 bp) using customized Perl scripts. PCR primer pairs were designed using primer3 online software (Supplementary, Table S1) to amplify each of the genome fragments by long-distance
PCR. Primer sequence specificity was confirmed through BLAST searches in GenBank.

Long-distance PCR mixtures (50 μl) consisted of 1 X PrimeSTAR GXL Buffer, 0.4 mM dNTPs, 0.20 pmol mL⁻¹ of each primer, 1.25 U mL⁻¹ PrimeSTAR GXL polymerase (TaKaRa Bio, Shiga, JP) and 4 μl template DNA. A 115V MultiGene OptiMax Thermal cycler (Labnet, Edison, NJ, USA) was used under the following thermal cycling conditions: 95 °C/1 min, 30 cycles at 98 °C/10 s, 60 °C/15 s, 72 °C/10 min, followed by 72 °C/10 min. PCR products were detected using 1% agarose gel electrophoresis and GelRed™ staining. We obtained individual, long-range PCR amplicons of approximately 20 kb to capture the complete WSSV-like genome for each strain, avoiding the host DNA. The amplicons were subsequently pooled in equimolar amounts for each of the nine WSSV strains and purified using the QIAEX® II Gel Extraction Kit (Qiagen).

Ion Torrent PGM library preparation and sequencing

The pooled WSSV amplicons from each of the nine strains were quantified using a Qubit fluorometer High-Sensitive Kit (Thermo Scientific, Carlsbad, CA, USA), and 100 ng were fragmented by sonication using a Bioruptor Sonication System, following the manufacturer's instructions, to obtain a DNA fragment size of 200–300 bp from each strain. To make the nine libraries using an Ion Plus Fragment Library Kit (Applied Biosystems, USA), the sonicated amplicon fragments were end-repaired, and the Ion Torrent barcode adapters X and P1 were incorporated using T4 DNA ligase, followed by purification with 2X AMPure-XP magnetic beads (Beckman Coulter, Pasadena, CA). The resulting libraries were size-selected using E-Gel at 2% (Invitrogen, USA) to recover 300–350 bp fragments, which were then amplified by PCR for 15 cycles and purified using AMPure-XP beads. The median fragment size and integrity of the final libraries were assessed using an Experion Automated Electrophoresis system (Bio-Rad Laboratories, Hercules, CA, USA), while library equalization, sample emulsion PCR, emulsion-breaking and the enrichment steps were performed using the Ion Xpress Template Kit (Applied Biosystems, USA). Enriched IonSpheres were loaded into an Ion 314 chip kit v2 (Applied Biosystems, USA). Base calls were generated using the Ion Torrent software suite v1.5, and individual sequence reads were filtered using the PGM software to remove low-quality and polyclonal sequences. The resulting 9 flowgram files (assembly) and FASTQ files (alignment) were used for subsequent analyses.

Sequence processing

Raw reads in fastq format were assessed with the FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Low quality reads (phred score Q < 20) were first filtered. Then those shorter than 50 nt were removed, and adapter sequences were trimmed using Cutadapt (Martin 2011) and Prinseq software (Schmieder and Edwards 2011). Reads were aligned into WSSV reference sequence WSSV-CN (AF332093) using the Bowtie 2 v2.0.2 program (Langmead and Salzberg 2012). Anvi’o
v2.3.1 software (Eren et al. 2015) was used to profile mapping results and visualise genomic bins. Analysis involved mapping the reads from each sequenced Mexican strain aligned into the WSSV reference sequence, WSSV-CN.

Draft genomes were created in the VirAmp galaxy-based system using the reference-guided AMOScmp assembler in MUMmer (Pop et al. 2004), which connects short contigs into longer ones by comparative mapping. The variation between the final assembled files and the WSSV-CN genome sequence was assessed by dot plot graph generated in JDotter (Brodie, Roper and Upton 2004). Global alignments were performed to detect insertions/deletions and sequence similarities across the different genomes. The WSSV draft genomes were annotated using Genome Annotation Transfer Utility (GATU) software (Tcherepanov, Ehlers and Upton 2006), individually based on the following sequences: WSSV-CN (China, AF332093, 534 CDS), WSSV-TH (Thailand, AF369029, 184 ORFs) and WSSV-MX08 (Mexico, KU216744, 169 ORFs) in order to make comparisons, since each one of the three sequenced WSSV genomes has its own gene or ORF nomenclature system. Average Nucleotide Identity (ANI) was calculated online using the ANI calculator (http://enveomics.ce.gatech.edu/ani).

Phylogenetic analysis

A phylogenetic tree of complete genomes was performed by the neighbour-joining method to infer the evolutionary relationships among 14 WSSV isolates. The isolates used were from China (WSSV-CN strain, access number AF332093), Thailand (WSSV-TH strain, access number AF369029), Taiwan (WSSV-TW strain, access number AF440570), Korea (WSSV-KR strain, access number JX515788), Mexico (WSSV-MX08 strain, access number KU216744) and nine sequenced WSSV strains from Sinaloa and Nayarit. Multiple sequence alignments of the whole genomes were created using Clustal (Thompson, Gibson and Higgins 2002). A rectangular phylogenetic tree with 1000 bootstrap replicates and percentage values was elaborated with PhyML in Smart mode Selection (Guindon and Gascuel 2003) with the FigTree V1.4.3 program.

Genome comparison

Multiple comparisons of the whole genome of the nine annotated WSSV genomes from strains of Sinaloa and Nayarit, and five public genomes, were performed with the BLAST Ring Image Generator (BRIG, Alikhan et al. 2011), a program that generates a circular map of genomic similarity using a central reference (in the present case, the WSSV-CN genome).

RESULTS

Genome assembly

The Ion PGM 314 sequencing of nine WSSV strains from Sinaloa and Nayarit generated 38.3 Mbp and 64.6 Mbp of total reads in two sequencing runs with a mean sequence length of 180 bp. The overall alignment rate for quality-processed reads was ~94.94%. The data associated with the strains analysed herein have been deposited to the NCBI Sequence Read Archive under BioProject accession number PRJNA413204. Individual sample accession numbers run from SAMN07757839 to SAMN07757847. The assembly statistics for the alignment-guided assemblies and connection of short contigs using the AMOScmp assembler are summarised in Table 1. Genome sizes ranged from 255 000–290 420 bp with 41% G + C content for each assembly. The complete nucleotide sequences annotated using GATU software were deposited in Genbank (Table 1).

Genome comparisons

The Anvi’o image revealed four distinct bins containing variability zones (single nucleotide variants detected) –two of them important– compared to the reference genome from China (WSSV-CN). These strains possessed two notable deletions: in seven of them, bin 2 (deletion zone) was 2.7 kb and corresponded to locations 130–133 kb of the reference genome; while bin 4 was a large deletion area of 10 kb that corresponded to locations 275–285 kb (with respect to the reference genome, WSSV-CN). The JP and GVE05 strains in bin 2 were 14 kb and 8.8 kb, respectively, and corresponded to location 130 kb of WSSV-CN. In bin 4, the JP and GVE strains showed the shortest deletions, approximately 4 kb and 5.7 kb, respectively; which corresponded to location 275 kb of WSSV-CN (Fig. 2).

The percentage of similarity among the nine WSSV genomes and the reference WSSV-CN was obtained using the Stretcher global alignment program. The DVI and AC2 strains had the highest percentage of similarity at 94.8% and 94.7%, respectively, while the LG strain showed the lowest similarity to WSSV-CN (84.1%) due to numerous deletions. The ANI pairwise index comparison between the reference genomes (WSSV-CN, WSSV-TH and WSSV-MX08) and each assembled genome strain showed a range of values from 99.67%–99.81%, with the genome GVE05 being the most outlying with respect to WSSV-TH and WSSV-MX08.

The phylogenetic analysis showed that the WSSV-CN genome was closely related to four strains from Sinaloa gathered in the Angostura and Guasave areas (LC1, LC10, DVI and ACF4). A strain from Sinaloa (ACF2) and a strain from the

### Table 1. Statistics from QUAST in the assembly, with AMOScmp, of white spot syndrome virus strains from Ion Torrent reads using WSSV-CN as a reference and Genbank accession number from annotated strains.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Input sequences</th>
<th>Contig number</th>
<th>Longest contig</th>
<th>Total length bp</th>
<th>Genbank accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>GVE05</td>
<td>42,728</td>
<td>92</td>
<td>33,650</td>
<td>272,607</td>
<td>MG432478</td>
</tr>
<tr>
<td>ACF2</td>
<td>40,695</td>
<td>28</td>
<td>50,581</td>
<td>289,870</td>
<td>MG432475</td>
</tr>
<tr>
<td>ACF4</td>
<td>53,558</td>
<td>47</td>
<td>47,166</td>
<td>277,297</td>
<td>MG432476</td>
</tr>
<tr>
<td>LC1</td>
<td>42,035</td>
<td>22</td>
<td>42,852</td>
<td>278,223</td>
<td>MG432481</td>
</tr>
<tr>
<td>LC10</td>
<td>55,400</td>
<td>22</td>
<td>42,757</td>
<td>270,274</td>
<td>MG432480</td>
</tr>
<tr>
<td>DVI</td>
<td>85,897</td>
<td>12</td>
<td>61,024</td>
<td>290,879</td>
<td>MG432477</td>
</tr>
<tr>
<td>AC1</td>
<td>41,097</td>
<td>23</td>
<td>32,988</td>
<td>263,174</td>
<td>MG432474</td>
</tr>
<tr>
<td>JP</td>
<td>22,770</td>
<td>18</td>
<td>59,167</td>
<td>283,858</td>
<td>MG432479</td>
</tr>
<tr>
<td>LG</td>
<td>37,525</td>
<td>33</td>
<td>54,448</td>
<td>257,675</td>
<td>MG432482</td>
</tr>
</tbody>
</table>
southern of Nayarit (LG) were clustered with WSSV-KR. The WSSV-TH strain, in contrast, was closely-related to the Sonora strain, WSSV-MX08, while WSSV-TW was seen to be closely related to strains from the northern and central areas of Nayarit (AC1 and JP, respectively), and one from Guasave (Sinaloa) isolated in 2005 (GVE05) (Fig. 3).

Comparative genomics

The BRIG map shows a zone of greater variability at the 20–40 kb position in the ACF4, LC1 and LC10 strains that corresponds to several hypothetical proteins and two RNA methyltransferases, according to the WSSV-MX08 annotation. The genome of the LG strain had the highest number of disrupted genes and lowest similarity when compared to the five public genomes. The highest variability was observed in the position from 60–100 kb, which corresponded to a flagellar protein, a ribonucleotide reductase, hypothetical proteins and unknown function proteins. The zone of deletions at position 130–133 kb of the sequenced strains corresponded to WSV237-WSV243 (according to the WSSV-CN annotation) and to ORF89 (according to the WSSV-TH annotation). The deletion zone corresponding to WSV237-WSV243 was longer in the JP and AC1 strains from Nayarit than in the others (Fig 4).

DISCUSSION

The genome lengths of the nine WSSV strains from different locations in Sinaloa and Nayarit ranged from 257,675–290,879 bp (LG and DVI strains, respectively), which highlighted the variability present in this virus. Previous reports have proposed that genome shrinkage in WSSV was an adaptive process that might have given the virus a replication advantage and enhanced its viral fitness and virulence (Marks et al. 2004; Zwart et al. 2010; Gao et al. 2014). Although the relationship between the genome size of WSSV and virulence has not been defined clearly, it has been suggested that isolates with high virulence may encode virulence-associated factors that may be absent or functionless in less virulent isolates (Li et al. 2017).

The variations in genome length were due to variable-length deletions at locations within several well-defined hypervariable genomic regions, including ORF23/24 (Lan, Lu and Xu 2002; Dieu et al. 2004; Marks et al. 2004; Tan et al. 2009). In seven of the nine Mexican WSSV strains (ACF2, ACF4, LC1, LC10, DVI, AC1, and LG), a ~10 kb deletion was evident in ORF23/24, as identified previously in strains from Vietnam, India and Korea (Dieu et al. 2004; Pradeep et al. 2008; Shekar, Pradeep and Karunasagar 2012; Chai et al. 2013). This similarity suggests a possible relationship between Asian and Mexican strains. Moreover, the deletions of 5.7 kb (GVE05 strain) and 4.6 kb (JP strain) seen in the same region (ORF23/24) are similar to those detected in a strain...
Figure 3. Inferred Phylogenetic tree based on full-length genome sequences of 14 WSSV strains. Bootstrap values on each branch indicate the percentage of trees in which that branch is present (1000 replicates). Strains isolated in northern of state of Sinaloa are underlined and Solid circles indicate strains of the state of Nayarit.

from Hainan, China (Tan et al. 2009). Interestingly, deletions of ∼8 (GVE05 strain) and ∼14 kb (JP strain) were also observed in the region, located at the position 130,000 (bp) with respect to the reference WSSV-CN genome sequence which refers to WSV234-WSV250, similar to reports from Li et al. (2017).

The nature of the identified deletions suggests genetic relatedness among Mexican and Asian WSSV strains, though large deletions like those found in strains GVE05 and JP have not been reported before. While viruses with large dsDNA genomes typically exhibit lower levels of genetic variation (Depledge et al. 2011), WSSV appears to possess genome regions that are redundant in terms of virus replication and virulence. However, until complete genome sequences are determined for more strains that differ in such factors as virulence, host preference and replication fitness, speculation on the roles that the more variable genome regions might play in these biological processes will continue (Ramos-Paredes et al. 2012; Chai et al. 2013; Li et al. 2017).

After assembling and analysing the genomes aligned to the reference genomes, we found that certain regions differ between sites and are highly variable. ORF118, ORF119 and ORF120 showed variations and a high number of SNPs in the WSSV genomes. Similarly, Van Hulten et al. (2001) found repeat regions in the WSSV-TH genome analysis, which were identified as homologous areas (hr). One of them (hr6) was located next to ORF120. The variability in ORF 118–120 found in our study could be used to classify new WSSV isolates, develop genetic maps and conduct genotyping and population analyses.

Inversion and new arrangements of ORF78–79 were found previously by Ramos-Paredes et al. (2012), suggesting new genotypes in WSSV from Sonora, Mexico. In contrast, in our analysis, ORF78–80 proved to be a conserved region (100% homology percentage) in most of the genomes once annotation to references was completed. These contradictory findings suggest a divergent origin between the strains analysed by Ramos-Paredes et al. (2012) and those analysed herein. Hence, additional research is necessary to determine a possible biological function for this region, as the characterisation of these genes has not been reported in the literature, and there is no homology with ORFs from other viruses.

Genomic changes that give rise to new variants through microevolution events have been observed in some DNA viruses, such as hepatitis B, human respiratory syncytial virus, African swine fever virus and Ostreid herpesvirus 1 (Stuyver et al. 2000; De Villiers et al. 2010; Agoti et al. 2015; Batista et al. 2015). In the case of WSSV, it may be that new arrangements and deleted regions reported previously, as well as the ones found in our work, could be part of an evolutionary process that responds to selection pressures derived from extensive and intensive production worldwide.

The full phylogenetic genomic analysis confirmed that the WSSV genomes of strains isolated in Mexico were interspersed with genomes from other countries, while also revealing a relatively rapid appearance of new variants. It has long been assumed that the mobility of organisms around the planet has been responsible for transferring certain pathogens from Asia to America (Lightner 2011). WSSV-CN and WSSV-TW have already been proposed as the possible origin of WSSV (Dieu et al. 2004; Pradeep et al. 2008), and our phylogenetic analysis showed that, in fact, six strains are closely-related to WSSV-CN and WSSV-KR, while the others are more closely-related to WSSV-TW. These results suggest that WSSV was introduced into Mexico from Asian countries through multiple events. Our findings further confirm that the WSSV-MX08 isolated from Sonora is closely related to WSSV-TH.

High genomic variation in the WSSV strains analysed could be due to the co-circulation of multiple genotypes, but it is important to note that a wide range of WSSV hosts—over 93 species of arthropods—has been reported (Sanchez-Paz 2010). These host species constitute a potential source of WSSV infection that...
could provide advantages for recombination events and thus increase the genomic variation of WSSV.

Genome sequencing allows for a more global understanding of viral pathogenesis. NGS technology is a highly-efficient way of sequencing whole viral genomes and performing accurate alignments of full-length viral genomes (Radford et al. 2012). Despite the small proportion of viral DNA compared to host DNA, careful capture of the viral nucleic acid of interest by amplification and purification produces quality results that can be sequenced by NGS (Beerenwinkel et al. 2012; Uribe-Convers et al. 2014). The use of NGS technologies makes it possible to assemble whole viral genomes and so make inferences on viral diversity in single deep-coverage NGS experiments that are far superior to those generated by other approaches, such as culture- and PCR-based methodologies (Depledge et al. 2011). The method employed in this study—based on capturing viral DNA by fragment amplification on long-PCRs—has the advantages of being sample-sparing and applicable to a wide range of viruses in which viral DNA exists in lower quantities. The WSSV genome captured by long-PCR amplification from approximately 20 kb generated in this study covered the complete WSSV genome in fragments with sequence overlap. It represents a robust optimization of
the information generated in 314 Ion Torrent Chip and barcoding, which makes it possible to perform such analyses in a time-efficient manner that further increases cost-effectiveness (Kvisgaard et al. 2013; Uribe-Convers et al. 2014). This study could also provide insights into the benefits of WSSV genome analysis in comparison to the current approach of genotyping WSSV based on sequencing variable regions, such as ORF75, ORF94 and ORF125. Our findings revealed differences among the several strains analysed that confirm the genomic diversity of the Mexican strains of WSSV and their provenance from at least two ancestors. The incorporation of nine new genomes with important differences will pave the way for new studies on short-term genome evolution and pathogenesis of WSSV.

Future research should consider to analyse the genome-wide transcription of the isolates identified during a replication cycle under controlled conditions, in order to compare the differential virulence observed in outbreaks to those expressed genomic elements (proteins, sense or antisense long non-coding RNAs) located in both conserved and variable regions of the WSSV, covering both strands of the genome.

SUPPLEMENTARY DATA

Supplementary data are available at FEMSLE online.

ACKNOWLEDGEMENTS

The lead author was financed by a graduate scholarship from Mexico’s Consejo Nacional de Ciencia y Tecnología (CONACYT) (Grant Number: 267040). The authors thank Zulía F. Nieves-López (UAN) and Juliisa Enciso-Ibarra (CIAD) for their technical support, Fíndaro Álvarez-Ruiz for donating the GVE05 strain, and Paul C. Kersey Johnson and Diana Fischer for editing services.

REFERENCES


Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMNet J 2011;17:10–12.


Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinf* 2002; Chapter 2:Unit 2.3.


