Journal of Invertebrate Pathology 122 (2014) 40–43

Contents lists available at ScienceDirect

Journal of Invertebrate Pathology

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

Short Communication

Genome sequence and organization analysis of *Heliothis virescens* ascovirus 3f isolated from a *Helicoverpa zea* larva

Yong-Lu Wei [a,b], Jue Hu [a,b], Shun-Ji Li [a,b], Zi-Shu Chen [a,b], Xiao-Wen Cheng [a,c,*], Guo-Hua Huang [a,b,*]

[a] Institute of Virology, Hunan Agricultural University, Changsha, Hunan 410128, China
[b] Hunan Provincial Key Laboratory for Biology and Control of Plant Diseases and Insect Pests, Hunan Agricultural University, Changsha, Hunan 410128, China
[c] Department of Microbiology, Miami University, 32 Pearson Hall, Oxford, OH 45056, USA

ARTICLE INFO

Article history:
Received 15 July 2014
Accepted 12 August 2014
Available online 20 August 2014

Keywords:
Genome organization
Repeat sequences
Gene transfer
Bro genes

ABSTRACT

The complete genome sequence of *Heliothis virescens* ascovirus 3f (HvAV-3f) was obtained. The HvAV-3f genome has a circular genome of 198,157 bp with a G + C content of 46.0%, and encodes 190 open reading frames (ORFs) longer than 69 amino acids. Two major homologous regions (*hrs*) and 29 'baculovirus repeat ORFs' (*bro*) were found in the genome. BLAST analyses revealed that three HvAV-3f genes were homologous to that of lepidopteran insects. Nine ORFs were unique to HvAV-3f, in which two ORFs showed significant levels of similarity to genes that have not been previously described for ascoviruses in the Genbank database.

1. Introduction

Ascovirus is the only genus recognized in the family *Ascoviridae* that can cause a chronic but ultimately fatal disease in the larvae of Noctuidae, Crambidae and Plutellidae (Hamm et al., 1986, 1998; Federici and Govindarajan, 1990; Cheng et al., 2000, 2005; Bigot et al., 2011). Based on virion morphology, DNA sequence information, host range and tissue tropism, five ascovirus species have been recognized by the International Committee on Taxonomy of Viruses, which are consisting of *Spodoptera frugiperda ascovirus 1a* (SfAV-1a), *Trichoplusia ni ascovirus 2a* (TnAV-2a), *Heliothis virescens ascovirus 3a* (HvAV-3a), and *Diadromus pulchellus ascovirus 4a* (DpAV-4a) and *Trichoplusia ni ascovirus 6a* (ex-TnAV-2c) (Bigot et al., 2011). Eighteen isolates or strains belonging to four ascovirus species have been reported in the world (Bigot et al., 2011; Huang et al., 2012a), in which five ascovirus isolates genomes that are SfAV-1a, TnAV-6a (*Trichoplusia ni ascovirus 6a*, previously named at its discovery TnAV-2c (Wang et al., 2006)), HvAV-3e, HvAV-3g, and DpAV-4a have been sequenced (Wang et al., 2006; Bideshi et al., 2006; Asgari et al., 2007; Bigot et al., 2009; Huang et al., 2012b). All these reported genomes have a double-stranded circular DNA ranging from 100 to 199 kbp in size (Xue and Cheng, 2011; Huang et al., 2012b). Based on the biological characteristics analysis, the species HvAV-3a is the most diverse and widely distributed from America to Asia and Australia (Hamm et al., 1998; Huang et al., 2012a), in which two isolates HvAV-3e and HvAV-3g genomic sequences were reported from Australia and Indonesia, respectively (Asgari et al., 2007; Huang et al., 2012b), but no HvAV-3a isolate from America is sequenced. To understand the diversity and, evolution and geographical differentiation of the HvAV-3a isolates, we report the complete genome of HvAV-3f which was a HvAV-3a isolate from USA, and make comparison for HvAV-3f with the other 5 previously published ascovirus genomes. Accordingly, the results should be helpful for providing insight into the origin and evolution of ascoviruses.

2. Materials and methods

HvAV-3f genomic DNA (Cheng et al., 2005) was sheared into fragment sizes of 400–600 bp by ultrasonication and sequenced by the Solexa genome analyzer at the Beijing Genome Institution, Shenzhen, China (BGI). A total of 6,244,397 clean pair-end (PE) reads were obtained. Genome assembly was performed using the Edena assembly software package (Hernandez et al., 2008). The

Abbreviations: *hrs*, major homologous regions; ORF, open reading frame; *bro*, baculovirus repeat ORFs; *iap*, inhibitor of apoptosis protein; PE, pair-end; BGI, the Beijing Genome Institution, Shenzhen, China; NCBI, the National Center for Biotechnology Information; ICTV, international committee on taxonomy viruses; SfAV, *Spodoptera frugiperda* ascovirus; TnAV, *Trichoplusia ni* ascovirus; HvAV, *Heliothis virescens* ascovirus; DpAV, *Diadromus pulchellus* ascovirus; kbp, kilo base pairs.

* Corresponding authors at: Institute of Virology, Hunan Agricultural University, Changsha, Hunan 410128, China.

E-mail addresses: Chengx@MiamiOH.edu (X.-W. Cheng), tineidaehgh@gmail.com (G.-H. Huang).

http://dx.doi.org/10.1016/j.jip.2014.08.003
0022-2011/

assembled contig representing the entire HvAV-3f genome sequence was confirmed by six restriction enzyme digestion profiles (*Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Xba*I and *Xho*I) that produced the predicted fragments (Cheng et al., 2005; Xue and Cheng, 2011; Huang et al., 2012a). ORFs with 50 aa and longer considered were identified through the GeneMarkS program (Besemer and Borodovsky, 2005), ORF finder (NCBI) and Geneious program (Drummond et al., 2010). ORFs longer than 210 bp with minimal overlap was considered to be putative genes. Homologies among ascovirus genomes were investigated by Basic Local Alignment Search Tool (BLAST) (blastp) in NCBI (Altschul et al., 1997), and repetitive regions were identified by the MIROPEATS program (Parsons, 1995). The sequences were deposited in NCBI's GenBank with the accession number KJ755191.

3. Results

3.1. General organization

The HvAV-3f genome was assembled into a circular contiguous sequence of 198,157 bp, which was very similar to the genome size (198 kbp) estimated by REN analysis (Cheng et al., 2005). The HvAV-3f genome size was slightly smaller than the HvAV-3g genome, but it is still currently the second largest ascovirus genome described (Huang et al., 2012b). A genomic map showing the organization of the ORFs longer than 210 nucleotides (nt) in the HvAV-3f genome is presented in Fig. 1. The genome had a G + C content of 46.0% and encoded 190 predicted ORFs, in which 107 are in the forward orientation and 83 in the reverse (Fig. 1). The coding region (166,217 bp) accounted for 83.88% of the total sequences. We annotated 75 of the ORFs (39.5%) with gene functionality predictions. Among the identified ORFs, 181 ORFs were related to genes reported for other ascoviruses (ORFs showing similarities with other ascoviruses), including SfAV-1a (61.9%), TnAV-6a (56.4%), HvAV-3e (93.4%), HvAV-3g (99.4%), DpAV-4a (22.7%). There were 32 ORFs that showed low identities with those from different parasitic protozoa (marked with asterisks in Supplementary Table 1). Twenty-five genes encoding for enzymes involved in gene transcription, DNA replication, and nucleotide metabolism were found in the HvAV-3f genome. In addition, two *hrs* were identified in the genome (Fig. 1, Supplementary Table 1). The summary of the project results are shown in Table 1.

3.2. Novel putative genes

In the 190 predicted ORFs, only 9 did not show any similarity to ascoviruses genes described previously. Among these 9 ORFs, ORF91 and ORF101 showed significant levels of similarity to genes in the Genbank database (March, 2014) that have not been described for ascoviruses. ORF91 showed about 48% amino acid (aa) sequence identity with significance (*E*-value of 3e−52) to ORF63 of *Agrotis segetum nucleopolyhedrovirus* (accession number NC_007921) (Jakubowska et al., 2006). On the other hand, the predicted protein of ORF101 exhibited 42% identity (6e−87) to ORF311 of *Choristoneura biennis entomopoxvirus* 'L' (accession number NC_021248) (Thézé et al., 2013). The functions of these two genes are unknown. In addition, ORF144 showed about 28% aa sequence identity to glycosyltransferase of *Zea mays*. Virus infection may affect the glycome either by regulating expression of host

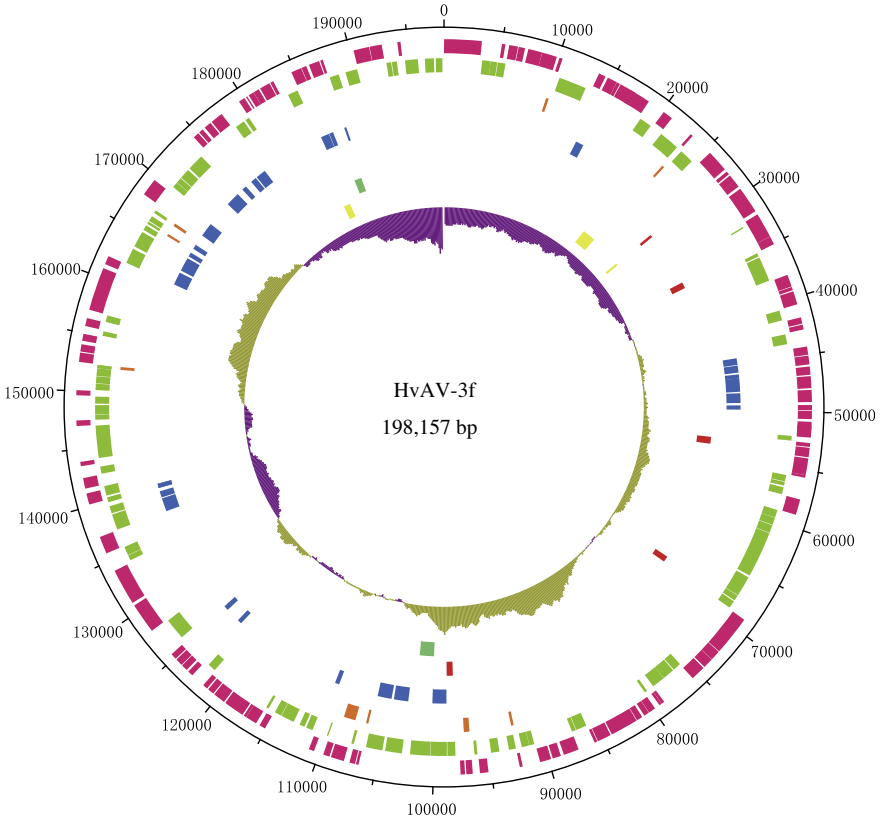

Fig. 1. Circular map of the HvAV-3f genome (198,157 bp). The outer scale is numbered clockwise in bp. Circles 1 and 2 (from outside to inside) denote coding DNA sequences (CDSs). Forward strands are in raspberry and reverse strands in spring green. The nine orange boxes in circle 3 represent the novel putative genes in HvAV-3f. Ocean boxes in circle 4 represent *Bro* genes. The boxes in red in circle 5 represent *iap*-like gene. The two turquoise boxes in circle 6 imply homologous regions (*hrs*). The boxes in circle 7 in yellow represent three regions that are homologous genes present in lepidopteran insects. Circle 8 represents the local variations of G + C content along the genome sequence. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)