Comparative analysis of a highly variable region within the genomes of *Spodoptera frugiperda* ascovirus 1d (SfAV-1d) and SfAV-1a

Jian-Li Xue and Xiao-Wen Cheng

Department of Microbiology, 32 Pearson Hall, Miami University, Oxford, OH 45056, USA

The recently discovered ascoviruses have a worldwide distribution. Here we report a new member of the family *Ascoviridae*, *Spodoptera frugiperda* ascovirus 1d (SfAV-1d) with a variable region in the genome. Restriction fragment length polymorphism, Southern hybridization and genome sequencing analyses confirmed that SfAV-1d and the earlier reported SfAV-1a are closely related but are not identical. The genome size of SfAV-1d is approximately 100 kbp, which is about 57 kbp smaller than SfAV-1a. The SfAV-1d genome has a major deletion of 14 kbp that corresponds to one of the inverted repeat (IR) regions of SfAV-1a. Cloning and sequencing revealed that the region flanking the deletion within the SfAV-1d genome is highly variable. In all the variants of this region, the whole IR region is missing, with 88.2% of the variants missing part of or the whole adjacent SfAV-1a ORF71, 94.1% missing part of or the whole of adjacent ORF72 and 64.6% missing part of or the whole of ORF73.

Ascoviruses are members of the family *Ascoviridae*. They are recently discovered insect-specific viruses that cause a chronic and ultimately fatal disease in lepidopteran larvae, most of which are noctuids (Federici et al., 2005). The enveloped virion of ascoviruses have a circular dsDNA genome with a size ranging from 119 to 186 kbp (Federici, 1983; Federici & Govindarajan, 1990; Cheng et al., 1999; Bideshi et al., 2006; Wang et al., 2006; Asgari et al., 2007; Bigot et al., 2009).

Of all the ascoviruses reported, four ascovirus species have been accepted by the International Committee on Taxonomy of Viruses (ICTV): *Spodoptera frugiperda* ascovirus (SfAV-1), *Trichoplusia ni* ascovirus (TnAV-2), *Heliothis virescens* ascovirus (HvAV-3) and *Diadromus pulchellus* ascovirus (DpAV-4) (Federici et al., 2005). Four ascovirus genomes have been sequenced and reported: TnAV-2c, SfAV-1a, HvAV-3e and DpAV-4a (Bideshi et al., 2005). The ascovirus genome is methylated and contains large interspersed repeats of 1–3 kbp (Bigot et al., 2000). Some of these repeats were identified as being baculovirus repeated ORF (*bro*), with 23 found in HvAV-3e, three in TnAV-2c, seven in SfAV-1a and 12 in DpAV-4a (Bideshi et al., 2006; Wang et al., 2006; Asgari et al., 2007; Bigot et al., 2009). Other than the *bro* genes, there are other repeated DNA regions distributed within the ascovirus genomes. In HvAV-3e, the repeat region encodes a conserved transposable (Asgari et al., 2007). Two homologous direct repeats (DR) were identified in the TnAV-2c genome and five DRs were identified in HvAV-3e, whereas 4.5 homologous inverted repeats (IR) were identified in the SfAV-1a genome (Bideshi et al., 2006; Wang et al., 2006; Asgari et al., 2007; Bigot et al., 2009).

A new ascovirus isolate, tentatively named SfAV-1d, was isolated from a *S. frugiperda* larva collected in a soybean field in South Carolina, USA, in 2002. In this study, we characterized the difference between SfAV-1d and the reported SfAV-1a. SfAV-1d was propagated in *S. frugiperda* larvae (Cheng et al., 2000). Viral DNA was purified by virion sucrose density-gradient centrifugation (Federici & Govindarajan, 1990) and cesium chloride–ethidium bromide gradient ultracentrifugation (Cheng et al., 1999). The other ascoviruses used in this study were SfAV-1a (a gift from Professor B. A. Federici, University of California, Riverside, California, USA), TnAV-2d and HvAV-3f isolated from *H. virescens* (Cheng et al., 2005).

The restriction fragment length polymorphism (RFLP) and Southern hybridization comparison between SfAV-1d and SfAV-1a showed that SfAV-1d is closely related to SfAV-1a but its genome size is smaller than that of SfAV-1a. In conventional agarose gel electrophoresis, the restriction fragment patterns of individual restriction endonuclease (REN) digestion of genomic DNA with *Hind* III and *Bam* HI of the two isolates showed that SfAV-1d is similar to SfAV-1a with seven *Hind* III and nine *Bam* HI co-migrating fragments, but three *Hind* III and seven *Bam* HI fragments showed different mobilities during agarose gel electrophoresis, thus suggesting that SfAV-1d is distinct from SfAV-1a (Fig. 1a). To analyse the relationship...
between SfAV-1d and SfAV-1a at the DNA level, an SfAV-1d genomic HindIII restriction library was constructed. SfAV-1d–HindIII-fragment library end sequencing showed that four of the seven HindIII REN fragments co-migrating with SfAV-1a displayed 99% identity with only a few single nucleotide polymorphisms (data not shown). The largest fragments resulting from HindIII and BamHI digestion of SfAV-1a were absent in the HindIII and BamHI digestion of SfAV-1d (Fig. 1a).

To confirm the similarities and differences between SfAV-1d and SfAV-1a further, pulse-field gel electrophoresis (PFGE) was conducted using a 1% agarose gel at 5 V cm⁻¹, 0.5–1.5 s interval at 15 °C with DNA PFGE marker (New England Biolabs) for 17 h to increase the resolution of large REN fragments. The three large HindIII fragments (>24.5 kbp) in SfAV-1a were not present in SfAV-1d. One (>24.5 kbp) and two (>29.9 kbp) large fragments of BamHI and PstI fragments of SfAV-1a, respectively, were not present in SfAV-1d (Fig. 1c). After the sizes of all the restriction fragments of each REN digestion of SfAV-1d were estimated and added up, the genome size of SfAV-1d was estimated to be approximately 100 kbp, which is about 57 kbp smaller than that of SfAV-1a (Bideshi et al., 2006). In order to test whether SfAV-1d is more closely related to SfAV-1a than to TnAV-2d and HvAV-3f, Southern hybridization was performed under high-stringency hybridization conditions by using Sau3AI-digested SfAV-1a genomic DNA, which was labelled with biotin (Phototope kit; New England Biolabs), as the probe. Strong hybridization occurred between SfAV-1a and SfAV-1d, confirming the high degree of homology between SfAV-1a and SfAV-1d. Little or no hybridization occurred between SfAV-1d and TnAV-2d or HvAV-3f (Fig. 1b). Collectively, these data suggest that the SfAV-1d is closely related to the reported SfAV-1a and that the difference between the two isolates is largely because of the deletion of DNA sequences from the SfAV-1d genome compared with that of SfAV-1a.

From the estimation of the SfAV-1d genome size based on the length of the restriction fragments, SfAV-1d is smaller than the published SfAV-1a genome. To confirm the deletion, a shotgun library of SfAV-1d genomic DNA was
constructed and sequenced (unpublished data). When these shotgun library sequences were aligned against the published SfAV-1a genome sequence using SEQUENCHER 4.8 (Gene Codes) at 80% identity, several deletions, including a major deletion of 14 kbp, were found (Fig. 2a). The largest deletion is from nt 79 979 to 93 882 of the SfAV-1a genome (Bideshi et al., 2006), which is the region with four non-coding IRs of SfAV-1a. Primers were designed upstream and downstream of the deletion region (DelF 5'-TCACCCATGTGAGGATCG-3' and DelR 5'-TCATCGCCGTCACAACAC-3') (Fig. 2a). The forward primer (DelF) anneals to the intergenic region between ORF70 and ORF71 (nt 79 918–79 935) of SfAV-1a; the reverse primer (DelR) anneals to ORF73 (nt 94 646–94 663) of SfAV-1a (Bideshi et al., 2006). The amplified products from SfAV-1d showed an expected major-product size of approximately 600 bp with some smears of DNA products in the range of 600–1500 bp, which suggest that different lengths of PCR products produced by these primers represent a variable region in the SfAV-1d genome at this locus. In SfAV-1a, six discrete DNA fragments of approximately 14, 10.5, 9, 5, 1.7 and 0.6 kbp were amplified. Weak-signal PCR products and smearing DNA were also observed between these discrete PCR products of SfAV-1a (Fig. 2b). Based on the published genome sequence of SfAV-1a, the 14 kbp PCR product is the expected size and these other DNA PCR products of SfAV-1a might also be from the same variable region of the SfAV-1a genome as found in SfAV-1d (Fig. 2b). The data suggest that wild-type SfAV-1a and SfAV-1d are a mixture of viruses containing different genomes with different lengths of DNA sequence at this locus.

Since the amplified PCR products from SfAV-1d genomic DNA using the primers were concentrated at 600 bp with DNA smears in the range of 600–1500 bp (Fig. 2b), PCR products in the size distribution of 600–1500 bp were gel purified and cloned. During the colony screening, different sizes of inserts were identified in the range of 600–1500 bp (Fig. 3a). Seventeen clones with inserts of different sizes of inserts were selected and sequenced. The sequence results confirmed that the deletion region of SfAV-1d is highly variable (Fig. 3b). The shortest sequence covers nt 79 918–80 294 and 94 542–94 663 of SfAV-1a, and is missing the whole IR region, ORF72 and part of ORF71; the longest sequence covers nt 79 918–80 202 and 93 901–94 663, and is missing the whole IR region as well as part of ORF71 and ORF72 (Bideshi et al., 2006). All 17 variants analysed are missing the whole IR region of SfAV-1a (Fig. 3b). Therefore, the longest deletion region of SfAV-1d that corresponds to nt 79 995–94 542 of SfAV-1a is considered to be the highly variable region of SfAV-1 (Fig. 3b). Among all these variants, 70.5% lack the whole of ORF72 and 17.7% lack the whole ORF71, which suggests that ORF71 and ORF72 may not be essential for SfAV-1 replication.

Since the discovery of ascovirus in the late 1970s in the south-eastern USA, ascoviruses have been reported from other parts of the world such as Indonesia, Australia and...
France (Carner & Hudson, 1983; Bigot et al., 1997; Cheng et al., 2000; Asgari, 2006). This suggests a worldwide distribution of ascoviruses. Interest in ascovirus research is rooted in the cytopathic effects of ascovirus on permissive insect cells resulting in the partitioning of the infected cells. Further genetic study showed that the virally encoded caspase of SfAV-1a is responsible for the cellular cleavage (Bideshi et al., 2005). One area, however, that has not been addressed well is the genes that are required for ascovirus replication and infection in cells, and its DNA replication strategy. Here, to the best of our knowledge, we, for the first time, present data showing highly variable DNA deletions in a region of the smallest SfAV-1d genome that correspond to the IR region of the SfAV-1a genome.

SfAV-1a was isolated from S. frugiperda in Georgia, USA, in 1982 (Hamm et al., 1986) and SfAV-1d was isolated from a single S. frugiperda larva in South Carolina. Geographically, SfAV-1a and SfAV-1d were isolated from the same region and theoretically they are closely related. Both conventional agarose-gel electrophoresis and PFGE analysis showed that SfAV-1d and SfAV-1a have similar restriction profiles, and Southern hybridization confirmed their relatedness (Fig. 1). The differences that contribute to the separation of SfAV-1d from SfAV-1a are within the deletions of repetitive sequences in SfAV-1a, since the genome sequence data obtained from SfAV-1d showed 99% identity with SfAV-1a at the DNA level (unpublished data). These DNA deletions in the SfAV-1d genome result in the formation of a smaller ascovirus genome, the smallest so far reported. Genome sequencing of SfAV-1d showed that the 14 kbp IR deletion in SfAV-1d corresponds to the largest 49 kbp HindIII fragment of SfAV-1a (Fig. 1 and unpublished data).

SfAV-1a was reported to have a genome size of 157 kbp, based on genome sequencing (Bideshi et al., 2006). Our PCR result showed that the 14 kbp IR region is a variable
region in the SfAV-1a genome (Fig. 2b). The reported size of SfAV-1a is probably based on the largest SfAV-1a genome with the longest 14 kbp IR region. Early estimates of the SfAV-1a genome by REN digestion showed a 140 kbp with the largest HindIII fragment to be 43.9 kbp (Fig. 6 in Federici et al., 1990). However the largest HindIII fragment from the genome sequence of SfAV-1a is 49.2 kbp (Bideshi et al., 2006), which further supports our discovery that the 14 kbp IR region of SfAV-1a is variable. In addition to the deletion of the IR region of SfAV-1a, the adjacent regions of the IR region are also missing in SfAV-1d, which may suggest that this region is neither essential for SfAV-1d nor for SfAV-1a replication in permissive cells. The largest clone of the deletion region of SfAV-1d is missing only the IRs and part of the adjacent ORF71 and ORF72 of SfAV-1a. The smallest clone is missing the IRs and the whole of ORF71 and ORF72, part of the SfAV-1a. ORF71 has no sequence similarity to known genes while ORF72 has high sequence similarity to the bro gene of baculovirus (Bideshi et al., 2006). It is reported that the IRs have different lengths and number of repeats among SfAV-1a, SfAV-1b and SfAV-1c, suggesting that only part of the IRs may be essential for SfAV-1 replication, and this might be the region that undergoes structural changes in converting between the linear and the circular configurations of the genome (Bideshi et al., 2006). This is supported by our data, although our study showed that since SfAV-1d misses the whole IR region, this region may not be essential for SfAV-1 replication.

Similar variable regions have been reported in the genome of baculovirus Anticarsia gemmatalis nucleopolyhedrovirus (AgMNPV) (Garcia-Maruniak et al., 1996). Similarities in the variable regions of genomes of ascovirus and baculovirus support the notion that ascovirus and baculovirus are closely related (Cheng et al., 2007). In addition, it has been reported that several viruses, including baculovirus and herpes simplex virus, use a rolling-circle replication mechanism to provide rapid genome replication in cells (Boehmer & Lehman, 1997; Oppenheimer & Volkman, 1997). This variable region of SfAV-1d and possibly SfAV-1a (nt 79993–94542) may serve as the origin of replication of the circular genome of ascovirus during rolling-circle replication inside the permissive insect cells. One possible explanation for the formation of this variable region is that during the cleavage of the concatenated ascovirus genomes by endonucleases, either encoded by the virus or by the cell, exonucleases from either the virus or the cell progressively degrade the ends of the linear genome of the ascovirus before the genome is circularized by ligation.

In conclusion, this study showed the high degree of sequence identity and the major difference between the SfAV-1a and SfAV-1d genomes, and furthermore estimated the total size of the SfAV-1d genome to be approximately 100 kbp, which makes it the smallest ascovirus genome found so far. A major deletion of 14 kbp in the IR region of SfAV-1a is found in SfAV-1d. The deleted region of the SfAV-1d genome is highly variable.

Acknowledgements

The authors would like thank Drs Yayi Kusumah and Gerald R. Carner for providing the newly isolated SfAV-1d. Dr Brian A. Federici is greatly appreciated for donating SfAV-1a. Undergraduates Andrew Kilianski, Samuel Murakami and Michael Howell are credited for viral DNA cloning and sequencing. The authors would like to thank Racheal Desmone for proofreading of this manuscript.

References


