Characterization of repetitive DNA regions and methylated DNA in ascovirus genomes

Yves Bigot,1,2 Karine Stasiak,1,2 Florence Rouleux-Bonnin1 and Brian A. Federici2

1 IRBI, Groupe d’Etude des Parasites Moléculaires, UPRESA CNRS 6035, Faculté des Sciences, Parc de Grandmont, 37200 Tours, France
2 Department of Entomology and Interdepartmental Graduate Program in Genetics, University of California, Riverside, CA 92521, USA

The accompanying phylogenetic study of large double-stranded DNA viruses based on their δ DNA polymerase genes suggests that ascoviruses (family Ascoviridae) and iridoviruses (family Iridoviridae) are closely related and may share a common ancestor. This relationship was unexpected because of marked differences between these viruses. Iridoviruses produce icosahedral virions and occur broadly among vertebrates and invertebrates, whereas ascoviruses typically produce reniform or bacilliform virions and are restricted to insect hosts, primarily lepidopterans. Detailed comparisons of these two virus types are not possible because fundamental information on the properties of the virions and their genomes is lacking, especially for ascoviruses. To facilitate further investigation of the putative evolutionary relationship between ascoviruses and iridoviruses, the genomes of representative viruses from each family were compared with respect to physical configuration, presence of DNA repeats and degree of DNA methylation. Genomes from Spodoptera frugiperda (SfAV1), Heliothis virescens (HvAV3) and Diadromus pulchellus (DpAV4) ascoviruses were all found to be circular and partially superhelical and to contain large interspersed repeats of 1–3 kbp. Mosquito (IV type 3), lepidopteran (IV type 6) and isopod (IV type 31) iridovirus genomes were all linear and lacked large regions of repetitive DNA. Ascovirus and iridovirus genomes were methylated and one, DpAV4, had the highest degree of methylation of any reported animal DNA virus. The major differences in the physical and biochemical characteristics of ascoviruses and iridoviruses reported here provide a foundation for further studies of their relatedness while making their possible close relationship and divergence during evolution of even greater interest.

Introduction

Viruses of the family Ascoviridae cause chronic, fatal diseases in insects of the order Lepidoptera (Federici et al., 2000). Ascovirus virions are large (400 by 100–150 nm) and enveloped, with complex symmetry, and they contain at least 20 proteins. The genome is a double-stranded (ds) DNA molecule ranging in size from 116 to 190 kbp, depending on the species (Bigot et al., 1997a; Cheng et al., 1999; Federici et al., 1990). The transmission of ascoviruses is unusual in that they are poorly infectious per os and appear to be transmitted among their lepidopteran hosts by parasitoid wasp vectors at oviposition (Hamm et al., 1985). Virus maintenance in host populations is therefore dependent upon their relationships with their vectors. In this regard, two types of relationships have been identified. In the first, the relationship is antagonistic. Parasitoid larvae emerging from eggs laid in the host typically die within a few days, with death resulting from infection of the host by virions transmitted at oviposition (Hamm et al., 1985). Current knowledge indicates that this is the most common type of relationship for ascoviruses and is characteristic of three of the four accepted species in the family, the Spodoptera frugiperda (SfAV1), Trichoplusia ni (TnAV2) and Heliothis virescens (HvAV3) ascoviruses, which all replicate in larvae of the family Noctuidae. In contrast to this, the Diadromus pulchellus (DpAV4) ascovirus has a symbiotic relationship with its hymenopteran vector, D. pulchellus. When...
Fig. 1. Summary of phylogenetic relationships of selected members of the Ascoviridae and Iridoviridae and status of knowledge of the physical properties of their genomic DNAs. (a) Putative relationships of several ascoviruses and iridoviruses based on a phylogenetic analysis of their δ DNA polymerases are shown on the left. The table on the right summarizes selected key physical and biochemical properties. The genome features of each virus genus are aligned with the viruses shown in the phylogenetic tree. The presence or absence of a feature is indicated by + or −; the absence of relevant data in the literature is indicated by ?. (b) Schematic representation of two types of genomic configuration for viruses with a circular restriction map. The first actually has a circular genome, whereas the second has a linear configuration that is circularly permuted at random, which results in a circular restriction map. In the latter, the grey arrows indicate different pairs of ends that can occur in a set of linear molecules that are circularly permuted at random. E1, E2 and E3 are hypothetical restriction sites defining the same restriction map in both conformations. Data on phylogenetic relationships are from Stasiak et al. (2000), whereas data on biochemical and physical properties are from Tidona & Darai (1997) (1), Schnitzler et al. (1987a, b) (2), Goorha & Murti (1982) (1, 2), Goorha et al. (1984) and Kaur et al. (1995) (3), Delius et al. (1984) (4), Williams (1994) (5), Bigot et al. (1997a) (6) and Cheng et al. (1999) (7).

oviposited with eggs in its normal host, the pupa of the leek moth, *Acrolepiopsis assectella*, DpAV4 improves the rate of successful parasitization by larvae of its wasp vector (Bigot et al., 1997b). In other hosts, however, the relationship is antagonistic as in other ascovirus–parasite relationships (Bigot et al., 1997b). The results of the accompanying study of ascovirus phylogenetics based on analyses of their δ DNA polymerase genes are consistent with these different ascovirus biologies. This study revealed two subgroups, in essence genera, within the family (Stasiak et al., 2000). One contained SfAV1, TnAV2 and HvAV3, the ascoviruses attacking noctuids, and the other DpAV4, the ascovirus that attacks the leek moth.

The unique pathobiology of ascoviruses makes their origin and phylogenetic relationship to other viruses of interest. Comparisons of ascovirus δ DNA polymerase genes with those of other large DNA viruses suggest that ascoviruses, which have bacilliform to reniform virions, and iridoviruses (family *Iridoviridae*), which have icosahedral virions, are closely related (Stasiak et al., 2000). These same studies revealed that viruses of the families *Asfaviridae* and *Phycodnaviridae*, which also have icosahedral virions, occur on the same DNA polymerase gene tree branch, but are less related to ascoviruses. Although the overall tree topology provides support for a close relationship between ascoviruses and iridoviruses, the markedly different shapes of their virions make this relationship questionable.

The putative close relationship of the families *Ascoviridae* and *Iridoviridae* is intriguing because, if verified, it would be the first case where viruses with different virion symmetries would
have been shown to have originated from a common ancestor. Moreover, a better understanding of the relationship between ascoviruses and iridoviruses could provide insights into how viruses evolved as well as clarify the origins and relationships of other large DNA virus families. At best, these remain poorly understood (Van Regenmortel et al., 2000; Domingo et al., 1999). It is difficult to address these problems, especially with respect to the ascoviruses and iridoviruses, because so little is known about the comparative properties of these viruses. For example, the physical configuration of the genome and the sequence of lymphocystis disease virus (LCDV), a vertebrate iridovirus, are known, but comparable knowledge of the invertebrate iridoviruses and ascoviruses is greatly lacking, as summarized in Fig. 1(a).

To begin to address this lack of knowledge, we studied some of the key physical and biochemical properties of a selected number of ascovirus and iridovirus genomes. In the present paper, using the genomes of SFAV1, HvAV3 and DpAV4 and iridoviruses from an isopod, a moth and two mosquitoes, we report on their physical configuration, the presence and features of repeated sequences and the presence of methylated DNA.

Methods

■ Viruses. The ascoviruses SFIAV1a, HvAV3c and DpAV4 and the iridoviruses turquoise mosquito iridovirus (T-MIV), regular mosquito iridovirus (R-MIV) and iridovirus type 31 (IV31) were obtained as described in Stasiak et al. (2000). Autographa californica multinucleocapsid nucleopolyhedrovirus (AcMNPV) strain L2 was used as a source of baculovirus genomic DNA.

■ Viral DNA purification. Ascovirus DNAs were purified as described previously (Federici et al., 1990). Iridovirus DNAs were purified by using the same procedure except that the sonication step was eliminated and the virions were prepurified by passing samples of crushed infected isopods or mosquito larvae through a nylon membrane filter (10 µm porosity, Millipore). Proteins were removed from the virus samples by treatment in 0.1% SDS, 200 µM EDTA, pH 9, 0.2% SDS, 200 µg/ml proteinase K for 2 h at 50 °C. During phenol and chloroform–isoamyl alcohol extraction, the organic phase was removed by pipetting. The aqueous phase was not pipetted to avoid shearing the DNA.

■ Wasp and DpAV4 host DNA. Total genomic DNA from males and females of D. pulchellus and from Acrolepiopsis assectella parasitized by D. pulchellus wasps or artificially infected by stinging with DpAV4 was obtained as described previously (Bigot et al., 1997b).

■ Southern blot hybridization. DNA purification, synthesis of labelled DNA probes and agarose gel electrophoresis followed standard procedures (Ausubel et al., 1994). Southern blots were hybridized by using conditions described in the accompanying paper (Stasiak et al., 2000).

■ DNA fragment cloning and sequencing. Fragments were purified and then sequenced as described in the accompanying paper (Stasiak et al., 2000). Fragments were cloned into the vector pBluescript SK+ (Stratagene). Two clones of each fragment were sequenced, one from each strand. The sequences reported here appear in the DDBJ/EMBL/GenBank sequence database under accession numbers AJ279828 and AJ279829.

■ Sequence analyses. The Infobiogen facilities (Dessen et al., 1990) were used for database searches, sequence alignments and calculations.

■ Pulse-field gel electrophoresis (PFGE). PFGE was done by using a CHEF-DR II PFGE system as described by the manufacturer (Biorad). Gels contained 1% LE agarose and 0.5% TBE (1× TBE is 89 mM Tris–HCl, 89 mM boric acid, 25 mM EDTA, pH 8.3). In order to separate fragments of between 5 and 500 kbp, electrophoresis was performed at 120 V, 14 °C, at an angle of 120° and a pulse ramping from 0.22 to 44 s for 16 h. A λ phage DNA ladder (New England Biolabs) was used to estimate DNA sizes. To obtain high resolution of undigested viral DNA, small quantities (1–2 ng) were loaded on the gels. Results were therefore detected by Southern blot hybridization with DNA of the virus species as probes.

■ CsCl equilibrium density gradient analyses. Viral DNA samples purified as described above were analysed by equilibrium density gradient centrifugation in CsCl gradients to determine whether genomic DNA occurred at different densities, indicative of different physical conformations from linear DNA molecules to supercoiled and relaxed circular DNA molecules. Standard conditions used for plasmid purification were followed (Ausubel et al., 1994). Briefly, 13.5 ml ultraclear ultracentrifugation tubes (Beckman) containing 200 µg viral DNA in a 1.6 g/cm³ CsCl gradient and 100 µg/cm³ ethidium bromide (EtBr) were centrifuged in a Ti 55.2 rotor (Beckman) at 38000 r.p.m. for 30, 38 or 48 h at 25 °C. Bands containing DNA were visualized under UV light (312 nm) at the end of the ultracentrifugation.

■ Detection of methylated bases. DNA was hydrolysed by the ‘four-hour-two-enzymes’ technique (Gehrke et al., 1984), which treats denatured nucleic acid successively with nuclease P1 and bacterial alkaline phosphatase for 4 h. The nucleosides were separated by high-speed HPLC with a Colospher Sphericol 80ODS2 column (25 cm long, 5 µm porosity). Elution was performed with 0.05 M K₂HPO₄, pH 4.4, 16% methanol at 20 °C at a flow rate of 1.2 ml/min. Detection was done at 254 nm with a Waters 996 photodiode array detector. Nucleosides were identified by comparing their retention times with those of standard synthetic nucleosides (Sigma). Amounts of deoxyctydine and 5-methyldoxcytidine were calculated by using the Millennium software (Waters).

Results

Physical properties of the genomes

Restriction endonuclease and PFGE results. Based on fragment patterns generated with BamHI, BglII, CiaI and NcoI, the IV31 genome was determined to be about 230 kbp (data not shown). This was in good agreement with a previous estimate based on EcoRI digests (Williams, 1994). A fragment of about 230 kbp was observed by Southern blotting of a pulse-field gel (Fig. 2). In addition, a smear of fragments ranging from 130 to 180 kbp was detected. Under other PFGE conditions (1% LE agarose gels in 0.5% TBE, with electrophoresis conditions at 200 V, 14 °C, an angle of 120° and a pulse ramping from 20 to 50 s for 26 h, which allowed separation of fragments between 50 and 600 kbp), this smear consisted of five to eight discrete molecules, suggesting that shorter genomes might be encapsidated in IV31 virions. Similar results were obtained...
The AcMNPV genome is 134 kbp (Ayres et al., 1994) and, as expected for a circular molecule, a large single DNA molecule was observed migrating faster than 134 kbp, at a position corresponding to 80–110 kbp (Fig. 2). These results provide additional evidence that these two ascoviruses have circular genomes.

CsCl ultracentrifugation. When measured by CsCl gradients, the density of double-stranded molecules of DNA depends on their tertiary structure. Linear or relaxed circular molecules have a mean density of 1.59 g/cm³, whereas superhelical circular molecules are of higher density. The increase in density depends upon the number of turns present in the circular molecule. The genome of AcMNPV has a superhelical circular configuration (Summers & Anderson, 1972). However, in CsCl equilibrium gradients of AcMNPV genomic DNA, relaxed circular and linear molecules occur in addition to superhelical circular molecules. The relaxed circular and linear forms result respectively from nicking and shearing of the superhelical molecules (Britten et al., 1974; Summers & Anderson, 1972). In our experiments, the linear and relaxed circular forms were separated easily from the superhelical forms in a CsCl–EtBr gradient (Fig. 3, left tube). As observed in previous studies of baculovirus DNA (Summers & Anderson, 1972), the band containing the superhelical DNA was less dense than that containing linear and relaxed circular molecules, indicating that the large superhelical circular molecules were fragile and frequently nicked during DNA purification.

In CsCl equilibrium density gradients of SfAV1a and HvAV3a genomic DNA, two bands were typically observed after 38 h centrifugation (Fig. 3). One of these had approximately the same density as the AcMNPV genome band containing the linear and relaxed circular molecules. Although less dense than the AcMNPV superhelical circular molecules, a second band of higher density was observed with SfAV1a and HvAV3c. The amount of DNA in this band increased the longer these large circular genomes were subjected to ultracentrifugation (30, 38 and 48 h at 38 000 r.p.m., 25 °C). This indicated that single-stranded DNA nicks occurred during the ultracentrifugation.

Characteristics of repetitive interspersed DNA

In previous studies, repetitive regions of DNA were reported in ascovirus genomes (Bigot et al., 1997a; Federici et al., 1990), but such regions have not been reported in iridoviruses. In the present study, therefore, a comparative search for the presence and location of repeated interspersed regions of DNA was undertaken with the genomes of SfAV1a, HvAV3a, IV31 and T-MIV by restriction enzyme analysis and Southern hybridization. The regions identified were then characterized.

In the SfAV1 genome, repeated fragments were found in HpaII (1 and 1-2 kbp), SspI (1-1 and 1-5 kbp) and XhoI (0-85 and 2-5 kbp) digests. Of these, the two HpaII and two SspI fragments were cloned and sequenced. Comparison of the sequences of the four fragments revealed that they were 98–100% identical, permitting construction of a 3-1 kbp
consensus fragment and its restriction map (Fig. 4a; accession no. AJ279828). In hybridization experiments, the two cloned HpaI fragments hybridized to two XhoI fragments in a Southern blot containing EcoRI, XhoI, and XbaI DNA digests from three SFAV1 isolates (Fig. 4b, c). In conjunction with the restriction map and hybridization data, these results indicated that the repeated motif in the SFAV1 genome was at least 0.85 kbp due to differences in the size of these repeats (Fig. 4b, lanes 1–2, 4–5 and 7–8). Other restriction digests with enzymes unable to cut within the repeats and dot-blot hybridization experiments indicated that five or six repeats were present in the SFAV1a genome. There appeared to be fewer repeats in the SFAV1b and SFAV1c genomes (data not shown).

In the HvAV3c genome, a 1-1 kbp repeat was identified in the Alul digest (Fig. 5a) and cloned and sequenced (accession...
Fig. 7. HPLC of nucleosides from hydrolysed DNA from different samples. (a) Synthetic nucleosides used as controls. 
(b)–(f) DNA from DpAV4 (b), A. assectella males (c), SfAV1 (d), T-MIV (e) and IV31 (f). Nucleosides peaks are labelled: 
dC, deoxycytidine; 5-metdC, 5-methyldeoxycytidine; dG, deoxyguanidine; dT, deoxythymidine; and dA, deoxyadenine. The 
bar at 2 min indicates an artifact resulting from sample injection. AU, Absorbance units.

no. AJ279829). Analysis of the sequence did not reveal any 
direct or inverted regions. This repeat was hybridized to 
Southern blots to define further its presence in the HvAV3c 
genome. The 1-1 kbp fragment hybridized to several other 
fragments of 0-2–1-8 kbp in the Sau3AI digest, with varying 
intensity (Fig. 5, lane 5). Because there was no Sau3AI site in 
the cloned repeat, the hybridization of fragments as large as 
1-1 kbp indicated that at least four repeats were present in the 
HvAV3c genome. Similarly, the hybridization of the 1-1 kbp 
repeat to as many as five fragments of less than 1-1 kbp, and 
the variation in the intensity of the signal, indicated that these 
repeats were polymorphic. This fragment also hybridized to 
two DraI fragments, of about 0.45 and 0.75 kbp (Fig. 5 b, lane 
1), and two SspI fragments, of about 1.4 and 1.7 kbp (Fig. 5 b, 
lane 6). There was only one DraI site and one SspI site in the 
sequence of this repeat, respectively at positions 293 and 484, 
which indicated that the size of this repeat in the HvAV3c 
genome was at least 1-1 kbp. In the EcoRV, HpaII and NdeII 
digests (Fig. 5 b, lanes 2–4), the 1-1 kbp fragment hybridized to 
only one or three fragments in each digest and these ranged 
from 2 to 30 kbp. In the NdeII digest, this indicated that the 
repeats were not distributed randomly in the 180 kbp HvAV3c 
genome, but were clustered primarily in the 30 kbp fragment. 
The 1-1 kbp AluI repeat did not hybridize in Southern 
blotting experiments to EcoRI digests of DNA from several 
isotypes of SfAV1, TnAV2 or HvAV3 (Fig. 5 c). Even when the 
blots were washed under conditions of low stringency, the 
AluI repeat hybridized only to the three HvAV3 isotypes (Fig. 
5 c, lanes 4, 6 and 7), indicating that this repeat occurred only 
in HvAV3.

In order to characterize further the 980 bp repeats identified 
previously in the DpAV4 genome (Bigot et al., 1997a), these 
repeats and their flanking regions were cloned and sequenced. 
Analysis of these sequences showed insertions of sequences 
that varied in size and sequence (Fig. 6). These insertions 
resulted in variation among the DpAV4 repeats that yielded 
sequence identities ranging from 72 to 94%.

Searches for repetitive DNA in the genomes of the
iridoviruses IV31 and T-MIV, with the same techniques used to search ascovirus genomes, did not reveal any large repetitive interspersed regions of DNA.

**Presence of 5-methyldeoxycytidine in ascovirus DNA**

Cytosine methylation of genomic DNA is a characteristic of several vertebrate iridoviruses (Willis & Granoff, 1980; Wagner et al., 1985). In order to determine whether ascovirus and invertebrate iridovirus genomes are also methylated, hydrolysed genomic DNAs from several representatives of these viruses were examined for the presence of 5-methyldeoxycytidine by HPLC with synthetic nucleosides as controls (Fig. 7). A very high level of 5-methyldeoxycytidine was observed in DpAV4 DNA extracted from A. assectella, with approximately 76% of the deoxycytidines being methylated (Fig. 7b). Because DpAV4 virions are difficult to purify due to their fragility, the possibility existed that the methylated DNA originated from contamination with pupal host DNA during the virion purification process. Previous studies indicated that approximately 5% of the DNA in DpAV4 DNA preparations was of host origin (Bigot et al., 1997a,b). Therefore, DNA isolated from non-parasitized, uninfected pupae of the lepidopteran host, A. assectella, was analysed. High levels of 5-methyldeoxycytidine were observed in these extracts, 36% in male DNA and 56% in female DNA (Fig. 7c). However, given the relatively low level of host DNA (5%) in DpAV4 DNA preparations, the data indicate that most of the 5-methyldeoxycytidine observed in the DNA isolated from infected hosts originated from DpAV4 DNA. In DNA isolated from DpAV4 and A. assectella, a sixth peak was present corresponding to an unidentified non-nucleoside molecule.

Analyses of T-MIV, IV31, Chilo iridovirus (CIV), SFAV1a and HvAV3c DNAs (data not shown for the last two) revealed the presence of 5-methyldeoxycytidine in the genomes of these viruses. However, the levels were low, ranging from 2.5 to 5% of all deoxycytidines (Fig. 7), levels much lower than those observed in DpAV4 or for the vertebrate iridoviruses.

Because methylation of ascovirus DNA is a potentially important aspect of ascovirus biology, the pattern of virus cytosine methylation in different DpAV4 hosts was examined. This was done by using restriction enzymes that are sensitive to the presence of 5-methyldeoxycytidine. Four types of DpAV4 DNA were examined, each corresponding to a different host or developmental stage of the viral genome. These were DpAV4 DNA from (i) D. pulchellus males, (ii) D. pulchellus females, (iii) 4-day-old A. assectella pupae parasitized by D. pulchellus and (iv) 4-day-old A. assectella pupae infected artificially by injecting them with DpAV4 virions (Bigot et al., 1997a,b). In males of D. pulchellus, where no virus replication has been detected, the DpAV4 genome is present in nuclei as an unencapsidated circular molecule (Bigot et al., 1997a). In D. pulchellus females, the virus replicates but the number of virions produced is small. In the two types of infected pupae, virions levels are high. Because these different sources of viral DNA produce highly variable amounts of DNA, the DNA extracts were examined by Southern blot hybridization with specific DpAV4 probes. The viral DNA used for the blots was digested with EcoRI and Clal and probed with a 5 kbp fragment containing the DpAV4 δ DNA polymerase gene (AI279812). (b) DpAV4 DNA from A. assectella pupae digested with MspI (lane 1) or HpaII (lane 2) and probed with a 6 kbp fragment of the DpAV4 genome (AI279815). In (a) and (b), the amount of DNA used for each sample was adjusted to obtain fragments that hybridized with similar intensities. Arrows identify restriction fragments present or absent in the different DNA samples. Final blot washes were in 0.5 x SSC, 65°C. Molecular masses are indicated to the left.

**Fig. 8.** Detection of methylated DpAV4 DNA by Southern blot analysis in wasp and caterpillar hosts. (a) Restriction digests from total DNA from male (lane 1) and female (lane 2) D. pulchellus wasps, A. assectella pupae at 4 days post-parasitization (lane 3) and A. assectella pupae inoculated with DpAV4 (lane 4). DpAV4 DNA in (a) was digested with EcoRI and Clal and probed with a 5 kbp fragment containing the DpAV4 δ DNA polymerase gene (AI279812). (b) DpAV4 DNA from A. assectella pupae digested with MspI (lane 1) or HpaII (lane 2) and probed with a 6 kbp fragment of the DpAV4 genome (AI279815). (a) and (b), the amount of DNA used for each sample was adjusted to obtain fragments that hybridized with similar intensities. Arrows identify restriction fragments present or absent in the different DNA samples. Final blot washes were in 0.5 x SSC, 65°C. Molecular masses are indicated to the left.
Discussion

The accompanying phylogenetic analysis based on δ DNA polymerase genes of large dsDNA viruses suggests that ascoviruses and iridoviruses may be closely related (Stasiak et al., 2000). These two virus types differ substantially in their biology and structure and thus, to facilitate further comparisons of their relatedness, we characterized several key physical and biochemical properties of representative ascovirus and iridovirus genomes to fill the knowledge gaps identified in Fig. 1. We determined that the mosquito, lepidoptera and isopod iridoviruses have linear genomes that lack large DNA repeats and contain low levels of methylation. In contrast, the ascoviruses were shown to have circular, partially supercoiled DNA genomes that contain DNA repeats of 1–3 kbp. Also, whereas the ascovirus genomes in general showed low levels of methylation, DpAV4 was an exception, with DNA that replicated in the caterpillar host having a level of methylation higher than that of any known DNA virus. The low level of superhelicity observed in ascovirus genomes in the present study (Fig. 3) was not detected in previous studies. This is probably because the method used to purify viral DNA in previous studies included sonication and substantial pipetting (Federici et al., 1990). Because proteins have a lower density (about 1·3 g/cm³) than DNA, it is unlikely that the second, upper layer observed was due to proteinase K-resistant proteins binding to the DNA. Had this occurred, it would have resulted in a second protein–DNA layer with a density of less than 1·59 g/cm³. In addition, superhelicity was not detected in the genome of DpAV4, in this case probably because the tubes used for equilibrium sedimentation (3 ml) were small (Bigot et al., 1997a). This apparently prevented the two molecule classes from separating from one another to an extent that they could be resolved. A possible explanation for the low level of supercoiling in ascoviruses is that the genome, although circular, does not have the same packing restrictions as, for example, baculovirus genomes. Many of the latter viruses, such as Trichoplusia ni granulovirus, assemble a DNA genome of 130–160 kbp into a narrow, rod-shaped nucleocapsid about 300 nm in length by 80 nm in diameter (Summers & Anderson, 1972). In contrast, the ascoviruses assemble a DNA genome of about the same size or slightly larger (130–180 kbp) in an inner particle that measures 350 nm in length by 130 nm at its maximum width (Federici et al., 1990).

Interspersed DNA repeats were previously reported in the genomes of SfAV1a (Federici et al., 1990) and DpAV4 (Bigot et al., 1997a) but were only partially characterized. In the present study, we found four to eight interspersed repeats in all ascovirus genomes, ranging from 960–980 bp to more than 3·8 kbp. None of these DNA repeats contained ORFs encoding more than 70 amino acids. Also, of the small ORFs detected, none shared any sequence similarity with repeats from other ascovirus species, beyond what would be expected randomly. This lack of sequence similarity among the DNA repeats of the different ascovirus species indicates that they do not correspond to a family of small repeated genes, such as those described for the ALI genes in the genome of the Melanoplus sanguinipes entomopoxvirus (Afonso et al., 1999).

Although their function remains unknown, the DNA repeats we have characterized might be useful as a taxonomic tool for ascoviruses. Characteristics we observed, such as their presence or absence as well as their length, position and polymorphism, especially in the SfAV1 and HvAV3 repeats, could be used for the rapid characterization of ascovirus isolates and determination of whether an isolate represents a new species or is affiliated to a known species. For example, analysis of the ascovirus δ DNA polymerase genes has shown that TnAV2 and HvAV3 are closely related and may possibly represent members of a large complex of variants of a single species. However, by using the HvAV3c 1·186 kbp Alul repeat as a probe, we were readily able to distinguish several HvAV3 variants from one another and from variants of TnAV2 (Fig. 5). Differences in length of the repeats in comparison with those reported previously (Bigot et al., 1997a) and internal polymorphisms ranging from 72 to 94% were also found in the DpAV4 genome (Fig. 6).

In the iridoviruses, terminal repeats at the 5′ and 3′ ends of each permuted genomic molecule have been described in LCDV, frog virus 3 and CIV, suggesting that this is a general feature of iridovirus genomes (Schnitzler et al., 1987a, b). However, we found no evidence of large interspersed repeats, such as those that occur in ascoviruses, in the invertebrate iridovirus genomes we examined. Aside from physical configuration, i.e. linear versus circular, this would appear to be another major difference between these two virus types.

For the third major characteristic we examined, cytosine methylation, we found significant variation within the ascoviruses and iridoviruses. In the invertebrate iridoviruses we studied, and most of the ascoviruses, we found only a low level of cytosine methylation. However, a very high level of methylation (76%) was detected in DpAV4, in fact the highest for any known virus. From previous reports, it is known that certain vertebrate viruses have a high level of cytosine methylation, in the range of 20% (Willis & Granoff, 1980; Wagner et al., 1985). In eukaryotic species in which
chromosomal DNA is methylated, viruses attacking them are known to have their genomes methylated (Karlin & Burge, 1995; Regev et al., 1998). The cytosine methylation in these viruses is due to the presence of a virus-encoded enzyme (Kaur et al., 1995) and it is therefore not surprising to find 5-methyldeoxycytidine in the genomes of these viruses. The function of genome methylation in these viruses is not certain but it is proposed to play a role in the regulation of gene expression (Goorha et al., 1984). Thus, it is possible that genome methylation is involved in regulating ascovirus gene expression, especially in DpAV4, which has a high level of methylation and varies in its pattern of replication depending on the host in which replication occurs.

In vertebrate genomes, the spontaneous deamination of 5-methyldeoxycytidine induces more frequent G–T transitions than other point mutations. As 95% of methylation occurs on CpG, this results in vertebrate genomes in an observed to expected CpG frequency ratio with a mean value of 0.2–0.5 (Jabbari et al., 1997). This phenomenon, the so-called CpG shortage, is also observed in most vertebrate viruses (Karlin & Burge, 1995). Dinucleotide data available from analysis of DpAV4 DNA (more than 35 kbp) does not reveal any CpG shortage. However, it does reveal a significant CpC and CpT shortage in eight sequences analysed (ranging from 1:5 to 12:2 kbp; accession nos AJ279812–AJ279815). Moreover, similar results were obtained with four SFAV1a sequences (3:1 and 8:9 kbp; accession nos AJ279828 and AJ279830; and two unregistered sequences of 1:1 and 1:2 kbp). Interestingly, although no CpG and CpC shortage is found in the available sequences of invertebrate iridoviruses like CIV, IV31 and TMIV, a weak but significant CpT shortage is found (accession nos L22300, M81388, AF003534 and AF83915 for CIV; unregistered sequences for IV31 and TMIV). Overall, these data suggest that cytosine methylation occurs in all ascovirus and iridovirus genomes. However, it appears that the specificity of dinucleotide methylation varies among different groups within a family. Thus, it appears that CpG is the most methylated dinucleotide in the vertebrate iridoviruses, whereas it is CpT in the invertebrate iridoviruses and CpC and the CpT in the ascoviruses.

Our purpose in undertaking this study was to characterize some of the key physical and biochemical properties of invertebrate ascovirus and iridovirus genomes in order to provide a firmer foundation for comparisons of their genomes. The results of the present study show that, for the properties we examined, these viruses are very different, whereas the accompanying data on the phylogenetic relatedness of several of the genes they have in common suggest that they are closely related (Stasiak et al., 2000). If these two virus types are closely related, this apparent paradox may be due to host shifts during evolution, after which there was rapid evolution that facilitated adaptation of the viruses to the biologies of their new hosts. Some evidence in support of this hypothesis comes from our accompanying study of ascovirus phylogenetics (Stasiak et al., 2000). In this study, based on an analysis of δ DNA polymerase genes, we showed that the family Ascoviridae contains two different types, with one species, DpAV4, being clearly distinguished from the other three, SFAV1, TNV2 and HvAV3. This clear phylogenetic separation of these ascoviruses based on the δ DNA polymerase gene corresponded with their biologies, an important trait being that the DpAV4 genome is capable of existing as an unencapsidated molecule in nuclei of its parasitoid vector, whereas this trait is not known for the other ascoviruses. Thus, further studies of these viruses, including studies based on the phylogenetic relatedness of structural protein genes, are now of even greater interest due to the results obtained in the present study. These studies have the potential to resolve the relatedness of these virus families and to contribute to our knowledge of the evolution of large dsDNA viruses.

We thank D. K. Bideshi and J. J. Johnson for their assistance throughout our investigations. This work was supported by grants from the CNRS (UPR 6035) and NATO to Y. Bigot, US National Science Foundation grant INT-9726818 to B. Federici and the Ministère de l’Education Nationale, de la Recherche et de la Technologie.

References


Y. Bigot and others


Received 31 March 2000; Accepted 16 August 2000