Available online at www.sciencedirect.com
ScienceDirect
Virology 354 (2006) 167–177
VIROLOGY
www.elsevier.com/locate/yviro

Sequence and organization of the *Trichoplusia ni* ascovirus 2c (*Ascoviridae*) genome☆

Lihua Wang [a], Jianli Xue [a], Craig P. Seaborn [a], Basil M. Arif [b], Xiao-Wen Cheng [a,*]

[a] *Department of Microbiology, 32 Pearson Hall, Miami University, Oxford, OH 45056, USA*
[b] *Laboratory for Molecular Virology, Great Lakes Forestry Center, 1219 Queen St. E., Sault Ste. Marie, Ontario, Canada P6A 2E5*

Received 4 May 2006; returned to author for revision 19 June 2006; accepted 22 June 2006
Available online 31 July 2006

Abstract

The complete *Trichoplusia ni* ascovirus 2c (TnAV-2c) genome sequence was determined. The circular genome contains 174,059 bp with 165 open reading frames (ORFs) of greater than 180 bp and two major homologous regions (*hrs*). The genome is quite A+T rich at 64.6%. Fifty-four ORFs had homologues in other insect viruses, such as ascoviruses, iridoviruses, baculoviruses and entomopoxviruses; 30 ORFs showed low identities with those from different parasitic protozoa and 12 ORFs were unique to TnAV-2c. TnAV-2c has 15 ORFs that could be grouped into six gene families. Three major conserved repeating sequences were identified and were interspersed in two regions. BLAST analyses revealed that there were 16 enzymes involved in gene transcription, DNA replication, and nucleotide metabolism. TnAV-2c has 12 and 25 ORFs sharing high identities with ascovirus and iridovirus homologues, respectively. The codon usage bias appears to be more similar to *Spodoptera frugiperda* ascovirus 1a than to iridoviruses.

Keywords: Ascovirus; *Trichoplusia ni* ascovirus 2c; Viral genome sequence; Viral genome organization; Repetitive sequences; Codon usage bias

Introduction

Ascoviruses (*Ascoviridae*) are a recently described group that contains double-stranded DNA genome of 115–180 kb (Federici et al., 1990b; Bigot et al., 1997a; Cheng et al., 1999). They were first isolated in the late 70s and early 80s. Their delayed discovery is attributed to the fact that the symptoms of infection by ascoviruses are not very pronounced in the field (Browning et al., 1982; Hamm et al., 1985, 1986; Federici et al., 1990b). In agricultural fields, ascoviruses have been isolated from economically important lepidopteran species such as *H. zea*, *Heliothis virescens*, *Spodoptera frugiperda* and *S. exigua* (Carner and Hudson, 1983; Hamm et al., 1985, 1986; Cheng et al., 2005). In addition to lepidopteran ascoviruses, a hymenopteran ascovirus was isolated from *Diadromus pulchellus* (Yponomeutidae) (Bigot et al., 1997b).

Isolates of ascoviruses were previously named following the traditional scheme of using the host's name. For example, ascovirus isolated from *S. frugiperda* was named SAV (Federici et al., 1990b). However, this often caused problems in the classification of ascoviruses, since most had a broad host range. Based on multiple criteria, such as natural and laboratory hosts, tissue tropism, virion morphology, restriction fragment length polymorphism (RFLP) and gene sequence analyses, four ascovirus species have been recognized by the International Committee on Taxonomy of Viruses (ICTV). These include *Spodoptera frugiperda ascovirus 1* (SfAV-1), *Trichoplusia ni ascovirus 2* (TnAV-2), *Heliothis virescens ascovirus 3* (HvAV-3) and *Diadromus pulchellus ascovirus 4* (DpAV-4) (Federici et al., 2000).

Ascovirus infection in permissive cells first causes invagination of the nuclear envelope and later dismantles the nuclear envelope and the infected cell becomes nucleus-free (Federici, 1983; Federici et al., 1991; Cheng et al., 2000). One of the unique features of infected larvae is the presence in the hemolymph of vesicles packed with virions. The vesicles are formed by the partitioning of the host cells in which the

☆ The sequences reported in this paper has been deposited in GenBank/NCBI with the accession number DQ517337.

* Corresponding author. Fax: +1 513 529 2431.
E-mail address: Chengx@muohio.edu (X.-W. Cheng).

0042-6822/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2006.06.029

Table 1
Predicted ORFs in TnAV-2c genome

TnAV-2c ORF	Position	aa[a]	BLAST best match [NCBI_GI[b]] score (bits)	Id%[c]	Comments
1	1>3078	1026	*Heliotis virescens ascovirus 3c* [21668320] 875	43	Delta DNA polymerase
2	3164>5596	810	Giardia lamblia ATCC 50803 [29250753] 59.7	29	Variant-specific surface protein
3	5637>5930	97	*Oryza sativa* [14587218] 47.8	39	Ring finger 1-like protein
4	5986>6618	210	*Entamoeba histolytica* [56465742] 43.5	21	Kinesin family protein
5	6661>7059	132	*Drosophila pseudoobscura* [54645260] 35.8	38	GA16575-PA
6	7190>8986	598	*Spodoptera exigua* NPV [9634305] 57.8	26	SeNPV ORF84
7	9325>11,643	772	*Spodoptera frugiperda ascovirus 1* [21668315] 119	28	Hypothetical protein
8	11,787>13,823	678	*Spodoptera frugiperda ascovirus 1* [21668316] 122	37	Ribonuclease III
9	13,825>14,673	282	Leishmania major [68124512] 59.7	58	Hypothetical protein
10	14,982<15,389	135	*Plasmodium yoelii yoelii* [83315811] 38.5	28	Hypothetical protein
11	15,458<17,692	744	*Plasmodium falciparum* 3D7 [4493986] 38.5	20	Hypothetical protein
12	17,762>19,084	440	*Plasmodium falciparum* 3D7 [23497387] 43.5	34	Hypothetical protein
13	19,088>19,426	112	*Marinobacter aquaeolei* VT8 [77953264] 32.0	37	NADH dehydrogenase I chain L
14	19,465>19,986	173	*Dictyostelium discoideum* [66803422] 89.4	34	Hypothetical protein
15	19,986>20,219	77			
16	20,292>20,486	64			
17	20,569>21,705	378	Invertebrate iridescent virus 6 [15078994] 173	31	282R
18	22,055>22,828	257	*Magnaporthe grisea* 70–15 [39958547] 34.7	27	Hypothetical protein
19	23,126>23,398	90	*Plasmodium chabaudi* [56518872] 31.6	37	Hypothetical protein
20	23,476<24,345	289	*Plasmodium yoelii yoelii* [82596836] 40.4	20	Hypothetical protein PY05890
21	24,536<25,762	408	*Amsacta moorei entomopoxvirus* (AmEPV) [9964423] 197	31	AMV109 hypothetical protein
22	25,982>26,437	151	*Kluyveromyces lactis* [39588930] 37.4	25	Hypothetical protein
23	26,573<27,889	438	*Agrotis segetum* GV [46309331] 75.1	25	ORF123, hypothetical protein
24	28,365<29,282	305	*Nitrosomonas europaea* [30180138] 38.1	28	Polymerase sigma subunits
25	29,347>29,832	161	*Caenorhabditis briggsae* [39598150] 35.4	24	Hypothetical protein CBG14804
26	29,936<30,976	346	*Kluyveromyces lactis* [50305125] 35.4	26	Hypothetical protein
27	31,317<32,093	258	*Helicoverpa armigera* NPV [18138388] 153	36	Bro-B
28	32,598<32,783	61			
29	33,505>34,173	222	*Clostridium bifermentans* [2292820] 57.4	34	Aegerolysin, hemolysin-like protein
30	34,211<34,747	178	*Mamestra configurata* NPV A [33331788] 84.3	36	SprT, XE-2
31	34,849<35,358	169	*Plasmodium yoelii yoelii* [83315596] 41.2	24	Hypothetical protein PY02913
32	35,847<37,310	487	*Homo sapiens* [33946285] 58.2	36	RING domain, baculoviral IAP repeat-containing 3
33	37,328<37,672	114	*Burkholderia thailandensis* E264 [83720038] 42.4	34	Metallo-beta-lactamase family protein
34	37,717>38,115	132	*Rattus norvegicus* [8393807] 43.5	24	Myosin heavy chain, polypeptide 7
35	38,232>38,864	210	*Plasmodium chabaudi* [50919195] 35.8	29	Conserved hypothetical protein
36	38,916<39,548	210	*Giardia lamblia* ATCC [71073432] 50.1	41	RING, hypothetical protein
37	39,520<40,044	174	*Oceanicaulis alexandrii* HTCC2633 [83945421] 33.1	57	Hypothetical protein
38	40,247<41,551	434	*Plasmodium falciparum* 3D7 [23509023] 38.5	21	Hypothetical protein
39	41,622>42,236	203	*Plasmodium berghei* strain ANKA [68075919] 37.0	32	Hypothetical protein
40	42,276>43,364	362	*Plasmodium falciparum* 3D7 [4493896] 47.8	24	Hypothetical protein, conserved
41	43,437<43,700	87	*Trypanosoma brucei* [25992512] 32.0	33	Histone acetyltransferase
42	43,881>46,382	833	Grouper iridovirus [56418272] 319	29	DNA-dependent RNA polymerase, 71L
43	46,379>49,312	977	Invertebrate iridescent virus 6 [15079007] 117	20	295L
44	49,497>49,754	85	*Drosophila melanogaster* [85725256] 33.9	31	CG34027-PA
45	49,957<50,196	79	*Euglena gracilis* [415790] 34.3	40	30S ribosomal protein S4
46	50,397<50,942	181	*Schistosoma japonicum* [56755413] 79.0	36	ELO, SJCHGC06698 protein
47	51,210>51,431	73			
48	51,415>51,780	121	*Gallus gallus* [50731773] 32.3	25	Hypothetical protein
49	51,797>52,231	144	*Haemophilus influenzae* [2909665] 33.9	30	Putative haemocin processing protein
50	52,203>52,484	93			
51	52,696>52,941	81	Invertebrate iridescent virus 6 [15079165] 40.0	25	454R
52	53,073>53,804	243	*Magnetococcus* [68246005] 36.6	31	ATP-binding region
53	53,860<54,510	216	*Dictyostelium discoideum* [66827053] 32.7	27	Hypothetical protein
54	54,620>55,234	204	*Plasmodium falciparum* 3D7 [23615199] 44.3	26	Hypothetical protein
55	55,506>55,745	79			
56	56,065>56,514	149	*Giardia lamblia* [71073373] 57.4	45	Hypothetical protein
57	56,693<57,247	184	*Cryptococcus neoformans* [50254545] 38.5	28	Hypothetical protein
58	57,313>60,252	979	Invertebrate iridescent virus 6 [15078891] 177	30	CAP10, 179R
59	60,322<60,666	114	Invertebrate iridescent virus 6 [15079112] 53.1	35	HMG-box, 401R
60	60,787>61,434	215	*Streptomyces coelicolor* A3 [10241793] 34.7	28	Putative ABC transport system integral membrane protein
61	61,549>62,184	211	*Plasmodium berghei strain* ANKA [68074813] 34.3	24	Hypothetical protein