JVI Journals.ASM.org

The Genome of the Nucleopolyhedrosis-Causing Virus from *Tipula oleracea* Sheds New Light on the *Nudiviridae* Family

Annie Bézier,[a] Julien Thézé,[a]* Frederick Gavory,[b] Julien Gaillard,[c] Julie Poulain,[b] Jean-Michel Drezen,[a] Elisabeth A. Herniou[a]

Institut de Recherche sur la Biologie de l'Insecte (IRBI), CNRS UMR 7261, Université François Rabelais, Tours, France[a]; Commissariat à l'Energie Atomique, Génoscope (Centre National de Séquençage, CEA), Evry, France[b]; Laboratoire de Biologie Cellulaire, Microscopie Electronique, Faculté de Médecine, Université François Rabelais, Tours, France[c]

ABSTRACT

A large double-stranded DNA (dsDNA) virus that produces occlusion bodies, typical of baculoviruses, has been described to infect crane fly larvae of the genus *Tipula* (Diptera, Tipulidae). Because of a lack of genomic data, this virus has remained unclassified. Electron microscopy of an archival virus isolated from *Tipula oleracea*, *T. oleracea* nudivirus (ToNV), showed irregularly shaped occlusion bodies measuring from 2 to 5 μm in length and 2 μm in middiameter, filled with rod-shape virions containing single nucleocapsids within a bilayer envelope. Whole-genome amplification and Roche 454 sequencing revealed a complete circular genome sequence of 145.7 kb, containing five direct repeat regions. We predicted 131 open reading frames, including a homolog of the polyhedrin gene encoding the major occlusion body protein of *T. paludosa* nucleopolyhedrovirus (NPV). BLAST searches demonstrated that ToNV had 21 of the 37 baculovirus core genes but shared 52 genes with nudiviruses (NVs). Phylogenomic analyses indicated that ToNV clearly belongs to the *Nudiviridae* family but should probably be assigned to a new genus. Among nudiviruses, ToNV was most closely related to the *Penaeus monodon* NV and *Heliothis zea* NV clade but distantly related to *Drosophila innubia* NV, the other nudivirus infecting a Diptera. Lastly, ToNV was found to be most closely related to the nuvidirus ancestor of bracoviruses. This was also reflected in terms of gene content, as ToNV was the only known exogenous virus harboring homologs of the *Cc50C22.6* and *27b* (*Cc50C22.7*) genes found in the nudiviral genomic cluster involved in bracovirus particle production.

IMPORTANCE

The *Nudiviridae* is a family of arthropod dsDNA viruses from which striking cases of endogenization have been reported (i.e., symbiotic bracoviruses deriving from a nudivirus and the endogenous nudivirus of the brown planthopper). Although related to baculoviruses, relatively little is known about the genomic diversity of exogenous nudiviruses. Here, we characterized, morphologically and genetically, an archival sample of the *Tipula oleracea* nudivirus (ToNV), which has the particularity of forming occlusion bodies. Comparative genomic and phylogenomic analyses showed ToNV to be to date the closest known relative of the exogenous ancestor of bracoviruses and that ToNV should be assigned to a new genus. Moreover, we revised the homology relationships of nudiviral genes and identified a new set of 32 core genes for the *Nudiviridae*, of which 21 were also baculovirus core genes. These findings provide important insights into the evolutionary history of large arthropod dsDNA viruses.

Crane fly larvae, such as *Tipula oleracea* Linnaeus (cabbage crane fly) or *Tipula paludosa* Meigen (European crane fly), can cause significant agricultural and horticultural damage (1). The need to control this insect of the order Diptera led to the identification of many infectious agents (1, 2), including viruses inducing polyhedrosis disease. Nuclear polyhedrosis is an arthropod-specific pathology typically associated with the *Baculoviridae*. Polyhedral occlusion bodies, containing rod-shaped viruses enclosing a large circular supercoiled double-stranded DNA (dsDNA) genome, are produced in the host nuclei. Nuclear polyhedrosis has been reported in a number of families in the order Diptera, including Culicidae (3–7), Chironomidae, Tipulidae, Sciaridae (8, 9), and possibly Phlebotomidae (10). With the exception of mosquito baculoviruses, this disease remains poorly documented in Diptera. *Culex nigripalpus Nucleopolyhedrovirus* (CuniNPV), infecting mosquito, has been defined as the *Deltabaculovirus* type species (11). The genome of CuniNPV (12) presents such sequence divergences with the other baculoviruses that until recently, it was difficult to determine the homology of all 37 baculovirus core genes (4, 13, 14).

Following Rennie's brief description (15), Smith and collaborators, in the 1950s, characterized morphologically the polyhedrosis from *T. paludosa* larvae (TpNPV) as being caused by a rod-shaped occluded virus replicating in the nuclei of blood cells (16–19). Later, Bergoin and collaborators further described TpNPV (20–22). The virus host range appears to be narrow, restricted to *T. paludosa* and possibly *T. oleracea*. Infected larvae show a marked pallor and delayed larval development before death (19).

Received 2 October 2014 **Accepted** 19 December 2014

Accepted manuscript posted online 24 December 2014

Citation Bézier A, Thézé J, Gavory F, Gaillard J, Poulain J, Drezen J-M, Herniou EA. 2015. The genome of the nucleopolyhedrosis-causing virus from *Tipula oleracea* sheds new light on the *Nudiviridae* family. J Virol 89:3008–3025. doi:10.1128/JVI.02884-14.

Editor: G. McFadden

Address correspondence to Annie Bézier, annie.bezier@univ-tours.fr.

* Present address: Julien Thézé, Laboratoire Ecologie et Biologie des Interactions (EBI), UMR CNRS 7267, Université de Poitiers, Poitiers, France.

doi:10.1128/JVI.02884-14

TABLE 1 Baculovirus, nudivirus, and hytrosavirus genome general features and whole-genome comparison with ToNV[a]

Virus name	Abbreviation	Accession no.	Genome size (bp)	No. of genes	AT%	CD%	Reference or source	Maximum score	Total score	% Query coverage	E value
Autographa californica MNPV	AcMNPV	NC_001623	133,894	156	59.3	97.2	85	29.9	789	1	0.022
Lymantria dispar MNPV	LdMNPV	NC_001973	161,046	164	42.5	87.5	86	77.6	1,307	2	2e−18
Cydia pomonella GV	CpGV	NC_002816	123,500	143	54.7	90.1	87	88.6	2,381	3	1e−30
Neodiprion sertifer NPV	NeseNPV	NC_005905	86,462	90	66.2	84.5	88	83.1	2,681	6	1e−30
Culex nigripalpus NPV	CuniNPV	NC_003084	108,252	109	49.1	91.2	12	74.4	739	1	1e−13
Heliothis zea NV-1	HzNV-1	AF451898	228,089	155	58.2	69.4	58	140	10,140	17	1e−107
Helicoverpa zea NV-2	HzNV-2	NC_004156	231,621	113	58.1	68.2	62	140	10,223	17	3e−104
Gryllus bimaculatus NV	GbNV	EF203088	96,944	98	72	93.6	63	126	12,781	14	2e−94
Oryctes rhinoceros NV	OrNV	EU747721	127,615	139	58.4	88.5	59	141	5,295	9	7e−71
Penaeus monodon NV	PmNV	KJ184318	119,638	115	65.5	95.6	64	220	10,757	18	2e−69
Glossina pallidipes SGHV	GpSGHV	NC_010356	190,032	160	72	86.5	32	42.8	3,187	6	8e−04
Musa domestica SGHV	MdSGHV	EU522111	124,279	108	56.5	90.9	33	62.9	1,131	2	5e−10
Tipula oleracea NV	ToNV	KM610234	145,704	131	74.5	85.7	This work				

[a] Maximum score, total score, query coverage, and E value data are those obtained by TBLASTX against genome sequences of the baculovirus, nudivirus, and hytrosavirus representative set presented in the table. MNPV, multiple nucleopolyhedrovirus; NPV, nucleopolyhedrovirus; GV, granulovirus; NV, nudivirus; SGHV, salivary gland hypertrophy virus; CD, coding density.

TpNPV polyhedra display unusual morphological, biological, and biochemical features compared to those of baculoviruses infecting Lepidoptera and Hymenoptera (19–21). The major occlusion body protein (MOBP) forming TpNPV polyhedra is 25.2 kDa in size (23, 24) and appears to be very distant from baculovirus polyhedrin/granulin based on their respective N-terminal amino acid sequences (24, 25).

Recently, viruses capable of infecting dipteran hosts have been characterized and classified into the *Hytrosaviridae* and *Nudiviridae* families (26–29), representing two new families related to the *Baculoviridae* (29–31). Hytrosaviruses cause salivary gland hypertrophy syndrome (SGH) and replicate within the nucleus of salivary gland cells of their dipteran hosts (26, 31). The genomes of two representative viruses, *Glossina pallidipes* and *Musca domestica* SGH viruses (GpSGHV and MdSGHV), were sequenced in 2008 (32, 33). Formerly described as nonoccluded baculoviruses (34), nudiviruses display wider host range than baculoviruses and hytrosaviruses, as they infect more insect orders as well as crustaceans (35, 36). Nudiviruses can infect all developmental stages and have varied tissue tropism. Depending on the host, the infection can be asymptomatic in both larvae and adults, lethal in larvae, and chronic in adults and can also induce malformations and sterility (35). These three viral families, which evolved over 310 million years ago (MYA) (30), share a number of structural and genomic features while retaining lineage specific characteristics (26, 28). In particular, nudiviruses can durably integrate into the genome of their hosts, as has occurred in braconid wasps ~103 MYA, leading to bracovirus symbiosis (37, 38), a phenomenon that has also occurred independently in the brown planthopper (39).

Given this recent enrichment in insect virus systematics, our aim was to characterize an archival *T. oleracea* virus (ToNV) preserved in the 1950s by Kenneth M. Smith as nuclear polyhedrosis from the cabbage crane fly. Although resembling TpNPV, it was unclear to which viral family ToNV might belong. Electron microscopy was performed on ToNV occlusion bodies (OBs) for morphological and structural characterization. After whole-genome amplification, the ToNV complete genome sequence was determined by next-generation sequencing (NGS) technology, which allowed the identification of the gene encoding the major occlusion body protein. Phylogenomic analyses on shared genes were used to resolve the relationships between invertebrate large dsDNA viruses and clarify the taxonomic affiliation of ToNV.

MATERIALS AND METHODS

Viral sample. The occluded ToNV stock solution was obtained from the historical insect virus collection held at the Natural Environment Research Council, Centre for Ecology and Hydrology (NERC-CEH, Wallingford, England). Kenneth M. Smith had deposited the virus referenced as "sample 35—*T. oleracea* NPV (Diptera)."

Electron microscopy. A 30-μl aliquot of the archival sample 35 (approximately 2.6×10^7 OBs) was fixed by incubation for 24 h in 4% paraformaldehyde and 1% glutaraldehyde in 0.1 M phosphate buffer (pH 7.2), washed in 1× phosphate-buffered saline (PBS), postfixed by incubation for 1 h with 2% osmium tetroxide, and dehydrated in a graded series of ethanol solutions. The scanning electron microscopy aliquot was then dried in hexamethyldisilazane. Dry pellets were sprinkled on adhesive carbon discs and sputter coated with platinum. Fixed viruses were examined with a Zeiss Ultra plus scanning electron microscope. For transmission electron microscopy, virus pellets were embedded in Epon resin (Sigma), which was allowed to polymerize for 48 h at 60°C. Ultrathin sections were cut, deposited on electron microscopy grids, and stained with 5% uranyl acetate and 5% lead citrate for examination under a JEOL 1011 transmission electron microscope.

Genome sequencing and assembly. DNA was purified from a small aliquot of the archival sample 35 using the DNeasy kit (Qiagen). To increase the amount of viral DNA for NGS sequencing, 30 individual reactions of whole-genome amplification were performed overnight from 3 ng of starting DNA material according to the Illustra TempliPhi Amplification kit (GE Healthcare Life Sciences) instructions. Reaction mixtures were then pooled and cleaned with the QIAamp DNA minikit (Qiagen). The DNA library was made from 5 μg amplified viral DNA using the GS FLX Titanium Rapid Library Preparation kit and then sequenced on a 454 GS-FLX Titanium platform (Roche Diagnostic). Over 52,000 single-paired reads were produced. The sequences were cleaned for quality and adaptors with the 454 v2.3 analysis package (Roche Diagnostic) prior to *de novo* assembly with Newbler v2.6 (40). Overlapping contigs were assembled by using Geneious 6 (41). Genome finishing was performed to resolve sequence ambiguities, like homopolymers and frameshifts, on the bases of PCR and Sanger sequencing (BigDye Terminator kit and ABI Prism 3100-Avant Genetic Analyzer, Life Technologies).

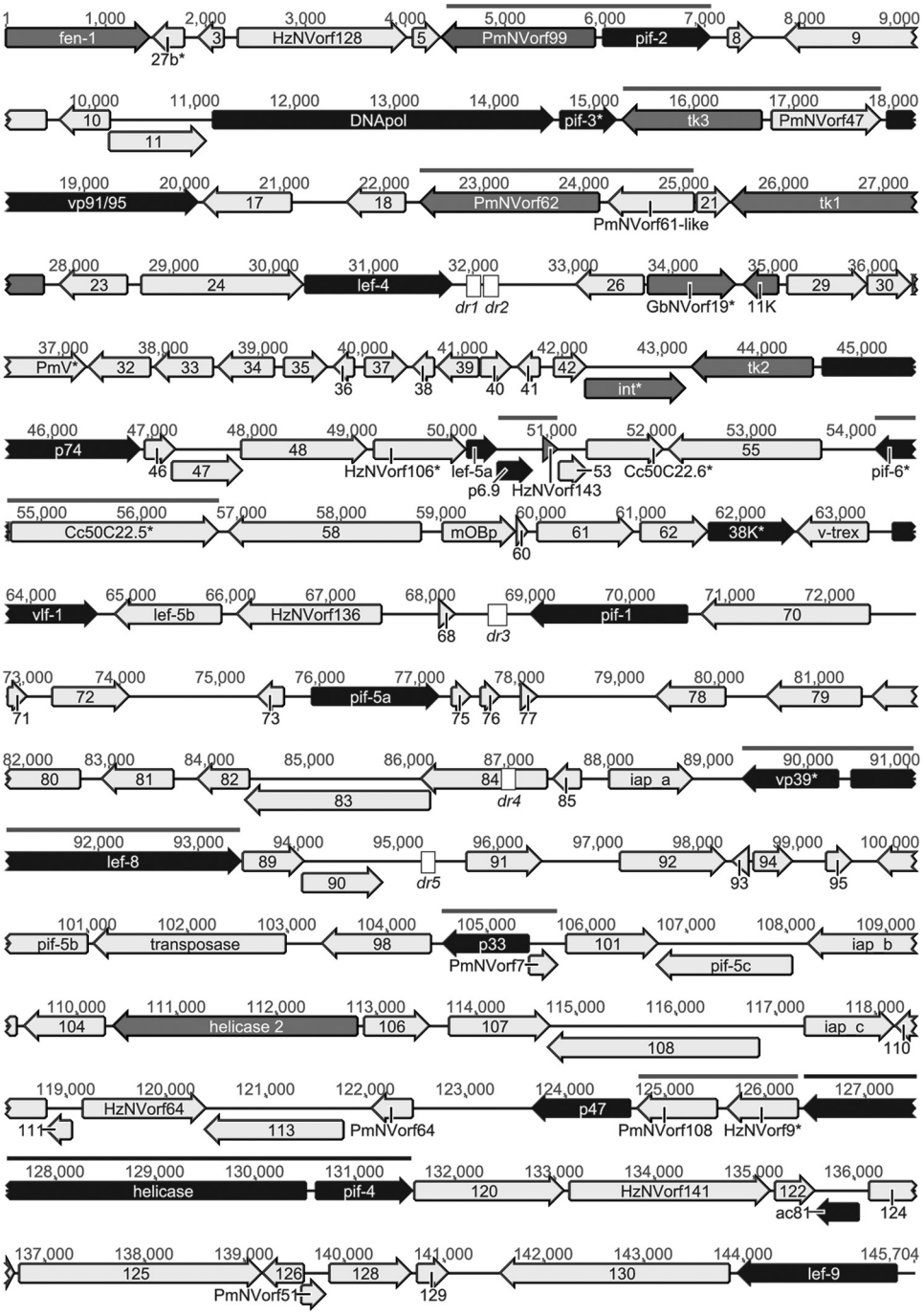


FIG 1 Linear map of the ToNV genome. Genes and their transcriptional directions are indicated as arrows. Black arrows, baculovirus and nudivirus shared core genes; gray arrows, nudivirus core genes. When a core gene belonged to a gene family (i.e., *lef-5*), only one gene was displayed in accentuated color. Thick black line, gene cluster conserved between baculoviruses and nudiviruses; thick gray lines, gene clusters conserved among several nudiviruses. Conserved genes were named based on their homologs in other virus genomes, but to simplify annotation, all "like" gene name extensions have been removed. ToNV ORFs with unknown function have been designated by their position in the genome (Table 2). *, genes belonging to the nudiviral cluster identified within *C. congregata* or *M. demolitor* wasp genomes (37, 38, 67). Direct repeat (*dr*) region position and name are indicated as boxes (see also Fig. 5 and Tables 3 and 4 for details).

Sequence analyses and genome annotation. Whole-genome similarities were estimated using TBLASTX (42, 43) against genome sequences of representative baculoviruses, nudiviruses, and hytrosaviruses (Table 1).

Open reading frames (ORFs) encoding proteins over 50 amino acids were predicted *de novo* using ORF Finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html) combined with FGENESV0 software from the SoftBerry platform. ORFs smaller than 50 amino acids were also taken into account when displaying clear similarity with other known viral gene products. When several ORFs could be predicted within the same region only the largest nonoverlapping ORFs were kept.

ORF similarities were identified using BLASTP, BLASTX, and/or TBLASTN (42, 43) against the NCBI nonredundant protein database (nr,

TABLE 2 ToNV putative ORFs and deduced protein features[a]

					Baculovirus, Nudivirus or Bracovirus best homolog						Homolog in Nudiviruses					
ORF	Gene name	Function/activity (by similarity)	Promoter	Size (aa)	Homolog	Source	Size (aa)	GI no.	E value	Feature	Hz NV-2	Hz NV-1	Pm NV	Gb NV	Or NV	Di NV
1	*fen-1*	DNA processing	TATA	479	ORF68	HzNV-1	652	22671535	2e-15	PIN domain-like	70	68	20	65	16	+
2	*27b-like**	particle component	E1	106	50C22.7	*C. congregata*	238	223587753	1e-004§$							
3	*ToNVorf003*	unknown	-	84												
4	*HzNVorf128-like*	unknown	L	561	HzNVORF128-like	*M. demolitor*	490	665800138	2e-008§		21	128	42			
5	*ToNVorf005*	unknown	L	90												
6	*PmNVorf99-like*	unknown	TATA	509	ORF99	PmNV	414	669192329	1e-006§		27	122	99	67	18	+
7	*pif-2*	ODV component	L	362	ORF15	PmNV	430	669192245	4e-60	SP; TM; Baculo_44	26	123	15	66	17	+
8	*ToNVorf008*	unknown	E1	83												
9	*ToNVorf009*	unknown	L	570												
10	*ToNVorf010*	unknown	TATA	167												
11	*ToNVorf011*	unknown	TATA	322						SP; TM						
12	*DNApol*	DNA processing	E1	1142	ORF18	HzNV-2	1136	370702964	4e-118	RNaseH-like; DNA/RNA pol	18	131	5	12	1	
13	*pif-3**	ODV component	L	193	ORF107	OrNV	204	213159375	8e-31	SP; TM; DUF666	53	88	93	3	107	
14	*tk3*	nucleotid metabolism	TATA	465	ORF115	HzNV-1	424	22671581	8e-27	P-loop NTPase	32	115	46	44	125	
15	*PmNVorf47-like*	unknown	E1	364	ORF47	PmNV	304	669192277	0.004§				47			
16	*vp91/95*	ODV component	L	742	ORF9	PmNV	675	669192239	1e-33	Chitin-bd_dom	89	46	9	2	106	+
17	*ToNVorf017*	unknown	L	294												
18	*ToNVorf018*	unknown	TATA	194												
19	*PmNVorf62-like*	unknown	L	597	ORF79	HzNV-2	145	365199622	4e-006§$	TM	79	+°	62	51	61	
20	*PmNVorf61-like*	unknown	E1	287	ORF61	PmNV	230	669192291	2e-06		49	93	61			
21	*ToNVorf021*	unknown	E1	110												
22	*tk1*	nucleotid metabolism	L	740	ORF65	PmNV	437	669192295	3e-24		85	51	65	17	137	
23	*ToNVorf023*	unknown	TATA	225												
24	*ToNVorf024*	unknown	TATA	538												
25	*lef-4*	transcription	L	494	LEF-4	*M. demolitor*	942	665800661	2e-40		43	98	91	96	42	
26	*ToNVorf026*	unknown	E1,L	226												
27	*GbNVorf19-like**	unknown	L	291	ORF19	GbNV	285	134303417	8e-20	SP; TM; a/b hydrolase; DUF2424	99	30	98	19	47	
28	*11K-like*	particle component	L	115	ORF41	OrNV	113	213159309	2e-005§	SP; TM	25	124	100	95	41	
29	*ToNVorf029*	unknown	L	265												
30	*ToNVorf030*	unknown	-	144												
31	*PmV-like**	particle component	TATA	272	PmV-like	*C. inanitus*	227	223587745	1e-007§	DUF3388	30	118	45			
32	*ToNVorf032*	unknown	TATA	205						SP; TM						
33	*ToNVorf033*	unknown	E1,L	193						SP; TM						
34	*ToNVorf034*	unknown	L	188						SP; TM						
35	*ToNVorf035*	unknown	E1	143												
36	*ToNVorf036*	unknown	E1,L	66												
37	*ToNVorf037*	unknown	-	140												
38	*ToNVorf038*	unknown	TATA	71												
39	*ToNVorf039*	unknown	TATA	137						Ds-RNA-bd						
40	*ToNVorf040*	unknown	E1,L	101												
41	*ToNVorf041*	unknown	TATA	74												
42	*ToNVorf042*	unknown	TATA	105												
43	*integrase**	DNA processing	TATA	331	INT-1	*M. demolitor*	328	665792519	8e-46	Phage_integrase; DNA_brk_join_enz	8	144	55	57	75	
44	*tk2*	nucleotid metabolism	TATA	404	ORF38	PmNV	314	669192268	6e-15	TM; P-loop NTPase	34	111	38	34	117	
45	*p74*	ODV component	E1,L	763	ORF72	PmNV	684	669192302	6e-81	TM; Baculo_p74_N; Baculo_p74	106	11	72	45	126	
46	*ToNVorf046*	unknown	TATA	99						TM						
47	*ToNVorf047*	unknown	TATA	233												
48	*ToNVorf048*	unknown	E1,L	416												
49	*HzNVorf106-like**	particle component	L	307	ORF37	HzNV-2	340	370702983	0.45		37	106	69			
50	*lef-5a*	transcription	TATA	103	ORF66	SfMNPV	276	125860188	8e-04	Baculo_LEF5_C; Zinc beta-ribbon	40	101	52	85	52	
51	*p6.9*	packaging/assembly	L	119	ORF73	GbNV	86	134303471	4e-013§	Arginine-rich and Serine-rich	+°	142	+°	73	+	
52	*HzNVorf143-like*	unknown	L	53	ORF53	PmNV	54	669192283	1e-04	SP	9	143	53	58	76	
53	*ToNVorf053*	unknown	L	93												
54	*Cc50C22.6-like**	unknown	TATA	254	Cc50C22.6-like	*M. demolitor*	228	607304520	6e-04							
55	*ToNVorf055*	unknown	L	509												
56	*pif-6**	ODV component	E1,L	143	ORF74	HzNV-1	154	22671541	4e-09	TM	64	74	88	55	72	
57	*Cc50C22.5-like**	unknown	-	690	Cc50C22.5-like	*M. demolitor*	764	607304496	3e-012§				87			
58	*ToNVorf058*	unknown	TATA	643												
59	*mOBp*	major OB protein	TATA	241	polH	TpNPV		#	6e-14§				1			
60	*ToNVorf060*	unknown	E1	36						SP; TM						
61	*ToNVorf061*	unknown	-	319						TM						
62	*ToNVorf062*	unknown	TATA	224						SP; TM						

(Continued on following page)

TABLE 2 (Continued)

ORF	Gene name	Function/activity (by similarity)	Promoter	Size (aa)	Baculovirus, Nudivirus or Bracovirus best homolog: Homolog	Source	Sizs (aa)	GI no.	E value	Feature	Homolog in Nudiviruses: Hz NV-2	Hz NV-1	Pm NV	Gb NV	Or NV	Di NV
63	*38K**	packaging/assembly	TATA	291	ORF108	HzNV-2	259	370703054	5e-18	viral_ppase	108	10	59	1	87	
64	*v-trex*	DNA processing	E1,L	235	ORF115	ApNPV	230	96979801	8e-27	RNaseH-like_dom; Exo_RNase_T						
65	*vlf-1*	packaging/assembly	TATA	385	VLF-1	*M. demolitor*	324	665796658	9e-27		28	121	56	80	30	
66	*lef-5b*	transcription	TATA	355	ORF66	sfMNPV		125860188	2.2							
67	*HzNVorf136-like*	unknown	E1	481	ORF14	HzNV-2		370702960	1e-006§		14	136				
68	*ToNVorf068*	unknown	L	52						SP; TM						
69	*pif-1*	ODV component	E1	527	ORF39	PmNV		669192269	6e-63	SP; TM; PIF	82	55	39	52	60	
70	*ToNVorf070*	unknown	E1,L	562												
71	*ToNVorf071*	unknown	E1	63						SP; TM						
72	*ToNVorf072*	unknown	E1,L	254												
73	*ToNVorf073*	unknown	L	87						SP						
74	*pif-5a*	ODV component	TATA	430	ORF76	HzNV-1		22671543	5e-38	TM; Baculo_E56	62	76	10	5	115	+
75	*ToNVorf075*	unknown	E1,L,HL	63												
76	*ToNVorf076*	unknown	TATA	62						TM						
77	*ToNVorf077*	unknown	E1,L	52												
78	*ToNVorf078*	unknown	L	229												
79	*ToNVorf079*	unknown	L	316						SP; TM						
80	*ToNVorf080*	unknown	TATA	388												
81	*ToNVorf081*	unknown	L	238												
82	*ToNVorf082*	unknown	TATA	172						SP; TM						
83	*ToNVorf083*	unknown	-	614												
84	*ToNVorf084*	unknown	TATA	421						SP; TM						
85	*ToNVorf085*	unknown	L	92												
86	*iap_a*	inhibition of apoptosis	L	277	ORF139	MacoNPV		20070018	0.001	ZF_RING_2; RING/U-box	12+15	135 +138	106	98	134	
87	*vp39**	packaging/assembly	E1	324	VP39	*M. demolitor*		665792481	6e-08	SP	52	89	22	64	15	+
88	*lef-8*	transcription	L	1001	LEF-8	*M. demolitor*		665800144	3e-139	LEF-8	51	90	23	49	64	
89	*ToNVorf089*	unknown	L	203						SP; TM						
90	*ToNVorf090*	unknown	L	265						TM						
91	*ToNVorf091*	unknown	E1,L	251						RNI-like						
92	*ToNVorf092*	unknown	TATA	353						F-box_dom						
93	*ToNVorf093*	unknown	L	54						TM						
94	*ToNVorf094*	unknown	E1,L	130												
95	*ToNVorf095*	unknown	E1	84												
96	*pif-5b*	ODV component	TATA	398	ORF10	PmNV		669192240	1e-38	TM; Baculo_E56	62	76	10	5	115	+
97	*transposase*	DNA processing	TATA	642	ORF39	PsunGV		285002314	3e-29	OrfB_Zn_ribbon						
98	*ToNVorf098*	unknown	L	362												
99	*p33*	packaging/assembly	E1,L	289	ORF13	HzNV-1		22671481	2e-08	TM; ERV_ALR	104	13	8	7	113	
100	*PmNVorf7-like*	unknown	TATA	92	ORF7	PmNV		669192237	4e-004§	SP; TM			7	59	79	
101	*ToNVorf101*	unknown	L	304												
102	*pif-5c*	ODV component	L	450	ORF76	HzNV-1		22671543	9e-08	TM	62	76	10	5	115	+
103	*iap_b*	inhibition of apoptosis	E1	395	ORF79	PiraGV		288804718	3e-17	IAP; BIR; ZF_RING_2	12+15	135 +138	106	98	134	
104	*ToNVorf104*	unknown	L	267						TM						
105	*helicase-2*	DNA processing	TATA	812	ORF79	PmNV		669192309	8e-27	P-loop NTPase	76	60	76 +79	46	108	
106	*ToNVorf106*	unknown	TATA	218												
107	*ToNVorf107*	unknown	L	336												
108	*ToNVorf108*	unknown	TATA	704						OTU						
109	*iap_c*	inhibition of apoptosis	E1,L	297	ORF118	ChocNPV		529153595	1e-45	IAP; BIR; ZF_RING_2	12+15	135 +138	106	98	134	
110	*ToNVorf110*	unknown	TATA	200												
111	*ToNVorf111*	unknown	L	85						SP; TM						
112	*HzNVorf64-like*	particle component	E1	408	ORF64	HzNV-1		22671531	2e-05		73	64	24			
113	*ToNVorf113*	unknown	E1	462						RNI-like						
114	*PmNVorf64-like*	unknown	TATA	139	ORF64	PmNV		160432011	5.00e-005§	SP; TM			64			
115	*p47*	transcription	L	333	p47	*C. congregata*		223587813	1e-35		63	75	14	69	20	+
116	*PmNVorf108-like*	unknown	TATA	265	ORF108	PmNV		669192338	1e-004§		101	28	108			
117	*HzNVorf9-like**	particle component	-	237	HzNVORF9-like1	*M. demolitor*		665792487	0.46		109	9	107			
118	***helicase***	**DNA processing**	**E1,L**	**1380**	**helicase**	***M. demolitor***		**665796869**	**1e-31**	**phage associated DNA primase**	**38**	**104**	**94**	**88**	**34**	+
119	***pif-4***	**ODV component**	**TATA**	**327**	**ORF96**	**PmNV**		**669192326**	**5e-27**	**SP; TM; Baculo_19**	**39**	**103**	**96**	**87**	**33**	
120	*ToNVorf120*	unknown	E1	497						Sec23_24_beta_S						
121	*HzNVorf141-like*	unknown	L	669	ORF141	HzNV-1		22671606	0.99$	DNA ligase/mRNA capping enzyme	10	141	48			
122	*ToNVorf122*	unknown	TATA	133												
123	*Ac81*	host interaction	TATA	151	ORF14	GbNV		134303412	1e-06	TM	96	33	86	14	4	
124	*ToNVorf124*	unknown	TATA	180						SP; TM						
125	*ToNVorf125*	unknown	E1,L	808												
126	*ToNVorf126*	unknown	TATA	136												
127	*PmNVorf51-like*	unknown	TATA	79	ORF51	PmNV		674547954	0.002§	SP; TM			51		63	
128	*ToNVorf128*	unknown	E1,HL	271												
129	*ToNVorf129*	unknown	E1,L	103												
130	*ToNVorf130*	unknown	TATA	765												
131	*lef-9*	transcription	E1	535	LEF-9	*M. demolitor*		665796783	1e-94	LEF-9	63	75	58	24	96	

(Continued on following page)

August 2014, at http://blast.ncbi.nlm.nih.gov/Blast.cgi) or against a local database using the BioEdit sequence alignment editor (44). This local database was composed of all the proteins of the previously determined representative baculovirus, nudiviruses, and hytrosaviruses (Table 1). Proteins from partially sequenced nudiviruses, i.e., *Penaeus vannamei* NV (PvNV, accession number DQ496179), *Macrobrachium rosenbergii* NV (MrNV, AFP33714), and *Drosophila innubia* NV (DiNV, JN680861 to JN680871), and the nudiviral proteins from *Cotesia congregata* bracovirus (BV) (CcBV, FM201559 to FM201576, FM877774, and FM212911 to FM212915), *Chelonus inanitus* BV (CiBV, FM201579 to FM201597, FN543427 to FN546858, and FN594617), and *Microplitis demolitor* BV (MdBV, JO913492 to JO979916 and JR139425 to JR139430) were also added. Sequences were further analyzed using probabilistic methods like HMMER (HMMER program using the nr sequence database and default parameters [http://hmmer.janelia.org/search/phmmer]) or HHpred (HHpred program using default parameters [http://toolkit.tuebingen.mpg.de/hhpred], pdb70_26Jul14 database). Finally, gene product sequence alignments were done using the MAFFT plugin in Geneious (best adapted algorithm strategy according to data size and BLOSUM-62 as score matrix) and manually refined if needed.

All ORFs were named according to their putative homologs when possible; otherwise, they were named based on their location in the genome. The final annotation was conducted using ARTEMIS software (45). Sequence coding density (CD) was measured as the ratio between the base number in the coding DNA sequences over the total base number. The presence of protein-specific domains and functional sites, as well as associated patterns and profiles, was defined using InterProScan database (http://www.ebi.ac.uk/Tools/pfa/iprscan/), Motif scan (http://myhits.isb-sib.ch/cgi-bin/motif_scan), and/or CD search (at the NCBI BLAST BLASTP platform).

Early and late baculovirus-like promoters were searched within 300 nucleotides upstream of each predicted ORF translation start codon (46, 47). Early promoters were defined as either a TATA box with the sequence TATA[AT][AT][AT] alone or associated 20 to 40 nucleotides downstream with either one of two initiator motifs, CA[TG]T (E1) or CGTGC (E2). Late promoters were defined as L ([ATG]TAAG) for the baculovirus late promoter and HL (TTATAGTAT) for the HzNV-1-specific late promoter (47, 48).

Repeat regions and imperfect palindromic motifs were searched with the etandem program (http://emboss.bioinformatics.nl/cgi-bin/emboss/etandem) and MEME program suite (49) with a cutoff score at 100 and a minimal size of 20 bp. Consensus motifs were visualized with WebLogo (50). Secondary structure predictions were obtained by using mfold software with default parameters (http://mfold.rna.albany.edu/?q=mfold/DNA-Folding-Form) (51).

Phylogenomic analysis. A phylogenomic approach was undertaken to position ToNV within the arthropod large dsDNA virus phylogeny (30). Amino acid multiple alignments were performed using the Clustal Omega program (52) on 37 nudivirus-related predicted homologs found in 16 other arthropod large dsDNA viruses, including the following: five nudiviruses, *Heliothis zea* NV-1 (HzNV-1), *Gryllus bimaculatus* NV (GbNV), *Oryctes rhinoceros* NV (OrNV), *Penaeus monodon* NV (PmNV), and the partially sequenced DiNV; five baculoviruses, *Autographa californica* MNPV (AcMNPV), *Lymantria dispar* MNPV (LdMNPV), *Cydia pomonella* GV (CpGV), *Neodiprion sertifer* NPV (NeseNPV), and CuniNPV; three bracoviruses (CcBV, CiBV, and MdBV); the endogenous nudivirus from *Nilaparvata lugens* (Hemiptera) (39); two hytrosaviruses (MdSGHV and GpSGHV); and one nimavirus, the *Penaeus monodon* white spot syndrome virus (WSSV-TH; AF369029), used as outgroup. As phylogenetic congruence tests did not show any conflicting phylogenetic signal between genes (data not shown), the 37-amino-acid (aa) multiple alignments were concatenated prior to phylogenetic reconstruction. A maximum likelihood (ML) phylogenetic analysis was done using RAxML (53) with the Whelan Goldman model of amino acid substitution and an estimated gamma distribution (WAG+G), selected using ModelGenerator (54) under the Akaike information criterion. Node supports in the ML tree were obtained from 100 bootstrap iterations.

ToNV oral infectivity. The infectivity of the archival sample 35 was checked by bioassay on *T. paludosa*. Adult females, identified by DNA barcoding (55, 56), were collected in Tours (France) and allowed to lay eggs individually. Larvae were reared on wet 1.5% agar medium at 25°C and 22°C under a 14-h light (8 μE) and 10-h dark photoperiod, respectively, and 60% relative humidity. They were fed with clean organic lettuce until they reached fourth instar. Then, crane fly larvae were individually infected using 10^5 or 10^6 OBs covering 0.5-mm-diameter lettuce leaf discs and monitored until death. To purify viral particles, insect cadavers were individually crushed in 0.5% SDS and then filtered through cheesecloth to remove large cellular debris. Successions of washes using 0.5% SDS, 0.1% SDS, and then water, associated with 13,500 × *g* centrifugation for 10 min, were applied to the sample until the OB pellet was sufficiently clean. Finally, the pellet was resuspended in 50 μl sterile water. DNA was purified with the QIAamp DNA minikit (Qiagen) from the collected *T. paludosa* adult females, to check their health status, and from a 5-μl aliquot of collected OBs according to the manufacturer's standard protocol for tissues, with the exception that the proteinase K digestion step was extended (90 min at 70°C) to ensure OB dissolution. *T. oleracea* nudivirus PCR diagnostic was then performed using *p74*-specific primers (p74-F, 5′-ACCACCCGTAGACGATAAAT-3′; p74-R, 5′-AACAGAACAGGCTAATGCTG-3′). Amplifications were performed on purified DNA using 0.5 pmol · μl^{-1} of each primer, 2.25 mM $MgCl_2$, 0.2 mM deoxynucleoside triphosphate (dNTP), and 1.5 unit Diamond *Taq* polymerase (Eurogentec) under a 35-cycle PCR program (95°C for 5 min; 35 cycles of 95°C for 60 s, 58°C for 60 s, 72°C for 90 s, and 72°C for 10 min).

Nucleotide sequence accession number. The ToNV genome sequence has been deposited under accession number KM610234 in the GenBank database (http://www.ncbi.nlm.nih.gov/GenBank/) (BioProject PRJNA261283).

RESULTS

General genome features. The assembly of ToNV sequence raw data produced a 145,704-bp-long contiguous circular genome with an ~127× average coverage. The ToNV genome size is in the range of other arthropod large dsDNA viruses (Table 1). ToNV

TABLE 2 (Continued)

[a] Protein functions were assigned based on protein similarity in baculovirus, nudivirus, bracovirus, and/or conserved domain identification. TATA, early promoter defined as either TATA[AT]T[AT] or TATA[AT]A[AT] alone; E1 and E2, TATA early promoter associated 20 to 40 nucleotides downstream with the initiator motifs CA[TG]T or CGTGC, respectively; L, baculovirus late promoters defined as [ATG]TAAG; HL, HzNV-1 late promoter defined as TTATAGTAT; GI no., GI accession number; "Feature" data include domains and functional sites as well as associated patterns and profiles defined using InterProScan database, Motif scan, or CD search; SP, signal peptide; TM, transmembrane regions; other domains are those defined by the previously determined tools. DiNV, *Drosophila innubia* NV; SfMNPV, *Spodoptera frugiperda* MNPV; ApNPV, *Antheraea pernyi* NPV; MacoNPV, *Mamestra configurata* NPV; PsunGV, *Pseudaletia unipuncta* granulovirus; PiraGV, *Pieris rapae* granulovirus; ChocNPV, *Choristoneura occidentalis* alphabaculovirus (see Table 1 for definitions of other virus name abbreviations). *, belongs to the nudiviral cluster characterized within *C. congregata* or *M. demolitor* wasp genomes (37, 38, 67); §, sequence similarity identified using local BLASTP or BLAST2 sequence comparison against a restricted protein database and global alignment for homolog confirmation (MAFFT under Geneious); $, sequence similarity identified using probabilistic methods (HMMER); +, gene present; °, gene not predicted in the genome of some nudiviruses but present; #, Rohrmann et al., 1981 (24). Boxes surrounding multiple ORF numbers indicate regions of microsynteny between ToNV and the other nudivirus genomes.

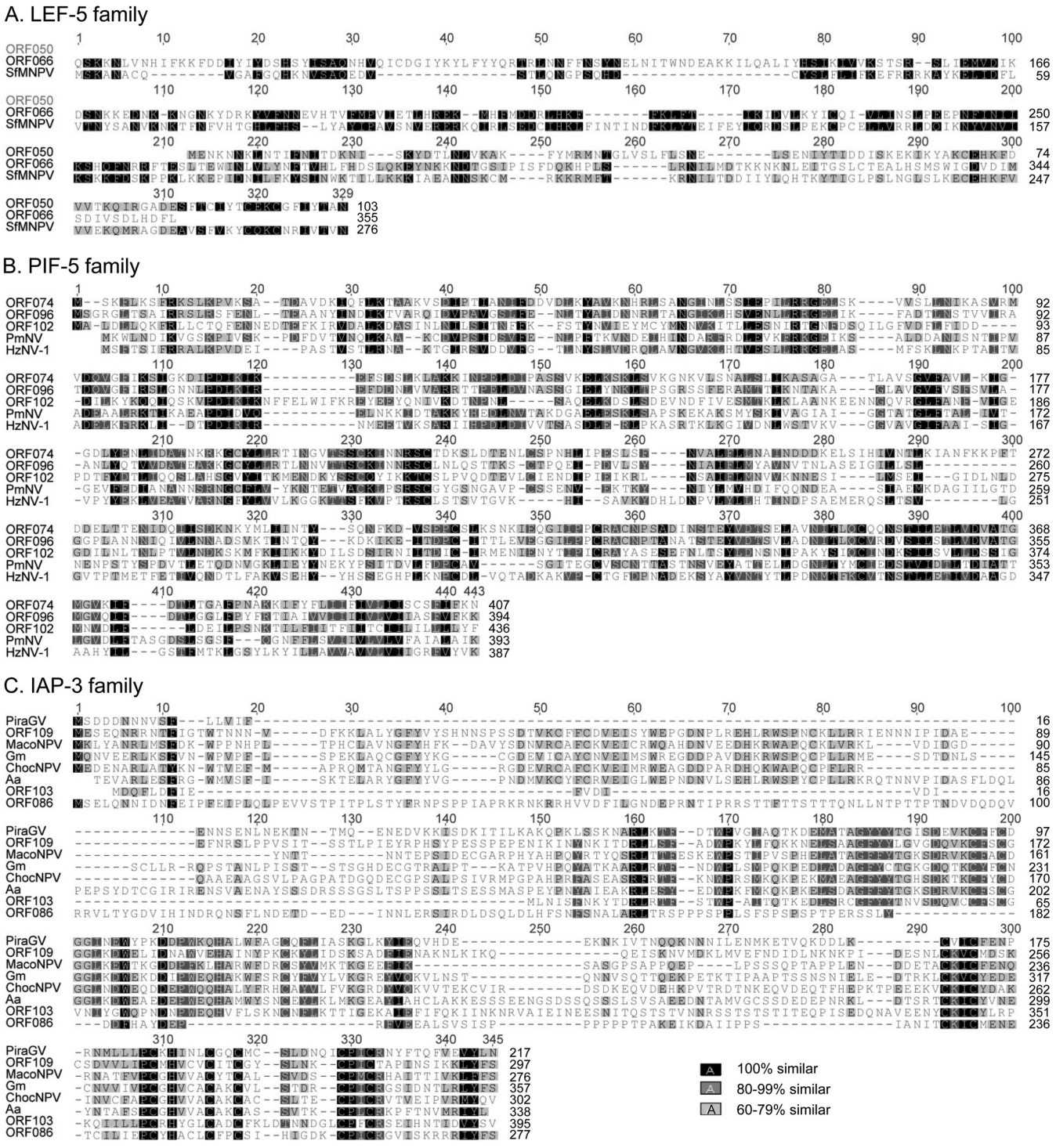


FIG 2 Amino acid sequence alignments of ToNV LEF-5 (ORF050 and ORF066), PIF-5 (ORF074, ORF096, and ORF102), and IAP-3 (ORF086, ORF103, and ORF109) multigenic family members with viral or insect homologs. (A) LEF-5 family; SfMNPV, *Spodoptera frugiperda* MNPV (YP_001036358); (B) PIF-5 family; HzNV-1, *Heliothis zea* NV-1 (AAN04370), and PmNV, *Penaeus monodon* NV (YP_009051848); (C) IAP-3 family; PiraGV, *Pieris rapae* GV (AGS18838); MacoNPV, *Mamestra configurata* NPV (NP_613222); Gm, *Galleria mellonella* (ACV04797); ChocNPV, *Choristoneura occidentalis* NPV (YP_008378620); Aa, *Aedes aegypti* (EAT39096). Alignments were generated using the MAFFT alignment plugin under Geneious and then eventually manually refined. Each multiple alignment is presented with a global numbering located above the aligned sequences and a specific numbering located on the right side of each considered sequence. The similarity color code shown on the bottom right applies to all parts of the figure.

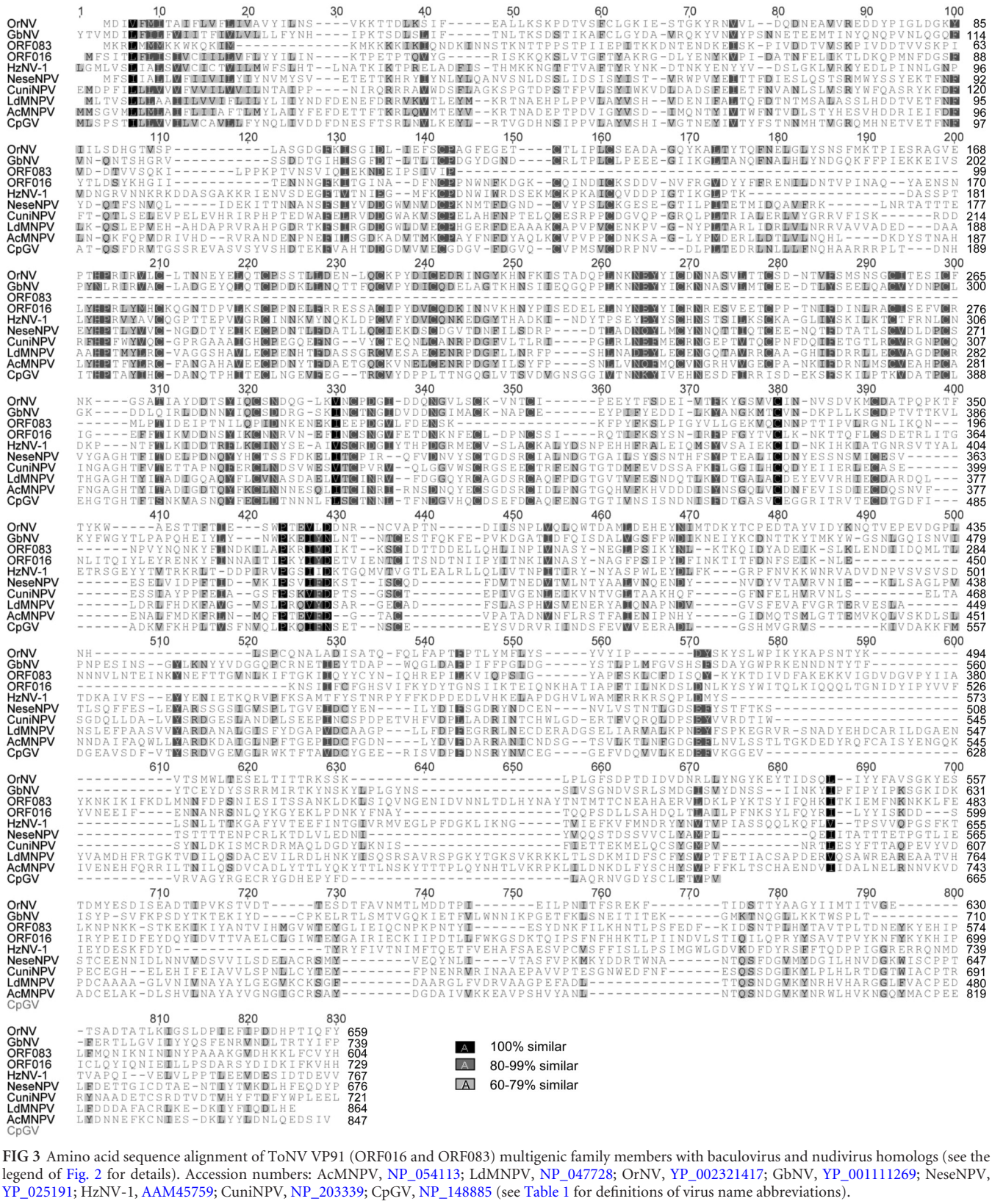


FIG 3 Amino acid sequence alignment of ToNV VP91 (ORF016 and ORF083) multigenic family members with baculovirus and nudivirus homologs (see the legend of Fig. 2 for details). Accession numbers: AcMNPV, NP_054113; LdMNPV, NP_047728; OrNV, YP_002321417; GbNV, YP_001111269; NeseNPV, YP_025191; HzNV-1, AAM45759; CuniNPV, NP_203339; CpGV, NP_148885 (see Table 1 for definitions of virus name abbreviations).

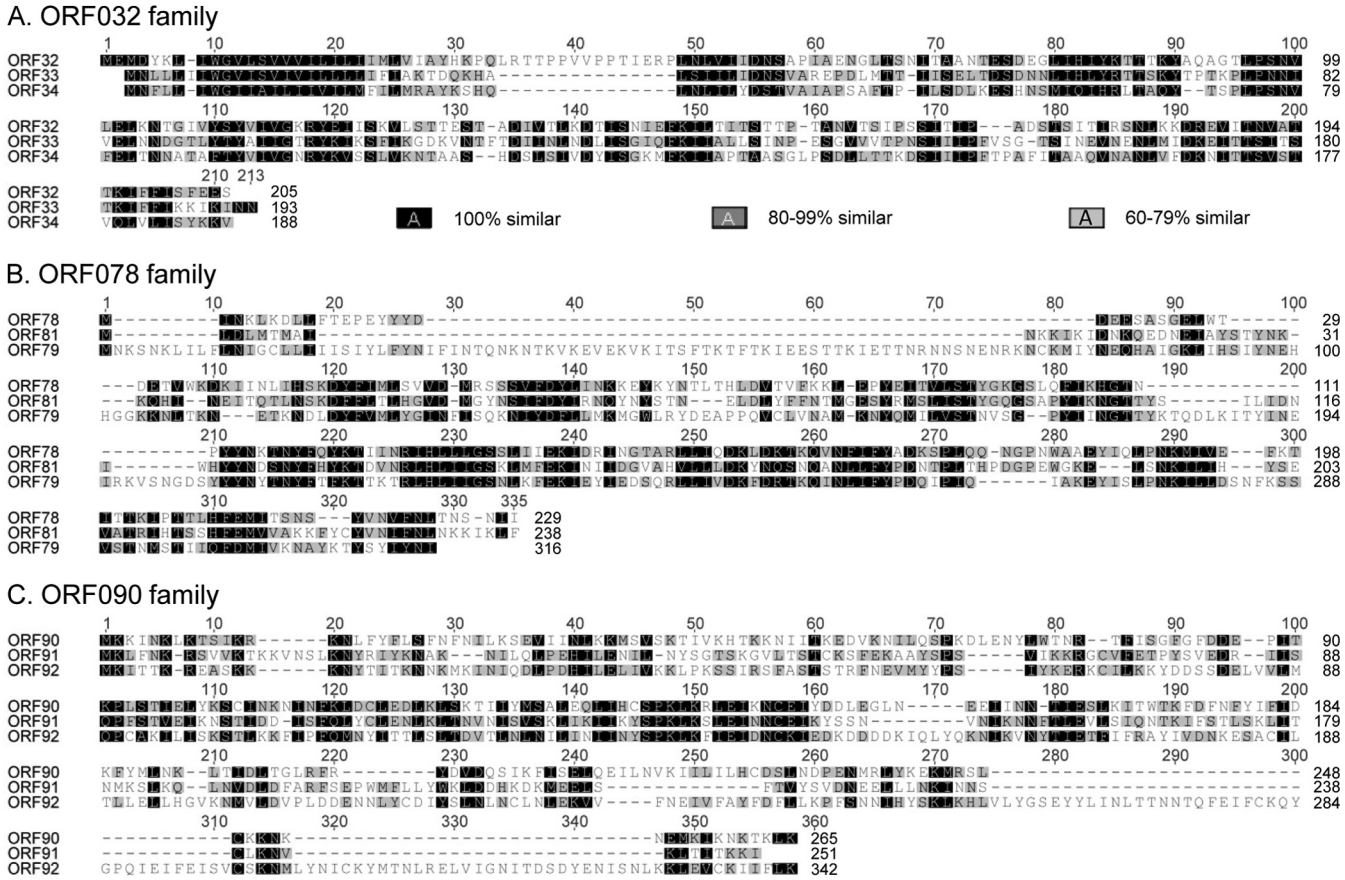


FIG 4 Amino acid sequence alignments of ToNV multigenic families with unknown function. (A) ORF032 family (ORF032, ORF033, and ORF034); (B) ORF078 family (ORF078, ORF079, and ORF081); (C) ORF090 family (ORF090, ORF091, and ORF092) (see legend of Fig. 2 for details).

base composition is highly AT biased, with AT making up 74.5% of the nucleotide content (Table 1). Whole-genome similarities searches revealed that ToNV is closer to HzNV-1 (E value, 1e−107) and PmNV (coverage, 18%) than to CuniNPV (E value, 1e−13; coverage, 1%) (Table 1). *De novo* gene prediction revealed that the ToNV genome carries 131 ORFs ranging from 111 to 4,143 bp in size (Fig. 1 and Table 2), representing a 85.7% coding density (one gene per ~1.11 kb), similar to baculoviruses, nudiviruses, and hytrosaviruses (Table 1). These ORFs were distributed evenly on both DNA strands: 53% clockwise and 47% counterclockwise (Fig. 1 and Table 2).

Gene content. Similarity searches in local and public databases were performed to assign a functional annotation to each ToNV predicted ORF. Among the 131 ORF products, ~24% presented known protein motifs, ~20% possessed a signal peptide, and ~30% had one to three transmembrane domains (Table 2). Fifty-seven ORFs displayed similarities with already known viral proteins, including those of baculoviruses, nudiviruses, or bracoviruses (Table 2). Assuming that viral homologs perform the same function, 7 ORFs were predicted to be involved in DNA processing, 3 in nucleotide metabolism, 6 in transcription, 5 in viral packaging, assembly, and release, 17 in viral structural proteins, including *per os* infectivity factors and other particle components, 1 in host interaction, 3 in apoptosis inhibition, and 15 with unknown function (Table 2). The remaining ORFs displayed either poor or no similarity with sequences available in public database (data not shown).

Notably, ToNV harbored several multigenic families (Table 2 and Fig. 2, 3, and 4). We found the following: two homologs of the late expression factor 5 (*orf050* = *lef-5a* and *orf066* = *lef-5b*), both most similar to *Spodoptera frugiperda* MNPV LEF-5 (Fig. 2A); three homologs of the *per os* infectivity factors 5 (*orf074* = *pif-5a*, *orf096* = *pif-5b*, and *orf102* = *pif-5c*), most similar to HzNV-1 or PmNV PIF-5 (Fig. 2B); three homologs of the inhibitor of apoptosis IAP-3 (*orf086* = *iap_a*, *orf103* = *iap_b*, and *orf109* = *iap_c*) similar to those of various organisms (*Galleria mellonella*, *Pieris rapae* GV, and *Aedes aegypti*, respectively) (Fig. 2C); two homologs of the viral structural protein VP91/95 (*orf016* and *orf083*), of which ORF016 appeared to be the likely ortholog of VP91/95 and ORF083 a distant paralog (Fig. 3). Moreover, several genes of unknown function could also be gathered in three families (ORF032, ORF078, and ORF090) (Fig. 4).

Promoter prediction. The presence of baculovirus early and late gene promoters (46, 47) was investigated within the 300 bp upstream of each ORF (Table 2). Early promoter motifs either corresponding to a TATA box alone (TATA) or in association with the initiator motif CA[TG]T (E1) were predicted, respectively, for 48 and 38 ORFs. Late baculovirus promoters (L) could be predicted for 57 ORFs. Eighteen ORFs displayed both E1 and L gene promoters. The HzNV-1 late promoter (48) was predicted

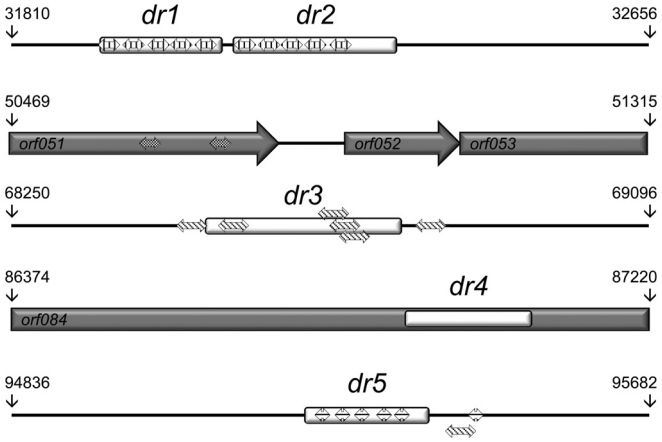


FIG 5 Direct repeat (*dr*) structures and imperfect palindromic motif (IPM) positions within the ToNV genome. The five regions including *dr* sequences are presented to scale. Positions within the genome are indicated (see Table 3 for details). *dr* sequences are represented by white boxes (see Table 3 for sequence details). IPMs are represented by small striped double arrows (IPM1, oblique lines; IPM2, vertical lines; IPM3, oblique grid; IPM4, horizontal lines; see Table 4 for details). Genes are represented by gray arrows or boxes, with their ORF numbers inside.

for two ORFs in combination with E1 and L (*orf075*) or only with E1 (*orf128*) promoters (Table 2). Moreover, a minimal TATA box (TATA[AT]) was predicted for five additional genes (*orf030*, *orf037*, *orf057*, *orf061*, and *orf117*). The transcriptional initiation small consensus sequence CA[TG]T was also predicted within the regions 300 bp upstream of all ORFs, except for *tk1*, *int*, and *orf080*.

Surprisingly and exactly opposite to what has been shown for AcMNPV (57), three of the four ToNV RNA polymerase subunit-encoding genes (*p47*, *lef-4*, and *lef-8*) were predicted to have a late promoter while the fourth subunit-encoding gene *lef-9* was predicted to have an early promoter. Similarly, the well-characterized *vp39*, which has a late promoter motif in baculoviruses (57), was predicted to have an early promoter in ToNV, a late promoter in HzNV-1 (58), and both early and late promoters in OrNV (59). Thus, prediction suggested that although transcriptional regulation mechanisms might be globally conserved, particular viral genes might be regulated in different manners.

The ToNV genome contains repeat regions with imperfect palindromic motifs. A common genome feature of the *Baculoviridae* is the presence of clusters of repeated imperfect palindromic AT-rich sequences called homologous repeat (*hr*) regions that serve as origins of viral DNA replication (60) and as transcriptional enhancers (61). Direct repeats (*dr*) are also common genome features in *Nudiviridae* (58, 59, 62–64) and *Hytrosaviridae* (32, 33). In the genome of ToNV, five different *dr* regions were detected, mainly outside coding regions, except for *dr4*, which was found within *orf084* (Fig. 1 and 5 and Table 3); all of these were predicted to be able to form hairpin structures (data not shown). Spanning from 160 to 262 bp, the *dr* regions contained two to seven repeats of specific consensus motifs (Table 3). With the exception of *dr4*, each *dr* also contained clusters of two to five small (22- to 41-bp) imperfect palindromic motifs (IPMs) (Table 4 and Fig. 5). The regions *dr1* and *dr2*, which were found located in tandem in the genome (Fig. 5), both contained clusters of five identical IPM2s (Tables 3 and 4). *dr3* and *dr5* also shared a minimal common 21-bp IPM (TGA[CG]TCAT[AC]G [TA]C[TG]ATGA[GC]TCA) present in six copies in each *dr* (data not shown). IPMs were not restricted to *dr* regions since IPM3 was found within *orf051* (Table 4 and Fig. 5). Finally, four *dr* regions (1, 2, 3, and 5) were AT-rich regions composed of direct repeats encompassing clusters of small imperfect, but not all identical, palindromic sequences, reminiscent of baculovirus *hr* regions.

ToNV shares 21 core genes with baculoviruses and nudiviruses. Core genes conserved between baculoviruses and nudiviruses were found in the genome of ToNV (Table 2 and Fig. 1). These genes are involved in DNA replication, transcription, virion structure, and infectivity. Twenty genes previously identified (28), including those coding for the DNA polymerase (DNApol) and the helicase, the four RNA polymerase subunits (P47, LEF-4, LEF-8, and LEF-9), and the late expression factor 5 (LEF-5), all the *per os* infectivity factors (P74, PIF-1, PIF-2, PIF-3, PIF-4 [19 kDa], PIF-5 [ODV-E56], PIF-6 [Ac68], and PIF-7 [VP91/95]), Ac81, the very late factor 1 (VLF-1), the sulfhydryl oxidase (P33), the viral phosphatase (38K), and the major capsid protein (VP39), were found in the ToNV genome. The later gene had been misidentified in PmNV and annotated as the 31K structural protein (64).

Furthermore, based on thorough comparative genomic analyses involving the reannotation of all available nudivirus genomes, an additional core gene that coded for the DNA-binding protein P6.9, a protein involved in viral packaging, assembly, and release, was identified (Fig. 6). The ORF142 (AAN04434) in HzNV-1 and ORF73 (YP_001111340) in GbNV had been predicted but not identified as P6.9. The OrNV homolog was identified as the C terminus of ORF22 (positions 176 to 243; YP_002321333), and its identification was strengthened by the synteny observed with GbNV (from OrNV ORF22 to ORF27 and from GbNV ORF72 to

TABLE 3 Direct repeat identified sequences in the ToNV genome[a]

Region name	Start position (bp)	Stop position (bp)	*dr* size (bp)	MM[b] size (bp)	No. of occurrences	Identity (%)	Consensus sequence
dr1	31,922	32,081	160	32	5	96.9	ATCACTTCTTATAACTAGCACTTTAATATTTC
dr2	32,095	32,311	217	31	7	86.6	ATATTTCATCACTTCTTATAACTAGCACTTA
dr3	68,604	68,765	262	131	2	92	AGTCTATATGAGTCAGAGTCTATGAGTCAGAGTCTATGAGTCATCGTC TATGAGTCATCGTCTATGAGTCATCGTCTATGAGTCATCATCGTCTATG AGTCATAAAAATTTCATTTTTTCCAAAAAAACTA
dr4	86,896	87,063	168	42	4	100	CTGATTCTTGTTCAACTGATGTTTGTTTTGTATTAATTTGTG
dr5	95,224	95,385	162	27	6	92	ATGACTCATAGACAATGACTCATGTAG

[a] Direct repeat (*dr*) regions were searched using etandem with a cutoff score of 100 and a minimal size of 20 bp.
[b] MM, minimal motif.

TABLE 4 Imperfect palindromic motifs[a]

Motif name	Size (bp)	No. of occurrences	Identity (%)	Palindromic sequence Logo
IPM1	41	7	97.56	
IPM2	30	10	96.66	
IPM3	29	2	100	
IPM4	22	6	95.45	

[a] Imperfect palindromic motifs (IPMs) were determined with the MEME program suite and a minimal size of 20 bp. Consensus motifs were visualized with WebLogo.

ORF78). P6.9 homologs were newly identified in HzNV-2 (positions 24,375 to 24,127, between ORF9 and ORF10) and in PmNV (positions 64,881 to 65,078, between ORF52 and ORF53). As ToNV ORF051 was identified as a P6.9 homolog, the set of core genes conserved between baculoviruses and nudiviruses was raised to 21 (shown in dark gray in Table 2 and in black arrows in Fig. 1).

ToNV presents nudiviral features, including 11 specific core genes. Eleven nudivirus core genes were found in the ToNV genome (shown in light gray in Table 2 and in gray arrows in Fig. 1). This included the *helicase-2*, the integrase (*int*), and the FEN-1/FLAP endonuclease (*fen-1*), three nudiviral thymidine kinases (*tk1*, *tk2*, and *tk3*), an *11K-like* homolog (62) and four other genes of unknown function, *orf006* (*PmNVorf99-like*), *orf019* (*PmNVorf62-like*), *orf027* (*GbNVorf19-like*), and *orf052* (*HzNVorf143-like*), that were homologous to HzNV-2 *orf27*, *orf79*, *orf99*, and *orf9* (62), respectively. Although the HzNV-2 ORF79 homolog had not been predicted in HzNV-1 (58), it was present between HzNV-1 ORF57 and ORF58 from position 79,463 to 79,900.

In contrast, all previously defined nudiviral core genes (28) were not identified in ToNV either because they were absent or because they were not monophyletic in phylogenetic analyses and therefore not orthologous (data not shown). The ribonucleotide reductases *rr1* and *rr2* as well as the thymidine kinase *tk4* (homologous to HzNV-1 *orf71* or HzNV-2 *orf67*) were absent in ToNV. ToNV features three inhibitor of apoptosis genes (*iap-3*) (i.e., *orf083*, *orf103*, and *orf109*), each deriving from various cellular origins (Fig. 2C), like other nudivirus *iap-3* genes. As no group of *iap-3* genes derived from a single viral common ancestor, they should not be considered core genes. Similarly, ToNV ORF121 displays a DNA ligase/mRNA capping enzyme domain and was phylogenetically related to HzNV-1 ORF141, HzNV-2 ORF10, and PmNV ORF48 but not to HzNV-1 ORF36, OrNV ORF121, or GbNV ORF38. DNA ligase and IAP-3 thus appear as core functions rather than core genes for nudiviruses.

Another common genome feature of baculoviruses is a conserved cluster of four core genes (*helicase*, *pif-4*, *38K*, and *lef-5*) found in synteny in all genomes (13). This cluster was not found in ToNV; only the *helicase* (*orf118*) and *pif-4* (*orf119*) were grouped as observed in all fully sequenced nudiviruses (28, 64) (Table 2 and Fig. 1). In addition, other nudivirus-specific gene microsyntenies were found in ToNV (Table 2 and Fig. 1).

Some ToNV genes are more similar to the endogenous nudivirus genes of bracoviruses. Bracoviruses evolved from the integration of a nudivirus in the genome of an ancestral parasitic wasp (37, 65). The genome of the wasps *Cotesia congregata* and *Microplitis demolitor* have retained a locus, comprising at least 10 genes from the original viral genome, termed the nudiviral cluster (37, 66, 67). ToNV displayed 12 of these nudivirus-like genes present in bracoviruses, although dispersed within its genome rather than as a cluster (Fig. 1 and Table 2). Surprisingly, from the 21 core

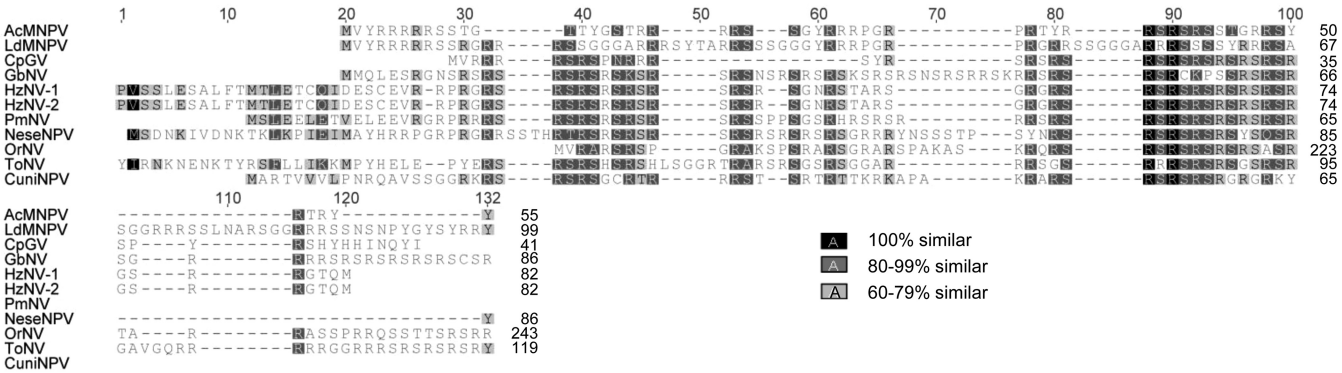
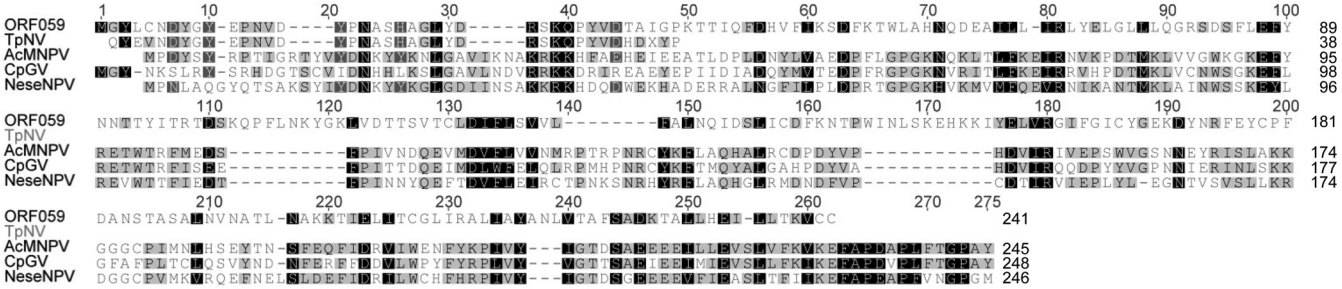
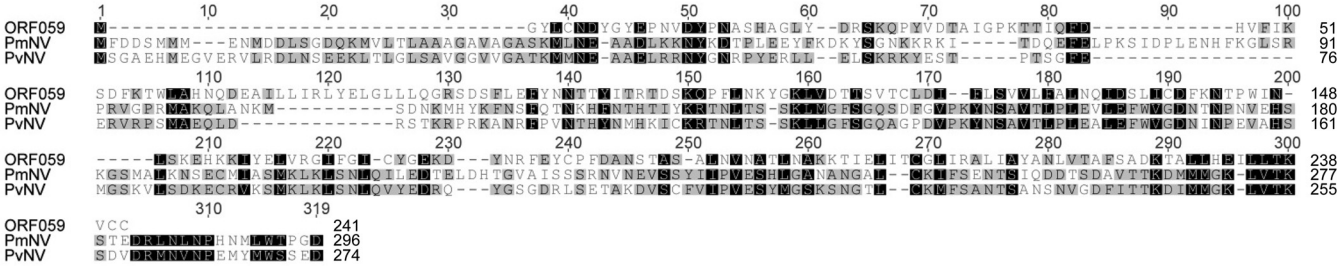


FIG 6 Amino acid sequence alignment of ToNV P6.9 (ORF051) with baculovirus and nudivirus homologs (see legend of Fig. 2 for details). Accession numbers: AcMNPV, NP_054130; LdMNPV, NP_047738; CpGV, NP_148870; GbNV, YP_001111340; HzNV-1, AAN04434; HzNV-2 (new prediction from positions 24,375 to 24,127), JN418988; PmNV (new prediction from positions 64,881 to 65,078), KJ184318; NeseNPV, YP_025143; OrNV, YP_002321333; CuniNPV, NP_203327 (see Table 1 for definitions of virus name abbreviations).

A. Polyhedrin/granulin homologous sequences



B. Major occlusion body protein nudiviral homologs



C. FEN-1 versus polyhedrin/granulin baculoviral sequences

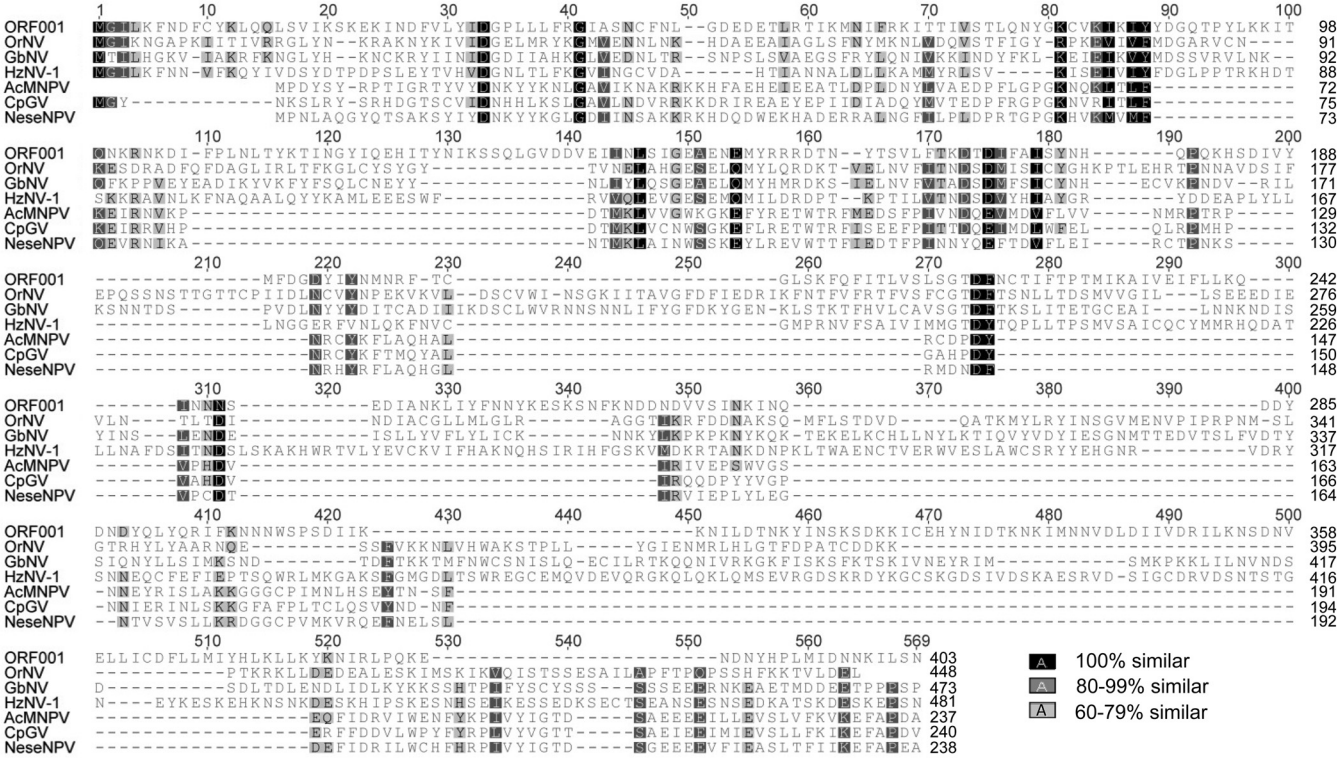


FIG 7 Amino acid sequence alignments of baculovirus polyhedrin/granulin sequences and presumed nudivirus homologs (see legend of Fig. 2 for details). (A) Alignment of ToNV ORF059 with TpNPV N-terminal sequence previously determined (24) and AcMNPV, CpGV, and NeseNPV polyhedrin/granulin sequences. (B) Alignment of ToNV ORF059 with *Penaeus* species nudiviruses major OB protein sequences. (C) Alignment of ToNV ORF001 with OrNV, GbNV, and HzNV-1 nudivirus FEN-1 homologs and AcMNPV, CpGV, and NeseNPV polyhedrin/granulin sequences. Accession numbers: AcMNPV, NP_054037; CpGV, NP_148785; NeseNPV, YP_025108; PmNV, ABY75164; PvNV, DQ496179; OrNV, YP_002321327; GbNV, YP_001111332; HzNV-1, AAN04362. TpNPV, *Tipula paludosa* NPV (24) (see Table 1 for definitions of other virus name abbreviations).

genes shared between baculoviruses and/or nudiviruses and bracoviruses, eight ToNV products showed higher similarity with bracovirus homologs than with their nudivirus or baculovirus next-best homologs (Table 2). Of note, these homologies between ToNV and bracovirus genes demonstrated that two genes found in the nudiviral cluster, *27b* (*Cc50C22.7*) and *Cc50C22.6*, were genuinely of viral origin. Likewise, a hypothetical protein at the locus tag K425_450 of the nudiviral cluster of the braconid wasp *M. demolitor* was identified as homologous to ToNV ORF027 and GbNV ORF19, and *Cc50C22.5* as homologous to PmNV ORF87. However, so far no ToNV homologs have been identified for *HzNVorf94-like* or *35a* (37, 65–67) or for the three newly identified MdBV putative nudiviral genes (locus tags K425_456, K425_459, and K425_461) (67). ToNV thus appears as an exogenous nudivirus most closely related to the ancestral nudivirus, whose genome integrated into the ancestor of the braconid wasps, thus producing bracoviruses.

ToNV *orf059* encodes a functional homolog of the baculovirus polyhedrin. Sequence alignment revealed that ORF059 displayed 81% sequence similarity (31/38 aa) with the previously published polyhedrin N-terminal sequence from TpNPV (24) (Fig. 7A). The protein encoded by *orf059* had a predicted molecular mass of 27.36 kDa, similar to that of its TpNPV homolog (23, 24). Multiple sequence alignments with baculovirus polyhedrin/granulin sequences revealed 43 conserved sites (Fig. 7A). The ToNV ORF059 sequence also displayed 58 conserved sites with two major occlusion body proteins (MOBPs) characterized from shrimp nudiviruses infecting *P. monodon* (64, 68) and *P. vannamei* (69) (Fig. 7B).

In contrast, when ToNV ORF059 similarity searches were made against OrNV ORF16, previously identified as the nudivirus polyhedrin (70), or against GbNV ORF65 and HzNV-1 ORF68 homologs, no significant E values (0.5, 0.66, and 0.54, respectively) were obtained. However, these genes were highly similar to ToNV ORF001 (E values between 3e−10 and 9e−23), which in turn was identified as a FEN-1/FLAP endonuclease by the sensitive HHpred method with a 98.7% probability. The GbNV, OrNV, HzNV-1, and PmNV homologs were also identified as FEN-1/FLAP endonucleases with high probability (97.8% to 99.1%). Lastly, when the nudivirus FEN-1/FLAP endonuclease sequences, including that of ToNV ORF001, were multiply aligned with baculovirus polyhedrin/granulin sequences, large gaps had to be introduced in the alignment, which revealed only 13 conserved sites (Fig. 7C), largely contrasting with the 43 sites conserved with ToNV ORF059 (Fig. 7A). Similarity searches and sequence alignments therefore revealed ToNV ORF059 as the most likely nudiviral homolog of baculovirus polyhedrin/granulin.

The ToNV genome encodes infectious occlusion bodies. To further characterize ToNV, particles from the archival sample 35 were observed by electron microscopy (Fig. 8) and infectivity was tested by bioassay. Electron microscopy observations showed that OBs measured from 2 to 5 μm in length and 2 μm in middiameter. The OB shape varied from a droplet form to all conceivable ellipsoid forms (Fig. 8A). OBs were filled with protein matrix and many rod-shaped virions (approximately 80 nm by 225 nm) (Fig. 8B to D) containing single nucleocapsids (approximately 40 nm by 160 nm) within a bilayer envelope (Fig. 8D).

To check whether the virus characterized in this study was still infectious after nearly 60 years of frozen storage, bioassays on healthy crane fly larvae were performed. *T. paludosa* larvae exposed with 10^5 to 10^6 OBs died between 4 and 15 days postinfection. Visible symptoms corresponded mainly to discoloration of larva epidermis, as previously described for TpNPV (19, 20). PCR diagnostics performed on viral particles purified from individual cadavers detected the presence of ToNV, thus confirming that the historical sample was still *per os* infectious (data not shown).

Arthropod dsDNA virus phylogenomy. Phylogenetic analyses were performed to determine to which virus ToNV was most closely related. Based on the concatenated alignment of 37 nudivirus-related genes from 18 dsDNA viruses, including four endogenous nudiviruses (three bracoviruses and the endogenous nudivirus of *N. lugens* [NlENV]) (Table 5), a highly supported phylogenetic tree was obtained by maximum likelihood analyses. The interrelationships between baculovirus, nudivirus, bracovirus, and hytrosavirus families were in accordance with previous results (30, 39, 64). ToNV clearly belonged to the *Nudiviridae* clade and not to the *Baculoviridae* (Fig. 9). The tree revealed that ToNV was distant from all other nudiviruses but more closely related to the HzNV-PmNV clade, to which bracoviruses were also related (30). The fruit fly virus DiNV was close to OrNV and belonged to the *Alphanudivirus* genus, from which NlENV originated (39). This indicated that two nudiviruses infecting the same host order (Diptera) could be distantly related.

DISCUSSION

To date, few *Nudiviridae* have been genetically characterized (27, 58, 59, 62–64). This study presents the first fully sequenced occluded nudivirus from a Diptera, the crane fly *T. oleracea*.

ToNV is closely related to the *T. paludosa* NPV based on the amino acid sequence identity between ToNV ORF059 and the N-terminal sequence of TpNPV MOBP (24) (Fig. 7A). Although diverging in sequence, these proteins appear to be both genetic and functional homologs of the baculovirus polyhedrin/granulin. In contrast, no convincing similarity could be identified between ToNV ORF059 and the previously labeled polh/gran of OrNV (ORF16) (70). The genetic bases of OrNV facultative OBs therefore remain to be determined. The relatedness between TpNPV and ToNV led us to redefine the virus TpNPV studied in the 1950s to 1970s (16–23) as a nudivirus (TpNV) instead of a baculovirus. Overall, this points out the phenotypic diversity of *Nudiviridae*, as within this family some viruses are transmitted as enveloped (*Tipula* NVs [21]) or nonenveloped (*Penaeus* NVs [64, 68, 71]) occlusion bodies, others as nonoccluded virions (*Penaeus* NVs [72], HzNV-1 [73], and GbNV [34, 74]) or as facultatively occluded virions (OrNV [34] and HzNV-2 [75]).

The phylogeny of the nudiviruses highlights two monophyletic clades, the OrNV/GbNV/DiNV clade, corresponding to the *Alphanudivirus* genus, and a second clade grouping ToNV with HzNVs and PmNV (30) (Fig. 9). As the phylogenetic distances between HzNVs, PmNV and ToNV are much larger than within the *Alphanudivirus*, their affiliation to a single genus (*Betanudivirus*) is therefore not as clear. If the proposal that a new genus should be created for PmNV (64) is taken forward, a fourth genus of exogenous nudiviruses would also need to be created for ToNV based on phylogenetic relationships (Fig. 9).

The ToNV data greatly improve comparative genomics analyses of nudiviruses and baculoviruses and allow the revision of core and accessory gene lists for nudiviruses within a new phylogenomic framework for arthropod large dsDNA viruses. Previous studies identify 33 nudivirus core genes (28, 62). The new data

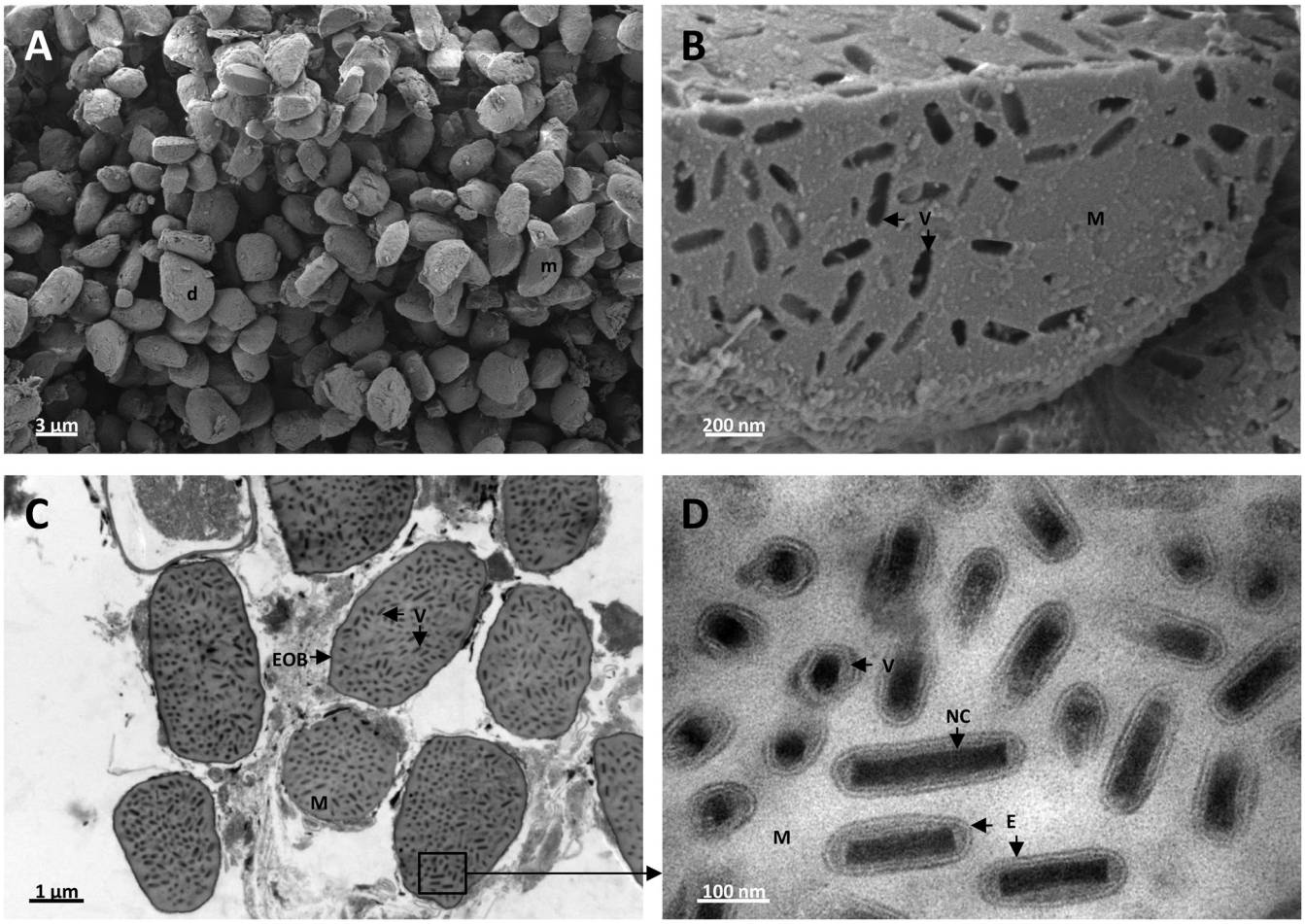


FIG 8 Occlusion bodies, virions, and nucleocapsids from *T. oleracea* purified nudivirus visualized by scanning (A and B) and transmission (C and D) electron microscopies. (A) Purified ToNV OBs shaped irregularly from droplet (d) to moon (m) by way of all conceivable ovoid and ellipsoid forms. (B) Section of typical OB enlarged image showing rod-shaped virion (V) prints inside the protein matrix (M). (C) Thin cross-section of enveloped occlusion bodies (EOB) filled with numerous virions (V) and protein matrix (M). (D) Enlarged image of protein matrix (M)-embedded virions (V) displaying, mainly in cross-sectional and longitudinal views, nucleocapsids (NC) surrounded by single bilayer membranes forming the envelopes (E).

show that *iap-3*, *ligase*, *rr1*, *rr2*, and *tk4* should rather be considered nudivirus accessory genes, probably acquired through horizontal transfers like in baculoviruses (76). Although deriving from diverse phylogenetic origins, the DNA ligase and IAP-3 are present in all sequenced nudiviruses and might therefore perform core nudivirus functions. Although the phylogenetic relationships between the different copies of *tk* genes are not well established, *tk1*, *tk2*, and *tk3* remain in the list of nudivirus core genes. Three new nudivirus core genes, *PmNVorf099-like* (*orf006*), *PmNVorf062-like* (*orf019*), and *11K-like* (*orf028*), had previously been overlooked during the annotation of at least one nudivirus genome. Recent developments in genomic comparison tools have improved the detection of gene orthologs, despite relatively distant sequences, and allowed the extension of the baculovirus core genome from 31 to 37 genes (14). However, for optimal results, it is necessary to carefully reannotate the genomes to see the emergence of the true common gene set shared between and within viral lineages. As a case in point, the detection of the *p6.9* gene in ToNV as well as in the other nudivirus genomes raises the pool of core genes shared by baculoviruses and nudiviruses to 21. This type of reanalysis could be applied to all large circular dsDNA viruses, which share a more or less restricted common gene set (26).

The genomes of large dsDNA viruses contain multiple gene families (77), particularly of accessory genes, which could reflect virus adaptation (77, 78), but also of some core genes. For example, many baculoviruses harbor two or three copies of *Ac66* and up to 16 *bro* genes (*Ac2*) (79). Two variants of *odv-e66* are described in several baculoviruses (NC_009011, NC_011616, NC_002169, NC_003529, and NC_004117) and in PmNV (ORF34 and ORF36) (64). The inhibitor of apoptosis gene family diversified in five lineages (*iap-1* to *iap-5*) in baculoviruses (79). This is also a diverse gene family in nudiviruses, as two *iap-3* genes are present in HzNVs (58, 62) and three in ToNV. All nudiviruses have two nonhomologous *helicase* genes: *DNA helicase*, belonging to the baculovirus/nudivirus core genes, and *helicase-2*, which is a nudivirus core gene (28), also present in both hytrosaviruses (26) and in some baculoviruses (79). The *integrase* superfamily is also quite diverse in nudiviruses, as it is represented by *vlf-1* (*HzNV-1 orf121-like*), *int* (*HzNV-1 orf144-like*), and *HzNV-1 orf140-like* homologs and found in even greater numbers in bracoviruses (80, 81). Lastly, ToNV is the first virus for which paralogs of the *lef-5*

TABLE 5 Sequences used for phylogenetic analyses[a]

	Nudiviruses						Baculoviruses					Endogenous nudiviruses				Hytrosaviruses		Nimavirus
Gene name	To NV	Hz NV-1	Pm NV	Gb NV	Or NV	Di NV	Ac MNPV	Ld MNPV	Cp GV	Nese NPV	Cuni NPV	Cc BV	Ci BV	Md BV	Nl ENV	Md SGHV	Gp SGHV	WSSV-TH
DNA polymerase	12	131	5	12	1		65	83	111	28	91				+	1	79	27
helicase	118	104	94	88	34	+	95	97	90	61	89			+	+			
helicase-2	105	60	79	46	108			50	126						+	104	74	
integrase	43	144	55	57	75								+	+	+			
fen-1	1	68	20	65	16	+								+	+	76		
tk1	22	51	65	17	137										+			
guaK				74	23	+									+			
p47	115	75	14	69	20	+	40	48	68	49	73	+		+	+			
lef-8	88	90	23	49	64		50	51	131	81	26	+	+	+	+	70	40	
lef-9	131	75	58	24	96		62	64	117	40	59			+	+	74	33	
lef-4	25	98	91	96	42		90	93	95	62	96		+	+	+	87	51	
vlf-1	65	121	56	80	30		77	86	106	45	18		+	+	+			
vp91/95	16	46	9	2	106	+	83	91	101	84	35		+	+	+			
vp39	87	89	22	64	15	+	89	92	96	89	24	+	+	+	+			
p33	99	13	8	7	113		92	94	93	24	14			+	+			
38K	63	10	59	1	87		98	99	88	59	87	+	+	+	+	73	44	
p74	45	11	72	45	126		138	27	60	50	74	+	+	+	+	39	1	72
pif-1	69	55	39	52	60		119	155	75	79	29		+	+	+	29	102	
pif-2	7	123	15	66	17	+	22	119	48	55	38		+	+	+	89	53	41
pif-3	13	88	93	3	107		115	143	35	69	46	+		+	+	106	76	
pif-4	119	103	96	87	33		96	98	89	60	90	+	+	+	+			
pif-6	56	74	88	55	72		68	80	114	41	58	+		+	+			
Ac81	123	33	86	14	4		81	89	103	48	106				+	108	78	
11K	28	124	100	95	41							+	+	+	+			
HzNVorf9	117	9	107									+	+	+				
HzNVorf64	112	64	24									+		+				
HzNVorf118	31	118	45									+	+	+				
HzNVorf128	4	128	42									+	+	+				
OrNVorf18	6	122	99	67	18	+									+			
OrNVorf19				68	19	+									+			
OrNVorf22				72	22	+									+			
OrNVorf25				76	25										+			
OrNVorf27				78	27	+									+			
OrNVorf44				97	44	+									+			
GbNVorf19	27	30	98	19	47										+			
OrNVorf54				83	54										+			
PmNVorf62	19	+	62	51	61										+			

[a] An ORF number, when available, or a "+" sign indicates sequences used in the analyses. Virus abbreviations and sequence accession numbers are those defined in Table 1, to which must be added the following: DiNV, *Drosophila innubia* NV (JN680861 to JN680871) (27); NlENV, *Nilaparvata lugens* endogenous NV (KJ566523 to KJ566588) (39); CcBV, *Cotesia congregata* BV (FM201559 to FM201576, FM877774, and FM212911 to FM212915) (37); CiBV, *Chelonus inanitus* BV (FM201579 to FM201597, FN543427 to FN546858, and FN594617) (37); MdBV, *Microplitis demolitor* BV (JO913492 to JO979916 and JR139425 to JR139430) (67, 81); WSSV-TH, *Penaeus monodon* white spot syndrome virus (AF369029).

and *vp91/95* core genes were found (11.1% and 22.1% identity between paralogs, respectively) (Fig. 2A and 3).

The genome of ToNV displays a number of sequence features that could be implicated in the regulation of DNA replication or gene expression. Five direct AT-rich repeat regions are predicted to form hairpin structures, similar to baculovirus *hr* and *dr* regions, which are associated with the functions of replication origin and/or of transcription enhancer (60, 61, 82–84). In the absence of transcriptomic data, gene regulation has to be inferred from comparative data with baculoviruses. Early and late baculovirus promoter motifs could be predicted for most ToNV ORFs. As shown in nudiviruses (58, 59) and in nudivirus-derived bracoviruses (37), the conservation of early, intermediate, and late baculovirus promoter motifs suggests that gene expression regulation has been globally retained during virus evolution and relies on both cellular and viral RNA polymerases (79). Based on promoter prediction, particular genes could be expressed at different times in different viruses, like the nudivirus *p51* gene, which displays an HL motif in HzNV-1 (48, 58) and an E1 motif in ToNV. However, a promoter motif, in itself, is not sufficient to predict the timing of gene expression, as shown by the first comprehensive transcriptomic analysis of baculovirus gene expression (57). Gene transcription regulation should therefore not be generalized based on such predictions, and predictions should be viewed with caution until a transcriptomic study is done.

Finally, as more genomic data become available, it appears that nudiviruses have led far more intricate relationships with their hosts than baculoviruses, even though they are roughly the same age (30). ToNV seems to be an intermediary between the *Alpha-*, *Beta-* and the newly proposed *Gammanudivirus* genera (29, 64), as

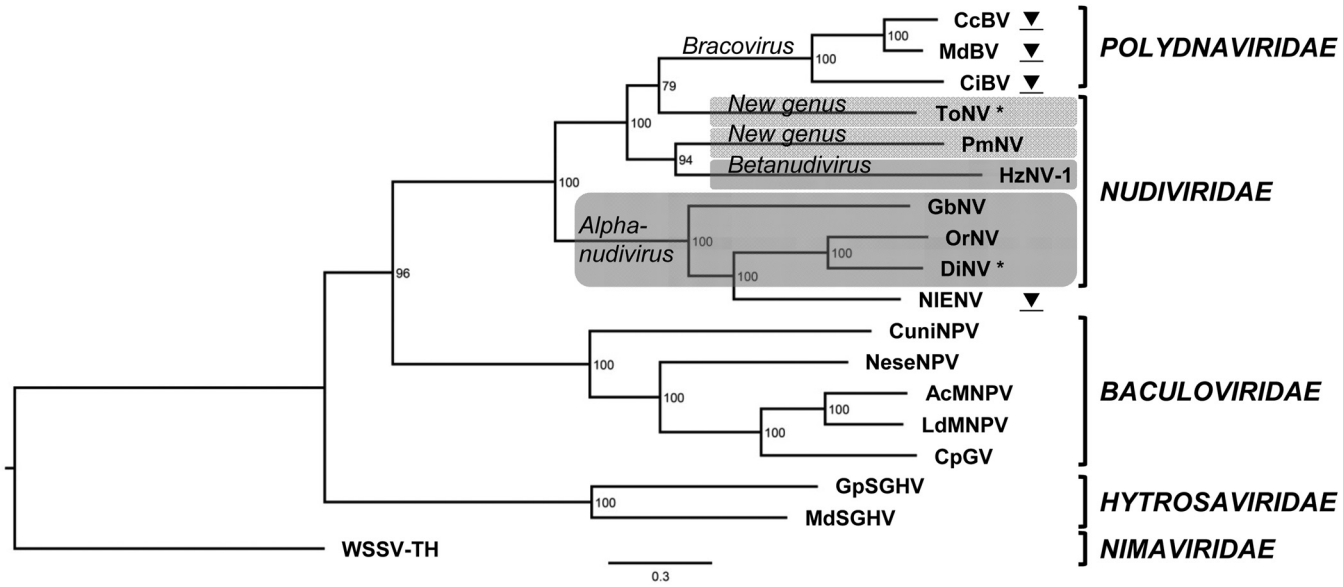


FIG 9 Arthropod large dsDNA virus phylogeny including ToNV. The tree was obtained by ML inference analysis of concatenated amino acid multiple alignments of 37 nudivirus-related genes (Table 5). Numbers on the nodes indicate ML nonparametric bootstrap supports (100 replicates). The white spot syndrome virus (WSSV-TH) was defined as an outgroup based on genome content. Virus families and genera are those approved by the International Committee on Taxonomy of Viruses (ICTV). The newly proposed and already approved nudivirus genera are shaded in gray. ▼, endogenous viral element derived from two independent nudivirus integration events, which led either to the bracovirus symbiosis or to the *Nilaparvata lugens* endogenous NV. *, diptera-infecting nudiviruses. Virus abbreviations, sequences, and accession numbers are those defined in Tables 1 and 5.

its genome displays genes specific to each genus. ToNV and PmNV are so far the only free nudiviruses with homologs to unknown genes previously identified within the *C. congregata* nudiviral cluster (37, 64, 66). Overall, this suggests that ToNV could belong to a new nudivirus genus. Surprisingly, two dipteran nudiviruses (ToNV and DiNV) clearly belong to different genera, which contrasts with the *Alphanudivirus* genus, which includes viruses infecting three insect orders. Nudiviruses also appear to be particularly prone to endogenization as recently shown with the sequences found in the planthopper genome, which derived from an *Alphanudivirus* (39). In an entirely independent event, the endogenization and domestication of a nudivirus more related to ToNV than to HzNVs or PmNV led to the evolution of bracoviruses (30, 65). The complex evolutionary history of the *Nudiviridae* will undoubtedly be revealed as their diversity is further explored. In this context, genomic data on the potential nudiviruses reported by Huger and Krieg to infect diverse insect orders would be of particular interest (34). The potential significance of nudivirus diversity will only be better understood when proteomic and functional data become available.

ACKNOWLEDGMENTS

This study was funded by European Research Council starting grant GENOVIR (205206) and supported by CEA-Genoscope (project AP2008: "The role of viruses in parasitoid wasp evolution").

This work has also been carried out with the technical support of the Genomic and Microscopy Facilities at Université François Rabelais.

We thank Max Bergoin and Aurélien Chateigner for helpful discussion, Karine Musset for her help in Sanger sequencing, and Philippe Roingeard for his expertise in electron microscopy.

REFERENCES

1. **Blackshaw RP, Coll C.** 1999. Economically important leatherjackets of grassland and cereals: biology, impact and control. Integr Pest Manage Rev **4:**143–160. http://dx.doi.org/10.1023/A:1009625724013.
2. **Taschereau E, Simard L, Brodeur J, Gelhaus J, Bélair G, Dionne J.** 2009. Seasonal ecology of the European crane fly (*Tipula paludosa*) and species diversity of the family Tipulidae on golf courses in Québec, Canada. ITS Res J **11:**681–693.
3. **Becnel JJ, White SE, Moser BA, Fukuda T, Rotstein MJ, Undeen AH, Cockburn A.** 2001. Epizootiology and transmission of a newly discovered baculovirus from the mosquitoes *Culex nigripalpus* and *C. quinquefasciatus*. J Gen Virol **82:**275–282.
4. **Moser BA, Becnel JJ, White SE, Afonso C, Gerald K, Shanker S, Almira E.** 2001. Morphological and molecular evidence that *Culex nigripalpus* baculovirus is an unusual member of the family *Baculoviridae*. J Gen Virol **82:**283–297.
5. **Andreadis TG, Becnel JJ, White SE.** 2003. Infectivity and pathogenicity of a novel baculovirus, CuniNPV from *Culex nigripalpus* (Diptera: Culicidae) for thirteen species and four genera of mosquitoes. J Med Entomol **40:**512–517. http://dx.doi.org/10.1603/0022-2585-40.4.512.
6. **Perera OP, Green TB, Stevens SM, White S, Becnel JJ.** 2007. Proteins associated with *Culex nigripalpus* nucleopolyhedrovirus occluded virions. J Virol **81:**4585–4590. http://dx.doi.org/10.1128/JVI.02391-06.
7. **de Araujo Coutinho CJPC, Alves R, Sanscrainte ND, DeBarros Pinto Viviani A, dos Santos PF, de Souza PAVM, de Carvalho-Mello IMVG, Becnel JJ.** 2012. Occurrence and phylogenetic characterization of a baculovirus isolated from *Culex quinquefasciatus* in São Paulo State, Brazil. Arch Virol **157:**1741–1745. http://dx.doi.org/10.1007/s00705-012-1372-1.
8. **Martignoni ME, Iwai PJ.** 1986. A catalogue of viral diseases of insects, mites, and ticks, 4th ed. USDA Forest Service General Technical Report PNW-195. USDA Forest Service, Portland, OR.
9. **Adams JR, McClintock JT.** 1991. *Baculoviridae*. Nuclear polyhedrosis viruses, p 87e226. *In* Adams JR, Bonami JR (ed), Atlas of invertebrate viruses. CRC Press, Boca Raton, FL.
10. **Poinar G, Jr, Poinar R.** 2005. Fossil evidence of insect pathogens. J Invertebr Pathol **89:**243–250. http://dx.doi.org/10.1016/j.jip.2005.05.007.
11. **Herniou EA, Arif BM, Becnell JJ, Blissard GW, Bonning B, Harrison R, Jehle JA, Theilmann DA, Vlak JM.** 2012. *Baculoviridae*, p 163–173. *In* King AMQ, Adams MJ, Carstens EB, Lafkowitz EJ (ed), Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, London, United Kingdom.
12. **Afonso CL, Tulman ER, Lu Z, Balinsky CA, Moser BA, Becnel JJ, Rock DL, Kutish GF.** 2001. Genome sequence of a baculovirus pathogenic for

Culex nigripalpus. J Virol **75:**11157–11165. http://dx.doi.org/10.1128/JVI.75.22.11157-11165.2001.
13. **Herniou EA, Olszewski JA, Cory JS, O'Reilly DR.** 2003. The genome sequence and evolution of *Baculoviruses*. Annu Rev Entomol **48:**211–234. http://dx.doi.org/10.1146/annurev.ento.48.091801.112756.
14. **Garavaglia MJ, Miele SAB, Iserte JA, Belaich MN, Ghiringhelli PD.** 2012. The *ac53*, *ac78*, *ac101*, and *ac103* genes are newly discovered core genes in the family *Baculoviridae*. J Virol **86:**12069–12079. http://dx.doi.org/10.1128/JVI.01873-12.
15. **Rennie J.** 1923. Polyhedral disease in *Tipula paludosa* MEIG. Proc R Soc Edinb Ser B **20:**265–268.
16. **Smith KM, Xeros N.** 1954. An unusual virus disease of a dipterous larva. Nature **173:**866–867. http://dx.doi.org/10.1038/173866a0.
17. **Smith KM.** 1955. Intranuclear changes in the polyhedrosis of *Tipula paludosa* (Diptera). Nature **176:**255. http://dx.doi.org/10.1038/176255a0.
18. **Smith KM.** 1956. The structure of insect virus particles. J Biophys Biochem Cytol **2:**301–306. http://dx.doi.org/10.1083/jcb.2.3.301.
19. **Smith KM.** 1967. The various types of insect viruses and the nuclear polyhedroses, p 8–40. *In* Smith KM (ed), Insect virology, 1st ed. Academic Press, New York, NY.
20. **Meynardier G, Ricou G, Bergoin M.** 1964. Virose à corps d'inclusion chez *Tipula paludosa* (Diptera) en France. Re Pathol Vég Entomol Agric Fr **43:**113–118.
21. **Bergoin M, Guelpa B.** 1977. Dissolution des inclusions du virus de la polyhedrose nucléaire du diptère *Tipula paludosa* MEIG. Etude ultrastructurale du virion. Arch Virol **53:**243–254.
22. **Revet BMJ, Guelpa B.** 1979. The genome of a baculovirus infecting *Tipula paludosa* (Meig.) (diptera): a high molecular weight closed circular DNA of zero superhelix density. Virology **96:**633–639.
23. **Croizier G, Croizier L.** 1977. Evaluation du poids moléculaire de la protéine des corps d'inclusion de divers baculovirus d'insectes. Arch Virol **55:**247–250.
24. **Rohrmann GF, Pearson MN, Bailey TJ, Becker RR, Beaudreau GS.** 1981. N-terminal polyhedrin sequences and occluded baculovirus evolution. J Mol Evol **17:**329–333.
25. **Rohrmann GF.** 1986. Polyhedrin structure. J Gen Virol **67:**1499–1513.
26. **Jehle JA, Abd-Alla AMM, Wang Y.** 2013. Phylogeny and evolution of *Hytrosaviridae*. J Invertebr Pathol **112:**S62–S67. http://dx.doi.org/10.1016/j.jip.2012.07.015.
27. **Unckless RL.** 2011. A DNA virus of *Drosophila*. PLoS One **6:**e26564. http://dx.doi.org/10.1371/journal.pone.0026564.
28. **Wang Y, Bininda-Emond ORP, Jehle JA.** 2012. Nudivirus genomics and phylogeny, p 33–52. *In* ML Garcia, V Romanowski (ed), Viral genomes—molecular structure, diversity, gene expression mechanisms and host-virus interactions. InTech, Rijeka, Croatia. http://dx.doi.org/10.5772/27793.
29. **Jehle JA, Burand J, Herniou EA, Harrison R, Arif B, Thielmann D, van Oers M, Becnel J.** 2013. Creation of a new Family *Nudiviridae* including two new genera and three species. Taxonomy Proposals http://talk.ictvonline.org/files/proposals/taxonomy_proposals_invertebrate1/m/inv04/4770.aspx.
30. **Thézé J, Bézier A, Periquet G, Drezen J-M, Herniou EA.** 2011. Paleozoic origin of insect large dsDNA viruses. Proc Natl Acad Sci U S A **108:**15931–15935. http://dx.doi.org/10.1073/pnas.1105580108.
31. **Abd-Alla AMM, Vlak JM, Bergoin M, Maruniak JE, Parker A, Burand JP, Jehle JA, Boucias DG, Hytrosavirus Study Group of the ICTV.** 2009. *Hytrosaviridae*: a proposal for classification and nomenclature of a new insect virus family. Arch Virol **154:**909–918. http://dx.doi.org/10.1007/s00705-009-0398-5.
32. **Abd-Alla AMM, Cousserans F, Parker A, Jehle JA, Parker N, Vlak JM, Robinson A, Bergoin M.** 2008. Genome analysis of a *Glossina pallidipes* salivary gland hypertrophy virus (GpSGHV) reveals a novel large double-stranded circular DNA virus. J Virol **82:**4595–4611. http://dx.doi.org/10.1128/JVI.02588-07.
33. **Garcia-Maruniak A, Maruniak JE, Farmerie W, Boucias DG.** 2008. Sequence analysis of a non-classified, non-occluded DNA virus that causes salivary gland hypertrophy of *Musca domestica*, MdSGHV. Virology **377:**184–196. http://dx.doi.org/10.1016/j.virol.2008.04.010.
34. **Huger AM, Krieg A.** 1991. *Baculoviridae*. Nonoccluded baculoviruses, p 287–319. *In* Adams JR, Bonami JR (ed), Atlas of invertebrate viruses. CRC Press, Boca Raton, FL.
35. **Burand JP.** 1998. Nudiviruses, p 69–90. *In* Miller LK, Ball LA (ed), The insect viruses. Plenum Press, New York, NY.
36. **Wang Y, Jehle JA.** 2009. Nudiviruses and other large, double-stranded circular DNA viruses of invertebrates: new insights on an old topic. J Invertebr Pathol **101:**187–193. http://dx.doi.org/10.1016/j.jip.2009.03.013.
37. **Bézier A, Annaheim M, Herbinière J, Wetterwald C, Gyapay G, Bernard-Samain S, Wincker P, Roditi I, Heller M, Belghazi M, Pfister-Wilhem R, Periquet G, Dupuy C, Huguet E, Volkoff AN, Lanzrein B, Drezen J-M.** 2009. Polydnaviruses of braconid wasps derive from an ancestral nudivirus. Science **323:**926–930. http://dx.doi.org/10.1126/science.1166788.
38. **Bézier A, Louis F, Jancek S, Periquet G, Thézé J, Gyapay G, Musset K, Lesorbe J, Lenoble P, Dupuy C, Gundersen-Rindal D, Herniou EA, Drezen J-M.** 2013. Functional endogenous viral elements in the genome of the parasitoid wasp *Cotesia congregata*: insights into the evolutionary dynamics of bracoviruses. Phil Trans R Soc B **368:**20130047. http://dx.doi.org/10.1098/rstb.2013.0047.
39. **Cheng R-L, Lou Y-H, Gu L-Z, Wang Z, Xu J-Y, Xu H-J, Zhang C-X.** 2014. The brown planthopper nudivirus DNA integrated in its host genome. J Virol **88:**5310–5318. http://dx.doi.org/10.1128/JVI.03166-13.
40. **Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380. http://dx.doi.org/10.1038/nature03959.
41. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.
42. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res **25:**3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.
43. **Altschul SF, Wootton JC, Gertz EM, Agarwala R, Morgulis A, Schäffer AA, Yu YK.** 2005. Protein database searches using compositionally adjusted substitution matrices. FEBS J **272:**5101–5109. http://dx.doi.org/10.1111/j.1742-4658.2005.04945.x.
44. **Hall TA.** 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser (Oxf) **41:**95–98.
45. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945. http://dx.doi.org/10.1093/bioinformatics/16.10.944.
46. **Xing K, Deng R, Wang J, Feng J, Huang M, Wang X.** 2005. Analysis and prediction of baculovirus promoter sequences. Virus Res **113:**64–71. http://dx.doi.org/10.1016/j.virusres.2005.04.016.
47. **Passarelli AL, Guarino LA.** 2007. Baculovirus late and very late gene regulation. Curr Drug Targets **8:**1103–1115. http://dx.doi.org/10.2174/138945007782151324.
48. **Guttieri MC, Burand JP.** 2001. Location, nucleotide sequence, and regulation of the *p51* late gene of the Hz-1 insect virus: identification of a putative late regulatory element. Virus Genes **23:**17–25. http://dx.doi.org/10.1023/A:1011166926225.
49. **Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS.** 2009. MEME Suite: tools for motif discovery and searching. Nucleic Acids Res **37:**W202–W208. http://dx.doi.org/10.1093/nar/gkp335.
50. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome Res **14:**1188–1190. http://dx.doi.org/10.1101/gr.849004.
51. **Zuker M.** 2003. Mfold web server for Nucleic acid folding and hybridization prediction. Nucleic Acids Res **31:**3406–3415. http://dx.doi.org/10.1093/nar/gkg595.
52. **Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins DG.** 2011.

Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol **7:**539. http://dx.doi.org/10.1038/msb.2011.75.

53. **Stamatakis A.** 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics **22:**2688–2690. http://dx.doi.org/10.1093/bioinformatics/btl446.
54. **Keane TM, Creevey CJ, Pentony MM, Naughton TJ, Mclnerney JO.** 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that *ad hoc* assumptions for choice of matrix are not justified. BMC Evol Biol **6:**29. http://dx.doi.org/10.1186/1471-2148-6-29.
55. **Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R.** 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol **3:**294–299.
56. **Ratnasingham S, Hebert PDN.** 2007. BOLD: the Barcode of Life Data System. Mol Ecol Notes **7:**355–364. http://dx.doi.org/10.1111/j.1471-8286.2007.01678.x.
57. **Chen Y-R, Zhong S, Fei Z, Hashimoto Y, Xiang JZ, Zhang S, Blissard GW.** 2013. The transcriptome of the baculovirus *Autographa californica* multiple nucleopolyhedrovirus in *Trichoplusia ni* cells. J Virol **87:**6391–6405. http://dx.doi.org/10.1128/JVI.00194-13.
58. **Cheng CH, Liu SM, Chow TY, Hsiao YY, Wang DP, Huang JJ, Chen HH.** 2002. Analysis of the complete genome sequence of the Hz-1 virus suggests that it is related to members of the *Baculoviridae*. J Virol **76:**9024–9034. http://dx.doi.org/10.1128/JVI.76.18.9024-9034.2002.
59. **Wang Y, van Oers MM, Crawford AM, Vlak JM, Jehle JA.** 2007. Genomic analysis of *Oryctes rhinoceros* virus reveals genetic relatedness to *Heliothis zea* virus 1. Arch Virol **152:**519–531. http://dx.doi.org/10.1007/s00705-006-0872-2.
60. **Pearson M, Bjornson R, Pearson G, Rohrmann G.** 1992. The *Autographa californica* baculovirus genome: evidence for multiple replication origins. Science **257:**1382–1384.
61. **Guarino LA, Summers MD.** 1986. Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. J Virol **60:**215–223.
62. **Burand JP, Kim W, Afonso CL, Tulman ER, Kutish GF, Lu Z, Rock DL.** 2012. Analysis of the genome of the sexually transmitted insect virus *Helicoverpa zea* nudivirus 2. Viruses **4:**28–61. http://dx.doi.org/10.3390/v4010028.
63. **Wang Y, Kleespies RG, Huger AM, Jehle JA.** 2007. The genome of *Gryllus bimaculatus* nudivirus indicates an ancient diversification of baculovirus-related nonoccluded nudiviruses of insects. J Virol **81:**5395–5406. http://dx.doi.org/10.1128/JVI.02781-06.
64. **Yang YT, Lee DY, Wang Y, Hu JM, Li WH, Leu JH, Chang GD, Ke HM, Kang ST, Lin SS, Kou GH, Lo CF.** 2014. The genome and occlusion bodies of marine *Penaeus monodon* nudivirus (PmNV, also known as MBV and PemoNPV) suggest that it should be assigned to a new nudivirus genus that is distinct from the terrestrial nudiviruses. BMC Genomics **15:**628. http://dx.doi.org/10.1186/1471-2164-15-628.
65. **Herniou EA, Huguet H, Thézé J, Bézier A, Periquet G, Drezen J-M.** 2013. When parasitic wasps hijacked viruses: genomic and functional evolution of polydnaviruses. Phil Trans R Soc B **368:**20130051. http://dx.doi.org/10.1098/rstb.2013.0051.
66. **Bézier A, Herbinière J, Lanzrein B, Drezen J-M.** 2009. Polydnavirus hidden face: the genes producing virus particles of parasitic wasps. J Invertebr Pathol **101:**194–203. http://dx.doi.org/10.1016/j.jip.2009.04.006.
67. **Burke GR, Walden KKO, Whitfield JB, Robertson HM, Strand MS.** 2014. Widespread genome reorganization of an obligate virus mutualist. PLoS Genet **9:**e1004660. http://dx.doi.org/10.1371/journal.pgen.1004660.
68. **Chaivisuthangkura P, Tawilert C, Tejangkura T, Rukpratanporn S, Longyant S, Sithigorngul W, Sithigorngul P.** 2008. Molecular isolation and characterization of a novel occlusion body protein gene from *Penaeus monodon* nucleopolyhedrovirus. Virology **381:**261–267. http://dx.doi.org/10.1016/j.virol.2008.08.036.
69. **Bonami JR, Bruce LD, Poulos BT, Mari J, Lightner DV.** 1995. Partial characterization and cloning of the genome of PvSNPV (=BP-type virus) pathogenic for *Penaeus vannamei*. Dis Aquat Org **23:**59–66. http://dx.doi.org/10.3354/dao023059.
70. **Wang Y, Bininda-Emonds ORP, van Oers MM, Vlak JM, Jehle JA.** 2011. The genome of *Oryctes rhinoceros* nudivirus provides novel insight into the evolution of nuclear arthropod-specific large circular double-stranded DNA viruses. Virus Genes **42:**444–456. http://dx.doi.org/10.1007/s11262-011-0589-5.
71. **Bonami JR, Aubert H, Mari J, Poulos BT, Lightner DV.** 1997. The polyhedral of the occluded baculoviruses of marine decapod crustacean: a unique structure, crystal organization, and proposed model. J Struct Biol **120:**134–145.
72. **Couch JA.** 1974. Free and occluded virus, similar to *Baculovirus*, in hepatopancreas of pink shrimp. Nature **247:**229–231.
73. **Granados RR, Nguyen T, Cato B.** 1978. An insect cell line persistently infected with baculovirus-like particle. Intervirology **10:**309–317.
74. **Huger AM.** 1985. A new virus disease of crickets (Orthoptera: Gryllidae) causing macronucleosis of fatbody. J Invertebr Pathol **45:**108–111.
75. **Raina AK, Adams JR, Lupiani B, Lynn DE, Kim W, Burand JP, Dougherty EM.** 2000. Further characterization of the Gonad-Specific Virus of corn eatworm, *Helicoverpa zea*. J Invertebr Pathol **76:**6–12. http://dx.doi.org/10.1006/jipa.2000.4942.
76. **Hughes AL, Friedman R.** 2003. Genome-wide survey for genes horizontally transferred from cellular organisms to baculoviruses. Mol Biol Evol **20:**979–987. http://dx.doi.org/10.1093/molbev/msg107.
77. **Thézé J, Takatsuka J, Li Z, Gallais J, Doucet D, Arif B, Nakai M, Herniou EA.** 2013. New insights into the evolution of *Entomopoxvirinae* from the complete genome sequences of four entomopoxviruses infecting *Adoxophyes honmai*, *Choristoneura biennis*, *Choristoneura rosaceana*, and *Mythimna separata*. J Virol **87:**7992–8003. http://dx.doi.org/10.1128/JVI.00453-13.
78. **Elde NC, Child SJ, Eickbush MT, Kitzman JO, Rogers KS, Shendure J, Geballe AP, Malik HS.** 2012. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. Cell **150:**831–841. http://dx.doi.org/10.1016/j.cell.2012.05.049.
79. **Rohrmann GF.** 2013. Baculovirus molecular biology, 3rd ed. National Library of Medicine (US), National Center for Biotechnology Information, Bethesda, MD.
80. **Drezen J-M, Herniou EA, Bézier A.** 2012. Evolutionary progenitors of bracoviruses, p 15–31. *In* Beckage NE, Drezen JM (ed), Parasitoid viruses symbionts and pathogens. CRC Press-Elsevier, San Diego, CA.
81. **Burke GR, Strand MR.** 2012. Deep sequencing identifies viral and wasp genes with potential roles in replication of *Microplitis demolitor* bracovirus. J Virol **86:**3293–3306. http://dx.doi.org/10.1128/JVI.06434-11.
82. **Lee H, Krell PJ.** 1994. Reiterated DNA fragments in defective genomes of *Autographa californica* nuclear polyhedrosis virus are competent for Ac-MNPV-dependent DNA replication. Virology **202:**418–429.
83. **Habib S, Hasnain SE.** 2000. Differential activity of two non-*hr* origins during replication of the baculovirus *Autographa californica* nuclear polyhedrosis virus genome. J Virol **74:**5182–5189. http://dx.doi.org/10.1128/JVI.74.11.5182-5189.2000.
84. **Cvetic C, Walter JC.** 2005. Eukaryotic origins of DNA replication: could you please be more specific? Semin Cell Dev Biol **16:**343–353. http://dx.doi.org/10.1016/j.semcdb.2005.02.009.
85. **Ayres MD, Howard SC, Kuzio J, Lopez-Ferber M, Possee RD.** 1994. The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology **202:**586–605.
86. **Kuzio J, Pearson MN, Harwood SH, Funk CJ, Evans JT, Slavicek JM, Rohrmann GF.** 1999. Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. Virology **253:**17–34.
87. **Luque T, Finch R, Crook N, O'Reilly DR, Winstanley D.** 2001. The complete sequence of the *Cydia pomonella* granulovirus genome. J Gen Virol **82:**2531–2547.
88. **Garcia-Maruniak A, Maruniak J, Zanotto PMA, Doumbouya AE, Liu J-C, Meritt TM, Lanoie JS.** 2004. Sequence analysis of the genome of the *Neodiprion sertifer* nucleopolyhedrovirus. J Virol **78:**7036–7051. http://dx.doi.org/10.1128/JVI.78.13.7036-7051.2004.