Chapter 8

Determination of Nucleopolyhedrovirus' Taxonomic Position

Yu-Shin Nai, Yu-Feng Huang, Tzu-Han Chen, Kuo-Ping Chiu and Chung-Hsiung Wang

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/66634

Abstract

To date, over 78 genomes of nucleopolyhedroviruses (NPVs) have been sequenced and deposited in NCBI. How to define a new virus from the infected larvae in the field is usually the first question. Two NPV strains, which were isolated from casuarina moth (*L. xylina*) and golden birdwing larvae (*Troides aeacus*), respectively, displayed the same question. Due to the identity of polyhedrin (*polh*) sequences of these two isolates to that of *Lymantria dispar* MNPV and *Bombyx mori* NPV, they are named LdMNPV-like virus and TraeNPV, provisionally. To further clarify the relationships of LdMNPV-like virus and TraeNPV to closely related NPVs, Kimura 2-parameter (K-2-P) analysis was performed. Apparently, the results of K-2-P analysis that showed LdMNPV-like virus is an LdMNPV isolate, while TraeNPV had an ambiguous relationship to BmNPV. Otherwise, MaviNPV, which is a mini-AcMNPV, also exhibited a different story by K-2-P analysis. Since K-2-P analysis could not cover all species determination issues, therefore, TraeNPV needs to be sequenced for defining its taxonomic position. For this purpose, different genomic sequencing technologies and bioinformatic analysis approaches will be discussed. We anticipated that these applications will help to exam nucleotide information of unknown species and give an insight and facilitate to this issue.

Keywords: nucleopolyhedroviruses, Kimura-2-parameter analysis, next-generation sequencing, bioinformatic analysis

1. Introduction

Baculoviruses are insect-specific viruses which have a large circular double-stranded DNA genome packaged in enveloped, rod-shaped nucleocapsid and occluded within a paracrystalline protein occlusion body (OB) [1, 2]. The family *Baculoviridae* has four genera, including

Alphabaculovirus, *Betabaculovirus*, *Gammabaculovirus* and *Deltabaculovirus*. Nucleopolyhedrovirus (NPV) is a member of *Alphabaculovirus* (lepidopteran-specific NPV) [3]; NPV replicates in the nucleus of the infected host cell and causes a disease of nuclear polyhedrosis. Epidemic outbreak of NPV may play a role in regulation of the host nature population [4]. Thereby, it is a potential agent for biological control with a number of eco-friendly benefits including high virulence and specificity against target insects, environmental safety and sustainable existence with target insects. Several baculoviruses showing promising results have been commercialized as biopesticides for the control of insect pests around the world [5]. For biotechnological applications, baculoviruses have been constructed as a eukaryotic protein expression vectors (baculovirus expression vector system (BEVS)) over the last 30 years and used to gene therapy trials. So far, many recombinant proteins have been expressed in insect cells by BEVS and contribute to human life [6].

To date, baculoviruses are known to infect more than 660 insect species; most of them are belonging to the order of Lepidoptera, Diptera and Hymenoptera [7, 8]. Baculoviruses exhibit genetic variations among species and its isolates [9]. Although a large number of baculoviruses in the nature, only a few have been well studied. To the best of our knowledge, a total of 78 fully sequenced genomes have been deposited in GenBank [10] and also several baculoviruses of whole genomes may soon be sequenced and deposited (**Table 1**). However, these published viral genomes represent only a small fraction and the genetic relationship among nucleopolyhedroviruses (NPVs) in the natural environment remains a puzzle.

Previously, Sanger sequencing was employed to sequence the viral genomic sequences cloned in plasmids. With the advances of sequencing technologies, next-generation sequencing (NGS) is becoming an important technology for large-scale viral genomic sequencing. The high cost of NGS and requirement of intensive bioinformatic analysis remain a hurdle for this application. In a word, NGS is an available tool to facilitate on the study of the genetic relationship of baculoviruses.

2. Identification of NPVs

Biochemical and biotechnology-based methods are the most common approaches employed to identify the NPVs. In most cases, more than one method is employed to compensate the pros and cons for each other. For example, restriction enzyme profiling of viral genomic DNA was used to reveal genetic variations among different isolates [97–99] and to distinguish one species from another between closely related viruses such as *Rachiplusia ou* (RoMNPV), AcMNPV, *Trichoplusia ni* (TnMNPV), *Galleria mellonella* (GmMNPV) [100, 101] and the MNPVs of *Spodoptera frugiperda* [102].

Polymerase chain reaction (PCR)-based methods were then established. These methods have been shown not only to be more sensitive and faster but also more reliable than restriction enzyme analysis for classifying baculoviral species [4, 103–105]. Multiple genetic markers (e.g., *egt*, *ac17*, *lef-2*, *polh*, *p35*, *pif-2*) could be used for the identification of baculoviruses [7, 106–109]. The *late expression factor 8* (*lef-8*), *late expression factor 9* (*lef-9*) and *polyhedrin*

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
Alphabaculovirus (Group I)	*Anticarsia gemmatalis* MNPV	AgMNPV-2D	NC_008520	132,239	36,623	29,338	29,513	36,765	44.5%	158	Sanger	PHRED/ ALIGNER	[11]
		AgMNPV-26	KR815455	131,678	36,411	29,288	29,405	36,574	44.6%	157	Roche 454 GS FLX	Geneious	[12]
		AgMNPV-27	KR815456	131,172	36,273	29,176	29,331	36,392	44.6%	157			
		AgMNPV-28	KR815457	130,745	36,185	29,018	29,242	36,300	44.6%	157			
		AgMNPV-29	KR815458	130,506	36,072	28,989	29,216	36,229	44.6%	157			
		AgMNPV-30	KR815459	130,741	36,195	29,011	29,173	36,362	44.5%	156			
		AgMNPV-31	KR815460	132,126	36,543	29,363	29,564	36,656	44.6%	158			
		AgMNPV-32	KR815461	131,494	36,341	29,234	29,384	36,535	44.6%	157			
		AgMNPV-33	KR815462	131,059	36,322	29,114	29,244	36,379	44.5%	157			
		AgMNPV-34	KR815463	131,543	36,435	29,233	29,383	36,492	44.6%	158			
		AgMNPV-35	KR815464	132,176	36,552	29,384	29,558	36,682	44.6%	159			
		AgMNPV-36	KR815465	131,216	36,293	29,127	29,270	36,526	44.5%	156			
		AgMNPV-37	KR815466	131,855	36,531	29,255	29,400	36,669	44.5%	156			
		AgMNPV-38	KR815467	130,740	36,194	29,012	29,172	36,362	44.5%	156			
		AgMNPV-39	KR815468	130,698	36,219	29,026	29,184	36,269	44.5%	157			
		AgMNPV-40	KR815469	132,180	36,542	29,409	29,583	36,646	44.6%	158			
		AgMNPV-42	KR815470	130,949	36,274	29,098	29,275	36,302	44.6%	157			
		AgMNPV-43	KR815471	132,077	36,539	29,369	29,528	36,641	44.6%	159			
	Antheraea pernyi NPV	AnpeNPV	NC_008035	126,629	29,513	34,041	33,664	29,406	53.5%	147	Sanger	ContigExpress9.1.0 + SeqMan5.0/ DNASTAR	[13]
	Autographa californica MNPV	AcMNPV	NC_001623	133,894	39,195	27,151	27,347	40,201	40.7%	156	Sanger	GCG package	[14]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
	Autographa californica MNPV-WP10	AcMNPV-WP10	KM609482	133,926	39,205	27,157	27,346	40,199	40.7%	151	Illumina HiSeq 2000	Newbler	[15]
	Bombyx mandarina NPV	BomaNPV	NC_012672	126,770	37,358	25,398	25,601	38,413	40.2%	141	Solexa GA	GENETYX-win Software + DNASTAR	[16]
	Bombyx mori NPV	BmNPV	NC_001962	128,413	37,747	25,828	26,056	38,782	40.4%	143	Sanger	DNASIS/PROSIS	[17]
	Catopsilia pomona NPV	CapoNPV	KU565883	128,058	38,938	25,348	25,444	38,328	39.7%	131	Roche 454 GS FLX+	GS *de novo* assembler	[10]
	Choristoneura fumiferana DEF MNPV	CfDEFMNPV	NC_005137	131,160	35,474	30,110	29,993	35,580	45.8%	149	Sanger	MacVector + Lasergene/ DNASTAR	[18]
	Choristoneura fumiferana MNPV	CfMNPV	NC_004778	129,593	32,224	32656	32,261	32,452	50.1%	146	Sanger	Gene Runner	[19]
	Choristoneura murinana NPV	ChmuNPV	NC_023177	124,688	31,408	30,986	31,370	30,924	50.0%	147	Roche 454	CLC Genomics Workbench	[20]
	Choristoneura occidentalis NPV	ChocNPV	NC_021925	128,446	32,108	31,905	32,481	31,952	50.1%	148	Roche 454 GS FLX	SeqMan Pro Lasergene/ DNASTAR	[21]
	Choristoneura rosaceana NPV	ChroNPV	NC_021924	129,052	33,309	31,261	31,425	33,057	48.6%	149			
	Condylorrhiza vestigialis MNPV	CoveMNPV	NC_026430	125,767	35,904	26,937	27,038	35,886	42.9%	138	Roche 454	Geneious + MIRA	[22]
	Dasychira pudibunda NPV	DapuNPV	KP747440	136,761	31,022	37,008	37,454	31,277	54.4%	161	Illumina MiSeq	Geneious	[23]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
	Ectropis obliqua NPV	EcobNPV	NC_008586	131,204	40,683	24,676	24,708	41,137	37.6%	126	Sanger	Genetyx-win	[24]
	Hyphantria cunea NPV	HycuNPV	NC_007767	132,959	36,031	30,039	30,465	36,424	45.5%	148	RISA-384	DNASIS	[25]
	Lonomia obliqua MNPV	LoobMNPV	KP763670	120,023	38,995	20,932	21,966	38,104	35.7%	134	Roche 454 GS FLX	Geneious	[26]
	Maruca vitrata MNPV	MaviMNPV	NC_008725	111,953	34,041	21,669	21,563	34,680	38.6%	126	Sanger	PHRED/PHRAP	[27]
	Orgyia pseudotsugata MNPV	OpMNPV	NC_001875	131,995	29,463	36,477	36,295	29,758	55.1%	152	Sanger	GCG package	[28]
	Philosamia cynthia ricini NPV	PhcyNPV	JX404026	125,376	28,966	33,461	33,809	29,140	53.7%	138	Sanger	N/A[1]	[29]
	Plutella xylostella MNPV	PlxyMNPV	NC_008349	134,417	39,437	27,303	27,396	40,281	40.7%	152	Sanger	Lasergene/ DNASTAR	[30]
	Rachiplusia ou MNPV	RoMNPV	NC_004323	131,526	39,674	25,630	25,793	40,429	39.1%	149	Sanger	Wisconsin package + Lasergene/ DNASTAR	[31]
	Thysanoplusia orichalcea NPV	ThorNPV	NC_019945	132,978	40,022	26,388	26,142	40,426	39.5%	145	Solexa GA	Edena	[32]
Alphabaculovirus (Group II)	*Adoxophyes honmai* NPV	AdhoNPV	NC_004690	113,220	36,505	20,025	20,328	36,362	35.6%	125	RISA-384	PHRED/PHRAP	[33]
	Adoxophyes orana NPV	AdorNPV	NC_011423	111,724	36,306	19,404	19,694	36,320	35.0%	121	Sanger	SeqMan II Lasergene/ DNASTAR	[34]
	Agrotis ipsilon MNPV	AgipMNPV	NC_011345	155,122	40,201	37,490	7,860	39,571	48.6%	163	Sanger	Lasergene/ DNASTAR	[35]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
	Agrotis segetum NPV	AgseNPV	NC_007921	147,544	40,237	33,200	34,247	39,860	45.7%	153	Sanger	Gap4	[36]
	Agrotis segetum NPV B	AgseNPV-B	NC_025960	148,981	40,490	33,698	4,371	40,422	45.7%	150	Roche 454	DNASTAR	[37]
	Apocheima cinerarium NPV	ApciNPV	NC_018504	123,876	41,223	20,865	20,449	41,332	33.4%	117	Sanger	SeqMan Pro Lasergene/ DNASTAR	unpublished
	Buzura suppressaria NPV	BusuNPV	NC_023442	120,420	37,568	22,152	22,142	38,558	36.8%	127	Roche 454 GS FLX	GS de novo assembler	[38]
	Chrysodeixis chalcites NPV	ChchNPV	NC_007151	149,622	45,151	29,304	29,060	46,107	39.0%	151	Sanger	Gap4	[39]
	Chrysodeixis chalcites SNPV	ChchSNPV-TF1-A	JX535500	149,684	45,090	29,324	29,133	46,137	39.1%	150	Roche 454	Newbler	[40]
		ChchSNPV-TF1-C	JX560539	150,079	45,146	29,384	29,096	46,447	39.0%	150			
		ChchSNPV-TF1-B	JX560540	149,080	44,989	29,152	28,987	45,952	39.0%	150			
		ChchSNPV-TF1-G	JX560541	149,039	45,075	29,136	28,869	45,958	38.9%	151			
		ChchSNPV-TF1-H	JX560542	149,624	45,162	29,285	29,034	46,143	39.0%	150			
	Clanis bilineata NPV	ClbiNPV	NC_008293	135,454	41,557	25,560	25,558	42,779	37.7%	129	Sanger	N/A	[41]
	Epiphyas postvittana NPV	EppoNPV	NC_003083	118,584	35,221	24,287	23,956	35,120	40.7%	136	Sanger	DNASTAR	[42]
	Euproctis pseudoconspersa NPV	EupsNPV	NC_012639	141,291	41,736	28,455	28,549	42,551	40.3%	139	Sanger	Wisconsin package + GENETYX-win	[43]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
	Helicoverpa armigera SNPV AC53	HaSNPV-AC53	NC_024688	130,442	39,121	25,389	25,606	40,326	39.1%	138	Ion Torrent PGM	CLC Genomics Workbench	[44]
	Helicoverpa armigera MNPV	HearMNPV	NC_011615	154,196	46,371	30,731	31,060	46,031	40.1%	162	Sanger	SeqMan 5.0/ DNASTAR	[45]
	Helicoverpa armigera NPV	HearNPV	NC_003094	130,759	39,345	25,340	25,552	40,522	38.9%	137	Sanger	Wisconsin package + Lasergene/ DNASTAR	[46, 47]
	Helicoverpa armigera NPV G4	HearNPV-G4	NC_002654	131,405	39,529	25,530	25,738	40,608	39.0%	135	Sanger	PHRED/PHRAP	[48]
	Helicoverpa armigera NPV NNg1	HearNPV-NNg1	NC_011354	132,425	39,754	25,791	26,054	40,826	39.2%	143	RISA-384	DNASIS	[49]
	Helicoverpa zea SNPV	HzSNPV	NC_003349	130,869	39,273	25,471	25,675	40,450	39.1%	139	Sanger	Wisconsin package + Lasergene/ DNASTAR	[50]
	Hemileuca sp. NPV	HespNPV	NC_021923	140,633	42,827	26,977	26,595	44,234	38.1%	137	Sanger	Wisconsin package + Lasergene/ DNASTAR	[51]
	Lambdina fiscellaria NPV	LafiNPV	NC_026922	157,977	45,363	34,616	34,350	43,648	43.7%	137	Roche 454	CLC Genomics Workbench	[52]
	Leucania separata NPV	LeseNPV	NC_008348	168,041	42,546	40,683	40,927	43,885	48.6%	169	MegaBACE 1000	DNASTAR	[53]
	Lymantria dispar MNPV	LdMNPV	NC_001973	161,046	34,229	46,226	46,331	34,260	57.5%	164	Sanger	GCG package	[54]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
	Lymantria dispar MNPV-27	LdMNPV-27	KP027546	164,158	35,020	47,133	47,118	34,887	57.4%	162	Illumina ZMiSeq	CLC Genomics Workbench	[55]
	Lymantria dispar MNPV-BNP	LdMNPV-BNP	KU377538	157,270	38,788	39,579	39,567	39,336	50.3%	154	Illumina MiSeq	Geneious	[56]
	Lymantria dispar MNPV-2161	LdMNPV-2161	KF695050	163,138	34,855	46,648	46,812	34,823	57.3%	174	Roche 454 GS Junior	SeqMan NGEN Lasergene/ DNASTAR	[9]
	Lymantria dispar MNPV-3029	LdMNPV-3029	KM386655	161,712	34,321	46,434	46,457	34,500	57.4%	163	Roche 454	Lasergene/ DNASTAR	[57]
	Lymantria dispar MNPV-45	LdMNPV-45	KU862282	161,006	34,234	46,192	46,314	34,264	57.5%	155	Illumina	CLC Genomics Workbench	[58]
	Lymantria dispar MNPV-3054	LdMNPV-3054	KT626570	164,478	35,151	47,119	47,140	35,068	57.3%	174	Roche 454 GS Junior	LaserGene/ DNASTAR	[59]
	Lymantria dispar MNPV-3041	LdMNPV-3041	KT626571	162,658	34,715	46,478	46,647	34,818	57.3%	178			
	Lymantria dispar MNPV-Ab-a624	LdMNPV-Ab-a624	KT626572	161,321	34,282	46,302	46,405	34,332	57.5%	176			
	Lymantria xylina MNPV	LyxyMNPV	NC_013953	156,344	36,207	41,674	41,933	36,530	53.5%	157	Sanger	PHRED/PHRAP	[60]
	Mamestra brassicae MNPV	MabrMNPV	NC_023681	152,710	46,042	30,311	30,604	45,753	39.9%	159	Roche 454	GS de novo assembler	[61]
	Mamestra configurata NPV-A	MacoNPV-A	NC_003529	155,060	45,336	32,160	32,463	45,101	41.7%	169	Sanger	Wisconsin package + Lasergene/ DNASTAR	[62]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
	Mamestra configurata NPV-B	MacoNPV-B	NC_004117	158,482	47,831	31,504	31,953	47,194	40.0%	168	Sanger	Sequencher 4.0	[63]
	Orgyia leucostigma NPV	OrleNPV	NC_010276	156,179	46,420	31,270	31,020	47,469	39.9%	135	Sanger	Agencourt BioScience	[64]
	Peridroma NPV	PespNPV	NC_024625	151,109	35,060	40,593	39,822	35,633	53.2%	139	Roche 454	CLC Genomics Workbench	[65]
	Perigonia lusca single NPV	PeluNPV	NC_027923	132,831	39,968	26,167	26,362	40,256	39.6%	145	Roche 454	Geneious	unpublished
	Pseudoplusia includens SNPV	PsinNPV	NC_026268	139,132	41,843	27,452	27,210	42,609	39.3%	141	Roche 454 GS FLX	MIRA	[66]
	Spodoptera exigua MNPV	SeMNPV	NC_002169	135,611	38,445	29,486	29,929	37,751	43.8%	139	Sanger	Wisconsin package + Lasergene/ DNASTAR	[67]
	Spodoptera frugiperda MNPV virus	SfMNPV	NC_009011	131,331	39,417	26,346	26,507	39,061	40.2%	143	Sanger	Lasergene/ DNASTAR	[68]
	Spodoptera litura MNPV	SpliMNPV-AN1956	JX454574	137,998	37,469	30,803	30,846	38,880	44.7%	132	Roche 454 GS Junior	LaserGene/ DNASTAR	[69]
	Spodoptera litura NPV	SpltNPV	NC_003102	139,342	39,180	29,691	29904	40,567	42.8%	141	MegaBACE1000	DNASIS + DNASTAR	[70]
	Spodoptera litura NPV II	SpltNPV-II	NC_011616	148,634	40,998	33,210	33,671	40,755	45.0%	147	n/a	N/A	unpublished
	Sucra jujuba NPV	SujuNPV	KJ676450	135,952	41,395	26,157	26,399	42,001	38.7%	131	Roche 454	GS de novo assembler	[71]
	Trichoplusia ni SNPV	TnSNPV	NC_007383	134,394	40,601	6,256	26,117	41,384	39.0%	145	Sanger	PHRED/PHRAP	[72]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
Betabaculovirus	*Adoxophyes orana granulovirus*	AdorGV	NC_005038	99,657	33,077	17,098	17,275	32,207	34.5%	119	Sanger	SeqMan II Lasergene/ DNASTAR	[73]
	Agrotis segetum granulovirus	AgseGV	NC_005839	131,680	41,892	25,179	23,953	40,656	37.3%	132	n/a	n/a	unpublished
	Clostera anastomosis GV isolate Henan	ClasGV-A	NC_022646	101,818	27,115	23,832	23,739	27,132	46.7%	122	Illumina GA	SOAPdenovo	[74]
	Clostera anastomosis granulovirus-B	ClasGV-B	KR091910	107,439	33,648	19,904	20,673	33,214	37.8%	123	Roche 454 GS FLX	Newbler	[75]
	Cnaphalocrocis medinalis GV	CnmeGV	NC_029304	111,246	36,021	19,756	19,385	36,084	35.2%	118	Roche 454 GS FLX	GS de novo assembler	[76]
	Cnaphalocrocis medinalis granulovirus	CnmeGV	KP658210	112,060	36,295	19,904	19,529	36,332	35.2%	133	PacBio RS II	HGAP2.2.0	[77]
	Choristoneura occidentalis GV	ChocGV	NC_008168	104,710	36,132	17,268	16,938	34,372	32.7%	116	Sanger	PHRED/PHRAP	[78]
	Clostera anachoreta granulovirus	ClanGV	NC_015398	101,487	28,188	22,554	22,523	28,222	44.4%	123	Illumina GA	SOAPdenovo	[79]
	Cydia pomonella granulovirus	CpGV	NC_002816	123,500	34,029	27,722	28,183	33,566	45.3%	143	Sanger	Wisconsin package + Lasergene/ DNASTAR	[80]
	Cryptophlebia leucotreta granulovirus	CrleGV	NC_005068	110,907	38,095	18,090	17,890	36,832	32.4%	128	Sanger	Lasergene/ DNASTAR	[81]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
	Diatraea saccharalis granulovirus	DisaGV	NC_028491	98,392	32,133	17,032	17,337	31,880	34.9%	125	Roche 454	Geneious	[82]
	Epinotia aporema granulovirus	EpapGV	NC_018875	119,082	35,524	24,984	24,403	34,171	41.5%	132	Roche 454 GS FLX	Newbler	[83]
	Erinnyis ello granulovirus	ErelGV	NC_025257	102,759	31,707	19,440	20,324	31,288	38.7%	130	Roche 454 GS FLX	Geneious	[84]
	Helicoverpa armigera granulovirus	HearGV	NC_010240	169,794	50,336	34,518	34,810	50,130	40.8%	179	Sanger	SeqMan Lasergene/ DNASTAR	[85]
	Plodia interpunctella granulovirus	PiGV	KX151395[2]	112,536	n/a	n/a	n/a	n/a	n/a	123	Roche 454 GS Junior	SeqMan NGEN Lasergene/ DNASTAR	[86]
	Phthorimaea operculella granulovirus	PhopGV	NC_004062	119,217	38,306	21,127	21,431	38,353	35.7%	130	Sanger	N/A	[87]
	Plutella xylostella granulovirus	PlxyGV	NC_002593	100,999	30,252	20,546	20,546	29,655	40.7%	120	DSQ-1000 L	GENETYX-win	[88]
	Pieris rapae granulovirus	PrGV	NC_013797	108,592	36,619	17,863	18,168	35,942	33.2%	120	Sanger	NN/A	[89]
	Pseudaletia unipuncta granulovirus	PsunGV	NC_013772	176,677	53,572	34,993	35,311	52,799	39.8%	183	n/a	N/A	unpublished
	Spodoptera frugiperda GV isolateVG008	SpfrGV	NC_026511	140,913	38,131	32,852	32,288	37,642	46.2%	146	Roche 454 GS FLX	Newbler	[90]

Genus	Virus	Virus Abbreviation	GenBank accession	Genome size (bp)	A	C	G	T	GC content	ORFs	Sequencing	Assembler	Reference
	Spodoptera litura granulovirus	SpliGV	NC_009503	124,121	38,360	23,813	24,377	37,571	38.8%	136	Sanger	N/A	[91]
	Xestia c-nigrum granulovirus	XcGV	NC_002331	178,733	53,166	36,079	36,627	52,861	40.7%	181	Sanger	DNASIS/PROSIS	[92]
Gammabaculovirus	*Neodiprion abietis* NPV	NeabNPV	NC_008252	84,264	28,292	13,948	14,177	27,847	33.4%	93	Sanger	PHRED/PHRAP	[93]
	Neodiprion lecontei NPV	NeleNPV	NC_005906	81,755	27,741	13,596	13,640	26,616	33.4%	89	Sanger	SeqMan Lasergene/ DNASTAR	[94]
	Neodiprion sertifer NPV	NeseNPV	NC_005905	86,462	29,158	14,444	14,745	28,115	33.8%	90	Sanger	Sequencher 4.1	[95]
Deltabaculovirus	*Culex nigripalpus* NPV	CuniNPV	NC_003084	108,252	26,623	27,228	27,839	26,562	50.9%	109	Sanger	CAP3	[96]

N/A: no information is available either in the paper or GenBank file.
The GenBank file with accession number KX1513952 is not available in GenBank website.

Table 1. List of sequenced baculoviruses genomes.

http://dx.doi.org/10.5772/66634

(*polh*) were found in a highly conserved genes among baculoviruses [110], therefore, used as targets for degenerating PCR to characterize lepidopteran NPVs through the amplification of the conserved regions from a variety range of baculoviruses [111–113]. The Kimura 2-parameter (K-2-P) distances between the aligned *polh*/*gran*, *lef-8* and *lef-9* nucleotide sequences were described by Jehle et al. for baculoviruses identification and species classification [3]. The K-2-P nucleotide substitution model from aligned nucleotide sequences were determined by using the pairwise distance calculation of MEGA version 3.0 applying the Kimura 2-parameter model [114].

Due to the higher cost of NGS for viral genome sequencing, it is frequently required to combine various approaches to cut down the cost but still ensure precision, e.g., PCR-based K-2-P analysis and NGS approach for identifying the potential new NPV species. Two NPVs were isolated from casuarina moth (*Lymantria xylina*) and golden birdwing larvae (*Troides aeacus*) collected from the fields, respectively, will be as representative cases for explanation in the following sections. We will focus on the characterization of these two potential new NPVs first and then the use of the sequences of three genes, *lef-8*, *lef-9* and *polyhedrin* of two NPV candidates was used to examine their taxonomic position by K-2-P analysis. Finally, we will focus on the genome sequencing technology and bioinformatic analysis on NPVs.

3. The identification of ambiguous NPVs

In this section, the discussion of molecular identification of NPV species based on K-2-P distance [3] is presented. Two new NPVs were used as examples in this study to reveal different issues regarding the classification of NPVs.

3.1. LdMNPV-like virus

The K-2-P distances, based on the sequences of three genes, between different viruses could mostly evaluate the ambiguous relationship among the NPVs. It was defined that distances less than 0.015 indicates that the two isolates are the same baculovirus species. On the other hand, the difference between two viruses is more than 0.05 should be considered as different virus species. For the distances between 0.015 and 0.05, complementary information is needed to determine whether these two viruses are of the same or different species [3, 9, 115].

A new multiple nucleopolyhedrovirus strain was isolated from casuarina moth, *L. xylina* Swinhoe, (Lepidoptera: Lymantriidae) in Taiwan. Since the *polyhedrin* sequence of this virus had high identity to *L. dispar* MNPV (98%), it was named LdMNPV-like virus [116]. To precisely clarify the relationship of three Lymantriidae-derived NPVs (LdMNPV-like virus, LdMNPV and LyxyMNPV [60]), the K-2-P of *polh*, *lef-8* and *-9* was performed. The distances between LdMNPV-like virus and LyxyMNPV exceeded 0.05 for each gene, *polh*, *lef-8*, or *lef-9* and also for concatenated *polh*/*lef-8*/*lef-9* (**Figure 1**). For LdMNPV-like virus and LdMNPV, not only the single *lef-8* and *lef-9* sequences but also concatenated *polh*/*lef-8*/*lef-9*, the distances were generally lower than 0.015, but only the *polh* sequence distance (0.016) exceeded slightly

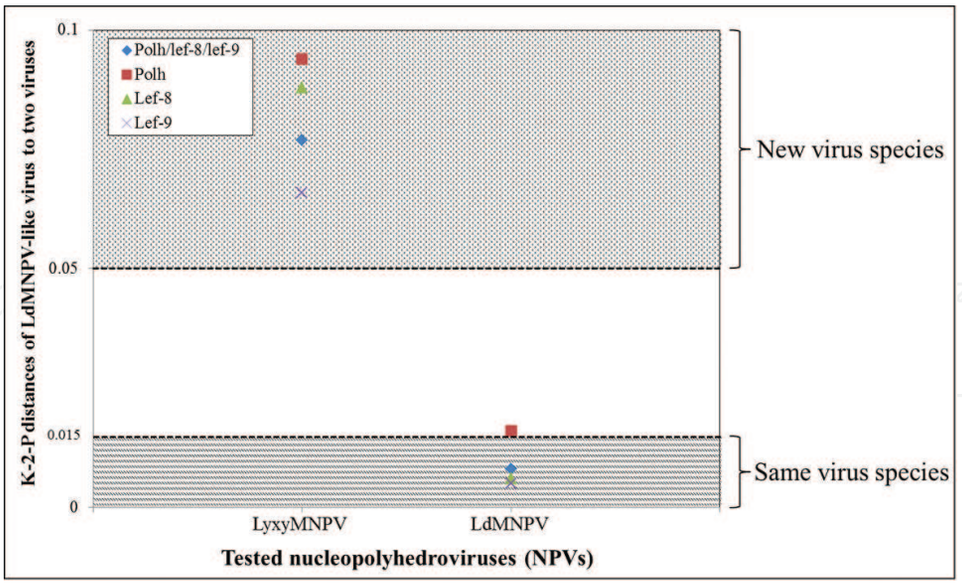

Figure 1. Pairwise K-2-P distances of the nucleotide sequences of *polh*, *lef-8* and *lef-9* and concatenated *polh/lef-8/lef-9* fragments of LdMNPV-like virus, LyxyMNPV and LdMNPV. Modified data reproduced with permission of the Elsevier [116].

0.015 (**Figure 1**). These results strongly suggested that LdMNPV-like virus is an isolate of LdMNPV. However, as indicated by our previous report, the genome of LdMNPV-like virus is approximately 139 Kb, due to large deletions compared to that of LdMNPV [116]. To further investigate the LdMNPV-like virus, a *Hin*dIII-*Pst*I fragment (7,054 nucleotides) was cloned, sequenced and compared to the corresponding region of LdMNPV. Nine putative ORFs (including seven with full lengths and two with partial lengths) and two homologous regions (hrs) were identified in this fragment (**Figure 2**) and those genes, in order from the 5′ to 3′ end, encoded part of *rr1*, *ctl-1*, *Ange-bro-c*, LdOrf151, LdOrf-152-like peptides, *Ld-bro-n*, two *Ld-bro-o* and part of LdOrf155-like peptides (**Table 2**). The physical map of *Hin*dIII-*Pst*I fragment of LdMNPV-like virus showed that the gene organization was highly conserved compared to the corresponding region of LdMNPV, although several restriction enzyme recognition sites were different. Additionally, the *ld-bro-o* gene in the LdMNPV-like virus was split into two ORF7 and ORF8, due to a point deletion in the downstream (+669) of ORF7 and this deletion causes a frameshift that results in the formation of a stop codon (TGA) after 73 bp. Afterward, ORF8 was overlapped with the last four base pairs (ATGA) in ORF7. The nucleotide identities of these genes were 96–100% homologous to those of LdMNPV, except ORF3 which was 68% homologous to *Ange-bro-c* and ORF7 and ORF8 showing low identities to *Ld-bro-o* (73% and 26%, respectively). The deduced amino acid sequences of these genes were similar to those of LdMNPV, with identities of 81–100%, except the similarity of ORF3 to *Ange-bro-c* was 70% and ORF7 and ORF8 also showed low similarity to *Ld-bro-o* (67% and 26%, respectively). These results imply that the LdMNPV-like and LdMNPV viruses are closely related but not totally identical.

Based on these results, LdMNPV-like virus has a genomic size significantly smaller than that of LdMNPV and LyxyMNPV and appears to be an NPV isolate distinct from LdMNPV or LyxyMNPV. Moreover, a gene, *ange-bro-c* of LdMNPV-like virus, was truncated into two ORF7

http://dx.doi.org/10.5772/66634

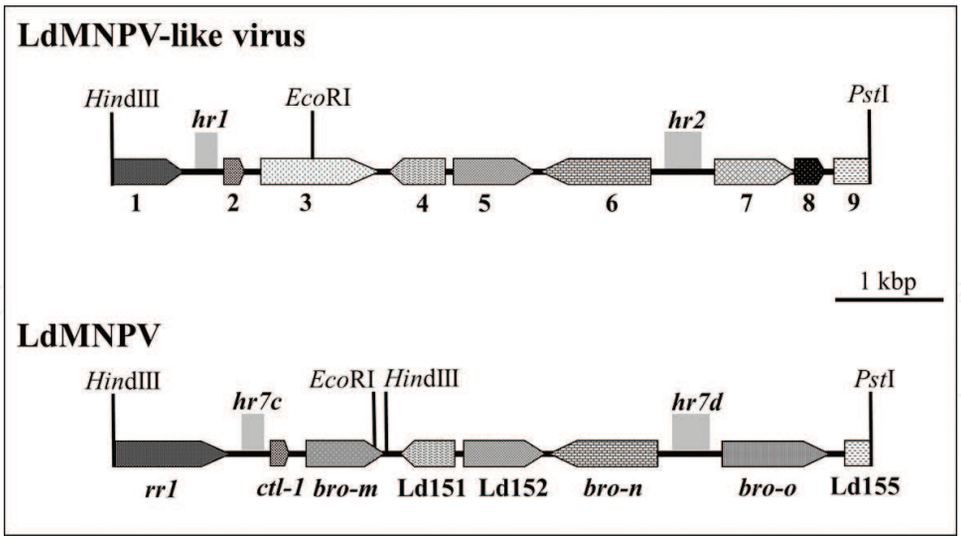

Figure 2. Comparison of relative restriction sites and gene locations in the LdMNPV-like virus *Hind*III-*Pst*I fragment with those of the corresponding LdMNPV fragment. Arrows denote ORFs and their direction of transcription. Gray boxes represent the homologous repeat regions (hrs). ORF homologues in the corresponding regions are drawn with the same patterns. Numbers below the arrows indicate the nine putative ORFs listed in **Table 2**.

and ORF8 and the sequence showed relatively low identity to that of LdMNPV (**Table 2**). Taken together, these results indicate that LdMNPV-like virus is a distinct LdMNPV strain with several novel features. Otherwise, LdMNPV-like virus and LdMNPV have distinct geographical locations (from subtropical and cold temperate zones, respectively) and are

No*	LdMNPV-like virus			LdMNPV§		
	Position†	Length		Name	Identity (%)	
		nt	aa		nt	aa
1	1 → 654	654	217	rr1	96	81
2	1063 → 1224	162	53	Ctl-1	100	100
3	1397 → 2473	1077	358	Ange-bro-c	68	70
4	2590 → 3596	504	168	LdOrf-151	99	98
5	3200 → 3952	753	251	LdOrf-152	99	99
6	4019 → 5026	1005	335	Ld-bro-n	93	91
7	5645 → 6391	744	248	Ld-bro-o	73	67
8	6388 → 6654	264	88	Ld-bro-o	26	26
9	6758 → 7054	297	99	LdOrf-155	100	100

†The directions of the transcripts are indicated by arrows.

§Reference from the genome of LdMNPV (Kuzio et al. [63])

*The nine potentially expressed ORFs are numbered in the order in which they occur in the LdMNPV-like virus genomic fragment from the 5′ to 3′ end. Two ORFs extend past this cloning site are printed in bold; only the N-terminus which contains 217 amino acids (654 nucleotides) and 99 amino acids (297 nucleotides) was examined.

Table 2. Comparison of the nucleotide (nt) and deduced amino acid (aa) sequences for putative ORFs in LdMNPV-like virus genomic fragment and their corresponding LdMNPV homologues.

distinct in genotypic and phenotypic characteristics and it also showed broad genetic variation among LdMNPV isolates [9].

3.2. An NPV isolate from *T. aeacus* larvae

A nucleopolyhedrosis disease of the rearing of the golden birdwing butterfly (*T. aeacus*) larvae was found and the polyhedral inclusion bodies (PIBs) were observed under light microscopy (**Figure 3**). PCR was performed to amply the *polh* gene by 35/36 primer set (**Figure 3**) to further confirm NPV infection [117, 118]. Therefore, this NPV was named provisionally TraeNPV. The three genes, *polh*, *lef-8* and *lef-9* of TraeNPV, were cloned and sequenced and then the K-2-P distances between the aligned single and concatenated *polh*, *lef-8* and *lef-9* nucleotide sequences were analyzed. The results indicated that TraeNPV belonged to the group I baculoviruses and closely related to BmNPV group. **Figure 4** showed that most of the distances between TraeNPV and other NPVs were between 0.015 and 0.050, whereas the distances for *polh* between TraeNPV, PlxyNPV, RoNPV and AcMNPV group exceeded 0.05. It should be noted that for all the concatenated *polh*/*lef-8*/*lef-9* sequences, the distances were apparently much more than 0.015 and even to 0.05. These results left an ambiguous situation of this NPV isolate; so far, we could conclude that TraeNPV neither belongs to BmNPV group nor AcMNPV group. More complementary information is needed to determine the viral species of TraeNPV.

In summary, K-2-P distances were employed to further clarify the relationship between closely related NPVs. We discussed two different cases analyzed by K-2-P. From the sequence data

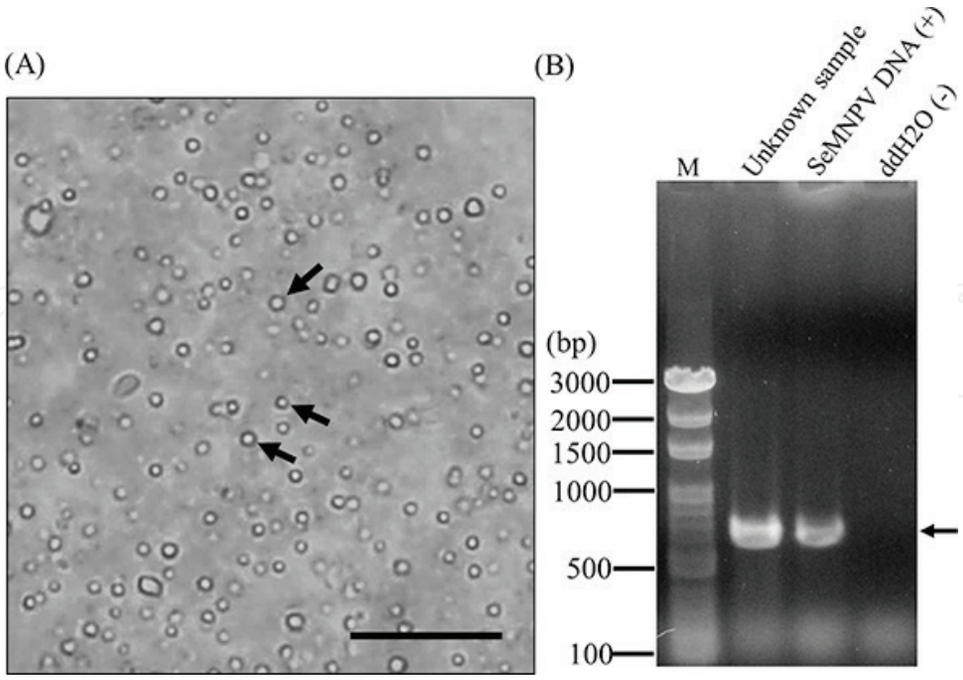

Figure 3. Identification of unknown NPV. (A) Light microscopy observation of liquefaction from the cadavers of *T. aeacus* larvae, scale bar = 20 μm. Black arrows indicated the polyhedral inclusion bodies (PIBs). (B) PCR detection of partial *polyhedrin* gene, M = 100 bp marker, (+) = positive control and (−) = negative control.

http://dx.doi.org/10.5772/66634

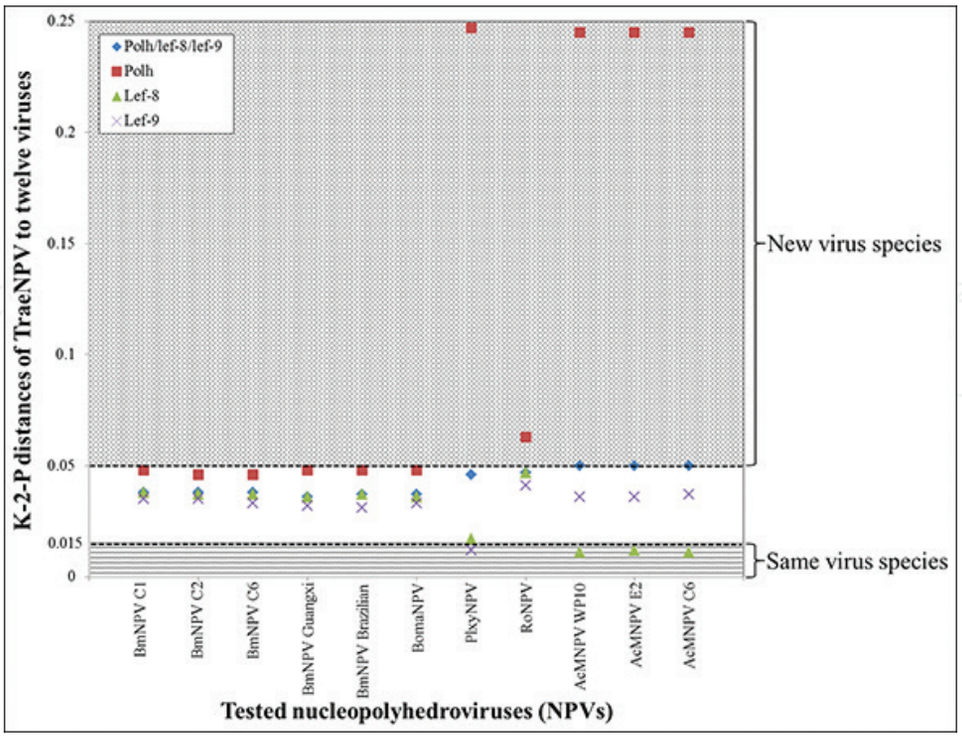

Figure 4. Pairwise Kimura-2-parameter distances of the nucleotide sequences of *lef-8*, *lef-9* and *polh* and concatenated *polh/lef-8/lef-9* fragments of TraeNPV and 12 viruses.

of LdMNPV-like virus, results strongly supported that LdMNPV-like virus is an isolate of LdMNPV. Since the RFLP profiles of the LdMNPV-like virus showed the genome of this isolate was deleted tremendously, this deletion also showed coordinately in our partial sequences of genomic DNA fragments and the results of K-2-P. The K-2-P distances between TraeNPV and BmNPV or AcMNPV were among 0.05 and 0.015. Anyway, we cannot define that this virus is a new species with the evidences of RFLP, part gene sequences and K-2-P results; therefore, it is necessary to get more data, especially the whole genome sequence of TraeNPV.

4. The importance of whole genome sequencing on baculoviruses

The rapidly growing mass of genomic data shifts the taxonomic approaches from traditional to genomically based issues. The K-2-P distance supported LyxyMNPV as a different viral species (K-2-P values = 0.067–0.088), even though they were still a closely relative species phylogenetically. But, "how different did LyxyMNPV and LdMNPV?" become another question. Thus, the whole genome sequence could provide deep information of this virus. For example, as the genomic data revealed, the most part of the ORF (151 ORFs) between LyxyMNPV and LdMNPV was quite similar while still have several different ORF exhibits or absent in LyxyMNPV, e.g., two ORFs were homologous to other baculoviruses and four unique ORFs were identified in the LyxyMNPV genome and LdMNPV contains 23 ORFs that are absent in LyxyMNPV [60]. Besides, there is a huge genomic inversion in LyxyMNPV compared to LdMNPV [60]. Another example is *Maruca vitrata* NPV (MaviNPV). All of the K-2-P distance-

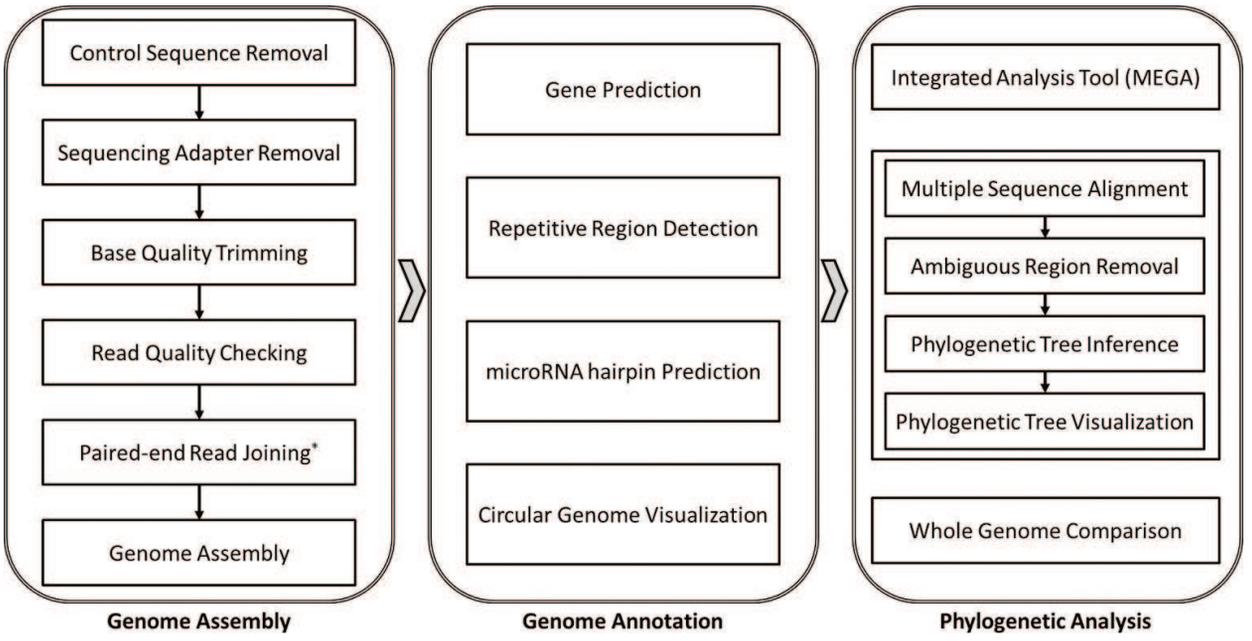

Figure 5. Pairwise Kimura-2-parameter distances of the nucleotide sequences of lef-8, lef-9 and polh and concatenated polh/lef-8/lef-9 fragments of MaviNPV and 12 viruses.

supported MaviNPV is quite different from other NPVs (K-2-P values = 0.092–0.237) (**Figure 6**). While the gene content and gene order of MaviNPV were highly similar to that of AcMNPV and BmNPV, through the genomic sequencing, it showed the 100% collinear to AcMNPV [27] and MaviNPV shared 125 ORFs with AcMNPV and 123 with BmNPV. The detailed information could only be captured after whole genome sequencing rather than partial gene

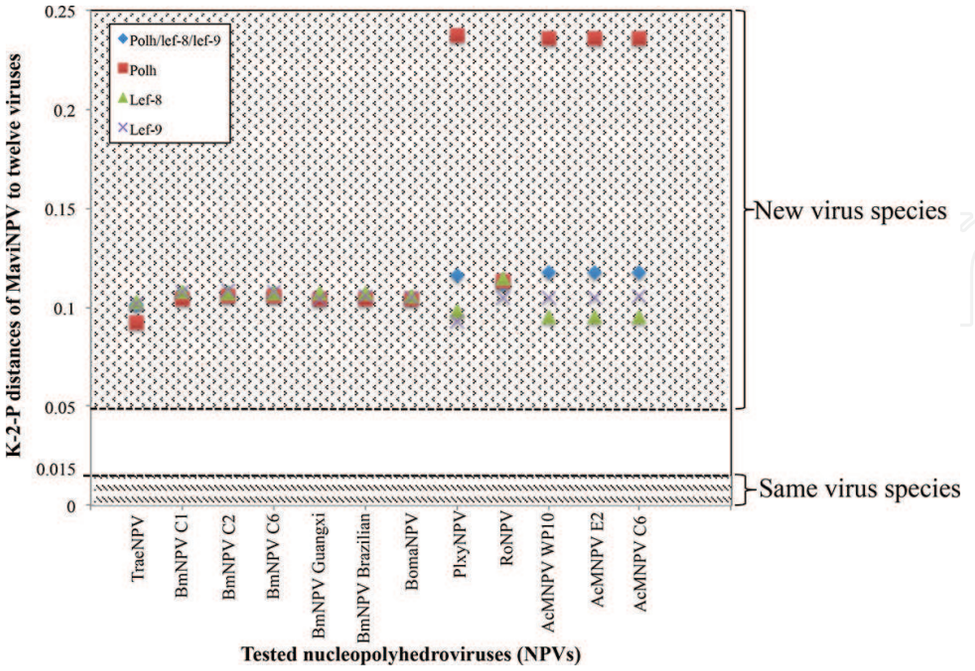

Figure 6. Common bioinformatic workflow for genome assembly and analysis.

http://dx.doi.org/10.5772/66634

sequences or other phylogenetic analyses. Sometimes, usage of K-2-P data may raise other problems, which we mentioned above; it seems LdMNPV-like virus and LdMNPV were the same viral species. While through the restriction enzyme profile and partial genomic data, we could identify that there are some deletion fragments and different gene contents within the LdMNPV-like virus genome. For the TraeNPV, most of the K-2-P values were ranged from 0.015 to 0.05; thus, whole genome sequencing could be one of the best ways to figure out this ambiguous state. The more detailed information we can get, the more deep aspect we can evaluate, e.g., the taxonomic problems and further evolutionary studies.

5. Genomic sequences of NPVs

5.1. Genome sequencing technology

Previous NPV genome sequencing employed three types of approaches: plasmid clone (or template) enrichment, NGS, or a combination of the two methods. Initially, the most common approach used restriction enzymes to fragmentize the viral genome into smaller pieces. Plasmid-based clone amplification was then employed to enrich templates for sequencing. Later, conventional Sanger sequencing and/or next-generation sequencing was employed for genome assembly. In addition, purely high-throughput sequencing-based approach from isolated viral genome was also employed [9, 15]. To date, next-generation sequencing technology plays an increasingly important role on viral genome assembly. Previous researches showed that Illumina HiSeq has superior performance in yield than 454 FLX [119–121]. Baculoviruses usually contain a novel homologous region (hr) feature, which comprises a palindrome that is usually flanked by short direct repeats located elsewhere in the genome [122]. Thereby, the shorter single-read length of Illumina sequencers might lead the difficulty during genome assembly. Further application of paired-end read sequencing method could certainly provide alternative for sequencing overlap the hrs in baculoviral genomes.

5.2. Bioinformatic analysis

Construction of a complete genome map is essential for future genomic investigations. Besides sequencing, bioinformatic approaches are also required for determining the order and content of the nucleotide sequence information for the viral genome of interest. In general, bioinformatic approaches can be separated into three consecutive steps: genome assembly, genome annotation and phylogenetic relationship inference (**Figure 5**).

5.2.1. Genome assembly

Sequence reads are the building blocks for genome sequencing and assembly. Thus, quality control of sequence reads plays a key role in determining the fidelity of a genome assembly. The procedure of read quality checking includes, but not limited to, the removal of unrelated sequences such as control sequences, adaptors, vectors, potential contaminants, etc., trimming of low-quality bases and selection of high-quality reads. The control sequences (e.g., PhiX control reads in Illumina sequencers, control DNA beads in Roche 454 sequencer) are routinely

used by sequencer manufacturers to evaluate the quality of each sequencing run. There are software applications made available to be utilized to identify and remove control sequences and low-quality bases. For NGS, sequencing adapters could be identified in reads if the fragment size is shorter than read length. Cutadapt [123] was implemented to trim the adapter sequences. Ambiguous bases or bases with lower-quality values can be removed by PRINSEQ [124] from either 5′ or 3′ end. NGS QC Toolkit [125] has programmed module to select high-quality reads. If paired-end technology was applied, paired-end reads could be joined by PANDAseq [126], PEAR [127], FLASH [128] and COPE [129], if a fragment size is shorter than read length.

Genome can be assembled from quality paired-end or single-end reads with de novo or reference-guided approaches. There are two standard methods known as the de Bruijn graph (DBG) approach and the overlap/layout/consensus (OLC) approach for de novo genome assembly. The idea of de Bruijn graph is to decompose a read into kmer-sized fragments with sliding window screening. Each kmer-sized fragment will be used to construct graph for longer path (e.g., contigs). Then, long-range paired reads can be utilized to build scaffolds from contigs with given insert size and read orientation. SOAPdenovo [130] is one of the DBG assembler that has an extreme speed by utilizing threads parallelization [131]. The OLC assembler starts by identifying all pairs of reads with higher overlap region to construct an overlap graph. The contig candidates are identified by pruning nodes to simplify the overlap graph. The final contigs are then output based on consensus regions. Additionally, Newbler [132] is a widely used OLC assembler distributed by 454 Life Sciences.

Reference-guided genome assembly is another solution for genome assembly if the genome of a closely related species is already available. For viral genome assembly, closely related species can be identified by mapping quality reads against sequenced viral genomes deposited in GenBank (http://www.ncbi.nlm.nih.gov/genome/viruses/) and select top-ranked species as the reference genome(s) to facilitate the assembly of the genome of interest. Reference-guided assembler is also called mapping assembler that the complete genome is generated by mapping quality reads with variant (single nucleotide polymorphism (SNP), insertion and deletion) identification. For example, MIRA (a computer program) [133] can create a reference-based assembly by detecting the difference between references.

During the assembly process, gap filling (or gap elimination) is conducted to resolve the undetermined bases either by bioinformatics or other approaches such as PCR and additional sequencing. Bioinformatic approaches normally use paired-end reads to eliminate gaps. PCR coupled with Sanger sequencing is a common approach to finalize the undetermined regions [134]. In addition, Sanger sequencing can also be used for genome validation and homologous region (hr) checking.

5.2.2. Genome annotation

Annotation determines the locations of protein-coding and noncoding genes as well as the functional elements in the genome. Glimmer [135], N-SCAN [136], NCBI ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/), GeneMark [137] and VIGOR [138] are gene prediction tools for identifying protein-codivng genes in the genome. Repetitive sequence regions were detected by

RepeatMasker (http://www.repeatmasker.org/). Viral microRNA candidate hairpins can be predicted by Vir-Mir [139]. The circular map of the viral genome was generated by CGView [140].

5.2.3. Phylogenetic analysis

Phylogenetic relationship inference reveals the evolutionary distances of various, especially closely related, species. MEGA [141] was the most widely used software suite that provides the sophisticated and integrated user interface for studying DNA and protein sequence data from species and populations. Alternatively, phylogenetic relationships among species based on the complete viral genomes or functional regions could also be estimated with Clustal Omega [142]. Clustal Omega was employed for multiple sequence alignment on the complete genomes and DNA fragments, respectively. ClustalW [143] was employed to do file format conversion of multiple sequence alignment. Ambiguously aligned positions were removed by using Gblocks version 0.91b [144, 145] under default settings. Phylogenetic tree inference could be constructed by hierarchical Bayesian method (e.g., MrBayes [146]) or maximum likelihood method (e.g., RAxML [147]) to estimate phylogeny [148]. Tree was depicted with FigTree version 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/). The divergence times of different species were estimated using BEAST version 1.8 or version 2.3.2 [149]. In addition, pairwise sequence identity was determined by BLASTN (NCBI BLAST Package) [150] to analyze sequence-level variation. Also, whole genome pairwise alignment can be done by LAGAN [151]. CGView comparison tool (CCT) [152] was used to represent the block similarity among different species. Mauve [153], one of the multiple genome alignment tools, can help us to visualize the consensus sequence blocks among distant-related species.

Up to 78 baculoviruses have been reported; most of baculoviruses have a narrow host range, only infect their homogenous hosts, such as BmNPV, SpltNPV, SpeiNPV, MaviNPV and so on; LyxyNPV can infect LD and LY cell lines, while AcMNPV has a wide host range; at least 40 hosts in vitro have be found. Therefore, a new baculovirus isolate needs to define its taxonomic position and to analyze its phylogenetic relationship with a known baculovirus member.

6. Conclusion

With the accomplishment of the sequencing technologies, more NPV genomes were sequenced. So far, more than 78 baculoviruses have been fully sequenced and based on the sequencing methods, we can divide into two parts, one is sequencing by Sanger method and another is sequencing by NGS method (**Table 1**). Among these sequenced genomes, 35 genomes were sequenced by Sanger method and 43 genomes were sequenced by NGS methods. It could be expected that whole genome sequencing by NGS method would get much common in this field; however, the upcoming metagenomic era is imperative that one remains aware of and careful about the shortcomings of the information presented about the organisms that are being sequenced and that these databases can oversee neither the correctness of the organismal identifications nor of the sequences entered into the databases.

The natural environment harbors a large number of baculoviruses. However, only a few of them have been sequenced and studied. A lot more information related to the genetic relationship of NPVs in the natural environment is needed to facilitate our understanding of these creatures. Though NGS technology has become an important technology for viral genomic sequencing, high cost of NGS for whole viral genome sequencing remains a barrier. To reduce the cost, it is necessary to evaluate whether the newly collected NPVs are suitable for whole genome sequencing or not. Alternatively, biochemical approaches and biological tools, such as PCR-based K-2-P analysis, can be good options to facilitate the process. As expected, all these applications are anticipated to help us reveal the genetic information of unknown species, so that more detailed insights of their genetic makeup and functional composition can be obtained to help us better understand the nature of these viruses. By using the powerful sequencing technique, the metagenomic progress (e.g., transcriptome analysis of insect host), new pathogen species in the natural environment would be easier to be found in the future. With the increase of new baculoviral genomic data, improvement of bioinformatic analysis methods and further validation of biological information would generate a group of genes, which connect to the viral host range and solve the contradiction situation in the baculoviral genomics.

Acknowledgements

This research was supported by Grant 105AS-13.2.3-BQ-B1 from the Bureau of Animal and Plant Health Inspection and Quarantine, the Council of Agriculture, Executive Yuan and Grant 103-2313-B-197-002-MY3 from the Ministry of Science and Technology (MOST).

Author details

Yu-Shin Nai[1*†], Yu-Feng Huang[2†], Tzu-Han Chen[2], Kuo-Ping Chiu[2] and Chung-Hsiung Wang[3]

*Address all correspondence to: ysnai@niu.edu.tw

1 Department of Biotechnology and Animal Science, National Ilan University, Yilan City, Taiwan

2 Genomics Research Center, Academia Sinica, Taipei, Taiwan

3 Department of Entomology, National Taiwan University, Taipei, Taiwan

† These authors contributed equally to this chapter.

References

[1] Takatsuka, J., *Lymantria mathura* nucleopolyhedrovirus: identification, occurrence and genetic diversity in Iwate Prefecture, Japan. J Invertebr Pathol, 2016. **138**: pp. 1-9.

[2] Boucias, D. and Pendland, J.C., Principles of insect pathology. 1998, Boston: Kluwer Aca demic Publishers. 537p.

[3] Jehle, J.A., et al., Molecular identification and phylogenetic analysis of baculoviruses from Lepidoptera. Virology, 2006. **346**(1): pp. 180-93.

[4] Herniou, E.A., et al., The genome sequence and evolution of baculoviruses. Annu Rev Entomol, 2003. **48**: pp. 211-34.

[5] Moscardi, F., Assessment of the application of baculoviruses for control of Lepidoptera. Annu Rev Entomol, 1999. **44**: pp. 257-89.

[6] Smith, G.E., Summers, M.D. and Fraser, M.J., Production of human beta interferon in insect cells infected with a baculovirus expression vector. Mol Cell Biol, 1983. **3**(12): pp. 2156-65.

[7] Mehrvar, A., R.R.J., Veenakumari, K., Narabenchi, G.B., Molecular and biological characteristics of some geographic isolates of nucleopolyhedrovirus of *Helicoverpa armigera* (Lep.: Noctuidae). J Entomol Soc Iran, 2008. **28**(1): pp. 39-60.

[8] Murhammer, D.W., Useful tips, widely used techniques and quantifying cell metabolic behavior. Methods Mol Biol, 2007. **388**: pp. 3-22.

[9] Harrison, R.L., Keena, M.A. and Rowley, D.L., Classification, genetic variation and pathogenicity of *Lymantria dispar* nucleopolyhedrovirus isolates from Asia, Europe and North America. J Invertebr Pathol, 2014. **116**: pp. 27-35.

[10] Wang, J., et al., Genome sequencing and analysis of *Catopsilia pomona* nucleopolyhedrovirus: a distinct species in group I Alphabaculovirus. PLoS One, 2016. **11**(5): p. e0155134.

[11] Oliveira, J.V., et al., Genome of the most widely used viral biopesticide: *Anticarsia gemmatalis* multiple nucleopolyhedrovirus. J Gen Virol, 2006. **87**(Pt 11): pp. 3233-50.

[12] Brito, A.F., et al., The pangenome of the *Anticarsia gemmatalis* multiple nucleopolyhedrovirus (AgMNPV). Genome Biol Evol, 2016. **8**(1): pp. 94-108.

[13] Nie, Z.M., et al., Complete sequence and organization of *Antheraea pernyi* nucleopolyhedrovirus, a dr-rich baculovirus. BMC Genomics, 2007. **8**: pp. 248.

[14] Ayres, M.D., et al., The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. Virology, 1994. **202**(2): pp. 586-605.

[15] Chateigner, A., et al., Ultra deep sequencing of a baculovirus population reveals widespread genomic variations. Viruses, 2015. **7**(7): pp. 3625-46.

[16] Xu, Y.P., et al., Comparative analysis of the genomes of *Bombyx mandarina* and *Bombyx mori* nucleopolyhedroviruses. J Microbiol, 2010. **48**(1): pp. 102-10.

[17] Gomi, S., Majima, K. and Maeda, S., Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. J Gen Virol, 1999. **80** (Pt 5): pp. 1323-37.

[18] Lauzon, H.A., et al., Gene organization and sequencing of the *Choristoneura fumiferana* defective nucleopolyhedrovirus genome. J Gen Virol, 2005. **86**(Pt 4): pp. 945-61.

[19] de Jong, J.G., et al., Analysis of the *Choristoneura fumiferana* nucleopolyhedrovirus genome. J Gen Virol, 2005. **86**(Pt 4): pp. 929-43.

[20] Rohrmann, G.F., Erlandson, M.A. and Theilmann, D.A., Genome sequence of an alphabaculovirus isolated from *Choristoneura murinana*. Genome Announc, 2014. **2**(1): e01135-13.

[21] Thumbi, D.K., et al., Comparative genome sequence analysis of *Choristoneura occidentalis* Freeman and *C. rosaceana* Harris (Lepidoptera: Tortricidae) alphabaculoviruses. PLoS One, 2013. **8**(7): p. e68968.

[22] Castro, M.E., et al., Identification of a new nucleopolyhedrovirus from naturally-infected *Condylorrhiza vestigialis* (Guenée) (Lepidoptera: Crambidae) larvae on poplar plantations in South Brazil. J Invertebr Pathol, 2009. **102**(2): pp. 149-54.

[23] Krejmer, M., et al., The genome of *Dasychira pudibunda* nucleopolyhedrovirus (DapuNPV) reveals novel genetic connection between baculoviruses infecting moths of the Lymantriidae family. BMC Genomics, 2015. **16**: p. 759.

[24] Ma, X.-C., et al., Genome sequence and organization of a nucleopolyhedrovirus that infects the tea looper caterpillar, *Ectropis obliqua*. Virology, 2007. **360**(1): pp. 235-46.

[25] Ikeda, M., et al., Gene organization and complete sequence of the *Hyphantria cunea* nucleopolyhedrovirus genome. J Gen Virol, 2006. **87**(Pt 9): pp. 2549-62.

[26] Aragao-Silva, C.W., et al., The complete genome of a baculovirus isolated from an insect of medical interest: *Lonomia obliqua* (Lepidoptera: Saturniidae). Sci Rep, 2016. **6**: p. 23127.

[27] Chen, Y.R., et al., Genomic and host range studies of *Maruca vitrata* nucleopolyhedrovirus. J Gen Virol, 2008. **89**(Pt 9): pp. 2315-30.

[28] Ahrens, C.H., et al., The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. Virology, 1997. **229**(2): pp. 381-99.

[29] Qian, H., et al., Analysis of the genomic sequence of *Philosamia cynthia* nucleopolyhedrin virus and comparison with *Antheraea pernyi* nucleopolyhedrin virus. BMC Genomics, 2013. **14**: p. 115.

[30] Harrison, R.L. and Lynn, D.E., Genomic sequence analysis of a nucleopolyhedrovirus isolated from the diamondback moth, *Plutella xylostella*. Virus Genes, 2007. **35**(3): pp. 857-73.

[31] Harrison, R.L. and Bonning, B.C., Comparative analysis of the genomes of *Rachiplusia ou* and *Autographa californica* multiple nucleopolyhedroviruses. J Gen Virol, 2003. **84**(Pt 7): pp. 1827-42.

[32] Wang, Y.S., et al., Genome of *Thysanoplusia orichalcea* multiple nucleopolyhedrovirus lacks the superoxide dismutase gene. J Virol, 2012. **86**(21): pp. 11948-9.

[33] Nakai, M., et al., Genome sequence and organization of a nucleopolyhedrovirus isolated from the smaller tea tortrix, *Adoxophyes honmai*. Virology, 2003. **316**(1): pp. 171-83.

[34] Hilton, S. and Winstanley, D., Genomic sequence and biological characterization of a nucleopolyhedrovirus isolated from the summer fruit tortrix, *Adoxophyes orana*. J Gen Virol, 2008. **89**(Pt 11): pp. 2898-908.

[35] Harrison, R.L., Genomic sequence analysis of the Illinois strain of the *Agrotis ipsilon* multiple nucleopolyhedrovirus. Virus Genes, 2009. **38**(1): pp. 155-70.

[36] Jakubowska, A.K., et al., Genome sequence of an enhancin gene-rich nucleopolyhedrovirus (NPV) from *Agrotis segetum*: collinearity with *Spodoptera exigua* multiple NPV. J Gen Virol, 2006. **87**(Pt 3): pp. 537-51.

[37] Wennmann, J.T., Gueli Alletti, G. and Jehle, J.A., The genome sequence of *Agrotis segetum* nucleopolyhedrovirus B (AgseNPV-B) reveals a new baculovirus species within the *Agrotis baculovirus* complex. Virus Genes, 2015. **50**(2): pp. 260-76.

[38] Zhu, Z., et al., Genome sequence and analysis of *Buzura suppressaria* nucleopolyhedrovirus: a group II Alphabaculovirus. PLoS One, 2014. **9**(1): p. e86450.

[39] van Oers, M.M., et al., Genome sequence of *Chrysodeixis chalcites* nucleopolyhedrovirus, a baculovirus with two DNA photolyase genes. J Gen Virol, 2005. **86**(Pt 7): pp. 2069-80.

[40] Bernal, A., et al., Complete genome sequences of five *Chrysodeixis chalcites* nucleopolyhedrovirus genotypes from a Canary Islands isolate. Genome Announc, 2013. **1**(5).

[41] Zhu, S.Y., et al., Genomic sequence, organization and characteristics of a new nucleopolyhedrovirus isolated from *Clanis bilineata* larva. BMC Genomics, 2009. **10**: p. 91.

[42] Hyink, O., et al., Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus. J Gen Virol, 2002. **83**(Pt 4): pp. 957-71.

[43] Tang, X.D., et al., Morphology and genome of *Euproctis pseudoconspersa* nucleopolyhedrovirus. Virus Genes, 2009. **38**(3): pp. 495-506.

[44] Noune, C. and Hauxwell, C., Complete genome sequences of *Helicoverpa armigera* single nucleopolyhedrovirus strains AC53 and H25EA1 from Australia. Genome Announc, 2015. **3**(5):e01083-15..

[45] Tang, P., et al., Genomic sequencing and analyses of HearMNPV--a new Multinucleocapsid nucleopolyhedrovirus isolated from *Helicoverpa armigera*. Virol J, 2012. **9**: p. 168.

[46] Zhang, C.X., Ma, X.C. and Guo, Z.J., Comparison of the complete genome sequence between C1 and G4 isolates of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus. Virology, 2005. **333**(1): pp. 190-9.

[47] Zhang, C.X. and Wu, J.C., Genome structure and the p10 gene of the *Helicoverpa armigera* nucleopolyhedrovirus. Acta Biochim Biophys Sinica, 2001. **33**(2): pp. 179-84.

[48] Chen, X., et al., The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. J Gen Virol, 2001. **82**(Pt 1): pp. 241-57.

[49] Ogembo, J.G., et al., Comparative genomic sequence analysis of novel *Helicoverpa armigera* nucleopolyhedrovirus (NPV) isolated from Kenya and three other previously sequenced *Helicoverpa* spp. NPVs. Virus Genes, 2009. **39**(2): pp. 261-72.

[50] Chen, X., et al., Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. J Gen Virol, 2002. **83**(Pt 3): pp. 673-84.

[51] Rohrmann, G.F., Erlandson, M.A. and Theilmann, D.A., The genome of a baculovirus isolated from *Hemileuca* sp. encodes a serpin ortholog. Virus Genes, 2013. **47**(2): pp. 357-64.

[52] Rohrmann, G.F., Erlandson, M.A. and Theilmann, D.A., Genome sequence of an alphabaculovirus isolated from the Oak Looper, *Lambdina fiscellaria*, contains a putative 2-kilobase-pair transposable element encoding a transposase and a FLYWCH domain-containing protein. Genome Announc, 2015. **3**(3): e00186-15.

[53] Xiao, H. and Qi, Y., Genome sequence of *Leucania seperata* nucleopolyhedrovirus. Virus Genes, 2007. **35**(3): pp. 845-56.

[54] Kuzio, J., et al., Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. Virology, 1999. **253**(1): pp. 17-34.

[55] Kabilov, M.R., et al., Complete genome sequence of a Western Siberian *Lymantria dispar* multiple nucleopolyhedrovirus isolate. Genome Announc, 2015. **3**(2).

[56] Rabalski, L., et al., An alphabaculovirus isolated from dead *Lymantria dispar* larvae shows high genetic similarity to baculovirus previously isolated from *Lymantria monacha* – An example of adaptation to a new host. J Invertebr Pathol, 2016. **139**: pp. 56-66.

[57] Harrison, R.L. and Rowley, D.L., Complete genome sequence of the strain of *Lymantria dispar* multiple nucleopolyhedrovirus found in the gypsy moth biopesticide Virin-ENSh. Genome Announc, 2015. **3**(1):e01407-14.

[58] Martemyanov, V.V., et al., The enhancin gene: one of the genetic determinants of population variation in baculoviral virulence. Dokl Biochem Biophys, 2015. **465**: pp. 351-3.

[59] Harrison, R.L., Rowley, D.L. and Keena, M.A., Geographic isolates of *Lymantria dispar* multiple nucleopolyhedrovirus: Genome sequence analysis and pathogenicity against European and Asian gypsy moth strains. J Invertebr Pathol, 2016. **137**: pp. 10-22.

[60] Nai, Y.S., et al., Genomic sequencing and analyses of *Lymantria xylina* multiple nucleopolyhedrovirus. BMC Genomics, 2010. **11**: pp. 116.

[61] Choi, J.B., et al., Complete genomic sequences and comparative analysis of *Mamestra brassicae* nucleopolyhedrovirus isolated in Korea. Virus Genes, 2013. **47**(1): pp. 133-51.

[62] Li, Q., et al., Sequence and organization of the *Mamestra configurata* nucleopolyhedrovirus genome. Virology, 2002. **294**(1): pp. 106-21.

[63] Li, L., et al., Complete comparative genomic analysis of two field isolates of *Mamestra configurata* nucleopolyhedrovirus-A. J Gen Virol, 2005. **86**(Pt 1): pp. 91-105.

[64] Thumbi, D.K., et al., Complete sequence, analysis and organization of the *Orgyia leucostigma* nucleopolyhedrovirus genome. Viruses, 2011. **3**(11): pp. 2301-27.

[65] Rohrmann, G.F., Erlandson, M.A. and Theilmann, D.A., A distinct group II alphabaculovirus isolated from a *Peridroma* species. Genome Announc, 2015. **3**(2):e00185-15.

[66] Craveiro, S.R., et al., The genome sequence of *Pseudoplusia* includens single nucleopolyhedrovirus and an analysis of p26 gene evolution in the baculoviruses. BMC Genomics, 2015. **16**: p. 127.

[67] WF, I.J., et al., Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. J Gen Virol, 1999. **80** (Pt 12): pp. 3289-304.

[68] Harrison, R.L., Puttler, B. and Popham, H.J., Genomic sequence analysis of a fast-killing isolate of *Spodoptera frugiperda* multiple nucleopolyhedrovirus. J Gen Virol, 2008. **89**(Pt 3): pp. 775-90.

[69] Breitenbach, J.E., et al., Determination and analysis of the genome sequence of *Spodoptera littoralis* multiple nucleopolyhedrovirus. Virus Res, 2013. **171**(1): pp. 194-208.

[70] Pang, Y., et al., Sequence analysis of the *Spodoptera litura* multicapsid nucleopolyhedrovirus genome. Virology, 2001. **287**(2): pp. 391-404.

[71] Liu, X., et al., Genomic sequencing and analysis of *Sucra jujuba* nucleopolyhedrovirus. PLoS One, 2014. **9**(10): p. e110023.

[72] Willis, L.G., et al., Sequence analysis of the complete genome of *Trichoplusia ni* single nucleopolyhedrovirus and the identification of a baculoviral photolyase gene. Virology, 2005. **338**(2): pp. 209-26.

[73] Wormleaton, S., Kuzio, J. and Winstanley, D., The complete sequence of the *Adoxophyes orana* granulovirus genome. Virology, 2003. **311**(2): pp. 350-65.

[74] Liang, Z., et al., Genomic sequencing and analysis of *Clostera anachoreta* granulovirus. Arch Virol, 2011. **156**(7): pp. 1185-98.

[75] Yin, F., et al., The complete genome of a New Betabaculovirus from *Clostera anastomosis*. PLoS One, 2015. **10**(7): p. e0132792.

[76] Zhang, S., et al., Genome sequencing and analysis of a granulovirus isolated from the Asiatic rice leafroller, *Cnaphalocrocis medinalis*. Virol Sin, 2015. **30**(6): pp. 417-24.

[77] Han, G., et al., Genome of *Cnaphalocrocis medinalis* granulovirus, the first Crambidae-infecting betabaculovirus isolated from rice leaffolder to sequenced. PLoS One, 2016. **11**(2): p. e0147882.

[78] Escasa, S.R., et al., Sequence analysis of the *Choristoneura occidentalis* granulovirus genome. J Gen Virol, 2006. **87**(Pt 7): pp. 1917-33.

[79] Liang, Z., et al., Comparative analysis of the genomes of *Clostera anastomosis* (L.) granulovirus and *Clostera anachoreta* granulovirus. Arch Virol, 2013. **158**(10): pp. 2109-14.

[80] Luque, T., et al., The complete sequence of the *Cydia pomonella* granulovirus genome. J Gen Virol, 2001. **82**(Pt 10): pp. 2531-47.

[81] Lange, M. and Jehle, J.A., The genome of the *Cryptophlebia leucotreta* granulovirus. Virology, 2003. **317**(2): pp. 220-36.

[82] Ardisson-Araujo, D.M., et al., A betabaculovirus-encoded gp64 homolog codes for a functional envelope fusion protein. J Virol, 2016. **90**(3): pp. 1668-72.

[83] Ferrelli, M.L., et al., Genome of Epinotia aporema granulovirus (EpapGV), a polyorganotropic fast killing betabaculovirus with a novel thymidylate kinase gene. BMC Genomics, 2012. **13**: p. 548.

[84] Ardisson-Araujo, D.M., et al., Genome sequence of Erinnyis ello granulovirus (ErelGV), a natural cassava hornworm pesticide and the first sequenced sphingid-infecting betabaculovirus. BMC Genomics, 2014. **15**: p. 856.

[85] Harrison, R.L. and Popham, H.J., Genomic sequence analysis of a granulovirus isolated from the Old World bollworm, *Helicoverpa armigera*. Virus Genes, 2008. **36**(3): pp. 565-81.

[86] Harrison, R.L., Rowley, D.L. and Funk, C.J., The complete genome sequence of *Plodia interpunctella* granulovirus: evidence for horizontal gene transfer and discovery of an unusual inhibitor-of-apoptosis gene. PLoS One, 2016. **11**(7): p. e0160389.

[87] Taha, A., et al., Comparative analysis of the granulin regions of the *Phthorimaea operculella* and *Spodoptera littoralis* granuloviruses. Virus Genes, 2000. **21**(3): pp. 147-55.

[88] Hashimoto, Y., et al., Sequence analysis of the *Plutella xylostella* granulovirus genome. Virology, 2000. **275**(2): pp. 358-72.

[89] Zhang, B.Q., et al., The genome of *Pieris rapae* granulovirus. J Virol, 2012. **86**(17): p. 9544.

[90] Cuartas, P.E., et al., The complete sequence of the first *Spodoptera frugiperda* Betabaculovirus genome: a natural multiple recombinant virus. Viruses, 2015. **7**(1): pp. 394-421.

[91] Wang, Y., et al., Genomic sequence analysis of granulovirus isolated from the tobacco cutworm, *Spodoptera litura*. PLoS One, 2011. **6**(11): p. e28163.

[92] Hayakawa, T., et al., Sequence analysis of the *Xestia c-nigrum* granulovirus genome. Virology, 1999. **262**(2): pp. 277-97.

[93] Duffy, S.P., et al., Sequence analysis and organization of the *Neodiprion abietis* nucleopolyhedrovirus genome. J Virol, 2006. **80**(14): pp. 6952-63.

[94] Lauzon, H.A., et al., Sequence and organization of the *Neodiprion lecontei* nucleopolyhedrovirus genome. J Virol, 2004. **78**(13): pp. 7023-35.

[95] Garcia-Maruniak, A., et al., Sequence analysis of the genome of the *Neodiprion sertifer* nucleopolyhedrovirus. J Virol, 2004. **78**(13): pp. 7036-51.

[96] Afonso, C.L., et al., Genome sequence of a baculovirus pathogenic for *Culex nigripalpus*. J Virol, 2001. **75**(22): pp. 11157-65.

[97] Miller, L.K. and Dawes, K.P., Restriction endonuclease analysis for the identification of baculovirus pesticides. Appl Environ Microbiol, 1978. **35**(2): pp. 411-21.

[98] Smith, G.E. and Summers, M.D., Analysis of baculovirus genomes with restriction endonucleases. Virology, 1978. **89**(2): pp. 517-27.

[99] Lee, H.H. and Miller, L.K., Isolation of genotypic variants of *Autographa californica* nuclear polyhedrosis virus. J Virol, 1978. **27**(3): pp. 754-67.

[100] Miller, L.K. and Dawes, K.P., Restriction endonuclease analysis to distinguish two closely related nuclear polyhedrosis viruses: *Autographa californica* MNPV and *Trichoplusia ni* MNPV. Appl Environ Microbiol, 1978. **35**(6): pp. 1206-10.

[101] Smith, G.E. and Summers, M.D., Restriction Maps of Five *Autographa californica* MNPV Variants, *Trichoplusia ni* MNPV and *Galleria mellonella* MNPV DNAs with Endonucleases SmaI, KpnI, BamHI, SacI, XhoI and EcoRI. J Virol, 1979. **30**(3): pp. 828-38.

[102] Loh, L.C., et al., Analysis of the *Spodoptera frugiperda* nuclear polyhedrosis virus genome by restriction endonucleases and electron microscopy. J Virol, 1982. **44**(2): pp. 747-51.

[103] de Moraes, R.R. and Maruniak, J.E., Detection and identification of multiple baculoviruses using the polymerase chain reaction (PCR) and restriction endonuclease analysis. J Virol Methods, 1997. **63**(1-2): pp. 209-17.

[104] Ernoult-Lange, M., et al., Characterization of the simian virus 40 late promoter: relative importance of sequences within the 72-base-pair repeats differs before and after viral DNA replication. J Virol, 1987. **61**(1): pp. 167-76.

[105] Woo, S.D., Rapid detection of multiple nucleopolyhedroviruses using polymerase chain reaction. Mol Cells, 2001. **11**(3): pp. 334-40.

[106] Wang, L.H., et al., Sequence analysis of the Bam HI-J fragment of the *Spodoptera litura* multicapsid nucleopolyhedrovirus. Acta Biochim Biophy Sinica, 2001. **33**(6): pp. 615-20.

[107] Pijlman, G.P., A.J. Pruijssers and Vlak, J.M., Identification of pif-2, a third conserved baculovirus gene required for per os infection of insects. J Gen Virol, 2003. **84**(Pt 8): pp. 2041-9.

[108] Herniou, E.A., et al., Use of whole genome sequence data to infer baculovirus phylogeny. J Virol, 2001. **75**(17): pp. 8117-26.

[109] Somasekar, S., Jayapragasam, M., Rabindra, R. J., Characterization of five Indian isolates of the nuclear polyhedrosis virus of *Helicoverpa armigera*. Phytoparasitica, 1993. **21**(4): pp. 333-7.

[110] Lange, M., et al., Towards a molecular identification and classification system of lepidopteran-specific baculoviruses. Virology, 2004. **325**(1): pp. 36-47.

[111] Acharya, A. and Gopinathan, K.P., Characterization of late gene expression factors lef-9 and lef-8 from *Bombyx mori* nucleopolyhedrovirus. J Gen Virol, 2002. **83**(Pt 8): pp. 2015-23.

[112] Crouch, E.A., et al., Inter-subunit interactions of the *Autographa californica* M nucleopolyhedrovirus RNA polymerase. Virology, 2007. **367**(2): pp. 265-74.

[113] Toprak, U., et al., Preoperative evaluation of renal anatomy and renal masses with helical CT, 3D-CT and 3D-CT angiography. Diagn Interv Radiol, 2005. **11**(1): pp. 35-40.

[114] Kumar, S., K. Tamura and Nei, M., MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform, 2004. **5**(2): pp. 150-63.

[115] Jose, J., et al., Molecular characterization of nucleopolyhedrovirus of three lepidopteran pests using late expression factor-8 gene. Indian J Virol, 2013. **24**(1): pp. 59-65.

[116] Nai, Y.S., et al., A new nucleopolyhedrovirus strain (LdMNPV-like virus) with a defective fp25 gene from *Lymantria xylina* (Lepidoptera: Lymantriidae) in Taiwan. J Invertebr Pathol, 2009. **102**(2): pp. 110-9.

[117] Chou, C.M., et al., Characterization of *Perina nuda* nucleopolyhedrovirus (PenuNPV) polyhedrin gene. J Invertebr Pathol, 1996. **67**(3): pp. 259-66.

[118] Wang, C.H., et al., Continuous cell line from pupal ovary of *Perina nuda* (Lepidoptera: Lymantriidae) that is permissive to nuclear polyhedrosis virus from *P. nuda*. J Invertebr Pathol, 1996. **67**(3): pp. 199-204.

[119] Sims, D., et al., Sequencing depth and coverage: key considerations in genomic analyses. Nat Rev Genet, 2014. **15**(2): pp. 121-32.

[120] Goodwin, S., McPherson, J.D. and McCombie, W.R., Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet, 2016. **17**(6): pp. 333-51.

[121] Luo, C., et al., Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. PLoS One, 2012. **7**(2): p. e30087.

[122] Garcia-Maruniak A.et al., A variable region of Anticarsia gemmatalis nuclear polyhedrosis virus contains tandemly repeated DNA sequences. Virus Res, 1996. 41:123-132.

[123] Martin, M., Cutadapt removes adapter sequences from high-throughput sequencing Reads. EMBnet.journal, 2011. **17**(1): pp. 10-12.

[124] Schmieder, R. and Edwards, R., Quality control and preprocessing of metagenomic datasets. Bioinformatics, 2011. **27**(6): pp. 863-4.

[125] Patel, R.K. and Jain, M., NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One, 2012. **7**(2): p. e30619.

[126] Masella, A.P., et al., PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics, 2012. **13**: pp. 31.

[127] Zhang, J., et al., PEAR: a fast and accurate illumina paired-end reAd mergeR. Bioinformatics, 2014. **30**(5): pp. 614-20.

[128] Magoc, T. and Salzberg, S.L., FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics, 2011. **27**(21): pp. 2957-63.

[129] Liu, B., et al., COPE: an accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. Bioinformatics, 2012. **28**(22): pp. 2870-4.

[130] Luo, R., et al., SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience, 2012. **1**(1): p. 18.

[131] Zhang, W., et al., A practical comparison of de novo genome assembly software tools for next-generation sequencing technologies. PLoS One, 2011. **6**(3): p. e17915.

[132] Margulies, M., et al., Genome sequencing in microfabricated high-density picolitre reactors. Nature, 2005. **437**(7057): pp. 376-80.

[133] Chevreux, B., et al., Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res, 2004. **14**(6): pp. 1147-59.

[134] Untergasser, A., et al., Primer3--new capabilities and interfaces. Nucleic Acids Res, 2012. **40**(15): p. e115.

[135] Salzberg, S.L., et al., Microbial gene identification using interpolated Markov models. Nucleic Acids Res, 1998. **26**(2): pp. 544-8.

[136] van Baren, M.J., Koebbe, B.C. and Brent, M.R., Using N-SCAN or TWINSCAN to predict gene structures in genomic DNA sequences. Curr Protoc Bioinformatics, 2007. **4**: p. Unit 4 8.

[137] Lukashin, A.V. and Borodovsky, M., GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res, 1998. **26**(4): pp. 1107-15.

[138] Wang, S., Sundaram, J.P. and Spiro, D., VIGOR, an annotation program for small viral genomes. BMC Bioinformatics, 2010. **11**: pp. 451.

[139] Li, S.C., Shiau, C.K. and Lin, W.C., Vir-Mir db: prediction of viral microRNA candidate hairpins. Nucleic Acids Res, 2008. **36**(Database issue): pp. D184-9.

[140] Stothard, P. and Wishart, D.S., Circular genome visualization and exploration using CGView. Bioinformatics, 2005. **21**(4): p. 537-9.

[141] Kumar, S., et al., MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform, 2008. **9**(4): pp. 299-306.

[142] Sievers, F., et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol, 2011. **7**: pp. 539.

[143] Thompson, J.D., Gibson, T.J. and Higgins, D.G., Multiple sequence alignment using ClustalW and ClustalX. Curr Protoc Bioinformatics, 2002. **2**: p. Unit 2 3.

[144] Talavera, G. and Castresana, J., Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol, 2007. **56**(4): pp. 564-77.

[145] Castresana, J., Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol, 2000. **17**(4): pp. 540-52.

[146] Ronquist, F., et al., MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol, 2012. **61**(3): pp. 539-42.

[147] Stamatakis, A., RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics, 2014. **30**(9): pp. 1312-3.

[148] Douady, C.J., et al., Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol Biol Evol, 2003. **20**(2): pp. 248-54.

[149] Drummond, A.J., et al., Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol, 2012. **29**(8): pp. 1969-73.

[150] Camacho, C., et al., BLAST+: architecture and applications. BMC Bioinformatics, 2009. **10**: p. 421.

[151] Brudno, M., et al., LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res, 2003. **13**(4): pp. 721-31.

[152] Grant, J.R., Arantes, A.S. and Stothard, P., Comparing thousands of circular genomes using the CGView comparison tool. BMC Genomics, 2012. **13**: pp. 202.

[153] Darling, A.C., et al., Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res, 2004. **14**(7): pp. 1394-403.