The Journal of Microbiology (2010) Vol. 48, No. 1, pp. 102-110
DOI 10.1007/s12275-009-0197-4

Comparative Analysis of the Genomes of *Bombyx mandarina* and *Bombyx mori* Nucleopolyhedroviruses

Yi-Peng Xu[1], Zheng-Pei Ye[2], Chang-Ying Niu[2], Yan-Yuan Bao[1], Wen-Bing Wang[3], Wei-De Shen[3], and Chuan-Xi Zhang[1]*

[1]*Ministry of Agriculture Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Zhejiang University, Hangzhou 310029, P. R. China*
[2]*Institute of Insect Resources, Huazhong Agricultural University, Wuhan 430070, P. R. China*
[3]*School of Life Science, Suzhou University, Suzhou 215123, P. R. China*

(Received June 26, 2009 / Accepted August 3, 2009)

The *Bombyx mandarina* nucleopolyhedrovirus (BomaNPV) S1 strain can infect the silkworm, *Bombyx mori*, but is significantly less virulent than *B. mori* nucleopolyhedrovirus (BmNPV) T3 strain. The complete nucleotide sequence of the S1 strain of BomaNPV was determined and compared with the BmNPV T3 strain. The circular, double stranded DNA genome of the S1 strain was 126,770 nucleotides long (GenBank accession no. FJ882854), with a G+C content of 40.23%. The genome contained 133 potential ORFs. Most of the putative proteins were more than 96% identical to homologs in the BmNPV T3 strain, except for *bro-a*, *lef-12*, *bro-c*, and *bro-d*. Compared with the BmNPV T3 strain, however, this genome did not encode the *bro-b* and *bro-e* genes. In addition, hr1 lacked two repeat units, while hr2L, hr2R, hr3, hr4L, hr4R, and hr5 were similar to the corresponding hrs in the T3 strain. The sequence strongly suggested that BomaNPV and BmNPV are variants with each other, and supported the idea that baculovirus strain heterogeneity may often be caused by variation in the hrs and *bro* genes.

***Keywords*:** *B. mandarina*, *B. mori*, nucleopolyhedrovirus, genome sequence comparison

The baculoviruses form a diverse group of viruses with large, double-stranded, circular DNA genomes ranging from 80 to 180 kb, which are packaged in enveloped, rod-shaped virions. At present, this family can be devided into four genera: Alphabaculovirus (lepidopteran-specific NPV), Betabaculovirus (lepidopteran-specific Granuloviruses), Gammabaculovirus (hymenopteran-specific NPV), and Deltabaculovirus (dipteran-specific NPV) (Jehle *et al*., 2006). During the life cycle, two types of baculoviral virions, occlusion-derived virions (ODVs) and budded virions (BVs), are produced. ODVs transmit infection through an oral route, establishing a primary infection of the midgut of the host, while BVs are responsible for systemic infection, spreading the virus from cell to cell within the host (Keddie *et al*., 1989). Currently more than forty-eight baculovirus genomes have been sequenced because of their importance for pest control and expression of recombinant proteins. Most baculoviruses infect insects of the orders Lepidoptera, Hymenoptera, and Diptera.

The wild silkworm, *Bombyx mandarina*, which is an insect pest of mulberry production, is commonly believed to have the same ancestor as the domesticated silkworm, *Bombyx mori*, based on similar genetic, morphological, and physiological characteristics (Kawaguchi, 1928; Astaurov *et al*., 1959; Yoshitake, 1968; Chikushi, 1972; Nakamura *et al*., 1999), although they have different chromosome numbers (*B. mandarina,* n=27; *B. mori* n=28). Therefore, the study of *B. mandarina* nucleopolyhedrovirus (BomaNPV), which is the natural pathogenic virus of *Bombyx mandarina*, will contribute to the understanding of the evolutionary relationship between BomaNPV and *B. mori* nucleopolyhedrovirus (BmNPV) and the controlling of the virus disease in the sericulture, and to the increased understanding of the baculovirus family. Currently little is known about BomaNPV, however. In this study, we sequenced the complete genome of BomaNPV and, in addition, performed a comparative analysis of the genomes of BomaNPV (S1 strain) and BmNPV (T3 strain). Furthermore, a polyhedral inclusion body (PIB) bioassay was carried out to compare the infectivity of BomaNPV and BmNPV T3 for *B. mori* larvae.

Materials and Methods

Virus and virus DNA

The BomaNPV was originally isolated from a diseased larva of *B. mandarina* in the Jiangsu Province of China. BomaNPV was plaque-purified in BmN cells and then propagated by infecting 5th instar larvae of the silkworm. The occlusion bodies were purified by sucrose-gradient centrifugation. Virus DNA was extracted from the purified occlusions as previously described (Ma *et al*., 2006).

Nucleotide sequence determination

The purified virus genomic DNA was sheared into 200- to 300-bp fragments by ultrasonication and sequenced by Solexa technology using the Solexa 1G Genome Analyzer in the Zhejiang-California International NanoSystems Institute, according to the manufacturer's protocols. Ambiguous regions, which contain hrs and *bro* genes, were amplified by PCR using specific primers (Table 1) and cloned into the pMD18-T vector (TaKaRa, China). All the clones were sequenced

* For correspondence. E-mail: chxzhang@zju.edu.cn; Tel: +86-571-8697-1697; Fax: +86-571-8697-1697

Table 1. Primers designed to amplify ambiguous regions of BomaNPV genome

Primers	Amplifying regions	Sequence (5′ → 3′)	Locus[a]
BRO-AF	*bro-a*	GTTGGAGCCATGTCGTTACTAC	20,780—20,801
BRO-AR		CGCTAATGATGCAAATGGCTTTC	22,031—22,053
hr2F	hr2L, *fgf*, and hr2R	GCACACGGACAAAGACGATCTG	22,379—22,400
hr2R		CAAACGCATACGAATTGATGGAC	24,376—24,398
hr3F	hr3	ACATGGCGTGTCCTGAAGAATTG	64,662—64,684
hr3R		CTTGAAACGCGTGTATGCGTCG	66,445—66,466
BRO-BF	*bro-b*	GGAATATATGCCACCGCATGCAC	76,465—76,487
BRO-BR		GCGACTGGCACAATTATCACGA	78,710—78,731
hr4LF	hr4L	TGCTTTCGACCGCATATCCCTT	86,155—86,176
hr4LR		ACCTGTCTCTGGCCTTTTCTAC	86,816—86,837
hr4RF	hr4R	CTTGAGAGGTTTGCGGTTGTTG	89,361—89,382
hr4RR		CGATACAACTACGAAGTGCTGTG	90,090—90,112
hr5F	hr5	GCACAATGTAACTAGTACACTCAAC	106,768—106,792
hr5R		GTCCACATTGTCGACTTGCTCT	107,755—107,776
hr1F	hr1	GAAGTCGTCGATAAAACTGACGC	123,623—123,645
hr1R		CTGTAAATAGTTGTGCCAACGC	124,442—124,463
BRO-EF	*bro-d* and *bro-e*	GGCATTAATCGCACCGGTTACA	124,794—124,815
BRO-ER		CTTTGAAGTGCAGCGCATCAAC	126,876—126,896

[a] Locus in the genomic sequence of BmNPV T3 (GenBank accession no. NC 001962).

using the primers M13F(-47) and M13R(-48).

DNA sequence analysis

The nucleotide composition of the genomic DNA and predicted ORFs were analyzed using Genetyx-win Software (Software Development Co. Ltd, Japan) and the DNASTAR software. Relevant homologous ORFs were checked with reference to the BmNPV T3 strain and AcMNPV C6. Proteins coded by ORFs were translated using the EditSeq program, and amino acid alignments were carried out by the MegAlign program.

B. mori larvae bioassay

The LC_{50} values of PIB were determined using per os infection. After being starved for 3 h, the newly molted fourth instar *B. mori* larvae (strain Lanzhou 5) were fed with mulberry leaves, to which different concentration of PIBs (10^9, 10^8, 10^7, 10^6, 10^5 PIB/ml) had been applied. Thirty larvae were used for each concentration, and the mortality was counted every 4 h.

The estimation of LC_{50} values, as well as the statistical analysis comparing the two viruses, were performed using the DPS data processing system for practical statistics (Tang and Feng, 2002), using Probit analysis (Finney, 1971).

Results and Discussion

Genome sequence analysis

The BomaNPV genome consisted of 126,770 nucleotides (GenBank accession no. FJ882854), with only 40.2% G+Cs. The size and G+C content most closely resembled the size and G+C content of the BmNPV T3 strain (128,413 kb, 40.4% G+Cs). The whole genome sequence of BomaNPV was 98.7% identical to that of BmNPV T3. There were 133 ORFs encoding predicted proteins of more than 60 amino acids, beginning with the *polyhedrin* gene. This genome also possessed eight ORFs encoding predicted products of fewer than 60 aa, ORF7a (53aa), ORF58a (60aa), ORF69a (56aa), ORF91a (59aa), ORF94a (56aa), ORF97a (57aa), ORF109a (57aa), and ORF110 (56aa). The sequences encoding proteins with more than 60aa accounted for 89.54% of the whole genome, and the percentage increased to 90.39% if predicted ORFs encoding proteins with fewer than 60aa were taken into account. The location of ORFs, and the size of the corresponding predicted proteins are shown in Table 2. As in BmNPV T3, seven homologous repeat regions (hrs) were also present in this genome, though hr1 showed some difference. This result suggested that BomaNPV and BmNPV T3 were variants with each other, which might be due to the close evolutionary relationship between *B. mandarina* and *B. mori*.

Comparison of the ORFs between BomaNPV and BmNPV T3

As shown in Table 2, the predicted polypeptides encoded by 133 BomaNPV ORFs were highly similar to those of the BmNPV T3 strain with identical orientation and order of the putative ORFs. Most of them showed more than 95% identity, with the exception of *bro-a* (89.6%), *lef-12* (93.2%), *bro-c* (95.0%), and *bro-d* (91.7%). Focusing on the small ORFs of fewer than 60aa, ORF110, and ORF69a should be paid more attention to. The ORF110 looked like its homolog in the BmNPV T3 strain, but without the 3′ terminus, while ORF69a was unique to BomaNPV, and highly homologous to the 3′ end of homologs of Ac-PNK/PNL（E (value=8e-15). However, excluding ORF69a, the other 7 small ORFs showed high degrees of identity with their counterparts in BmNPV T3. In BomaNPV, we could not find an ORF corresponding to ORF22a of BmNPV T3. In addition, extensive nucleotide changes, including substitutions, deletions and insertions, were observed in 102 ORFs, especially *arif-1*, *bro-a*, *lef-12*, *bro-c*, and *bro-d*. What's more, the *bro-b* and *bro-e* genes were absent in BomaNPV when compared to BmNPV T3.

Table 2. Characteristics of putative ORFs of BomaNPV and BmNPV T3

BomaNPV					BmNPV T3 strain			Amino acid identity (%)
ORF	Name	Position	Length (aa)	Predicted *Mr*(×10-3)	ORF	Name	Length (aa)	
1	*polyhedrin*	1→738	245	28.8	1	*polyhedrin*	245	100.0
2	*orf1629*	768←2396	542	60.7	2	*orf1629*	542	99.1
3	*pk1*	2395→3219	274	32.4	3	*pk1*	275	97.8
4	*Ac11*	3245←4267	340	39.8	4	*Ac11*	340	98.2
5	*Ac13*	4602←5597	331	39.3	5	*Ac13*	331	99.4
6	*lef-1*	5476←6289	270	31.1	6	*lef-1*	270	99.6
7	*egt*	6404→7924	506	57.1	7	*egt*	506	99.4
7a	*unknown*	7937→8098	53	6.2	7a	*unknown*	53	100.0
8	*bv/odv-e26*	8064→8753	229	26.2	8	*bv/odv-e26*	229	98.7
9	*Ac17*	8722→9354	210	24.0	9	*Ac17*	210	99.0
10	*Ac18*	9384←10454	356	41.4	10	*Ac18*	356	98.3
11	*Ac19*	10456→10788	110	12.5	11	*Ac19*	110	100.0
12	*arif-1*	10975←12231	418	47.8	12	*arif-1*	440	98.8
13	*pif-2*	12268→13416	382	43.8	13	*pif-2*	382	98.4
14	*f*	13519→15540	673	77.8	14	*f*	673	99.1
15	*pkip*	15570←16079	169	19.4	15	*pkip*	169	98.8
16	*dbp*	16119←17072	317	36.7	16	*dbp*	317	100.0
17	*Ac26*	17148→17537	129	14.5	17	*Ac26*	129	99.2
18	*iap1*	17539→18408	289	33.7	18	*iap1*	292	97.9
19	*lef-6*	18413→18934	173	20.4	19	*lef-6*	173	100.0
20	*Ac29*	19053←19268	71	8.6	20	*Ac29*	71	100.0
21	*Ac30*	19323←20741	472	55.7	21	*Ac30*	472	99.4
22	*bro-a*	20776←21768	330	37.0	22	*bro-a*	317	89.6
					22a	*unknown*	54	94.4
23	*sod*	21992→22447	151	16.3	23	*sod*	151	98.0
	hr2L					hr2L		
24	*fgf*	23377→23925	182	20.7	24	*fgf*	182	98.4
	hr2R					hr2R		
25	*Ac34*	24337←24984	215	24.8	25	*Ac34*	215	99.1
26	*ubiquitin*	25005→25238	77	8.7	26	*ubiquitin*	77	100.0
27	*39k*	25291←26124	277	31.5	27	*39k*	277	100.0
28	*lef-11*	26118←26456	112	13.2	28	*lef-11*	112	98.2
29	*bv-e31*	26419←27072	217	25.5	29	*bv-e31*	217	100.0
30	*p43*	27140←28228	362	43.4	30	*p43*	362	99.4
31	*p47*	28236←29435	399	47.3	31	*p47*	399	99.0
32	*lef-12*	29440→29973	177	20.6	32	*lef-12*	183	93.2
33	*gta*	30049→31569	506	59.1	33	*gta*	506	99.2
34	*Ac43*	31583→31819	78	89.4	34	*Ac43*	78	98.7
35	*Ac44*	31800→32195	131	15.0	35	*Ac44*	131	99.2
36	*Ac45*	32197→32784	195	22.7	36	*Ac45*	193	99.5
37	*odv-e66*	32769→34880	703	79.2	37	*odv-e66*	702	98.9
38	*ets*	34976←35245	89	10.5	38	*ets*	89	100.0
39	*lef-8*	35491←38124	877	101.7	39	*lef-8*	877	99.8
40	*Ac51*	38151→39110	319	37.8	40	*Ac51*	319	99.4
41	*Ac52*	39101←39685	194	23.3	41	*Ac52*	194	100.0
42	*Ac53*	39687→40106	139	16.9	42	*Ac53*	139	100.0
42a	*lef-10*	40103→40339	78	8.5	42a	*lef-10*	78	98.7
43	*vp1054*	40197→41294	365	42.0	43	*vp1054*	365	100.0
44	*Ac55*	41376→41609	77	8.6	44	*Ac55*	77	98.7
45	*Ac56*	41611→41865	84	9.9	45	*Ac56*	84	100.0
46	*Ac57*	42119→42604	161	19.1	46	*Ac57*	161	99.4
47	*Ac58/59*	42622←43137	171	20.2	47	*Ac58/59*	171	98.8
48	*Ac60*	43149←43400	83	9.8	48	*Ac60*	83	98.8

Table 2. Continued

BomaNPV					BmNPV T3 strain			Amino acid identity (%)
ORF	Name	Position	Length (aa)	Predicted *Mr*(×10-3)	ORF	Name	Length (aa)	
49	*fp25*	43549←44193	214	25.3	49	*fp25*	214	99.1
50	*lef-9*	44297→45769	490	56.3	50	*lef-9*	490	99.6
51	*Ac63*	45829→46296	155	18.6	51	*Ac63*	155	98.7
52	*gp37*	46374←47258	294	33.8	52	*gp37*	294	98.3
53	*dnapol*	47388←50357	989	114.8	53	*dnapol*	986	99.3
54	*Ac66*	50366→52783	805	93.4	54	*Ac66*	805	99.1
55	*lef-3*	52786←53943	385	44.9	55	*lef-3*	385	99.2
56	*odv-nc42*	53962→54366	134	15.8	56	*odv-nc42*	134	100.0
57	*Ac69*	54344→55132	262	30.4	57	*Ac69*	262	99.2
58	*iap2*	55281→56030	249	28.7	58	*iap2*	249	99.6
58a	*Ac72*	56089→56271	60	7.0	58a	*Ac72*	60	98.3
59	*Ac73*	56282←56581	99	11.5	59	*Ac73*	99	96.0
60	*Ac74*	56578←57381	267	30.9	60	*Ac74*	268	99.3
61	*Ac75*	57399←57800	133	15.5	61	*Ac75*	133	100.0
62	*Ac76*	57819←58076	85	9.6	62	*Ac76*	85	100.0
63	*vlf-1*	58092←59231	379	44.3	63	*vlf-1*	379	99.7
64	*Ac78*	59237←59569	110	12.7	64	*Ac78*	110	99.1
65	*Ac79*	59572←59886	104	12.2	65	*Ac79*	104	100.0
66	*gp41*	59889←61100	403	44.8	66	*gp41*	403	99.0
67	*Ac81*	61090←61794	234	27.0	67	*Ac81*	234	99.6
68	*Ac82*	61640←62185	181	20.1	68	*Ac82*	181	98.9
69	*p95*	62151→64664	837	95.6	69	*p95*	839	99.5
	hr3					hr3		
69a	*unknown*	65475←65645	56	6.8				
70	*vp15*	66184→66564	126	15.1	70	*vp15*	126	100.0
71	*cg30*	66569←67372	267	30.7	71	*cg30*	267	99.3
72	*vp39*	67375←68418	347	38.8	72	*vp39*	350	98.3
73	*lef-4*	68437→69834	465	54.0	73	*lef-4*	465	99.4
74	*Acf91*	69831←70295	154	17.3	74	*Acf91*	154	99.4
75	*p33*	70331←71110	259	30.9	75	*p33*	259	100.0
76	*Ac93*	71109→71594	161	18.4	76	*Ac93*	161	100.0
77	*odv-e25*	71603→72289	228	25.5	77	*odv-e25*	228	99.6
78	*dnahel*	72328←75996	1222	143.6	78	*dnahel*	1222	99.8
79	*odv-e28*	75983→76531	182	20.9	79	*odv-e28*	182	99.5
					80	*bro-b*	239	
80	*bro-c*	76619→77596	325	36.8	81	*bro-c*	318	95.0
81	*38k*	77741←78703	320	38.0	82	*38k*	320	99.7
82	*lef-5*	78638→79435	265	31.1	83	*lef-5*	265	100.0
83	*p6.9*	79432←79629	65	8.1	84	*p6.9*	65	98.5
84	*p40*	79671←80759	362	41.7	85	*p40*	362	98.6
85	*p12*	80779←81156	125	13.7	86	*p12*	123	99.2
86	*p45*	81137←82300	387	45.5	87	*p45*	387	99.5
87	*vp80*	82326→84407	693	80.0	88	*vp80*	692	99.4
88	*he65*	84430←85299	289	34.3	89	*he65*	289	97.9
	hr4L					hr4L		
89	*Ac106/107*	85933→86682	249	28.9	90	*Ac106/107*	249	99.2
90	*Ac108*	86683←87000	105	11.8	91	*Ac108*	105	100.0
91	*Ac109*	87015←88190	391	45.0	92	*Ac109*	391	100.0
91a	*Ac110*	88214←88393	59	7.1	92a	*Ac110*	59	100.0
92	*Ac111*	88442←88645	67	8.2	93	*Ac111*	67	100.0
	hr4R					hr4R		
93	*Ac114*	89320←90594	424	49.4	94	*Ac114*	424	100.0
94	*pif-3*	90616←91230	204	23.0	95	*pif-3*	204	99.0

Table 2. Continued

BomaNPV					BmNPV T3 strain			Amino acid identity (%)
ORF	Name	Position	Length (aa)	Predicted *Mr*(×10-3)	ORF	Name	Length (aa)	
94a	*Ac116*	91238←91408	56	6.5	95a	*Ac116*	56	94.6
95	*Ac117*	91344→91631	95	10.9	96	*Ac117*	95	100.0
96	*pif-1*	91763→93346	527	59.8	97	*pif-1*	527	99.1
97	*Ac120*	93354→93602	82	9.5	98	*Ac120*	82	98.8
97a	*Ac121*	93705→93878	57	6.7	98a	*Ac121*	57	94.7
98	*Ac122*	93771←93956	61	7.1	99	*Ac122*	61	98.4
99	*gcn2/Pk2*	93990←94667	225	26.1	100	*gcn2/Pk2*	225	99.1
100	*Ac124*	94851→95585	244	28.2	101	*Ac124*	244	98.0
101	*lef-7*	95606←96289	227	26.7	102	*lef-7*	227	96.5
102	*chitinase*	96279←97934	551	61.7	103	*chitinase*	552	99.1
103	*v-cath*	97983→98954	323	37.0	104	*v-cath*	323	99.1
104	*gp64/67*	99071←100663	530	60.6	105	*gp64/67*	530	99.8
105	*p24*	100790→101377	195	21.8	106	*p24*	195	100.0
106	*gp16*	101405→101725	106	12.1	107	*gp16*	106	100.0
107	*pp34*	101787→102734	315	35.3	108	*pp34*	315	99.7
108	*ac132*	102737→103399	220	25.2	109	*ac132*	220	100.0
109	*alk-exo*	103427→104689	420	48.5	110	*alk-exo*	420	99.5
109a	*unknown*	104733←104906	57	6.6	110a	unknown	57	98.2
*110	*unknown*	104805→104975	56	7.9	111	unknown	70	98.2
111	*p35*	105151→106050	299	34.9	112	*p35*	299	100.0
	hr5					hr5		
112	*p26*	106931→107653	240	27.3	113	*p26*	240	98.8
113	*p10*	107726→107938	70	7.6	114	*p10*	70	98.6
114	*p74=pif*	108025←109962	645	74.1	115	*p74=pif*	645	99.7
115	*me53*	110193←111551	452	52.7	116	*me53*	451	99.1
116	*ie-0*	111828→112613	261	30.0	117	*ie-0*	261	100.0
117	*bv/odv-nc50*	112628→114058	476	55.4	118	*bv/odv-nc50*	476	100.0
118	*odv-e18*	114066→114371	101	10.4	119	*odv-e18*	101	100.0
119	*odv-ec27*	114386→115258	290	33.5	120	*odv-ec27*	290	100.0
120	*Ac145*	115273→115560	95	11.0	121	*Ac145*	95	100.0
121	*Ac146*	115555←116160	201	23.0	122	*Ac146*	201	99.0
122	*ie-1*	116226→117980	584	66.9	123	*ie-1*	584	99.7
123	*odv-e56*	118069←119196	375	41.3	124	*odv-e56*	375	98.7
124	*Ac149*	119225←119545	106	12.3	125	*Ac149*	106	98.1
125	*Ac150*	119514→119861	115	13.3	126	*Ac150*	115	100.0
126	*ie-2*	119894←121162	422	48.7	127	*ie-2*	422	99.1
127	*pe38*	121593→122522	309	36.0	128	*pe38*	309	99.4
128	*Ac154*	122623→122856	77	8.9	129	*Ac154*	77	100.0
	hr1					hr1		
129	*ptp*	123593→124099	168	19.2	130	*ptp*	168	97.0
130	*bro-d*	124096←125142	348	40.3	131	*bro-d*	349	91.7
					132	*bro-e*	241	
131	*Ac4*	125215←125670	151	17.6	133	*Ac4*	151	98.0
132	*Ac5*	125699→126028	109	12.4	134	*Ac5*	109	99.1
133	*lef-2*	126009→126641	210	23.8	135	*lef-2*	210	99.5

Overlapping genes and intergenic regions

Overlapping genes are found most commonly in rapidly evolving genomes with high mutation rates such as those of bacteria, mitochondria, bacteriophage and viruses. The presence of overlapping genes or regions has been hypothesized to be related not only to genome size minimization, but also to regulatory mechanisms of gene expression, both at the level of expression and at the level of protein-protein interaction (Normark *et al*., 1983; Fukuda *et al*., 1999; Krakauer, 2000; Johnson and Chisholm, 2004). The genome of BomaNPV contained 29 overlapping regions involving 52 genes. These include 6 "end-on" (→←) overlapping pairs (ORF40-41, 42-42a, 42a-43, 73-74, 83-84, 97a-98, 120-121, 129-130), 7 "head-on" (←→) overlapping pairs (ORF2-3, 68-69, 75-76, 78-79,

Table 3. Comparisons of the sizes of four insertions in BomaNPV and BmNPV T3

Virus	The insertion size (bp)				
	p47—lef-12	*he65—orf89(90)*	*he65*—hr4L	hr1	hr1—*ptp*
BomaNPV	4	633	365	383	327
BmNPV T3 strain	59	489	218	592	133

81-82, 94a-95, 124-125) and 16 "uni-directional" (→→) overlapping pairs (ORF5-6, 7a-8, 8-9, 27-28, 28-29, 34-35, 36-37, 42-42a, 42a-43, 56-57, 59-60, 66-67, 67-68, 85-86, 101-102, 132-133), according to the classification defined by Fukuda *et al*. (1999). Disppointingly, they showed no significant difference from the BmNPV T3 strain.

Comprising promoters, enhancers, and other regulatory elements, intergenic regions play an important role in the transcription of genes (Pinschewer *et al*., 2005; López and Franze-Fernández, 2007; Nakagawa *et al*., 2008). Among the baculoviruses, the intergenic regions of LsNPV occupied 17.2% of genome, the highest known so far (Xiao and Qi, 2007). In BomaNPV, this number was 9.6%, the same as BmNPV T3. Most of the intergenic regions were similar to those in the BmNPV T3 strain; however, the intergenic region between *p47* and *lef-12* was 4 bp in BomaNPV but 49 bp in BmNPV, and that between *he65* and *orf89* (*Bm90*) was 633 bp in BomaNPV but 489 bp in BmNPV (Table 3). In addition, the size of the insertion between hr1 and *ptp* of BomaNPV was significantly larger than that of BmNPV T3 (327 bp vs. 133 bp), while the sizes of hr1 of BomaNPV and BmNPV T3 were reversed (383 bp vs. 592 bp) (Table 3). These results implied that the utilization rate of genomic nucleotides by BomaNPV was similar to that of the BmNPV T3 strain. It appears that it may have been difficult for the genomes to diverge, considering that BmNPV and BomaNPV may have coevolved with *Bombyx* for a long time.

Arif-1

Arif-1, actin-rearrangement-inducing factor, was able to induce actin rearrangement in Tn-368 cells (Roncarati and Knebel-Mörsdorf, 1997). It was found to colocalize with F-actin at the plasma membrane, but didn't play a significant role in the propagation of budded viruses (BVs) as analyzed by mutants (Dreschers *et al.*, 2001). The proline-rich C-terminus was essential for transporting and/or anchoring Arif-1 at the plasma membrane. Compared with the T3 strain, the Arif-1's counterpart in BomaNPV lacked a Pro-rich sequence of 22 amino acids (PVPTAPVKPPTPPVPTAHVPTP) (Fig. 1), which might be relevant to its function of localization (Pancio *et al*., 2000; Gan *et al*., 2004; Kane *et al*., 2004). This extra Pro-rich residue couldn't be found in baculoviruses other than the T3 strain, suggesting that this sequence was specific for T3. When the structure of Arif-1 of BomaNPV was analyzed by PredicProtein (http://www.predictprotein.org/), four membrane helices were detected at the N-terminus and the C-terminus of about 20aa was a long loop with a NLS (Nuclear Localization Signal) site (QRKFKERK). Besides, lots of Protein-Protein

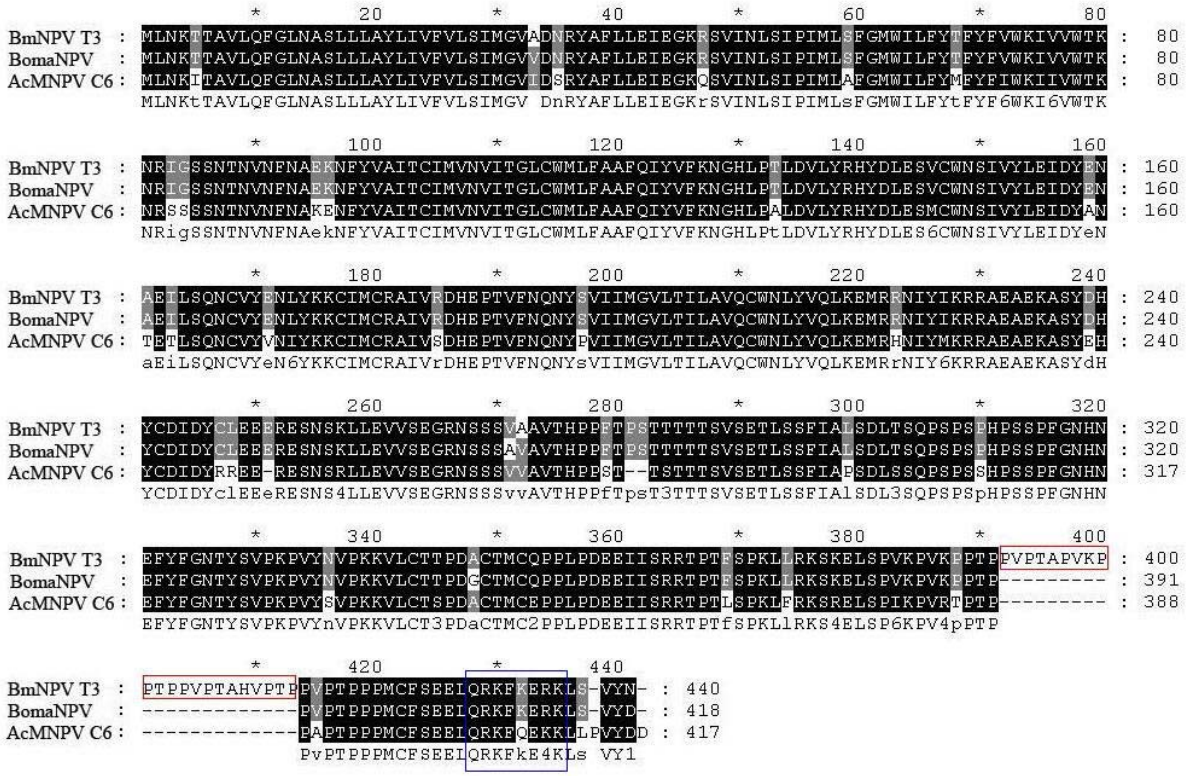

Fig. 1. Comparison of Arif-1 in BomaNPV, BmNPV T3 strain, and AcMNPV C6.

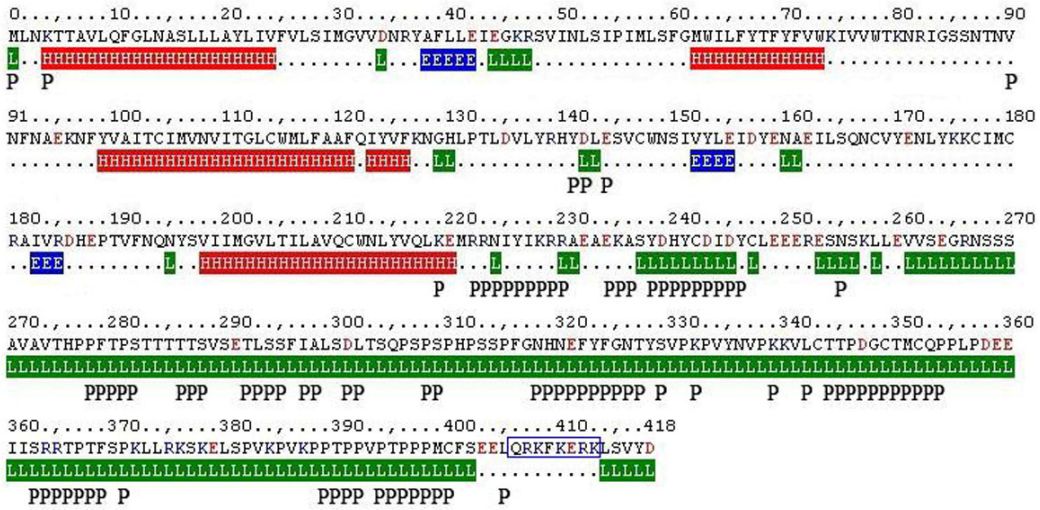

Fig. 2. Analysis of the structure of BomaNPV Arif-1.
H represents the Helix, E represents the Sheet, L represents the Loop, QRKFKERK in blue frame represent the NLS (Nuclear Localization Signal), meeting the rule of [QMN]R[RK]xKx[RK][RK]

interaction sites were found at the C-terminus of Arif-1 (Figs. 1 and 2). These comfirmed the importance of the proline-rich C-terminus for Arif-1 localizing or transporting.

Baculovirus repeated ORFs (*bro* genes)

The presence of *bro* genes is a notable feature in baculoviruses, with various copies present in the baculovirus genomes identified so far. As reported by (Kang *et al*., 1999), all five of the *bro* genes in BmNPV T3 were described as early genes, and they might be involved in some other important viral functions rather than in host range, according to the results of a mutant analysis achieved by deleting *bro-a*, *bro-c*, *bro-b* or *bro-e*. According to further research by Kang *et al*. (2006), BmNPV BRO proteins (especially BRO-A and BRO-C) contained a nucleic acid binding activity involved in nucleosome organization in infected cells and might function as laminin binding proteins that could influence host DNA replication (Zemskov *et al*., 2000; Kang *et al*., 2003). Furthermore, BRO proteins were identified as nucleocytoplasmic shuttling proteins utilizing the CRM1-mediated nuclear export pathway for there was a leucine-rich region in the N-terminal region that functions as a CRM1-dependent nuclear export signal (NES) (Kang *et al*., 2006). There were three *bro* genes (ORF22, ORF31, and ORF131) present in the genome of BomaNPV, referred to as *bro-a*, *bro-c*, and *bro-d*, respectively, which showed strong homology to the corresponding genes of the BmNPV T3 strain. The sizes of the proteins, however, are different, with those from BomaNPV being 330 aa, 325 aa, and 348 aa, respectively, while those of the BmNPV T3 strain are 317 aa, 318 aa, and 349 aa. Compared with the other genes, *bro-a* and *bro-d* showed the highest and the lowest identity, respectively, to their homologs in the BmNPV T3 strain. According to the protein classification defined by Kuzio *et al*. (1999), all three *bro* genes (*bro-a*, *bro-c*, and *bro-d*) belonged to Group I. While *bro-a* fell into Subgroup B with *bro-c*, *bro-d* was placed in Subgroup A based on further analysis by Kang *et al*. (1999). Significantly, however, *bro-b* and *bro-e*, belonging to Subgroup B, were absent in BomaNPV, in contrast to BmNPV T3. Similarly, only three *bro* genes were found in the BmNPV SC7 isolate, but each of them belonged to one of the three subgroups present in the BmNPV T3 strain (Table 4) (López Ferber *et al*., 2001). Maybe different *bro* genes of different subgroups have diverse functions. To date, *bro-d* is the only *bro* gene generally present in baculovirus genomes. Its homolog was first found in the AcMNPV genome as ORF2, and was the only copy in that genome, indicating that *bro-d* gene may play a special role in the virus life cycle. Notably, *bro-d* was found to be essential for viral growth in BmN cells, while *bro-a* and *bro-c* might complement each other (Kang *et al*., 1999, 2006). However, the significance of the presence of BRO-B and BRO-E remains an open question. They were found only in the cytoplasm of BmN cells infected with BmNPV, while BRO-A/C and BRO-D localized in the cytoplasm and the nucleus (Kang *et al*., 1999).

Baculovirus homologous repeat regions (hrs)

The baculovirus hrs have been implicated both as origins of DNA replication and transcriptional enhancers (Rodems and Friesen, 1993), but their functional significance is still unknown. It has been demonstrated that deletion of each single hr, or two hrs, fails to affect virus replication in cell culture (Carstens and Wu, 2007). There were seven hrs (hr2L, hr2R, hr3, hr4L, hr4R, hr5, hr1) in the BomaNPV genome,

Table 4. Distribution of *bro* genes in BomaNPV, BmNPV T3, and BmNPV SC7

Virus	Subfamily[a]		
	Subgroup A	Subgroup B	Subgroup C
BomaNPV	*bro-d*	*bro-a*, *bro-c*	
BmNPV T3	*bro-d*	*bro-a*, *bro-c*	*bro-b*, *bro-e*
BmNPV SC7	*bro-III*	*bro-II*	*bro-I*

[a] Subfamily was refered to Kang *et al*. (1999)

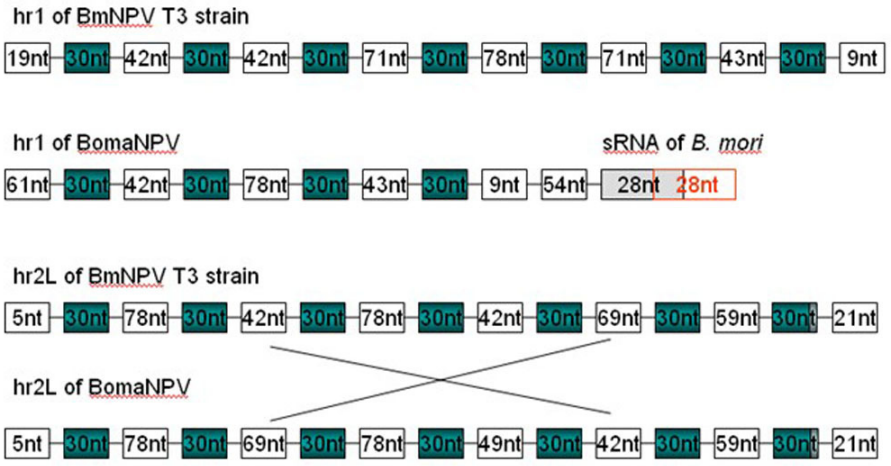

Fig. 3. Comparisons of hr1 and hr2L of BomaNPV S1 and BmNPV T3.
30nt Green boxes represent the core palindrome, 42nt white boxes represent insertion sequence between palindromes and the sequences in boxes containing same number of nucleotides have exceeding identity.

located at different sites, as in the BmNPV T3 genome. All hrs showed strong identity to those of the BmNPV T3 strain in nucleotide analysis [hr2L (90.9%), hr2R (99.3%), hr3 (96.3%), hr4L (94.4%), hr4R (98.1%), hr5 (95.6%), hr1 (92.4%)]. The second and fifth insertion sequences between palindromes of hr2L, however, seem to have experienced shifts between the BmNPV T3 strain and BomaNPV (Fig. 3). In addition, hr1 of BomaNPV lacked two repeat units, compared with that of the BmNPV T3 strain. Interestingly, there was an extra inserted sequence of 194 bp following the end of hr1 in the BomaNPV but not in the T3 strain. This insert possesses a special sequence (CTGTTATAAGACGGCCCTGTACCCTTTACTG CTGACA). The 28 nucleotides as shown in the frame and in grey were 100% identical to two ovarian small RNAs in *B. mori*. These RNAs were believed to cooperate in the regulation of transposon activity (Kawaoka *et al*., 2008), and might be involved in defense against viruses. This special sequence might have been acquired from the host by this baculovirus when it coevolved with its host, and may confer a survival advantage on the virus.

Infectivity of PIB in *B. mori* larvae: BomaNPV compared to BmNPV T3

To investigate whether BomaNPV has greater pathogenicity than BmNPV T3, the LC_{50} of the PIB in *B. mori* larvae was examined. The data revealed that the LC_{50} of BomaNPV was 18 fold higher than that of the BmNPV T3 strain (Table 5). This result showed that the BmNPV T3 strain was more infectious to the domesticated silkworm than was BomaNPV. This might be due to the coevolution of the BmNPV T3 strain, but not BomaNPV, with *B. mori* contributing an advantage to the BmNPV T3 strain.

In conclusion, the complete BomaNPV genome was determined and compared with that of the BmNPV T3 strain. The sequence and biological data suggested strongly that BomaNPV S1 and BmNPV T3 are variants with each other. Regarding the differences between BomaNPV and BmNPV T3, the hr1 and *bro* genes should be highlighted, especially the *bro* genes, for their close correlation with the evolution of baculoviruses. These results supported the idea that baculovirus strain heterogeneity may be often caused by SNPs in the whole genome and viriation in the hrs and bro genes (Zhang *et al*., 2005). The result of PIB bioassays indicated that BmNPV T3 was more infectious than BomaNPV for the domesticated silkworm. Understanding the basis for this difference will require further analysis of the changes in the 102 putative genes that diverge between the two strains.

Acknowledgements

This project was supported by the grants from the National Natural Science Foundation of China (30271008 and 30570074).

References

Astaurov, B.L., M.D. Golysheva, and I.S. Rovinskaya. 1959. Chromosome complex of Ussuri geographical race of *Bombyx mandarina* M. with special reference to the problem of the origin of the domesticated silkworm, *Bombyx mori*. *Cytology* 1, 327-332.

Carstens, E.B. and Y. Wu. 2007. No single homologous repeat region

Table 5. Concentration mortality of BomaNPV and BmNPV T3 for fourth instar *B. mori* larvae

Virus	Regression equation	χ^2 value	*P* value	LC_{50} (PIB/ml)	95% confidence limit (PIB/ml)		LC_{50} Ratio	95% confidence interval
					Lower	Upper		
BomaNPV	y=0.5482×+0.6345	0.7219	0.6970	91810393	21980125	2554219106	0.0515	0.0093-0.2860[a]
BmNPV T3 strain	y=1.2613×–3.4185	0.7155	0.3611	4725461	2140349	8398903		

[a] If the LC_{50} ratio in the 95% confidence interval includes value 1, the model is determined to be statistically significant.

is essential for DNA replication of the baculovirus *Autographa californica* multiple nucleopolyhedrovirus. *J. Gen. Virol.* 88, 114-122.

Chikushi, H. 1972. Genes and genetic stocks of the silkworm, pp. 1-6. Keigaku Publications, Tokyo, Japan.

Dreschers, S., R. Roncarati, and D. Knebel-Morsdorf. 2001. Actin rearrangement-inducing factor of baculoviruses is tyrosine phosphorylated and colocalizes to F-actin at the plasma membrane. *J. Virol.* 75, 3771-3778.

Finney, D.J. 1971. Probit Analysis. Cambridge University Press, Cambridge, UK.

Fukuda, Y., T. Washio, and M. Tomita. 1999. Comparative study of overlapping genes in the genomes of *Mycoplasma genitalium* and *Mycoplasma pneumoniae*. *Nucleic Acids. Res.* 27, 1847-1853.

Gan, X., Z. Ma, N. Deng, J. Wang, J. Ding, and L. Li. 2004. Involvement of the C-terminal proline-rich motif of G protein-coupled receptor kinases in recognition. *J. Biol. Chem*. 279, 49741-49746.

Jehle, J.A., G.W. Blissard, B.C. Bonning, J.S. Cory, E.A. Herniou, G.F. Rohrmann, D.A. Theilmann, S.M. Thiem, and J.M. Vlak. 2006. On the classification and nomenclature of baculoviruses: a proposal for revision. *Arch. Virol.* 151, 1257-1266.

Johnson, Z.I. and S.W. Chisholm. 2004. Properties of overlapping genes are conserved across microbial genomes. *Genome Res.* 14, 2268-2272.

Kane, L.P., M.N. Mollenauer, and A. Weiss. 2004. A proline-rich motif in the C terminus of Akt contributes to its localization in the immunological synapse. *J. Immunol.* 172, 5441-5449.

Kang, W.K., N. Imai, M. Suzuki, M. Iwanaga, S. Matsumoto, and E.A. Zemskov. 2003. Interaction of *Bombyx mori* nucleopolyhedrovirus BRO-A and host cell protein laminin. *Arch. Virol.* 148, 99-113.

Kang, W., M. Kurihara, and S. Matsumoto. 2006. The BRO proteins of *Bombyx mori* nucleopolyhedrovirus are nucleocytoplasmic shuttling proteins that utilize the CRM1-mediated nuclear export pathway. *Virology* 350, 184-191.

Kang, W., M. Suzuki, E. Zemskov, K. Okano, and S. Maeda. 1999. Characterization of baculovirus repeated open reading frames (*bro*) in *Bombyx mori* nucleopolyhedrovirus. *J. Virol.* 73, 10339-10345.

Kawaguchi, E. 1928. Zytologische Untersuchungen am Seidenspinner und seinen Verwandten. I. Gametogenese von *Bombyx mori* L. und *B. mandarina* M. und ihrer Bastar. Z. *Zellforsch. Mikrosk. Anat.* 7, 519-552.

Kawaoka, S., N. Hayashi, S. Katsuma, H. Kishino, Y. Kohara, K. Mita, and T. Shimada. 2008. *Bombyx* small RNAs: genomic defense system against transposons in the silkworm, *Bombyx mori. Insect Biochem. Mol. Biol*. 38, 1058-1065.

Keddie, B.A., G.W. Aponte, and L.E. Volkman. 1989. The pathway of infection of *Autographa californica* nuclear polyhedrosis virus in an insect host. *Science* 243, 1728-1730.

Krakauer, D.C. 2000. Stability and evolution of overlapping genes. *Evolution* 54, 731-739.

Kuzio, J., M.N. Pearson, S.H. Harwood, C.J. Funk, J.T. Evans, J.M. Slavicek, and G.F. Rohrmann. 1999. Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar. Virology* 253, 17-34.

López Ferber, M., O. Argaud, L. Croizier, and G. Croizier. 2001. Diversity, distribution, and mobility of *bro* gene sequences in *Bombyx mori* nucleopolyhedrovirus. *Virus Genes* 22, 247-254.

López, N. and M.T. Franze-Fernández. 2007. A single stem-loop structure in Tacaribe arenavirus intergenic region is essential for transcription termination but is not required for a correct initiation of transcription and replication. *Virus Res.* 124, 237-244.

Ma, X.C., H.J. Xu, M.J. Tang, Q. Xiao, J. Hong, and C.X. Zhang. 2006. Morphological, phylogenetic, and biological characteristics of *Ectropis obliqua* single-nucleocapsid nucleopolyhedrovirus. *J. Microbiol.* 44, 77-82.

Nakagawa, H., M. Tategu, R. Yamauchi, K. Sasaki, S. Sekimachi, and K. Yoshida. 2008. Transcriptional regulation of an evolutionary conserved intergenic region of CDT2-INTS7. *PLoS. ONE.* 3, e1484.

Nakamura, T., Y. Banno, T. Nakada, S.K. Nho, M.K. Xü, K. Ueda, T. Kawarabata, Y. Kawaguchi, and K. Koga. 1999. Geographic dimorphism of the wild silkworm, *Bombyx mandarina*, in a chromosome number and the occurrence of a retroposon-like insertion in the arylphorin gene. *Genome* 42, 1117-1120.

Normark, S., S. Bergström, T. Edlund, T. Grundström, B. Jaurin, F.P. Lindberg, and O. Olsson. 1983. Overlapping genes. *Ann. Rev. Genet.* 17, 499-525.

Pancio, H.A., N. Vander Heyden, and L. Ratner. 2000. The C-terminal proline-rich tail of human immunodeficiency virus Type 2 Vpx is necessary for nuclear localization of the viral preintegration complex in nondividing cells. *J. Virol.* 74, 6162-6167.

Pinschewer, D.D., M. Perez, and J.C. de la Torre. 2005. Dual role of the lymphocytic choriomeningitis virus intergenic region in transcription termination and virus propagation. *J. Virol.* 79, 4519-4526.

Rodems, S.M. and P.D. Friesen. 1993. The hr5 transcriptional enhancer stimulates early expression from the *Autographa californica* nuclear polyhedrosis virus genome but is not required for virus replication. *J. Virol*. 67, 5776-5785.

Roncarati, R. and D. Knebel-Mörsdorf. 1997. Identification of the early actin-rearrangement-inducing factor gene, *arif-1*, from *Autographa californica* multicapsid nuclear polyhedrosis virus. *J. Virol.* 71, 7933-7941.

Tang, Q. and M. Feng. 2002. DPS Data Processing System for Practical Statistics. Science Press, Peking, China.

Xiao, H. and Y. Qi. 2007. Genome sequence of *Leucania seperata* nucleopolyhedrovirus. *Virus Genes* 35, 845-856.

Yoshitake, N. 1968. Phylogenetic aspects on the origin of Japanese race of the silkworm, *Bombyx mori* L. *J. Sericult. Sci. Jpn.* 37, 83-87.

Zemskov, E.A., W. Kang, and S. Maeda. 2000. Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolhedrovirus BRO proteins. *J. Virol.* 74, 6784-6789.

Zhang, C.X., X.C. Ma, and Z.J. Guo. 2005. Comparison of the complete genome sequence between C1 and G4 isolates of the Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus. *Virology* 333, 190-199.