Journal of General Virology (1998), **79**, 2841–2851. Printed in Great Britain

Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome

Z. H. Hu,[1,2] B. M. Arif,[3] F. Jin,[2] J. W. M. Martens,[1]† X. W. Chen,[2] J. S. Sun,[2]‡ D. Zuidema,[1] R. W. Goldbach[1] and J. M. Vlak[1]

[1] Department of Virology, Wageningen Agricultural University, Binnenhaven 11, 6709 PD Wageningen, The Netherlands

[2] Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan 430071, People's Republic of China

[3] Forest Pest Management Institute, Sault Ste Marie, Ontario, Canada P6A 5M7

The genome organization of the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus (BusuNPV) was largely elucidated and compared to those of other baculoviruses. A detailed physical map was constructed for the restriction enzymes *Bam*HI, *Bgl*I, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Xba*I and *Xho*I. The 120·9 kbp viral genome was cloned as restriction fragments into a plasmid library from which about 43·5 kbp of dispersed sequence information was generated. Fifty-two putative open reading frames homologous to those of other baculoviruses were identified and their location in the genome of BusuNPV was determined. Although the gene content of BusuNPV is similar to that of *Autographa californica* multiple-nucleocapsid nucleopolyhedrovirus, *Bombyx mori* nucleopolyhedrovirus and *Orgyia pseudotsugata* multiple-nucleocapsid nucleopolyhedrovirus, the gene order is, however, significantly different from that observed in the other viruses, which have a high degree of collinearity. A new approach (GeneParity-Plot) was developed to represent the differences in gene order among baculoviruses when limited sequence information is available and to take advantage of the high degree of gene conservation. The data obtained show that BusuNPV is a distinct baculovirus species and the analyses suggest that gene distribution along baculovirus genomes may be used as a phylogenetic marker.

Introduction

The *Baculoviridae*, a diverse family of more than 600 viruses, contains two genera, the genus *Nucleopolyhedrovirus* (NPV) and the genus *Granulovirus* (GV) (Murphy *et al.*, 1995). The virions are rod-shaped and contain a circular, double-stranded DNA genome of 90–160 kbp depending on the species. Baculoviruses are pathogenic for arthropods, particularly insects of the order Lepidoptera. The current interest in the molecular biology of these viruses is fostered by their potential as alternatives to chemical insecticides in the control of agricultural and forest insect pests and also by their successful use as vectors for the expression of foreign proteins. To date, the complete sequences of the genomes of three baculoviruses, that of *Autographa californica* multiple-nucleocapsid nucleopolyhedrovirus (AcMNPV), *Bombyx mori* nucleopolyhedrovirus (BmNPV) and *Orgyia pseudotsugata* multiple-nucleocapsid nucleopolyhedrovirus (OpMNPV) have been determined (Ayres *et al.*, 1994; S. Maeda, GenBank accession number L33180; Ahrens *et al.*, 1997). The genome of AcMNPV is 133 894 bp and potentially encodes 154 proteins, that of BmNPV is 128 413 bp and contains 136 putative genes, and the OpMNPV genome is 131 990 bp and contains 152 putative genes. Genomic comparisons have indicated that, even though small inversions and insertions (or deletions) are observed, these three baculovirus genomes have a similar gene content and a similar gene arrangement along the genome.

As more genomic sequences become available from other baculoviruses, their molecular biology will be better understood and their phylogenetic relationships and evolution will be more accurately determined. Until now, only limited DNA

Author for correspondence: Just Vlak.
Fax +31 317 484820. e-mail just.vlak@medew.viro.wau.nl

† **Present address:** Department of Endocrinology and Reproduction, Faculty of Medicine, Erasmus University, 3000 DR Rotterdam, The Netherlands.

‡ **Present address:** National Research Institute for Family Planning, Beijing 100081, People's Republic of China.

0001-5652

sequence information has been reported from single-nucleocapsid (S) NPVs and GVs. Consequently, our understanding of the gene content and organization of these viruses is rudimentary in comparison to that of MNPVs. In this paper, we present a study elucidating the genome organization of an SNPV infecting *Buzura suppressaria* Guenée (Lepidoptera: *Geometridae*). The virus (BusuNPV) was originally isolated from this major pest of tea plantations in China and it has since been used to control infestations (Xie *et al*., 1979). So far, about 63 NPVs have been described from the insect family *Geometridae* (Adams & McClintock, 1991), but molecular characterization of these viruses has been limited. A study based on the amino acid sequences of baculovirus ecdysteroid UDP-glucosyltransferase (*egt*) genes has indicated that BusuNPV belongs to a phylogenetic group within the family *Baculoviridae*, which is distinct from that encompassing AcMNPV, BmNPV and OpMNPV (Hu *et al*., 1997). Sequence analysis of a 3·2 kbp *Hin*dIII fragment located at map unit (*mu*) 26·6–29·4 and the *p10* gene region (*mu* 94·0–95·3) revealed that the genome organization of BusuNPV might differ considerably from those of other baculoviruses investigated so far (Hu *et al*., 1998; van Oers *et al*., 1998).

In order to investigate the gene organization of BusuNPV and evaluate the phylogenetic status of this virus, a detailed physical map of the viral genome was constructed. A genomic library was made by cloning the viral DNA as restriction fragments into plasmid vectors and the partial or total nucleotide sequence of the inserts was determined. The sequence data were used to identify open reading frames (ORFs) and potential genes by comparison with databases. A genetic map describing the order of ORFs in the genome of BusuNPV was generated and compared to those of AcMNPV, BmNPV and OpMNPV using a novel approach (GeneParityPlot).

Methods

■ **Virus and DNA.** The virus used in this study was originally isolated from an infestation by *B. suppressaria* larvae of a tea plantation in Hubei province in the People's Republic of China (Xie *et al*., 1979). The virus was propagated in these larvae and occlusion bodies were purified by differential and rate-zonal centrifugation. The DNA was isolated directly from purified occlusion bodies by dissolution in 0·1 M Na_2CO_3, 0·01 M EDTA and 0·17 M NaCl, followed by proteinase K and SDS treatment, phenol–chloroform extraction and dialysis (Caballero *et al*., 1992). The purity of the DNA was determined spectrophotometrically.

■ **Restriction endonuclease analysis and cloning of viral DNA fragments in plasmid vectors.** Viral DNA was digested with various restriction enzymes (GIBCO/BRL) and the fragments were separated by electrophoresis in 0·6, 0·7 and 0·8% agarose gels at 45 V (1·5 V/cm) for 14–20 h. Lambda DNA digested with *Bam*HI/*Eco*RI/*Hin*dIII was used as a molecular size standard.

Viral DNA restriction fragments were cloned into the plasmid vector pTZ19R by standard techniques (Sambrook *et al*., 1989). Firstly, the shotgun method was used to generate a library of clones. Secondly, restriction fragments, absent from the shotgun library, were separated in agarose gels, purified by the freeze–squeeze method, and cloned individually into pTZ19R. Restriction enzyme analysis and Southern blot hybridization were performed to confirm the authenticity and the location of the cloned fragments on the physical map of the BusuNPV genome.

■ **Construction of the physical map of the BusuNPV genome.** The order of the restriction fragments on the viral genome was derived from single and double digestion of the cloned fragments with various enzymes, as well as by Southern blot hybridization. Sequence data of the cloned fragments were also used to establish and confirm the detailed map.

■ **DNA sequencing and computer analysis.** Plasmid DNA for sequencing was purified via Qiagen columns. Partial sequencing was conducted on both ends of the cloned fragments with the universal forward and reverse primers. Certain regions of the genome were selected for complete sequencing of both strands using either a series of overlapping clones containing nested deletions or by the 'sequence walking' method with custom synthesized primers. The T7 DNA polymerase sequence system (Promega) was used for manual sequencing and automatic sequencing was carried out at the Sequencing Facility in the Department of Molecular Biology of the Wageningen Agricultural University (Wageningen, Netherlands) or at the Core Facility at Queens University (Kingston, Ontario, Canada). The generated sequences were analysed with UWGCG computer programs (release 9.0). The DNA and the deduced amino acid sequences were compared with the updated GenBank/EMBL, SWISSPROT and PIR databases using FASTA and BLAST. For studying baculovirus gene homology, the amino acid sequences of the selected homologues were aligned with the PileUp program of GCG. The resulting alignments were then imported into GeneDoc software from which the amino acid identities were calculated. The nucleotide sequences of the complete ORFs of BusuNPV have been deposited in GenBank and assigned the following accession numbers: X70844 (polyhedrin gene; Hu *et al*., 1993); U61154 (*egt*; Hu *et al*., 1997); AF058928 (homologue of *Ac111*); AF060564 (late expression factor 2 gene); AF045936 (homologue of *Ac117*, conotoxin-like gene, inhibitor of apoptosis-like gene, superoxide dismutase gene; Hu *et al*., 1998); AF058929 (cathepsin gene); and AF034410 (*p26*, *p10*; van Oers *et al*., 1998). The partial sequences are available upon request.

■ **Gene order in the baculovirus genome.** In order to facilitate comparisons of gene orders among BusuNPV, AcMNPV, BmNPV and OpMNPV, only those gene homologues identified so far in all genomes were selected. The identified BusuNPV ORFs were provisionally numbered 1–52 (Fig. 3). To allow a computational comparison, the BusuNPV ORF homologues of AcMNPV, BmNPV and OpMNPV (Table 2; first four columns) were renumbered manually starting with the polyhedrin gene as number 1 and renumbering the other ORFs according to their sequential occurrence on the respective linearized genomes from left to right (Table 2; last eight columns). With these reassigned numbers as input data (see Table 2), the Chart Program packaged in Microsoft Excel software was then used to obtain a graphic comparison. The input data are sorted in ascending order according to the gene order of one virus (e.g. AcMNPV). By choosing 'line' as chart type the GeneParityPlot is obtained showing the gene order as a line. The X-axis then represents the gene order of one given virus and the Y-axis the gene order of the other virus. This diagonal line (reference) is obtained when a given virus is on both the X-axis and Y-axis. From such GeneParityPlots, the difference in gene organization between two baculoviruses can be conveniently compared.

If the gene arrangement of a given virus is collinear with that of AcMNPV, then a straight parity line will appear. A gene or gene cluster

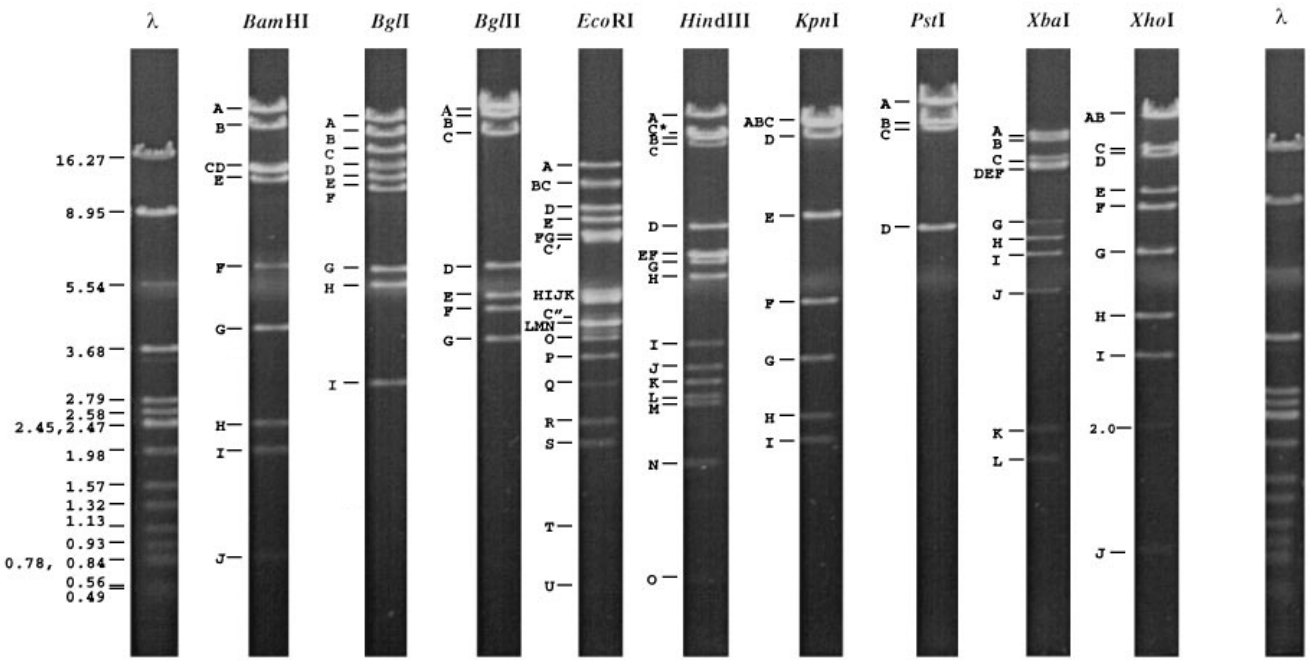

Fig. 1. BusuNPV DNA digested with *Bam*HI, *Bgl*I, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Xba*I and *Xho*I and electrophoretically separated in 0·6% agarose. The fragments are named alphabetically, starting with A for the largest fragment. Submolar bands, such as *Hin*dIII-C*, were also assigned. Lambda DNA digested with *Bam*HI/*Eco*RI/*Hin*dIII was used as molecular marker with sizes indicated in kbp.

Table 1. Size of restriction endonuclease fragments (in kbp) of BusuNPV DNA

Numbers in bold indicate that the fragment has been cloned in a plasmid vector and numbers that are underlined indicate that the size of the fragment was confirmed by sequence analysis.

Fragment	*Bam*HI	*Bgl*I	*Bgl*II	*Eco*RI	*Hin*dIII	*Kpn*I	*Pst*I	*Xba*I	*Xho*I
A	40·3	29·7	47·1	**13·9**	32·4	27·4	62·8	19·5	28·5
B	25·4	22·2	32·6	11·9	20·2	26·1	26·6	18·3	28·2
C	13·9	16·6	21·5	**11·7**	**18·1**	25·3	23·5	**14·0**	16·1
D	13·6	14·5	**6·2**	9·2	**8·1**	20·8	**7·8**	13·2	15·0
E	**12·0**	12·2	5·2	**8·8**	**6·5**	8·6	**0·2**	**13·2**	9·8
F	6·3	11·5	4·5	7·7	**6·4**	**4·8**		**12·7**	8·7
G	**4·2**	6·0	**3·8**	**7·3**	**6·1**	**3·4**		**7·8**	6·3
H	**2·4**	5·2		**5·2**	**5·8**	**2·4**		6·9	4·2
I	**2·0**	3·0		**5·2**	**3·6**	**2·1**		**6·3**	3·3
J	**0·8**			**5·1**	**3·2**			4·9	0·8
K				4·9	**2·9**			**2·2**	
L				4·4	**2·7**			1·9	
M				4·4	2·6				
N				4·2	**1·6**				
O				3·9	**0·6**				
P				**3·5**	**0·1**				
Q				2·9					
R				**2·4**					
S				**2·1**					
T				1·1					
U				0·6					
V				**0·4**					
W				**0·1**					

that is shifted to a different location will appear as a dot or a small line parallel to the parity line, with the distance from the diagonal dependent on the location. The direction of transcription of genes was also taken into account to identify inversions. When the direction of one ORF is the same as that of its homologue in AcMNPV, an open symbol was used and when the direction is opposite, a solid symbol was used. Therefore, a typical inversion will be a solid line perpendicular to the diagonal. The resulting 'GeneParityPlot' shows the clustering of genes and their spatial distribution without knowing the entire sequence.

Results and Discussion

Restriction endonuclease analysis of BusuNPV DNA

Digestion of BusuNPV with *Bam*HI, *Bgl*I, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Xba*I and *Xho*I, produced 10, 9, 7, 23, 16, 9, 5, 12 and 10 fragments, respectively. Restriction profiles of these digestions of BusuNPV genome DNA are shown in Fig. 1 and sizes of the fragments are summarized in Table 1. These sizes were estimated from the original agarose gel and adjusted according to the data obtained from the mapping and sequencing experiments. The total size of the BusuNPV was determined to be 120·9 kbp which is about 8–13 kbp smaller than the genomes of AcMNPV, OpMNPV and BmNPV (Ayres *et al.*, 1994; S. Maeda, GenBank accession number L33180; Ahrens *et al.*, 1997).

A few submolar bands, such as *Bam*HI-F, *Eco*RI-Q and *Xba*I-G were observed in the restriction profiles indicating that the virus isolate used in this study was not homogeneous, but contained major and minor genotypic variants (Fig. 1). Comparison with a baculovirus isolate from *Buzura thibtaria* indicated that many of the submolar bands observed for BusuNPV were equimolar in ButhNPV (data not shown). *In vivo* cloning techniques (Smith & Crook, 1988) or plaque purification techniques using cultured cells (King & Possee, 1992) are required to isolate and characterize the genotypic variants of BusuNPV in more detail.

Restriction map of BusuNPV DNA

With the aid of a genomic library of clones containing overlapping restriction fragments and additional extensive cross hybridization experiments, we were able to generate a physical map encompassing 103 restriction sites for *Bam*HI, *Bgl*I, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Xba*I and *Xho*I for the major genotype (Fig. 2). Some fragments were submolar as a consequence of the presence of multiple genotypes in the isolate. Fragments *Eco*RI-Q and *Eco*RI-T were less than equimolar. A subpopulation of genotypes lacked an *Eco*RI site between these two fragments (Fig. 2) and formed a submolar

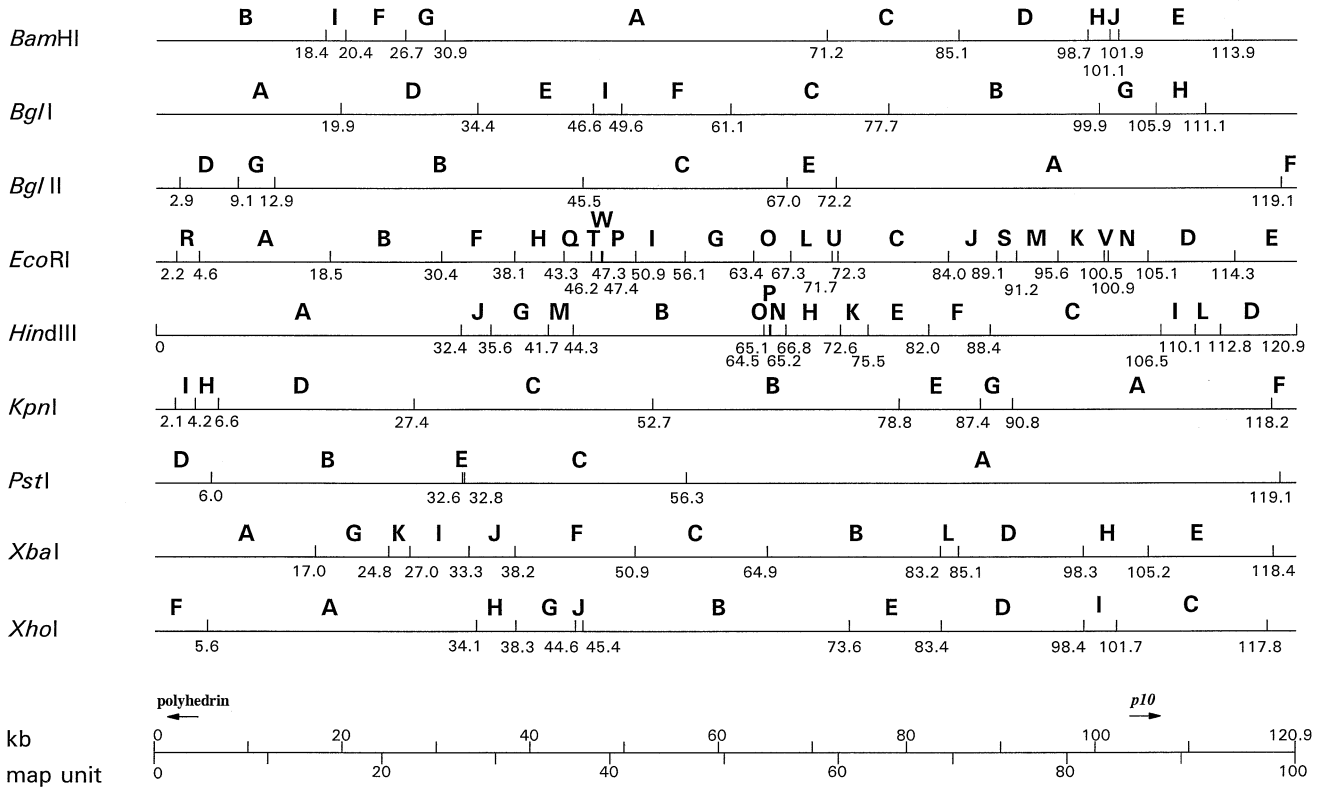

Fig. 2. Linearized physical map of BusuNPV DNA with restriction sites for *Bam*HI, *Bgl*I, *Bgl*II, *Eco*RI, *Hin*dIII, *Kpn*I, *Pst*I, *Xba*I and *Xho*I. The restriction sites are indicated in kbp from the zero point. The genome size in kbp and map units is indicated on a scale at the bottom. The location and the direction of transcription of polyhedrin and *p10* are shown by arrows.

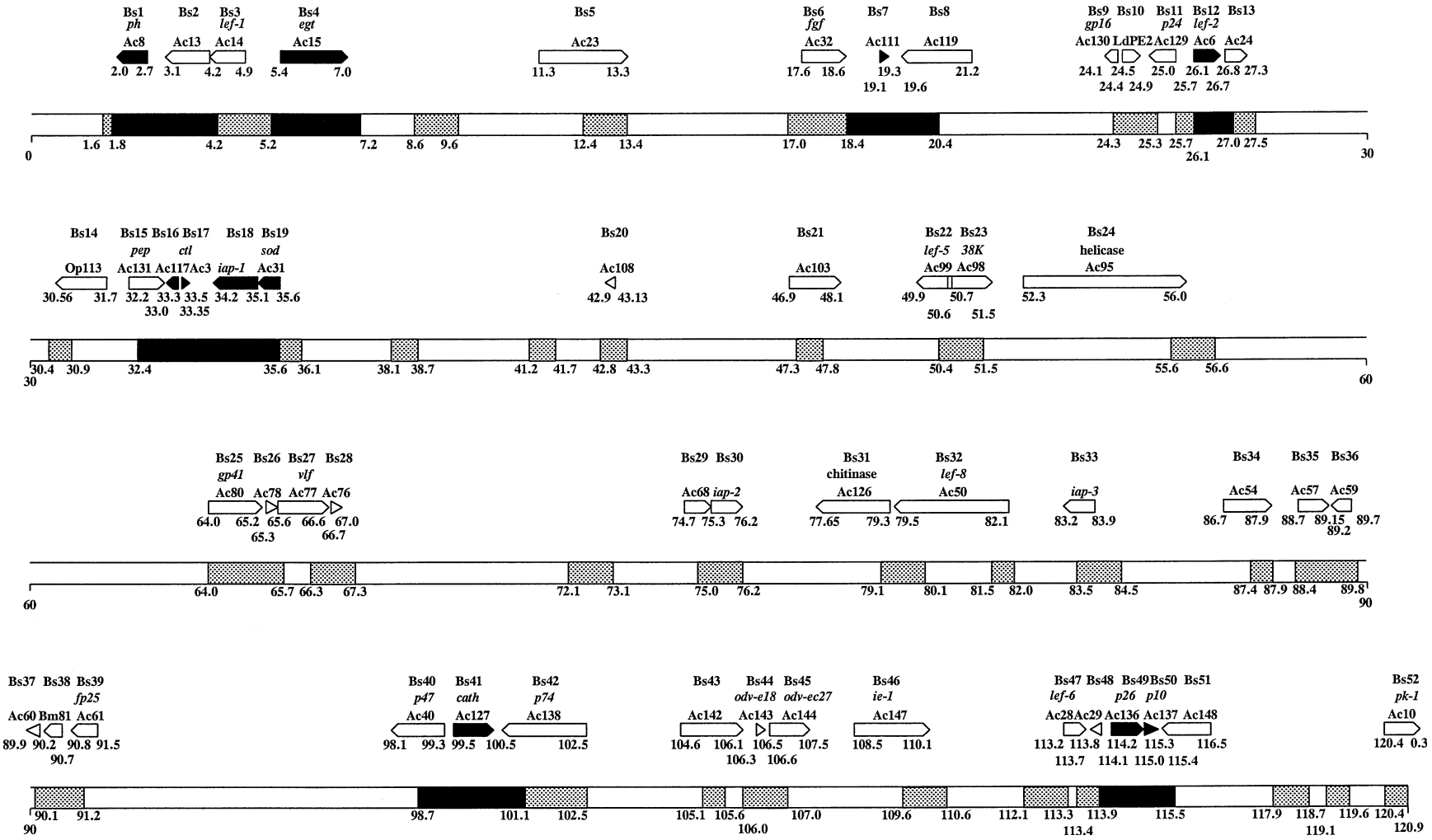

Fig. 3. Genomic organization of the BusuNPV DNA. The upper part of each panel represents the location of putative ORFs; open arrows indicate the direction of transcription; completely sequenced ORFs are shown as solid arrows. The location of the ORFs, given in map units (*mu*), is approximated. Names assigned to ORFs are referenced in the text and the homologues of AcMNPV ORFs (or other baculovirus when there is no homologue from AcMNPV) are shown. The location (in *mu*) of the 3′ and 5′ ends of the ORF were approximations based on the size of baculovirus homologues. The lower part of each panel represents an overview of the sequence information; the black region indicates that the section has been sequenced from both strands; and the shaded region indicates a sequence of one strand.

Table 2. ORFs identified in BusuNPV and their homologues in AcMNPV, BmNPV and OpMNPV

The first four columns represent the actual gene numbering derived for BusuNPV from Fig. 3 and for AcMNPV, BmNPV and OpMNPV from Ayres *et al.* (1994), S. Maeda (GenBank accession number L33180) and Ahrens *et al.* (1997), respectively. The last eight columns represent the assigned ORF number after transformation and the direction of transcription of the ORFs in their respective genomes.

ORF						Order				Direction			
Busu	Ac	Bm	Op	Name	Abbreviation	Ac	Bm	Op	Busu	Ac	Bm	Op	Busu
1	8	1	3	Polyhedrin	*ph*	1	1	1	1	<	>	<	<
2	13	5	12			3	3	3	2	<	<	<	<
3	14	6	13	Late expression factor	*lef-1*	4	4	4	3	<	<	<	<
4	15	7	14		*egt*	5	5	5	4	>	>	>	>
5	23	14	21			6	6	6	5	>	>	>	>
6	32	24	27	Fibroblast growth factor	*fgf*	11	11	7	6	<	>	>	>
7	111	93	112			29	29	29	7	<	<	<	>
8	119	97	119			31	31	31	8	>	>	>	<
9	130	107	128		*gp16*	35	35	35	9	>	>	>	<
10				LdMNPV PEP ORF2									>
11	129	106	127	Capsid protein	*p24*	34	34	34	10	>	>	>	<
12	6	135	6	Late expression factor	*lef-2*	45	45	2	11	>	>	<	>
13	24	15	44			7	7	11	12	<	<	>	>
14			113									>	<
15	131	108	129	Calyx protein	*pep*	36	36	36	13	>	>	>	>
16	117	96	117			30	30	30	14	>	>	>	<
17	3		136	Conotoxin	*ctl-1*					>	>		>
18*				iap-like gene	*iap-1*								<
19	31	23	29	Superoxide dismutase	*sod*	10	10	8	15	>	>	<	<
20	108	91	108			28	28	28	16	<	<	<	<
21	103	87	104			27	27	27	17	<	<	<	>
22	99	83	100	Late expression factor	*lef-5*	26	26	26	18	>	>	>	<
23	98	82	99		*38K*	25	25	25	19	<	<	<	>
24	95	78	96	Helicase		24	24	24	20	<	<	<	>
25	80	66	83		*gp41*	23	23	23	21	<	<	<	>
26	78	64	81			22	22	22	22	<	<	<	>
27	77	63	80	Very late factor	*vlf-1*	21	21	21	23	<	<	<	>
28	76	62	79			20	20	20	24	<	<	<	>
29	68	56	73			18	18	18	25	>	>	>	>
30*	71	58	74	iap-like gene	*iap-2*	19	19	19	26	>	>	>	>
31	126	103	124	Chitinase	*chiA*	32	32	32	27	<	<	<	<
32	50	39	54	Late expression factor	*lef-8*	13	13	13	28	<	<	<	<
33*				iap-like gene	*iap-3*								<
34	54	43	58			14	14	14	29	>	>	>	>
35	57	46	61			15	15	15	30	>	>	>	>
36	59		62							<	<		<
37	60	48	63			16	16	16	31	<	<	<	<
38		81									>		<
39	61	49	64	Few polyhedra	*25K*	17	17	17	32	<	<	<	<
40	40	31	45		*p47*	12	12	12	33	<	<	<	<
41	127	104	125	Cathepsin	*cath*	33	33	33	34	>	>	>	>
42	138	115	134		*p74*	39	39	39	35	<	<	<	<
43	142	118	139			40	40	40	36	>	>	>	>
44	143	119	140	ODV protein	*odv-e18*	41	41	41	37	>	>	>	>
45	144	120	141	ODV protein	*odv-ec27*	42	42	42	38	>	>	>	>
46	147	123	145	Major transactivator	*ie-1*	43	43	43	39	>	>	>	>
47	28	19	40		*lef-6*	8	8	10	40	>	>	<	>
48	29	20	39			9	9	9	41	<	<	>	<
49	136	113	132		*p26*	37	37	37	42	>	>	>	>
50	137	114	133	Fibrillin	*p10*	38	38	38	43	>	>	>	>
51	148	124	146	ODV protein	*odv-e56*	44	44	44	44	<	<	<	<
52	10	3	1	Protein kinase	*pk-1*	2	2	45	45	>	>	>	>

* The *iap* genes of BusuNPV are numbered according to their location in the BusuNPV genome.

band (Q + T) of about 4 kbp. A similar situation existed for *Eco*RI-C (11·7 kbp) which, in a subpopulation of genotypes, has an additional *Eco*RI site giving fragments *Eco*RI-C′ and *Eco*RI-C″ of 7·3 and 4·4 kbp in size, respectively. Fragments *Hin*dIII-C and *Hin*dIII-I are also less than equimolar, but they form a contig (*Hin*dIII-C*) in a subpopulation of genotypes. Fragments *Xba*I-G, *Eco*RI-B and *Bam*HI-F were submolar, but this may be explained by the observation that these fragments cover a genomic region (*mu* 19) which is highly variable and contains multiple sequence repeats in an intergenic region (data not shown). All submolar fragments were either found in our plasmid library and or identified by hybridization. The origin of a minor molar *Xho*I fragment of 2·0 kbp was not further investigated.

Since the entire polyhedrin gene (*ph*) is contained within the fragment *Hin*dIII-A (Hu *et al*., 1993), the linearized map was drawn in such a way that the zero point is located at the beginning of this fragment, as proposed by Vlak & Smith (1982). This also positioned *p*10 (van Oers *et al*., 1998) to the right-hand end of the map. The linearized map and its orientation reported in this paper differ from the preliminary map published previously (Liu *et al*., 1993). Fragment *Hin*dIII-J (Fig. 2) has been previously assigned *Hin*dIII-I (Hu *et al*., 1998).

Sequence analyses of BusuNPV DNA

In order to generate sufficient information to decipher the gross gene content and arrangement in the genome of BusuNPV, approximately 43·5 kbp of the genome was sequenced. A total of 52 putative ORFs and genes homologous to those of other baculoviruses were identified by comparison with databases. These 52 ORFs were provisionally named Bs1–Bs52 from left to right according to their sequential location on the physical map (Fig. 3). Of these 52 ORFs, 11 are completely sequenced, including *ph*, *egt*, late expression factor 2 gene (*lef*-2), conotoxin-like peptide gene (*ctl*), superoxide dismutase gene (*sod*), cathepsin gene (*cath*), *p*26 and *p*10.

Comparison of the overall gene content of divergent baculoviruses provides an initial assessment of the genes that are retained in all genomes and thus are likely to be essential for virus multiplication and survival. This comparison may also provide insight into the evolutionary origin of baculoviruses and their relatedness with the family *Baculoviridae*. Sequence analyses of AcMNPV, BmNPV and OpMNPV have so far revealed a total of 184 different baculoviral ORFs (Ayres *et al*., 1994; S. Maeda, GenBank accession number L33180; Ahrens *et al*., 1997). Among these ORFs, 119 are shared by the three genomes, 21 are present in two of the three genomes and 44 are unique to individual genomes (14 ORFs for AcMNPV, 4 ORFs for BmNPV and 26 ORFs for OpMNPV). This indicates a high degree of gene conservation.

Table 2 shows the 52 ORFs of BusuNPV and their homologues in AcMNPV, BmNPV and OpMNPV. It is interesting to note that genes previously thought to be unique for OpMNPV (Op113) or BmNPV (Bm81), were also found in BusuNPV, as Bs14 and Bs38, respectively. A homologue of the ORF2 in the polyhedral envelope protein (PEP) region of *Lymantria dispar* MNPV (Bjornson & Rohrmann, 1992) was also found in BusuNPV as Bs10. Three inhibitor of apoptosis-like genes (*iap*) were identified in BusuNPV and they are numbered according to their relative location on the linearized BusuNPV genome (Fig. 3). So far, the sequence information indicates that *iap*-2 of BusuNPV belongs to the lineage of *iap*-2 of AcMNPV, BmNPV and OpMNPV, whereas the grouping of BusuNPV *iap*-1 and *iap*-3 is not clear (data not shown). Therefore, only the counterparts of *iap*-2 are shown in Table 2.

The 52 BusuNPV ORFs occupied about 70% of the sequenced region. Combined with the information from AcMNPV, BmNPV and OpMNPV, this suggests an extensive conservation of gene content in the NPV genomes. The remainder of the sequenced BusuNPV regions consists of intergenic sequences between ORFs, as well as some potential ORFs that do not have significant homology to any sequence in GenBank. If the sequence information reported here represents a random distribution, it can be predicted that the unique ORFs occupy less than 25% of the BusuNPV genome. The high conservation of the gene content indicates that the NPVs have evolved from an ancestor whose genetic make-up is not that dissimilar to that of the present day baculoviruses. In fact, this conservation is what defines a baculovirus including the genes essential for DNA replication, expression, assembly of progeny particles and occlusion at the end of the replication cycle. Some of these characteristic baculovirus genes have also been identified in the BusuNPV genome, such as immediate early gene 1 (*ie-1*), helicase gene, several of the late expression factor genes (*lef-1*, *lef-2*, *lef-5*, *lef-6*, *lef-8*), and the 'very late factor' gene (*vlf-1*), as well as some structural genes known as *gp41*, *odv-e18*, *odv-ec27*, *odv-e56*, *ph*, *p10* and *pep* (Fig. 3; Table 2).

Apart from the above characteristic genes, baculoviruses also contain the so-called auxiliary genes. These genes are not essential for replication, but some provide the virus with selective advantages in nature (see O'Reilly, 1997 for review). Auxiliary genes, such as *egt*, *ctl*, *iap*, *sod*, *cath* and the chitinase gene (*chiA*), were also found in BusuNPV. EGT delays larval moulting and allows the virus to produce large numbers of progeny particles. The *cath* and the *chiA* genes are needed for larval liquefaction and thus aid dissemination of the virus in nature (Hawtin *et al*., 1997). It is plausible that an ancestral baculovirus contained some of the auxiliary genes and that the present baculovirus survived through evolution partly because of the advantages conferred by their encoded proteins. The conservation of these auxiliary genes in baculovirus genomes suggests that they are important to the virus life-cycle.

Baculovirus gene homology

Comparison of gene homology among different baculoviruses provides an initial measure of gene conservation and an

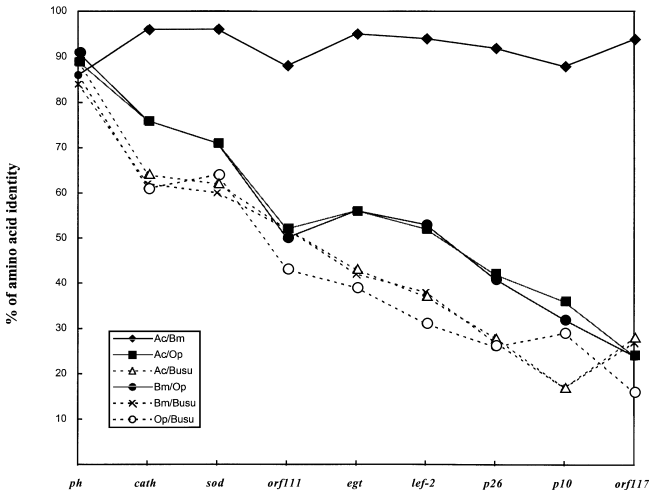

Fig. 4. Pairwise homology of nine ORFs of AcMNPV (Ac), BmNPV (Bm), OpMNPV (Op) and BusuNPV (Bs). The homology is expressed as amino acid sequence identity. Gene nomenclature is according to Ayres *et al.* (1994).

insight into the function of these genes. Furthermore, based on gene homology, such comparisons may also have implications for the diversity of baculoviruses. Among the 52 ORFs of BusuNPV, 11 ORFs have been completely sequenced from both strands. These include Bs1 (*ph*), Bs4 (*egt*), Bs7 (*Ac111*), Bs12 (*lef-2*), Bs16 (*Ac117*), Bs17 (*ctl*), Bs18 (*iap-1*), Bs19 (*sod*), Bs41 (*cath*), Bs49 (*p26*) and Bs50 (*p10*). Nine of the above 11 ORFs were used to study the gene homology among BusuNPV, AcMNPV, BmNPV and OpMNPV. The *ctl* was not included in the analysis because it does not exist in BmNPV. Similarly, the *iap* genes were not used because of the multiple occurrence in all four genomes.

Pairwise comparison of amino acid sequence identity of the products of nine selected genes revealed two kinds of conservation patterns (Fig. 4). For the closely related baculoviruses, such as AcMNPV and BmNPV, the homology is high for all the genes. Hence, when all genes are aligned, a conserved pattern with constantly high amino acid identity appeared. This pattern is characteristic of closely related baculoviruses and distinguishes these from the others. When less related viruses are compared, it is obvious that the extent of gene homology differs among individual genes. For example, the amino acid sequence identity ranks from in excess of 84% for polyhedrin to no higher than 28% for the *orf117* gene product. It is plausible that the extent of homology is related to the function of different genes. The pairwise comparison is useful to identify closely related baculoviruses when sequence information of genes that are not too highly conserved is available. In the case of OpMNPV and *Perina nuda* (Penu) MNPV, for example, the amino acid sequence identity is about 95% for the *p10* gene product (van Oers *et al.*, 1994) and 97% for the *lef-2* gene product (X. W. Chen, personal communication). Therefore, it is very likely that these two viruses are closely related.

Although the genome organization of OpMNPV is quite similar to that of AcMNPV and BmNPV as compared to BusuNPV (see below), gene homology data show that OpMNPV is relatively distant from AcMNPV and BmNPV. The homology of BusuNPV ORFs to their counterparts in other viruses (dashed lines in Fig. 4) is slightly lower than the homology of OpMNPV ORFs to their counterparts in AcMNPV and BmNPV. This suggests that gene order and gene homology are two independent parameters in the study of baculovirus phylogeny.

Distinct gene arrangement in the BusuNPV genome

While the gene organization in AcMNPV, BmNPV and OpMNPV is basically similar, we investigated whether the organization in the BusuNPV genome is collinear with these viruses. The homologues of 45 ORFs which exist in all four genomes were chosen for comparison in the GeneParityPlot analysis (see Methods). After transformation, the resulting relative order number and direction of the ORFs were determined (listed in Table 2). The comparison of gene arrangement of the selected ORFs is shown in Fig. 5. Although the 45 ORFs represent only about 30% of the total ORFs in the baculovirus genome, the comparison of BmNPV and OpMNPV with AcMNPV (Fig. 5*a*, *b*) well represents the actual collinearity of these genomes (Ahrens *et al.*, 1997). The inversion in OpMNPV (Op10–Op21) in comparison to the AcMNPV genome (Ahrens *et al.*, 1997) is reflected in Fig. 5 (*b*) as being perpendicular to the parity line. It is clear that the gene arrangement in BusuNPV is quite different from that in AcMNPV, BmNPV and OpMNPV, as evidenced from the dispersal of individual contig ORFs. This strongly suggests that BusuNPV is sufficiently distinct to receive a species status in the genus *Nucleopolyhedrovirus* (Murphy *et al.*, 1995).

Attempts to find conserved gene clusters in the viral genomes may provide an initial insight into the evolution of baculovirus genomes. The method used in this paper provides a new approach to identify conserved gene clusters among baculovirus genomes. Despite differences in the gene arrangement, certain genes that are clustered in the genomes of AcMNPV, BmNPV and OpMNPV, also remained together in the BusuNPV. Eight potential clusters have been identified so far (Fig. 5*c*): 1, *Ac13-14-15-23*; 2, *Ac28-29*; 3, *Ac50-54-57-60-61*; 4, *Ac68-71*; 5, *Ac76-77-78-80-95-98-99-103*; 6, *Ac129-130*; 7, *Ac137-138*; and 8, *Ac138-142-143-144-147*. The numbers refer to the original ORFs in AcMNPV (Ayres *et al.*, 1994). These clusters can be refined when more sequence information of BusuNPV becomes available. Since the genome of OpMNPV is largely collinear with that of AcMNPV, the GeneParityPlot pattern of BusuNPV remains similar when the OpMNPV order is used as a reference instead of that of AcMNPV (Fig. 5*d*).

In addition to the common gene clusters described above, BusuNPV contains certain clusters which are also found in some other baculoviruses but not in AcMNPV, OpMNPV or

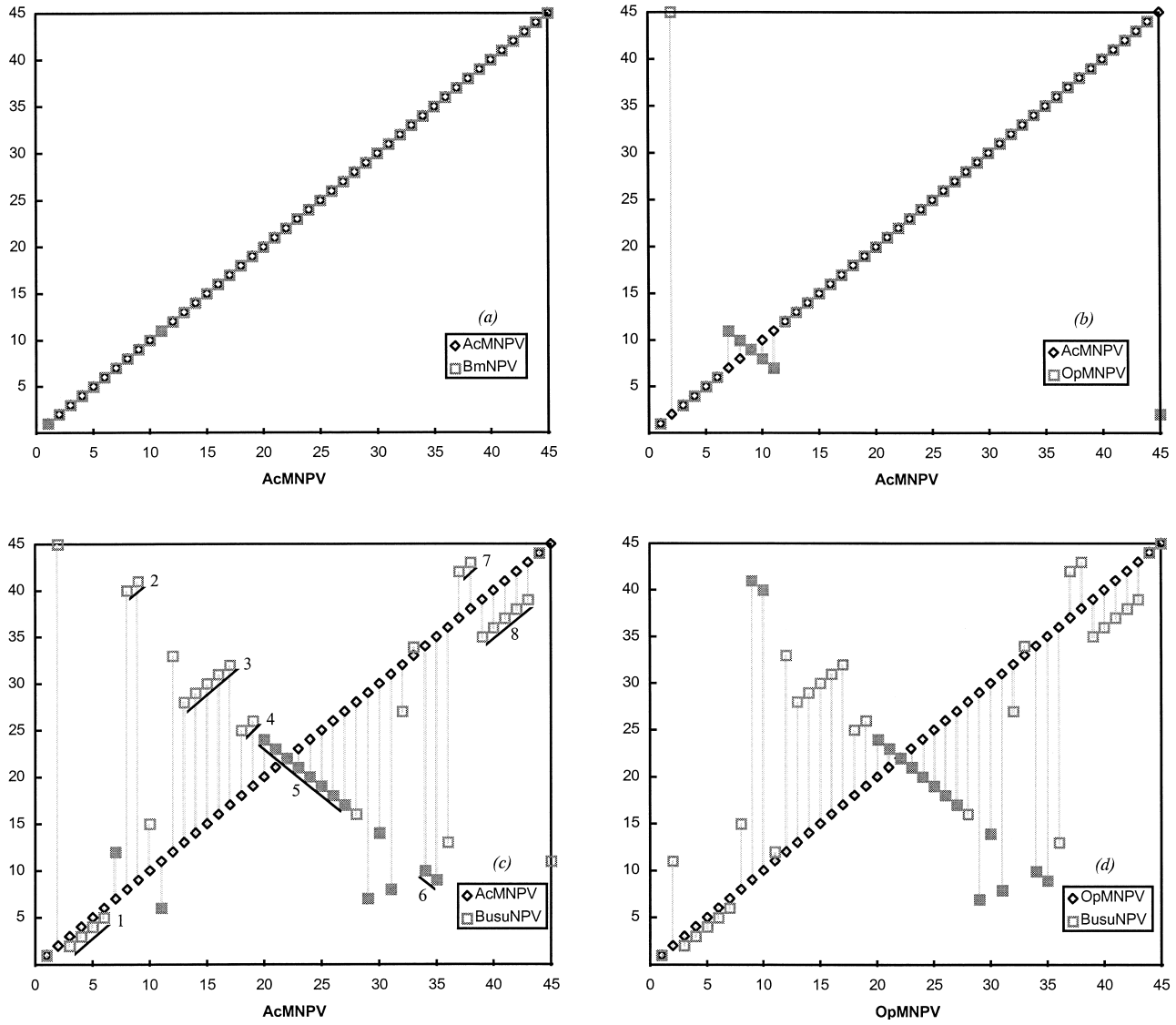

Fig. 5. Graphic representation of the collinearity of baculovirus genomes obtained by GeneParityPlot analysis (see Methods and Table 2). (*a*) BmNPV versus AcMNPV; (*b*) OpMNPV versus AcMNPV; (*c*) BusuNPV versus AcMNPV; (*d*) BusuNPV versus OpMNPV. Eight putative gene clusters of the BusuNPV genome, which are similar to those of AcMNPV, were numbered and underlined. Solid symbols indicate that the direction of transcription of an ORF is opposite to that of its homologues in AcMNPV, whereas open symbols indicate the same direction of transcription.

BmNPV. For example, the cluster encompassing Bs47(*lef-6*)-Bs48(*Ac29*)-Bs49(*p26*)-Bs50(*p10*) (Fig. 3), which is absent in the genome of AcMNPV, OpMNPV and BmNPV, also exists in the genome of *Spodoptera exigua* (Se) MNPV (van Strien *et al.*, 1997). This indicates that such an arrangement might have already existed before SeMNPV and BusuNPV diverged into different species. Another gene cluster, Bs31(*chiA*)-Bs32(*lef-8*) (Fig. 3), is also present in the genome of *Helicoverpa zea* (Hz) SNPV (GenBank U67265). The absence of *Ac79* between Bs25 (*Ac80*; *gp41*) and Bs26 (*Ac78*) in cluster 5, is also observed in *Spodoptera frugiperda* (Sf) MNPV (GenBank accession number U14725) and HzSNPV (GenBank accession number L04747). It is noteworthy that the cathepsin and chitinase genes, normally positioned in a head-to-head arrangement in all baculovirus genomes characterized so far, are located distally from each other in BusuNPV. This confirms a recent study (Hawtin *et al.*, 1997) that the function of these two genes, involved in larval liquefaction, is not dependent on their proximal position in many baculovirus genomes.

Gene arrangements of baculovirus genomes as phylogenetic markers

The study of molecular evolution of baculoviruses has so far been based on the phylogeny of single genes. Zanotto *et al.*

(1993), for example, constructed a phylogenetic tree based on polyhedrin. Accordingly, most lepidopteran NPVs fall into two large groups, Group I and Group II. A similar grouping was observed using other baculovirus genes, such as *egt* (Chen *et al.*, 1997). Some of the NPVs mentioned in this paper such as AcMNPV, BmNPV and OpMNPV, belong to Group I, whereas SeMNPV, SfMPNV and HzSNPV (Cowan *et al.*, 1994) are positioned in Group II. BusuNPV is placed in Group II based on the studies of its polyhedrin and EGT sequences (Hu *et al.*, 1993, 1997).

While the genomes of AcMNPV, BmNPV and OpMNPV have a collinear gene arrangement (Fig. 5), it is not known if this holds true for the other baculoviruses in Group I. It appears, however, that the genomic organization of another Group I virus, *Christoneura fumiferana* MNPV (CfMNPV), is basically similar to that of OpMNPV (B. M. Arif, personal communication). Therefore, it is likely that all the baculoviruses belonging to Group I have a similar gene arrangement.

BusuNPV has a quite different gene arrangement to that seen in Group I viruses. A few of the different arrangements are shared by some of the Group II viruses. Thus, gene arrangement may be used as a marker for the phylogenetic study of baculoviruses as proposed by van Strien *et al.* (1997). When the phylogenetic trees based on single genes were evaluated by bootstrap analysis, it was shown that the clustering of Group II is far less stable in comparison to Group I (Cowan *et al.*, 1994; Hu *et al.*, 1997; Chen *et al.*, 1997). In other words, Group II contains viruses that are more diverse than those in Group I. Therefore, a collinear arrangement in the genomes of Group II viruses may be less likely. This is supported by the fact that the gene arrangement of SeMNPV, at least over a 20 kbp region, is quite different from that of BusuNPV as well as from Group I viruses (van Strien *et al.*, 1997). By applying the GeneParityPlot method presented in this paper, it can be expected that other types and degrees of collinearity could be found when data of more genomes become available.

The mechanisms governing re-arrangement of baculovirus genes remain unclear. It is conceivable that a certain gene cluster might be in some way important for virus replication. For example, the region adjacent to the helicase gene appears to have a highly conserved gene order in baculoviruses as diverse as NPVs and GVs (Heldens *et al.*, 1998). The extensive gene re-arrangement in other regions of the BusuNPV genome apparently did not have adverse effects on virus replication and survival in the natural host. Since the baculovirus family comprises more than 600 members occurring in a variety of arthropod orders and families (Murphy *et al.*, 1995), it is likely that extensive gene re-arrangement has taken place in other species. The gene arrangement in baculoviruses may be, in addition to gene homology, a reflection of their evolutionary history. As more data on the genomic organization of baculoviruses become available, it should be possible to study the phylogeny based on gene arrangement (van Strien *et al.*, 1997) as has been shown with herpesviruses (Hannenhalli *et al.*, 1995). Phylogeny based on gene order is independent of that based on single genes because the selection pressure for the sequence conservation of a certain gene is more linked to its structure–function relationship. Therefore, studies of genome arrangement can explore the ancestral history of baculoviruses from a different point of view.

We are grateful to Dr J. A. Jehle and Dr E. A. van Strien for critical review of the manuscript. We thank Dr S. Maeda for providing the order of BmNPV ORFs prior to its publication. This research was supported in part by the National Natural Science Foundation of China (NSFC, grant 39370031) and a joint grant from the Chinese Academy of Sciences (CAS) and the Royal Netherlands Academy of Sciences (KNAW). Z. H. Hu was a recipient of an EC Marie Curie fellowship, a special grant from the chairman of CAS and a sandwich PhD grant from the Wageningen Agricultural University.

References

Adams, J. R. & McClintock, J. T. (1991). Nuclear polyhedrosis viruses of insects. In *Atlas of Invertebrate Viruses*, pp. 87–204. Edited by J. R. Adams & J. R. Bonami. Boca Raton, FL: CRC Press.

Ahrens, C. H., Russell, R. L. Q., Funk, C. J., Evans, J. T., Harwood, S. H. & Rohrmann, G. F. (1997). The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **229**, 381–399.

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M. & Possee, R. D. (1994). The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202**, 586–605.

Bjornson, R. M. & Rohrmann, G. F. (1992). Nucleotide sequence of the polyhedron envelope protein gene region of the *Lymantria dispar* nuclear polyhedrosis virus. *Journal of General Virology* **73**, 1499–1504.

Caballero, P., Zuidema, D., Santiago-Alvarez, C. & Vlak, J. M. (1992). Biochemical and biological characterization of four isolates of *Spodoptera exigua* nuclear polyhedrosis virus. *Biocontrol Science and Technology* **2**, 145–157.

Chen, X., Hu, Z., Jehle, J. A., Zhang, Y. & Vlak, J. M. (1997). Analysis of the ecdysteroid UDP-glucosyltransferase gene of *Heliothis armigera* single nucleocapsid baculovirus. *Virus Genes* **15**, 219–225.

Cowan, P., Bulach, D., Goodge, K., Robertson, A. & Tribe, D. E. (1994). Nucleotide sequence of the polyhedrin gene region of *Helicoverpa zea* nucleocapsid nuclear polyhedrosis virus: placement of the virus in lepidopteran nuclear polyhedrosis group II. *Journal of General Virology* **75**, 3211–3218.

Hannenhalli, S., Chappey, C., Koonin, E. V. & Pevzner, P. A. (1995). Genome sequence comparison and scenarios for gene arrangements: a test case. *Genomics* **30**, 299–311.

Hawtin, R. E., Zarkowska, T., Arnold, K., Thomas, C. J., Gooday, G. W., King, L. A., Kuzio, J. & Possee, R. D. (1997). Liquefaction of *Autographa californica* nucleopolyhedrovirus-infected insects is dependent on the integrity of virus encoded chitinase and cathepsin genes. *Virology* **238**, 243–253.

Heldens, J. G. M., Liu, Y., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1998). A highly conserved genomic region in baculoviruses: sequence analysis of an 11.3 kbp DNA fragment (m.u. 46.5–55.1) of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus. *Virus Research* **54**, 187–198.

Hu, Z. H., Liu, M. F., Jin, F., Wang, Z. X., Li, M. J., Liang, B. F. & Xie, T. E. (1993). Nucleotide sequence of the *Buzura suppressaria* single

nucleocapsid nuclear polyhedrosis virus polyhedrin gene. *Journal of General Virology* **74**, 1617–1620.

Hu, Z. H., Broer, R., Westerlaken, J., Martens, J. W. M., Jin, F., Jehle, J. A., Wang, L. M. & Vlak, J. M. (1997). Characterization of the ecdysteroid UDP-glucosyltransferase gene of a single nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria*. *Virus Research* **47**, 91–97.

Hu, Z. H., Arif, B. M., Sun, J. S., Chen, X. W., Zuidema, D., Goldbach, R. W. & Vlak, J. M. (1998). Genetic organization of the *Hin*dIII-I region of the single-nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria*. *Virus Research* **54**, 71–82.

King, L. A. & Possee, R. D. (1992). *The Baculovirus Expression System: A Laboratory Guide*. London: Chapman & Hall.

Liu, M., Hu, Z., Liang, B., Jin, F., Li, M. & Xie, T. (1993). Physical mapping of *Buzura suppressaria* nuclear polyhedrosis virus genome. *Archives of Virology* **128**, 357–362.

Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A. & Summers, M. D. (editors) (1995). *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses*. Vienna & New York: Springer-Verlag.

O'Reilly, D. R. (1997). Auxiliary genes of baculoviruses. In *The Baculoviruses*, pp. 267–300. Edited by L. K. Miller. New York: Plenum Press.

Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Smith, I. R. L. & Crook, N. E. (1988). *In vivo* cloning of baculovirus genotypes. *Virology* **166**, 240–244.

van Oers, M. M., Flipsen, J. T. M., Reusken, C. B. E. M. & Vlak, J. M. (1994). Specificity of baculovirus p10 functions. *Virology* **200**, 513–523.

van Oers, M. M., Hu, Z. H., Arif, B. M., van Strien, E., van Lent, J. W. M. & Vlak, J. M. (1998). The single-nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria* encodes a P10 protein. *Journal of General Virology* **79**, 1553–1562.

van Strien, E. A., Zuidema, D. & Vlak, J. M. (1997). The genetic organization of a 20 kb DNA segment of *Spodoptera exigua* nucleopolyhedrovirus: gene order as a marker for baculovirus phylogeny. In *Characterization of the Spodoptera exigua baculovirus genome: structural and functional analysis of a 20 kb fragment*, pp. 87–115. E. A. van Strien, PhD thesis, Wageningen Agricultural University, The Netherlands.

Vlak, J. M. & Smith, G. E. (1982). Orientation of the genome of *Autographa californica* nuclear polyhedrosis virus: a proposal. *Journal of Virology* **41**, 1118–1121.

Xie, T. E., Peng, H. Y., Gong, H. Z. & Liu, Y. L. (1979). Identification and isolation of *Buzura suppressaria* nuclear polyhedrosis virus. *Acta Virologica Sinica* (special issue), 11–20.

Zanotto, P. M. de A., Kessing, B. D. & Maruniak, J. E. (1993). Phylogenetic relationships among baculoviruses, evolutionary rates and host associations. *Journal of Invertebrate Pathology* **62**, 147–164.

Received 28 April 1998; Accepted 7 July 1998