ANNOTATED SEQUENCE RECORD

Genomic sequence analysis of *Helicoverpa armigera* nucleopolyhedrovirus isolated from Australia

Huan Zhang · Qing Yang · Qi-Lian Qin · Wei Zhu · Zhi-Fang Zhang · Yi-Nü Li · Ning Zhang · Ji-Hong Zhang

Received: 15 May 2013/Accepted: 4 July 2013/Published online: 29 September 2013 © Springer-Verlag Wien 2013

Abstract The complete genomic sequence of *Helicoverpa armigera* nucleopolyhedrovirus from Australia, HearNPV-Au, was determined and analyzed. The HearNPV-Au genome was 130,992 bp in size with a G + C content of 39 mol% and contained 134 predicted open reading frames (ORFs) consisting of more than 150 nucleotides. HearNPV-Au shared 94 ORFs with AcMNPV, HearSNPV-G4 and SeMNPV, and was most closely related to HearSNPV-G4. The nucleotide sequence identity between HearNPV-Au and HearSNPV-G4 genome was 99 %. The major differences were found in homologous

H. Zhang and Q. Yang contributed equally to the work.

Electronic supplementary material The online version of this article (doi:10.1007/s00705-013-1823-3) contains supplementary material, which is available to authorized users.

H. Zhang · Q. Yang · Q.-L. Qin · W. Zhu · N. Zhang · J.-H. Zhang (\boxtimes)

State Key Laboratory of Integrated Management of Pest Insects and Rodents, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China e-mail: zhangjh@ioz.ac.cn

H. Zhang e-mail: zhanghuan@ioz.ac.cn

Q. Yang e-mail: qingyang316@126.com

Q. Yang Henan Jiyuan Baiyun Industry Co., Ltd, Henan 459002, China

Z.-F. Zhang (\boxtimes) · Y.-N. Li Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing 100081, China e-mail: zhifangzhang@yahoo.com regions (*hrs*) and baculovirus repeat ORFs (*bro*) genes. Five *hrs* and two *bro* genes were identified in the Hear-NPV-Au genome. All of the 134 ORFs identified in HearNPV-Au were also found in HearSNPV-G4, except the homologue of ORF59 (*bro*) in HearSNPV-G4. The sequence data strongly suggested that HearNPV-Au and HearSNPV-G4 belong to the same virus species.

Introduction

Being pathogenic to many insects, baculoviruses are frequently used as bio-insecticides to control the size of pest populations in nature. It is well established that baculovirus populations exhibit large amounts of genotypic variation, which may have multiple origins, such as geographical or temporal differences, a different host, or even an individual host [1-8]. The cotton bollworm *Helicoverpa armigera*, is a serious global pest that is responsible for economic losses to over 60 cultivated crops and is resistant to chemical insecticides. In China, HearNPV (family Baculoviridae, genus Alphabaculovirus) has been commercialized and extensively used on cotton fields since 1994 [9, 10]. High levels of genetic variation have also been found within HearNPV populations [6-8]. Now, genomes of five Heli*coverpa* spp. NPVs, including HearSNPV-C1 (China) [11], HearSNPV-G4 (China) [12], HearNPV-NNg1 (Kenya) [13], HzSNPV (USA) [14], and HearMNPV (China) [15], have been sequenced. The gene content and arrangement of HearMNPV were distinct from the other four NPVs, and those four NPV genomes shared very high nucleotide sequence identity except for the homologous regions (hr) and the baculovirus repeat ORFs (bro). In this study, we sequenced and analyzed the complete genome of another HearNPV, HearNPV-Au, which isolated from Australia.

Materials and methods

The HearNPV-Au used in this study was supplied by Tri-Delta Chemicals Pty Ltd. (Australia) and Henan Jiyuan Baiyun Industry Co., Ltd. (China). Polyhedra of HearNPV-Au were propagated in *H. armigera* larvae and purified by washing with 1 % SDS and distilled water multiple times with centrifugation. The purified polyhedra were solubilized in 0.7 ml alkaline solution (0.1 M Na₂CO₃, 0.1 M NaCl, 0.005 M EDTA, pH 8.0) at 37 °C for 1 h. The pH was adjusted to 7.0 with 0.1 M HCl, 5 μ l of 20 mg/ml proteinase K was added and the sample was incubated at 55 °C for 3 h. The genomic DNA was extracted with phenol and chloroform, precipitated with 100 % ethanol, and washed with 70 % ethanol.

A random genomic library of HearNPV-Au was constructed according to the "partial filling-in" method as described previously with minor modifications [16]. Viral DNA fragments ranging from 1.5 to 5.0 kbps were cloned into the *Sal*I site of the pUC19 vector. A total of 464 recombinant plasmids were prepared for sequencing using a BigDye Terminator v3.1 Cycle Sequencing Kit (ABI) on a Genetic Analyzer 3130XL (ABI). The combined sequence generated from these clones represented sixfold genomic coverage. Additional sequences for conformation of ambiguous regions and for filling in of gaps in the assembled sequence were obtained from sequencing of PCR products. All of the sequences were assembled into contigs using SeqMan from the DNASTAR 7.0 software package.

ORFs were defined using ORF Finder (http://www.ncbi. nlm.nih.gov/gorf/gorf.html). DNA and protein comparisons were performed using BLAST (http://blast.ncbi.nlm.nih. gov/Blast.cgi) or Vector NTI Advance Suite v8.0. Multiple sequences were aligned in Clustal X and displayed in GeneDot. Promoter motifs present upstream of the putative ORFs were screened as described previously [17].

Results and discussion

During the assembly of the genome sequence, we found some nucleotide variability. Based on a longest assembled sequence (13.5 kb), the rate of nucleotide variability was 0.19 %. We just picked the predominant nucleotide when nucleotide variability occurred. This confirmed that Hear-NPV populations exhibit genotypic variation.

The HearNPV-Au genome was 130,992 bp in size (GenBank accession no. JN584482), similar to those of HzSNPV (130,869 bp, GenBank accession no. AF334030) and HearSNPV-G4 (131,405 bp, GenBank accession no. NC002654), with a G + C content of 39 mol%. There were 134 predicted ORFs consist of more than 150

nucleotides. The HearNPV-Au genome shared 94 ORFs with the AcMNPV, HaSNPV-G4 and SeMNPV genomes. Homologues of these 94 ORFs were chosen for the GeneParityPlot analysis. The comparison showed that HearNPV-Au and HearSNPV-G4 were completely co-linear and identical in their gene arrangement (Fig. 1). The comparisons between HearNPV-Au and AcMNPV, SeM-NPV were in agreement with results reported previously [12]. A comparison of the locations and predicted amino acid sequences of the 94 ORFs between HearNPV-Au and the other three baculovirus genomes (Table 1) also indicated that HearNPV-Au was most closely related to HearSNPV-G4.

The nucleotide sequence identity between HearNPV-Au and HearSNPV-G4 genomes was 99 %, and the major differences were found in the *hrs* and *bro* genes.

Characterized by the presence of multiple imperfect palindrome sequences, hrs may function in gene replication, transcription, recombination and rearrangement events [11, 18–20]. Both the number and location of hrs from the



Fig. 1 GeneParityPlot comparison of HearNPV-Au with AcMNPV, HearSNPV-G4 and SeMNPV. Homologous ORFs are plotted based on their relative location in the genome. The horizontal and vertical axes indicate the relative position of each ORF

Table 1 Putative ORFs identified in HearNPV-Au
--

ORF	Name	Position	Length (aa)	Promoter ^a	Homologous ORFs			Amino acid sequence identity to homologues (%)		
					AcMNPV	HearSNPV- G4	SeMNPV	AcMNPV	HearSNPV- G4	SeMNPV
1	Polyhedrin	$1 \rightarrow 741$	246	E, L, e	8	1	1	86	100	87
2	orf1629	738 ← 1979	413	Ν	9	2	2	27	99	33
3	pk-1	$1994 \rightarrow 2797$	267	Ν	10	3	3	40	99	55
4	Hoar	2920 ← 5181	753	E*		4	4		93	28
5		$5377 \rightarrow 5556$	59	Ν		5			97	
6	Hzorf480	$5724 \rightarrow 6575$	283	E*		6			99	
7		$6787 \leftarrow 6942$	51	Ν		7			98	
8	i.e0	$6930 \rightarrow 7787$	285	Ν	141	8	138	32	99	35
9	p49	$7804 \rightarrow 9210$	468	L	142	9	137	50	99	56
10	odv-e18	$9221 \rightarrow 9466$	81	L	143	10	136	75	99	60
11	odv-e27	$9481 \rightarrow 10335$	284	L	144	11	135	50	99	57
12		$10381 \rightarrow 10659$	92	L	145	12	134	48	100	58
13		$10686 \leftarrow 11297$	203	Ν	146	13	133	30	100	32
14	i.e1	$11339 \rightarrow 13306$	655	E*, e	147	14	132	34	99	30
15	odv-e56	13359 ← 14423	354	L	148	15	6	51	100	50
16	me53	$14584 \rightarrow 15663$	359	E*, L	139	16-17	7	24	99	33
17		15666 → 15833	55	L		18			100	
18		15886 ← 16167	93	E*		19			96	
19	p74	16188 → 18254	688	Ν	138	20	131	53	99	55
20	p10	18308 ← 18571	87	L	137	21	130	18	100	51
21	p26	18654 ← 19457	267	E, L	136	22	129	35	99	43
22		$19570 \rightarrow 19773$	67	E*	29	23	128	32	100	48
23	lef-6	19849 ← 20412	187	Ν	28	24	127	32	99	50
24	Dbp	20426 ← 21397	323	Е	25	25	126	32	100	50
25		$21617 \rightarrow 22018$	133	Ν	26	26	125	42	100	36
	hr1	22019 — 24339								
26		24340 ← 25107	255	E*	34	27	124	37	99	51
27	ubiquitin	$24947 \rightarrow 25198$	83	L	35	28	123	75	100	78
28	-	25262 → 25768	168	E*		29			100	
29	Lese25-like protein	25788 → 26360	190	L		30			98	
30	39 k/pp31	26419 ← 27354	311	Ν	36	31	120	40	100	33
31	lef-11	27320 ← 27703	127	Ν	37	32	119	39	100	51
32		27672 ← 28388	238	Ν	38	33	118	52	100	63
33		$28620 \rightarrow 29699$	359	E*		34			99	
34	p47	29767 ← 31005	412	e	40	35	115	54	99	61
35		$31078 \rightarrow 31749$	223	E*	41	36		32	100	
36		$31835 \rightarrow 32077$	80	L	43	37	113	30	100	31
37	lef-8	32074 ← 34779	901	Ν	50	38	112	62	99	67
38		$34832 \rightarrow 35410$	192	L	51	39	111	31	99	36
39		35551 → 35703	50	L		40			96	
40	Chitinase	35711 ← 37480	589	Ν	126	41	19	66	98	62
41		37524 ← 38066	180	E*	52	42	109	26	100	27
42		38184 → 38594	136	E, L	53	43	108	43	100	56
43		38601 ← 39737	378	e, L		44	107		99	35
44		39745 ← 39972	75	E*, L		45			100	
45	lef-10	$39932 \rightarrow 40147$	71	Ν	53a	46	106	38	100	56
46	vp1054	$40020 \rightarrow 41075$	351	E, e	54	47	105	41	99	49
47		$41195 \rightarrow 41401$	68	Ν	55	48	104	35	100	50
48		$41402 \rightarrow 41596$	64	L	56	49	103	37	98	53

	ORF	Name	Position	Length (aa)	Promoter ^a	Homologous ORFs			Amino acid sequence identity to homologues (%)		
9 41875 - 42390 171 E.L 57 50 102 42 99 43 50 4235 - 42301 88 L 60 51 101 46 100 65 52 Fp 4423 - 4406 217 L 61 53 98 63 98 70 53 167 4423 - 46101 519 F 62 55 97 64 98 71 55 1647 4438 - 46101 519 F 62 55 97 64 98 72 57 g77 47990 - 4829 279 L 64 58 25 60 98 72 78 Bro 49990 - 51(22 519 L 61 51 51 63 61 74 99 43 61 5372 5372 5479 24 N 69 61 89 41 99 61 63						AcMNPV	HearSNPV- G4	SeMNPV	AcMNPV	HearSNPV- G4	SeMNPV
50 4441 + 4293 160 N 59 51 101 46 100 63 51 4295 + 43201 38 L 60 53 98 63 98 70 53 4428 + 44423 61 E 54 97 64 99 72 54 16/9 4454 + 44014 51 E 54 16 46 98 77 55 Cambrain 44188 + 44104 51 E 54 16 46 98 77 57 gr37 47990 + 48839 279 L 64 58 25 60 99 63 663 5128 5 5005 236 E 105 61 34 100 46 610 5372 5 750 236 E 10 66 61 91 25 99 63 610 6373 5 5700 75 L 66 67 93 46 90	49		$41875 \rightarrow 42390$	171	E, L	57	50	102	42	99	43
51 42035 - 42001 88 L 60 52 100 42 100 56 52 <i>Fp</i> 44313 - 44066 217 L 61 53	50		42441 ← 42923	160	Ν	59	51	101	46	100	63
52 <i>P</i> p 4413 - 44066 217 L 61 53 98 63 98 70 53 4423 - 44123 61 E 54 67 64 99 72 55 Cableysin 46188 - 47919 367 N 127 56 16 46 98 77 57 gp37 47990 + 48829 279 L 64 58 25 60 99 63 bn2 48800 - 49979 - - 61 34 100 - 74 60 ip2 51629 549 N 60 34 199 61 61 - 5476 - 53065 236 F* 105 61 34 199 62 62 - 5716 - 5316 133 N 68 64 90 47 99 62 63 lef3 5716 133 N 66 66 91 25 99 52 64 - 5618 - 5632 379 N <t< td=""><td>51</td><td></td><td>42935 ← 43201</td><td>88</td><td>L</td><td>60</td><td>52</td><td>100</td><td>42</td><td>100</td><td>56</td></t<>	51		42935 ← 43201	88	L	60	52	100	42	100	56
53 4428 44428 61 E 54 67 64 98 72 54 60 4518 4710 519 E 62 55 97 64 98 72 57 9727 47900 48529 270 L 64 58 97 64 98 72 78 870 49800 48830 4997 - 57 60 97 60 - 98 61 6130 51630 5162 549 N 61 34 100 - 61 - 64 99 61 61 53 53 63 89 41 99 62 63 64 91 25 99 61 63 63 53 63 53 75 L 66 62 92 29 61 63 63 63 63 63 63 63 63 63 <td< td=""><td>52</td><td>Fp</td><td>43413 ← 44066</td><td>217</td><td>L</td><td>61</td><td>53</td><td>98</td><td>63</td><td>98</td><td>70</td></td<>	52	Fp	43413 ← 44066	217	L	61	53	98	63	98	70
54 6j-9 44545 - 46104 519 E 62 55 07 64 99 72 55 Cahepain 46188 + 47291 367 N 127 56 16 46 98 47 57 g37 47990 - 4829 29 L 64 58 25 60 99 63 hr2 4830 - 04079	53		$44238 \rightarrow 44423$	61	Е		54			98	
55 Cambragin 44588 - 47391 367 N 127 56 16 46 98 47 56 gp47 47990 - 48829 279 L 64 58 25 60 99 61 58 Mr 49830 - 4997 - 60 - 96 - 51 - 96 - 51 - 96 - 51 - 5130 - 52384 - 70 60 - 96 - 61 - 34 100 - 61 - 61 - 34 100 - 61 61 31 99 48 61 61 51 51 61 61 51 70 90 61	54	lef-9	$44545 \rightarrow 46104$	519	Е	62	55	97	64	99	72
56 47332 47390 48829 279 L 57 57 99 63 h2 48830 -9979 - - 64 58 25 60 99 63 h2 48830 -9979 - 60 - 96 - 76 58 Bro 9980 -5163 256 E* 105 61 34 100 60 iap2 53172 -5394 250 E, L 71 62 88 31 99 36 61 53172 -5396 R 66 62 22 9 9 61 64 -6343 1020 e 65 67 93 44 90 63 64 -6438 1020 e 65 67 94 43 100 91 67 16438 1020 e 65 67 93 100 63	55	Cathepsin	46188 ← 47291	367	Ν	127	56	16	46	98	47
57 gp37 47990 - 48830 279 L 64 58 25 60 99 63 hr2 48830 - 49979 - 60 -<	56		47332 ← 47919	195	L		57			99	
hr2 4830 - 4979 8 Bro 49980 - 51629 549 N 60 96 hr3 51629 5285 - 105 61 34 100 59 hc53 5285 - 5392 250 E, L 71 62 88 31 99 62 60 inp-2 53172 - 53790 274 N 60 63 89 41 99 62 61 50372 - 54796 274 N 60 63 90 47 90 62 63 lef3 5433 - 5879 785 L 66 66 92 29 99 61 64 640 648 90 43 53 10 64 90 42 100 64 64 62873 + 6313 85 L 76 70 95 40 100 64 70 62873 + 6313 85 L 76 71	57	gp37	47990 ← 48829	279	L	64	58	25	60	99	63
58 Bro 4980 - 5163 549 N 60 96 hr3 51630 - 52384 - 105 61 34 100 60 iap-2 53172 - 53294 250 E, L 71 62 88 31 99 36 61 53972 - 54796 274 N 63 89 41 99 62 63 64 54765 - 55166 33 N 68 64 90 47 99 62 64 5635 - 58790 785 L 66 62 29 99 61 67 H20K - 56183 1020 c 65 67 93 46 99 59 66 9273 - 61330 85 L 76 70 95 40 100 64 69 9174 6425 - 64700 10 L 78 72 81 42 90 65 71 gp1 6425 -		hr2	48830 — 49979								
hr3 51630 - 52384 59 hc65 52385 - 53095 236 E* 105 61 34 100 61 53972 - 54796 274 N 69 63 89 41 99 48 61 53972 - 54796 274 N 69 63 89 41 99 48 62 54765 - 55166 133 N 68 66 92 29 99 61 64 56433 - 58790 785 L 66 66 92 29 99 61 65 DAA pol 58821 - 6183 1020 e 65 67 93 46 99 67 66 62484 62874 + 6287 127 E, L 75 69 94 23 100 63 70 6287 + 66130 85 L 76 70 95 40 100 57 70 64428 + 64760 110 L <td>58</td> <td>Bro</td> <td>$49980 \rightarrow 51629$</td> <td>549</td> <td>Ν</td> <td></td> <td>60</td> <td></td> <td></td> <td>96</td> <td></td>	58	Bro	$49980 \rightarrow 51629$	549	Ν		60			96	
99 he65 52385 - 53095 236 E* 105 61 34 100 60 iap-2 53172 - 53924 250 E, L 71 62 88 31 99 363 61 53972 - 54796 133 N 68 64 90 47 99 62 63 lef-3 55186 - 56325 379 N 67 65 91 25 99 32 64 58821 - 6183 1020 e 65 67 93 46 99 59 66 6190 - 62418 152 E, L 74 68 26 100 39 67 H20R534 6344 + 6287 101 L 78 69 94 23 100 64 62873 + 63130 85 L 76 70 95 40 100 43 71 gp41 6429 + 65797 322 E*, L 74 78 34 99 45 73 632 + 67002 241 N 81 <td< td=""><td></td><td>hr3</td><td>51630 — 52384</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></td<>		hr3	51630 — 52384								
60 $i \mu p^2$ $53172 + 53924$ 250 E, L 71 62 88 31 99 36 61 $33972 + 5396$ 274 N 69 63 89 41 99 42 63 $l q^{c3}$ $55186 + 56325$ 379 N 67 65 91 25 99 61 65 $DN \rho ol$ $58821 + 6833$ 1020 e 65 67 93 46 99 59 66 -6090 62418 152 E, L 74 68 26 100 300 67 H_{CRF384} $62484 + c267$ 127 E, L 75 69 94 23 100 37 70 $64428 + c4700$ 110 L 77 71 82 74 99 67 71 $gp41$ $6432 + c6700$ 225 e^{c} 81 74 79 52 100 633 72 $gp41$ $6432 + c6700$	59	he65	52385 → 53095	236	E*	105	61		34	100	
61 53972 + 54796 274 N 69 63 89 41 99 48 62 54765 + 55166 133 N 68 64 90 47 99 62 64 56433 + 58790 785 L 66 66 92 29 99 61 65 DAA pol 58821 + 6183 1020 e 65 67 93 46 99 82 66 600 62484 + 62867 127 E, L 74 68 20 100 43 67 H20RF384 62484 + 62867 127 E, L 76 70 94 23 100 64 70 63371 + 64415 414 L 78 72 81 42 100 43 71 gp41 6482 + 6707 322 E*, L 80 73 80 59 100 59 72 6525 + 67002 225 e 82 75 78 34 99 45 75 cg30 6952 - 9	60	iap-2	53172 ← 53924	250	E, L	71	62	88	31	99	36
$ \begin{array}{ccccccccccccccccccccccccccccccccccc$	61	*	53972 ← 54796	274	N	69	63	89	41	99	48
63 lef^3 55186 + 56325 379 N 67 65 91 25 99 32 64 5643 + 58700 785 L 66 66 92 29 99 61 65 DNA pol 58821 + 61883 1020 e 65 67 93 46 99 59 67 HzORF384 62484 + 62867 127 E, L 75 69 94 23 100 64 69 vlf-1 63171 + 64415 414 L 77 71 82 74 99 67 70 6422 + 66707 322 E*, L 80 73 80 59 100 59 71 gp41 6482 + 66707 322 E*, L 80 73 80 59 100 59 74 vp91capsid 66932 + 6932 816 L 83 76 77 40 99 45 75 63 6093 + 7051 233 81 75 43 100 53	62		54765 ← 55166	133	Ν	68	64	90	47	99	62
64 56433 - 58790 785 L 66 66 92 29 99 61 65 DNA pol 58821 - 61883 1020 e 65 67 93 46 99 59 66 61960 - 62418 152 E, L 74 68 26 100 39 68 62873 - 63130 85 L 76 70 95 40 100 64 69 ylf-I 63171 - 64415 414 L 77 71 82 74 99 67 70 64428 - 64700 110 L 78 72 81 42 100 59 72 65727 - 66452 241 N 81 74 79 52 100 59 73 66325 - 67002 225 e 82 75 78 34 99 45 75 cg30 69510 - 70361 283 E*, L 88 77 76 27 100 33 76 vp32caprisid 70350 - 771331	63	lef-3	55186 → 56325	379	Ν	67	65	91	25	99	32
66 DNA pol 5821 -> 61883 1020 e 65 67 93 46 99 59 66	64	5	56433 ← 58790	785	L	66	66	92	29	99	61
66 61960 \leftarrow 62418 152 E, L 74 68 26 100 67 $H_cORF384$ 62484 \leftarrow 62867 127 E, L 75 69 94 23 100 39 68 62873 \leftarrow 63130 85 L 76 70 95 40 100 64 69 ψ [f-1 63171 \leftarrow 64415 414 L 77 71 82 74 99 67 70 64428 \leftarrow 64760 110 L 78 72 81 42 100 43 71 gp/41 64829 \leftarrow 65797 322 E*, L 80 73 80 59 100 66 73 66932 \leftarrow 66932 241 N 81 74 79 52 100 63 75 cg30 69510 \leftarrow 7131 293 N 89 78 75 43 100 53 76 ψ 93/2casid 70450 \leftarrow 7131 293 N 92 80 73 53 100 63 76	65	DNA pol	$58821 \rightarrow 61883$	1020	e	65	67	93	46	99	59
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	66	I	61960 ← 62418	152	E, L	74	68		26	100	
68 62873 + 63130 85 L 76 70 95 40 100 64 69 $vff-I$ 63171 + 64415 414 L 77 71 82 74 99 67 70 64428 + 64760 110 L 78 72 81 42 100 43 71 gp41 64829 + 65797 322 E* L 80 73 80 59 100 66 73 63325 + 67002 225 e 82 75 78 34 99 45 74 $vp91capsid$ 66932 + 69382 816 L 83 76 77 40 99 45 75 $cg30$ 09510 + 70361 283 8 77 76 27 83 100 53 76 $vp39capsid$ 09510 + 71331 293 N 89 78 75 43 100 53 78 74056 + 73532 254 N 93 81 72 55 100 63	67	HzORF384	62484 ← 62867	127	E. L	75	69	94	23	100	39
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	68		62873 ← 63130	85	Ĺ	76	70	95	40	100	64
7064428 \leftarrow 64760110L787281421004371 $gp41$ 64829 \leftarrow 65797322E*, L80738059100597265727 \leftarrow 66452241N81747952100667366325 \leftarrow 6702225e82757834994574 $vp91capsid$ 6932 \leftarrow 6032816L83767740994575 $cg30$ 69510 \leftarrow 70361283E*, L887776271003376 $vp39capsid$ 70450 \leftarrow 71331293N897875431005377 $lef-4$ 71330 \leftarrow 7215461N907974449950787268 \leftarrow 73532254N9280735310059797354 \rightarrow 74022162N938172551006380 $adv-e25$ 74068 \rightarrow 74760230 e, L 948271421006381 $-74792 \leftarrow$ 75289165L8368983182Helicase75308 \leftarrow 7969123 e, L 95847042994783 $-9026 \rightarrow$ 79547173N96856948996384906668571321N	69	vlf-1	63171 ← 64415	414	L	77	71	82	74	99	67
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	70		64428 ← 64760	110	L	78	72	81	42	100	43
a_{1} <	71	gn41	64829 ← 65797	322	– E*. L	80	73	80	59	100	59
7366325 \leftarrow 67002225e82757834994574 $vp91capsid$ 66932 \rightarrow 69382816L83767740994575 $cg30$ 69510 \leftarrow 70361283E*, L887776271003376 $vp33capsid$ 70450 \leftarrow 71331293N897875431005377lef-471330 \rightarrow 72715461N9079744499507872768 \leftarrow 73532254N9280735310059797354 \rightarrow 74022162N938172551006380 odv -e2574068 \rightarrow 74760230e, L94827142100638174792 \leftarrow 75289165L8368983182Helicase75308 \leftarrow 790691253e, L9584704299478379066 \leftarrow 80571321N9685694899638479066 \leftarrow 80571321N98866745995385lef-580467 \rightarrow 8144315N998766431005188 $ap_{2}, ap_{2}, ap_{$	72	5 r	65727 ← 66452	241	N	81	74	79	52	100	66
74 $vp91capsid$ 66932816L83767740994575 $cg30$ 69510 \leftarrow 70361283E*, L887776271003376 $vp39capsid$ 70450 \leftarrow 71331293N897875431005377 $lef.4$ 71330 \rightarrow 72715461N9079744499507872568 \leftarrow 73532254N928073531006380 $odv-e25$ 74068 \leftarrow 74760230e, L94827142100638174792 \leftarrow 75289165L8368983182Helicase75308 \leftarrow 790691253e, L9584704299478379026 \rightarrow 79547173N9685694899638479066 \leftarrow 80571321N98866745995385 $lef.5$ 80467 \rightarrow 81414315N998766431005188 $p6.9$ 81802 \leftarrow 82911369L1018964401005188 $p2957 \leftarrow$ 83325122E, L1029063291003989 $p80capsid$ 84553 \rightarrow 86370605L104926127992990 $vp80capsid$ <	73		$66325 \leftarrow 67002$	225	e	82	75	78	34	99	45
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	74	vp91capsid	$66932 \rightarrow 69382$	816	L	83	76	77	40	99	45
0070450 \leftarrow 71331293N897875431005377lef-471330 \rightarrow 72715461N9079744499507872768 \leftarrow 73532254N92807353100597973534 \rightarrow 74022162N938172551006380odv-e2574068 \rightarrow 74760230e, L94827142100638174792 \leftarrow 75289165L8368983182Helicase75308 \leftarrow 790691253e, L9584704299478379026 \rightarrow 79547173N9685694899638479606 \leftarrow 80571321N98866745995385lef-580467 \rightarrow 81414315N998766431005186p6.981408 \leftarrow 81737109E, L100886543100518882957 \leftarrow 83325122E, L102906329100398983325 \leftarrow 84458377L103916250996090 <i>vp80capsid</i> 84553 \rightarrow 86370605L10492612799299186367 \rightarrow 8654358N11093	75	cg30	69510 ← 70361	283	E*. L	88	77	76	27	100	33
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	76	vn39cansid	70450 ← 71331	293	N	89	78	75	43	100	53
7872768 \leftarrow 73532254N92807353100597973534 \rightarrow 74022162N938172551006380odv-e2574068 \rightarrow 74760230e, L94827142100638174792 \leftarrow 75289165L8368983182Helicase75308 \leftarrow 790691253e, L9584704299478379026 \rightarrow 79547173N9685694899638479606 \leftarrow 80571321N98866745995385lef-580467 \rightarrow 81414315N998766431005186p6.981408 \leftarrow 81737109E, L100886543100518882957 \leftarrow 83325122E, L102906329100398983255 \leftarrow 84458377L103916250996090vp80capsid84553 \rightarrow 86370605L10492612799299186367 \rightarrow 8654358N11093603210064928658 \rightarrow 87643361N109945952100579387688 \rightarrow 8797294N108955847 <t< td=""><td>77</td><td>lef-4</td><td>$71330 \rightarrow 72715$</td><td>461</td><td>Ν</td><td>90</td><td>79</td><td>74</td><td>44</td><td>99</td><td>50</td></t<>	77	lef-4	$71330 \rightarrow 72715$	461	Ν	90	79	74	44	99	50
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	78		72768 ← 73532	254	Ν	92	80	73	53	100	59
80 $odv-e25$ 74068 \rightarrow 74760230e, L94827142100638174792 \leftarrow 75289165L8368983182Helicase75308 \leftarrow 790691253e, L9584704299478379026 \rightarrow 79547173N9685694899638479606 \leftarrow 80571321N98866745995385lef-580467 \rightarrow 81414315N998766431005186p6.981408 \leftarrow 81737109E, L100886543100518781802 \leftarrow 82911369L101896440100518882957 \leftarrow 83325122E, L102906329100398983325 \leftarrow 84458377L103916250996090vp80capsid84553 \rightarrow 86370605L10492612799299186367 \rightarrow 8654358N1109360321006492928658 \rightarrow 87643361N10994595210057939766688039 \leftarrow 90057672L469657/114439942/3394odv-e6688039 \leftarrow 90057672L<	79		73534 → 74022	162	Ν	93	81	72	55	100	63
81 82 82 83 $74792 \leftarrow 75289$ $75308 \leftarrow 79069$ 1253 84 165 1253 1253 1253 1253 1253 1253 1253 1253 1253 	80	odv-e25	74068 → 74760	230	e, L	94	82	71	42	100	63
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	81		74792 ← 75289	165	L		83	68		98	31
83 $79026 \rightarrow 79547$ 173N96856948996384 $79606 \leftarrow 80571$ 321N98866745995385lef-5 $80467 \rightarrow 81414$ 315N998766431005186p6.9 $81408 \leftarrow 81737$ 109E, L1008865431006787 $81802 \leftarrow 82911$ 369L1018964401005188 $82957 \leftarrow 83325$ 122E, L1029063291003989 $83325 \leftarrow 84458$ 377L103916250996090 $vp80capsid$ $84553 \rightarrow 86370$ 605L104926127992991 $86367 \rightarrow 86543$ 58N1109360321006492 $86558 \rightarrow 87643$ 361N1099459521005793 $87688 \rightarrow 87972$ 94N1089558471005194 $odv-e66$ $88039 \leftarrow 90057$ 672L469657/114439942/3395 $90078 \leftarrow 90908$ 276L97569960	82	Helicase	75308 ← 79069	1253	e. L	95	84	70	42	99	47
8479606 \leftarrow 80571321N98866745995385lef-580467 \rightarrow 81414315N998766431005186p6.981408 \leftarrow 81737109E, L100886543100678781802 \leftarrow 82911369L101896440100518882957 \leftarrow 83325122E, L102906329100398983325 \leftarrow 84458377L103916250996090 $vp80capsid$ 84553 \rightarrow 86370605L10492612799299186367 \rightarrow 8654358N110936032100649286558 \rightarrow 87643361N109945952100579387688 \rightarrow 8797294N1089558471005194odv-e6688039 \leftarrow 90057672L469657/114439942/339590078 \leftarrow 90908276L97569960	83		79026 → 79547	173	N	96	85	69	48	99	63
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	84		79606 ← 80571	321	Ν	98	86	67	45	99	53
86 $p6.9$ $81408 \leftarrow 81737$ 109E, L1008865431006787 $81802 \leftarrow 82911$ 369L1018964401005188 $82957 \leftarrow 83325$ 122E, L1029063291003989 $83325 \leftarrow 84458$ 377L103916250996090 $vp80capsid$ $84553 \rightarrow 86370$ 605L104926127992991 $86367 \rightarrow 86543$ 58N1109360321006492 $86558 \rightarrow 87643$ 361N1099459521005793 $87688 \rightarrow 87972$ 94N1089558471005194 $odv-e66$ $88039 \leftarrow 90057$ 672L469657/114439942/339590078 \leftarrow 90908276L97569960	85	lef-5	$80467 \rightarrow 81414$	315	Ν	99	87	66	43	100	51
8781802 \leftarrow 82911369L101896440100518882957 \leftarrow 83325122E, L102906329100398983325 \leftarrow 84458377L103916250996090 $vp80capsid$ 84553 \rightarrow 86370605L10492612799299186367 \rightarrow 8654358N110936032100649286558 \rightarrow 87643361N109945952100579387688 \rightarrow 8797294N1089558471005194odv-e6688039 \leftarrow 90057672L469657/114439942/339590078 \leftarrow 90908276L97569960	86	p6.9	81408 ← 81737	109	E, L	100	88	65	43	100	67
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	87	I ····	81802 ← 82911	369	Ĺ	101	89	64	40	100	51
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	88		82957 ← 83325	122	E. L	102	90	63	29	100	39
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	89		83325 ← 84458	377	Ĺ	103	91	62	50	99	60
91 $86367 \rightarrow 86543$ 58N1109360321006492 $86558 \rightarrow 87643$ 361N1099459521005793 $87688 \rightarrow 87972$ 94N1089558471005194 $odv-e66$ $88039 \leftarrow 90057$ 672L469657/114439942/3395 $90078 \leftarrow 90908$ 276L97569960	90	vp80capsid	84553 → 86370	605	L	104	92	61	27	99	29
92 $86558 \rightarrow 87643$ 361 N 109 94 59 52 100 57 93 $87688 \rightarrow 87972$ 94 N 108 95 58 47 100 51 94 $odv-e66$ $88039 \leftarrow 90057$ 672 L 46 96 $57/114$ 43 99 $42/33$ 95 $90078 \leftarrow 90908$ 276 L 97 56 99 60	91	I I som	86367 → 86543	58	Ν	110	93	60	32	100	64
93 $87688 \rightarrow 87972$ 94N1089558471005194 $odv-e66$ $88039 \leftarrow 90057$ 672 L4696 $57/114$ 439942/3395 $90078 \leftarrow 90908$ 276L97569960	92		$86558 \rightarrow 87643$	361	N	109	94	59	52	100	57
94 odv-e66 88039 \leftarrow 90057 672 L 46 96 57/114 43 99 42/33 95 90078 \leftarrow 90908 276 L 97 56 99 60	93		87688 → 87972	94	N	108	95	58	47	100	51
$95 90078 \leftarrow 90908 276 L 97 56 99 60$	94	odv-e66	$88039 \leftarrow 90057$	672	L	46	96	57/114	43	99	42/33
	95		$90078 \leftarrow 90908$	276	L		97	56	.0	99	60
hr4 90909 — 93506		hr4	90909 — 93506								

Table 1 continued

ORF	Name	Position	Length (aa)	Promoter ^a	Homologous ORFs			Amino acid sequence identity to homologues (%)		
					AcMNPV	HearSNPV- G4	SeMNPV	AcMNPV	HearSNPV- G4	SeMNPV
96		93508 → 94107	199	E, L	115	98	50	43	100	46
97		$94111 \rightarrow 94467$	118	Ν		99			98	
98	Parg	$94563 \rightarrow 96095$	510	E, L		100	52		99	27
99		$96174 \rightarrow 96935$	253	L	106/107	101	53	47/34	99	57
100		$96950 \rightarrow 97282$	110	Ν		102			100	
101	iap-3	97340 ← 98146	268	E*, L		103	110		99	35
102		98143 ← 98298	51	Ν		104			100	
103	Bro	98409 ← 99914	501	L		105			99	
104	Sod	$100082 \rightarrow 100561$	159	L	31	106	48	75	98	69
105		$100568 \rightarrow 101941$	457	e, L		107			99	
106		101994 ← 102572	192	E, e		108			99	
107		$102742 \rightarrow 103098$	118	E*		109			100	
108		$103109 \rightarrow 103375$	88	L	117	110	47	33	100	37
109		$103443 \rightarrow 105029$	528	Е	119	111	36	47	99	44
110		$105026 \rightarrow 105262$	78	L		112			100	
111	Fgf	105285 ← 106190	301	E*	32	113	38	27	100	33
112	alk-exo	106318 ← 107604	428	Е	133	114	41	41	99	41
113		107624 ← 108013	129	L	19	115	42	30	100	31
	hr5	108013 — 110818								
114		109693 ← 110619	308	E*		115a			100	
115		$110820 \rightarrow 111035$	71	E*	111	116		36	100	
116	lef-2	111151 ← 111867	238	E*	6	117	12	40	98	45
117	p24capsid	$112229 \rightarrow 112975$	248	L	129	118	10	32	99	55
118	gp16	$113037 \rightarrow 113327$	96	L	130	119	9	26	100	31
119	Calyx/pep	$113379 \rightarrow 114401$	340	e, L	131	120	46	27	99	43
120		$114480 \rightarrow 114944$	154	E*	63	121		26	100	
121		$115075 \rightarrow 115665$	196	E*, L		122			98	
122	38.7kd	115709 ← 116878	389	Ν	13	123	13	31	99	33
123	lef-1	116880 ← 117617	245	Ν	14	124	14	38	99	47
124		117592 ← 118020	142	E, L		125			92	
125	Egt	$118165 \rightarrow 119712$	515	E, e, L	15	126	27	44	99	52
126		$119912 \rightarrow 120490$	192	Ν		127			100	
127		$120441 \rightarrow 121241$	266	E, L	17	128	29	33	99	30
128		121322 ← 124165	947	L		129	30		99	28
129	pkip-1	$124571 \rightarrow 125080$	169	Ν	24	130	32	24	97	39
130	arif-1	125147 ← 125944	265	Ν	21	131	34	31	99	30
131		$126205 \rightarrow 127356$	383	L	22	132	35	61	99	66
132		127397 ← 129430	677	E*, L	23	133	8	26	99	39
133		129572 ← 130117	181	E*		134			99	
124			105	F *		105			07	

The direction of the ORF in the HearNPV-Au genome is indicated by an arrow

^a The presence of the following conserved promoter motifs is indicated: E, early promoter motif TATA-box (TATAWAW) within 120 bp upstream region of the initiation codon; W = A/T; number of E, 21. E*, early promoter motif (TATA-box followed by CAKT motif 40 bp downstream) within 120 bp upstream of the initiation codon; K = G/T; number of E*, 26. e, enhancer-like element (CGTGC) within 210 bp upstream of the initiation codon; number of e, 13. L, late promoter motif DTAAG within 120 bp upstream of the initiation codon; D = A/T/G; number of L, 64. N, no TATA-box, enhancer-like element or late promoter motif present within 120 bp upstream region of the ORF; number of N, 41

genomes of HearNPV-Au and HearSNPV-G4 were identical. Two types of repeats, type A and type B, were found in each of the five *hrs* in HearNPV-Au (supplementary material). The sequence identities of the five *hrs* between HearNPV-Au and HearSNPV-G4 were 95.8 %, 99.8 %, 98.9 %, 87.7 %, 99.9 % respectively, which showed that hr4 was more variable. Besides, hr1 contained a 58-bp insertion and hr4 contained a 289-bp insertion compared with isolate G4, and neither insertion contained a type A or type B repeat. This suggested that hrs were less conserved than ORF regions in the *Helicoverpa* spp. NPVs.

Typically with multiple copies per genome, bro genes may function in nucleic acid binding, nucleosome association, and nucleocytoplasmic shuttling activity; may influence baculovirus genome diversity; and are involved in recombination between baculovirus genomes [21-25]. HearNPV-Au encodes two bro genes, named bro-a (ORF58) and bro-b (ORF103) based on their gene order. HearSNPV-G4 encodes three bro genes, named bro-a (ORF59), brob (ORF60), bro-c (ORF105). HearNPV-Au bro-a, consisting of 1650 bp, shared 94.7 % nucleotide sequence identity (96 % amino acid identity) with HearSNPV-G4 bro-b and contained a 66-bp insertion compared with HearSNPV-G4 bro-b. HearNPV-Au bro-b, consisting of 1506 bp, shared 99.7 % nucleotide sequence identity (99 % amino acid identity) with HearSNPV-G4 bro-c. All 134 ORFs identified in HearNPV-Au were also found in HearSNPV-G4, and HearNPV-Au lacked only the homologue of another bro gene, ORF59, in HearSNPV-G4, which was the major reason for the size difference between these two genomes. This might have been caused by genetic variation in recombination, suggesting that bro genes might play a role in gene exchange and evolution in different geographic locations.

Conclusions

HearNPV-Au shared 99 % sequence identity with HearNPV-G4, lacking only ORF59 (*bro*). The sequence data strongly suggest that HearNPV-Au and HearSNPV-G4 belong to the same virus species, *Helicoverpa armigera* nucleopolyhedrovirus. Whether these minor differences in the genome affect pathogenicity and host range needs to be determined.

Acknowledgments This work was supported by the grants from the "863" Project (2011AA10A204), the International Cooperation Project (2011-G4) and the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-G-16).

References

- Gettig RR, McCarthy WJ (1982) Genotypic variation among wild isolates of *Heliothis* spp. nuclear polyhedrosis viruses from different geographical regions. Virol 117:245–252
- Parnell M, Grzywacz D, Jones KA, Brown M (2002) The strain variation and virulence of granulovirus of diamondback moth (*Plutella xylostella* Linnaeus, Lep., Yponomeutidae) isolated in Kenya. J Invertebr Pathol 79:192–196

- Graham RI, Tyne WI, Possee RD, Sait SM, Hails RS (2004) Genetically variable nucleopolyhedroviruses isolated from spatially separate populations of the winter moth *Operophtera brumata* (Lepidoptera: Geometridae) in Orkney. J Invertebr Pathol 87:29–38
- Cory JS, Green BM, Paul RK, Hunter-Fujita F (2005) Genotypic and phenotypic diversity of a baculovirus population within an individual insect host. J Invertebr Pathol 89:101–111
- van Oers MM, Vlak JM (2007) Baculovirus genomics. Curr Drug Targets 8:1051–1068
- Rowley DL, Popham HJR, Harrison RL (2011) Genetic variation and virulence of nucleopolyhedroviruses isolated worldwide from the heliothine pests *Helicoverpa armigera*, *Helicoverpa zea*, and *Heliothis virescens*. J Invertebr Pathol 107:112–126
- 7. Baillie VL, Bouwer G (2012) High levels of genetic variation within *Helicoverpa armigera* nucleopolyhedrovirus populations in individual host insects. Arch Virol 157:2281–2289
- Baillie VL, Bouwer G (2012) High levels of genetic variation within core *Helicoverpa armigera* nucleopolyhedrovirus genes. Virus Genes 44:149–162
- Zhang GY, Sun XL, Zhang ZX, Zhang ZF, Wan FF (1995) Production and effectiveness of the new formulation of Helicoverpa virus pesticide-emulsifiable suspension. Virol Sinica 10:242–247
- Qin QL, Cheng QQ, Zheng JF, Chen XZ, Zhang ST, Li X, Miao L, Zhang H (2008) Production and application of *Helicoverpa armigera* nucleopolyhedrovirus bio-pesticides trademarked KEYUN on large scale. Biotechnol Bull pp 467–470
- Zhang CX, Ma XC, Guo ZJ (2005) Comparison of the complete genome sequence between C1 and G4 isolates of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus. Virol 333:190–199
- Chen XW, Ijkel WFJ, Tarchini R, Sun XL, Sandbrink H, Wang HL, Peters S, Zuidema D, Lankhorst RK, Vlak JM, Hu ZH (2001) The sequence of the *Helicoverpa armigera* single nucleocapsid nucleopolyhedrovirus genome. J Gen Virol 82:241–257
- Ogembo JG, Caoili BL, Shikata M, Chaeychomsri S, Kobayashi M, Ikeda M (2009) Comparative genomic sequence analysis of novel *Helicoverpa armigera* nucleopolyhedrovirus (NPV) isolated from Kenya and three other previously sequenced *Helicoverpa* spp. NPVs. Virus Genes 39:261–272
- 14. Chen XW, Zhang WJ, Wong J, Chun G, Lu A, McCutchen BF, Presnail JK, Herrmann R, Dolan M, Tingey S, Hu ZH, Vlak JM (2002) Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. J Gen Virol pp 673–684
- Tang P, Zhang H, Li YN, Han B, Wang GZ, Qin QL, Zhang ZF (2012) Genomic sequencing and analyses of HearMNPV-a new multinucleocapsid nucleopolyhedrovirus isolated from *Helicoverpa armigera*. Virol J 9:168
- Chen Y, Lin X, Yi YZ, Lu YY, Zhang ZF (2009) Construction and application of a baculovirus genomic library. Z Naturforsch C 64:574–580
- Xiao H, Qi Y (2007) Genome sequence of *Leucania seperata* nucleopolyhedrovirus. Virus Genes 35:845–856
- Guarino LA, Gonzalez MA, Summers MD (1986) Complete sequence and enhancer function of the homologous DNA regions of *Autographa-californica* nuclear polyhedrosis-virus. J Virol 60:224–229
- Pearson M, Bjornson R, Pearson G, Rohrmann G (1992) The Autographa-californica baculovirus genome-evidence for multiple replication origins. Science 257:1382–1384
- Theilmann DA, Stewart S (1992) Tandemly repeated sequence at the 3' end of the *ie-2* gene of the baculovirus Orgyia-pseudotsugata multicapsid nuclear polyhedrosis-virus is an enhancer element. Virol 187:97–106

- 21. Bideshi DK, Renault S, Stasiak K, Federici BA, Bigot Y (2003) Phylogenetic analysis and possible function of bro-like genes, a multigene family widespread among large double-stranded DNA viruses of invertebrates and bacteria. J Gen Viro 84:2531–2544
- Zemskov EA, Kang WY, Maeda S (2000) Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. J Virol 74:6784–6789
- 23. Kang WK, Imai N, Suzuki M, Iwanaga M, Matsumoto S, Zemskov EA (2003) Interaction of *Bombyx mori* nucleopolyhedrovirus BRO-A and host cell protein laminin. Arch Virol 148:99–113
- 24. Kang WK, Kurihara M, Matsumoto S (2006) The BRO proteins of *Bombyx mori* nucleopolyhedrovirus are nucleocytoplasmic shuttling proteins that utilize the CRM1-mediated nuclear export pathway. Virol 350:184–191
- Li LL, Li OJ, Willis LG, Erlandson M, Theilmann DA, Donly C (2005) Complete comparative genomic analysis of two field isolates of *Mamestra configurata* nucleopolyhedrovirus-A. J Gen Virol 86:91–105