UDC 577

Extreme evolutionary stability of conserved non-protein coding element of baculovirus genome

V. E. Makarenko, I. M. Kikhno, V. I. Kashuba

Institute of Molecular Biology and Genetics, NAS of Ukraine
150, Akademika Zabolotnoho Str., Kyiv, Ukraine, 03680
i.m.kikhno@imbg.org.ua

Aim. To estimate a level of evolutionary stability of the baculovirus conserved non protein-coding element (CNE) playing an essential role in baculovirus pathogenesis. **Methods.** NCBI-BLAST was applied to identify the orthologous sequences in genomes of 50 alphabaculoviruses. The orthologous sequences were aligned by using ClustalW software. Alistat tool was applied to obtain the pairwise percent identity (%ID) matrix data. **Results.** An average pairwise ID value (73 %) calculated for the 1225-member sample was shown to be comparable with those of coding parts of the most conserved baculovirus genes *polh* and *pif2* as well as with those of the *polh* and *p18* promoter regions which are the most conserved representatives of a small group of the evolutionary stable promoter regions of the alphabaculovirus late genes. **Conclusion.** CNE is one of the most conserved elements of the alphabaculovirus genome.

K e y w o r d s. Baculovirus, conserved non-coding element

Introduction

The recent advances in genome-wide studies have revealed the abundance of extended non-coding regions that are conserved at the sequence level across eukaryotic genomes and that are believed to be related to basic processes in a eukaryotic cell [1]. Some of these regions known by several different names (conserved non-coding element, CNE is one among them) have been shown to encode the genes of ncRNA or the 5` and 3` untranslated regions (UTRs) of mRNAs. Others appear to be true non-coding sequences, the functional significance of which is still poorly understandable. Herewith, at least some of them were shown to be involved in transcriptional regulation of the key developmental genes in metazoan genomes [2].

The amenability of RNA-coding CNEs to negative selection reflected in the CNE sequence conservation can be explained by the necessity of strictly defined structure of their RNA products for functional efficiency and/or molecule stability. In contrast, none of the available hypotheses explaining amenability of true non-coding CNEs to selective pressure can be found quite satisfactory [3]. The point is that the vast majority of known non-coding elements of the genome (transcription and translation regulators, origins of replication, sites of recombination, scaffold attachment sites, *etc.*) recognized as platforms for the protein trans-factors performing particular DNA trans-actions have been found to evolve too fast to be detected on the basis of their overall sequence conservation in the genomes of evolutionarily distant organisms, and only short nucleotide stretches (5-10bp), targets for the protein trans-factors, may be expected to constitute their conserved parts. Herewith, regardless of the function that can be attributed to CNEs, the length of these sequences (up to several hundred bps) is much bigger than necessary to specify the binding of a single protein

trans-factor or even several cooperatively acting proteins.

Pairwise comparison of eukaryotic genomes results in the identification of a variety of CNEs characterized by vastly different levels of the sequence conservation. While the conservation of some of these CNEs appears to be a result of negative selection, the conservation of others may simply reflect the short evolutionary distance or slow rate of neutral divergence among the species compared [4]. Therefore, only the sequences showing an extreme level of conservation in comparison with other genome elements can be reliably identified as CNEs which are likely under evolutionary constraint. Some of the most conserved CNEs may exhibit up to 100 % sequence identity (for instance, ultraconserved CNEs in mammals [5]), and some of those may be shared by so distantly related lineages as invertebrates and vertebrates [6].

Non protein-coding element essential for productive infection has been revealed recently in the genomes of insect pathogens, large dsDNA viruses of family *Baculoviridae* [7]. The newly discovered element was shown to express extended homology (152–156 bp) across the representatives of genus *Alphabaculovirus* and, accordingly it was called CNE. The alphabaculovirus CNE is composed of 7 blocks of absolutely conserved nucleotides interspaced by relatively variable nucleotide stretches of conserved length. CNE of *Autographa californica* nucleopolyhedrovirus (AcMNPV) was found to overlap with short ORF and three ncRNA genes [8], it includes three previously described protein-binding motifs [9] and sequence that activates minimal promoter of adjacent gene *ie-2* [10]. The transfection-infection assays using CNE knock out AcMNPV demonstrated that an additional not-yet-defined essential function is also attributed to CNE [7].

The present research was aimed at estimation of the degree of CNE sequence conservation measured as percentage of pairwise nucleotide identity (%ID) and comparison with that of other functional elements of alphabaculovirus genome. An average pairwise percent ID (av %ID) of CNE across 1225 al-

phabaculovirus pairs was found to be 73 %. The established fact that this value greatly exceeds the conservation levels of the vast majority of baculovirus genomic functional elements, both coding and non-coding, strongly suggests that CNE represents the essential genome element whose overall sequence appears to be under strong negative selection

Materials and methods

The nucleotide sequences of CNE, *polh* and *pif2* were extracted from 50 alphabaculovirus genomes available in GenBank database by using NCBI-BLAST.

[The] Members of each orthologous group were aligned by using ClustalW software [11]. [The] %ID matrix table was built by Alistat software [Eddy S. SQUID – C function library for sequence analysis 2005. [http://selab.wustl.edu/cgi-bin/selab.pl?mode=software#squid] 1225-member %ID samples resulted from pairwise comparison of 50 orthologs were used for the av %ID calculation and subsequent statistical analysis.

Results

1. CNE has evolved very slowly over the evolutionary time

[The] Phylogenetic studies suggest that the representatives of genus *Alphabaculovirus* are pretty diverse [12]. It is of interest if the level of virus diversification is high enough to consider their common element CNE as a sequence of extreme conservation.

Earlier for the 27 particular alphabaculovirus pairs the evolutionary relationships were inferred on the basis of values of average amino acid identity (AAI) across all orthologs shared by two genomes [13–22] (Table S1). CNEs of all members of these pairs were subjected to [the] pairwise %ID analysis. It was found that three most unrelated pairs (LyxyMNPV- AcMNPV, LyxyMNPV- OpMNPV and LyxyMNPV – MaviMNPV) are characterized by 33 % AAI and 67–68 % CNE ID. The 33 % AAI value points that the pairwise amino acid ID of the majority of orthologous proteins shared by the mem-

Table S1. **List of 27 virus pairs and their AAI and CNE ID percentage value**

N	Virus pair	Average aminoacid identity for shared ORFs (%)	CNE identity (%)	Reference
1	HearSNPV-G4-HzSNPV	99	96	[13]
2	HearMNPV- MacoNPV-B	98	98	[14]
3	AcMNPV-BmMNPV	93	90	[15]
4	AcMNPV-RoMNPV	91	94	[16]
5	HearMNPV-MacoNPV-A	90	92	[14]
6	CfDEFMNPV- AgMNPV	88	92	[17]
7	LyxyMNPV- LdMNPV	74	86	[18]
8	CfDEFMNPV- EppoNPV	69	84	[17]
9	CfDEFMNPV- OpMNPV	68	83	[17]
10	CfDEFMNPV- CfMNPV	68	81	[17]
11	HearMNPV- AgseNPV	58	86	[14]
12	AcMNPV- OpMNPV	56	80	[19]
13	AcMNPV - CfDEFMNPV	55	80	[17]
14	OrleNPV - EcobNPV	52	79	[20]
15	OrleNPV - ChchNPV	48	81	[20]
16	HearSNPV-G4- SeMNPV	46	76	[22]
17	AcMNPV- OrleNPV	43	78	[20]
18	AcMNPV - LdMNPV	42	69	[21]
19	HearSNPV-G4- LdMNPV	41	68	[22]
20	HearSNPV-G4- AcMNPV	41	75	[22]
21	HearSNPV-G4- BmNPV	41	78	[22]
22	HearMNPV-HearSNPV-G4	41	76	[14]
23	LyxyMNPV-SeMNPV	37	73	[18]
24	CfDEFMNPV- HearSNPV-G4	35	72	[22]
25	LyxyMNPV- AcMNPV	33	67	[18]
26	LyxyMNPV- OpMNPV	33	67	[18]
27	LyxyMNPv - MaviMNPV	33	68	[18]

Standard abbreviations of full virus names are used in the table

bers of aforementioned pairs is close to so called "twilight zone" (region between 30 % and 20 % ID where homology may exist but cannot be reliably assumed [23]). This means that the level of divergence between members of the pair is as high as possible to provide their relatedness detection on the basis of protein homology search. Accordingly, the nucleotide similarity search tools are poorly applicable for the reliable detection of the orthologous genes for the members of these virus pairs due to the fact that the sequence similarity degrades more rapidly at the DNA than at the amino acid level [24]. Herewith, in spite of so large divergence between the species, the level of their CNE ID (67–68 %) is still significant for the nucleotide-based similarity detection. Taking into account that the baculovirus replication is characterized by a high degree of recombination events, there is theoretical probability that such high level of CNE %ID appears to be the result of a horizontal transfer of CNE between distantly related baculovi-

ruses. However, it is not the case with the CNE of viruses above: an analysis of relationship between pairwise AAIs and CNE %IDs (Fig.1) suggests that the distribution of variables follows the linear relationship, and there is no any outlier on the plot indicative of CNE transfer between highly unrelated alphabaculoviruses.

Thus, based upon the data above, it can be suggested that CNE evolved very slowly from the early stages of the alphabaculovirus evolution.

2. CNE conservation level is comparable with those of the most conserved coding regions

In order to quantify the relative CNE conservation value we compared the degree of CNE conservation, calculated as an av % ID for the 1225-member sample resulted from pairwise comparison of 50 CNE , with those of the genes coding regions of polh (the most conserved baculovirus protein) and pif2 (the most conserved protein among 34 core proteins shared by all baculoviruses) [25]. In the work presented here we have not checked for, nor removed recombinants, assuming that av %ID values cannot be affected significantly by the presence of some recombinants in the 1225-member sample.
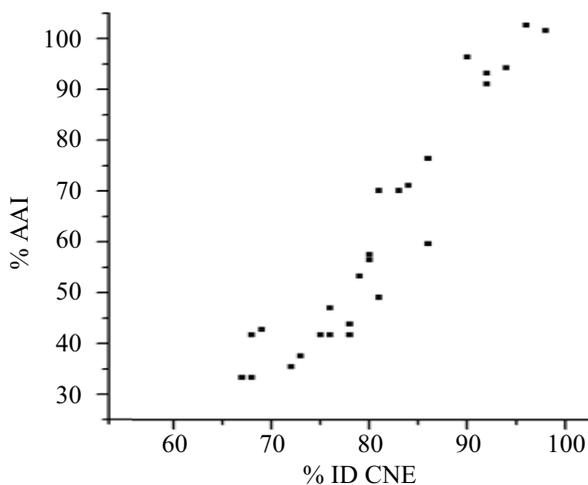
CNE, *polh* and *pif2* sequences were extracted from each of 50 alphabaculovirus genomes available in the GeneBank database and subsequently aligned.

1225 %ID values displayed in the %ID matrix table were subjected to statistical analysis. Box and whisker plots were used to visualize the distribution ranges of %IDs for each orthologous group (Fig. 2). The dataset of each group of orthologs was found to follow a nearly normal distribution ($p < 0.01$). The pairwise IDs values calculated for 1225 virus pairs ranged from 62 % to 100 % (av %ID 73 %) for the CNE, from 79 % to 100 % (av %ID 88 %) for the *polh,* and from 43 % to 100 % (av %ID 65 %) for the *pif2*.

The comparison of av %ID of the CNE and referenced genes showed that CNE tends to be more conserved than *pif2* and less conserved than *polh*. Consequently, CNE can be defined as an exceedingly conserved genomic element, the conservation degree of which is comparable with that of the coding regions of the most conserved baculovirus protein genes.
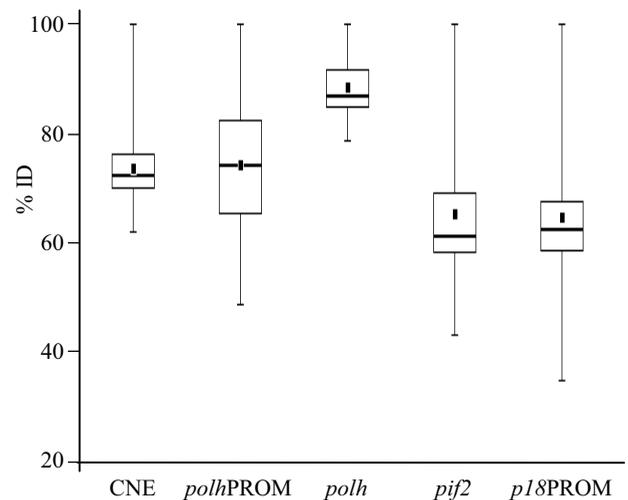

Fig. 2. Box and whisker plot for distribution of ID values for CNE sequence compared with those of the most conserved elements of baculovirus genome. All data were estimated for 1225 alphabaculovirus pairs from 50-member sample. Small black squares indicate average %IDs, horizontal line within boxes display median %IDs. *Polh* promoter region, *polh* coding region, *pif2* coding region and *p18* promoter region are marked as *polh*-PROM, *polh, pif2* and *p18*PROM respectively.

Fig. 1. Relationship between average amino acid identity (AAI) and CNE identity (ID) in alphabaculovirus genomes. The dot plot was created on the basis of 27 virus pairs dataset listed in TanbleS1.

3. CNE is quite unique among baculovirus non-protein coding elements in respect of its evolutionary stability

Besides CNE, the baculovirus set of the known non protein-coding functional elements includes the promoters of different temporal regulation, *hrs* (homologous repeat regions, that play roles of enhancers and origins of replication), so called non-*hr* origins of replication (Rewieved in [26]), 5` and 3` UTRs of mRNAs, ncRNA genes (ncRNA of unknown function [8] and microRNA precursor genes [27] are among them). It is well established that the known alphabaculovirus non-*hr* origins are not alignable [28], as well as *hrs* originating from the genomes of relatively diverse alphabaculoviruses do not express extended homology, although they contain short conserved palindromic segment, a binding site for ie1 [29]. The nucleotide sequences of baculovirus genes of miRNA precursors were found not being conserved even among closely related baculoviruses [30].

The question whether CNE is unique among the alphabaculovirus non protein-coding elements in respect of extended sequence conservation arises from these data. To clarify this issue we focused on the search for the functional non protein-coding sequences of extended length shared by the alphabaculovirus genomes.

The promoter regions of 44 late and very late genes shared by all alphabaculoviruses were chosen as a class of elements subjected to the analysis. The baculovirus late and very late genes are specific in the conserved motif (A/T/G)TAAG, which serves as a transcriptional initiator [8]. The 12–18-bp region encompassing the TAAG has been shown as a minimal promoter determinant for the basal transcription from late [31] and very late promoters [32]. AcMNPV was used as a reference virus to define the exact extension of the promoter region (5`UTR in conjunction with overlapping minimal promoter) of each gene because the genuine transcription start sites of all AcMNPV genes were determined [8]. The regions extracted for the analysis from the AcMNPV genome were bounded by a nucleotide upstream from the translation start codon and by the 30-bp nucleotide stretch upstream from the transcription start site TAAG. The promoter regions less than 40 bp in length were excluded, and finally 37 of 44 regions were chosen for the analysis, including the regions of *polh, lef-9, ODV-e18 , Ac81, lef-5, vlf-1, 38K, gp41, p33, alk-exo, p6.9, lef-2, Ac68, vp91, vp59, p143, pp34, ODV-e25, Ac106, Ac1313, pp31, Ac53, 25K, Ac75, Ac76, p18, ODV-e28, Ac102, F-protein, ODV-c42, me53, p78/83, Ac19, calyx, pkip, Ac60*. The recently published phylogenetic tree was used as a guide to infer the evolutionary distances within alphabaculoviruses [33], and 10 viruses belonging to diverse evolutionary lineages were chosen for the first round of analysis: AcMNPV, EppoNPV, LeseNPV, AgMNPV, OrleNPV, LdMNPV, HearSNPV, TnSNPV, MacoNPV-B, SeMNPV. The promoter regions of each of 37 genes were extracted from each of 10 above genomic sequences. The lengths of extracted sequences were similar to the lengths of their AcMNPV orthologs if their TAAG was found at the position close to that of TAAG site in the AcMNPV promoter region. If the TAAG was absent in the above position the 5`end of the extracted sequence was extended up to the first upstream TAAG.

The multiple sequence alignment was used to assess the conservation levels of 37 promoter regions. The consensus sequences derived from the alignments served as the main criteria of conservation. A strong tendency to the extended sequence conservation was clearly displayed by the alignments of 9 promoter regions (*polh, ODV-e18, pp34, ODV-e25, pp31, 25K, p18, Ac102, calyx*): long sequence stretches (from 40 to 140bp) enriched by absolutely conserved nucleotides were visualized. The most conserved of them assigned to the *polh* promoter region and the longest of them assigned to the *p18* promoter region were chosen for the subsequent aligning using a sample from 50 viruses. The multiple alignments demonstrated that the conserved nucleotide stretch of the *polh* promoter region is represented by full 5`UTR, whereas the conserved nucleotide

stretch of the *p18* promoter region includes not only the full 5`UTR but also a dozen of conserved nucleotides located just upstream the TAAG.

The %ID datasets of the promoter regions of both genes were found to be distributed nearly normally ($p<0.01$). ID values of the *polh* promoter region varied from 49 % to 100 % (av %ID 74 %) over sequence lengths ranging from 39 to 42 bp. ID values of *p18* varied from 35 % to 100 % (av %ID 65 %) over sequence lengths ranging from 129 to 145 bp (Fig.2). The comparison of avID values allows the suggestion that CNE index of this value is nearly the same as that of the *polh* promoter region and higher than that of the *p18* promoter region.

Discussion

In this study the relationship between the CNE conservation degree and the degree of virus pair divergence has been analyzed. This preliminary analysis revealed the tendency to strong evolutionary stability of CNE. A comparison of the CNE conservation level with those of the nucleotide sequences of genes for the structural proteins under strong evolutionary constraint (polh and pif2) and with those of the most conserved promoter regions of the late genes (*polh* and *p18*) was done to investigate the observed tendency further. The CNE sequences, gene coding sequences and gene promoter region sequences were extracted from 50 alphabaculovirus genomes available in NCBI database, aligned and subjected to pairwise comparison. Resulting data allowed us to recognize the alphabaculovirus CNE as an element of extreme conservation across the 152–156 bases, being more highly conserved than the vast majority of both coding and non-coding sequences in the alphabaculovirus genomes. CNE is unique among the previously studied baculovirus non protein-coding elements (*hrs*, non-*hr* oris, genes of miRNA precursors) because the latter do not exhibit a tendency of sequence conservation between distantly related alphabaculoviruses (see references in Results, section3).

Of the particular interest are the results of promoter region analysis conducted within a framework of

this study: nine baculovirus late genes (*polh, ODV-e18, pp34, ODV-e25, pp31, 25K, p18, Ac102, calyx*) were found to be associated with the promoter regions of relatively high levels of sequence conservation. ID values of the two most conserved of them (the *polh* promoter region including 5'UTR and *p18* promoter region including 5'UTR and dozen nucleotides upstream from the transcription start site) were found to be comparable with the CNE ID value. An extreme conservation of some baculovirus 5`-UTRs is quite explicable: 5`UTRs may represent RNA regions that adopt a selectively constrained spatial structure providing molecule stability and/or interaction with proteins.

In contrast, it is difficult to predict whether the revealed phenomenon of extreme conservation of the alphabaculovirus CNE specifies its similarity with some eukaryotic CNEs or it indicates the uniqueness of some viral element providing the alphabaculovirus-specific function(s).

Three alternative (but not mutually exclusive) hypotheses may be suggested to explain the above phenomenon.

1. "The overlap hypothesis". The CNE represents the region of overlapping several non-coding elements of extended length. The overlapping elements may evolve slowly due to the increasing deleterious effect of mutations affecting more than one element. However, this assumption is valid for the explanation of overall high conservation degree of the CNE but the following question remains unanswered: why is the length of quite variable CNE blocks kept unaltered (i.e. are any insertions or deletions within variable regions not allowed)?

2. "The secondary structure hypothesis". The CNE nucleotide sequence conservation is dictated by significance of the secondary structure of this element at either DNA or RNA level. This assumption is supported by the fact of CNE enrichment in the dyad symmetry elements that specifies CNE as a sequence potentially capable of the DNA conformational changes [34], and by the fact that as shown by the AcMNPV transcriptome analysis three ncRNA are likely overlap CNE in the AcMNPV genome [8].

136

It should be noted, however, that conserved parts of the overlapping ncRNAs represented by CNE were shown not to be essential for the AcMNPV life cycle [7] and this put the necessity of their extreme conservation under the question. In addition, the hypothesis can be considered strictly valid only if the CNE-overlapping ncRNAs would be identified in all other alphabaculovirus genomes.

3. "The module hypothesis". CNE appears to be the module, consisting of several conserved motifs, which bind cooperatively acting proteins. These modules, within which the particular distances between protein-binding sites provide interaction of bound proteins, were predicted by bioinformatic approaches in eukaryotic genomes [35]. Herewith, such organization has been observed only in very few cases of transcriptional regulatory elements [36], among them the most studied is the enhancer of interferon-β gene that rather represents [a] particular case as it consists of numerous non spaced overlapping motifs [37]. While [the] c4-c5 CNE conserved block was shown to be the only which bind[s] proteins at 40 hours post infection [9], c2, c3 and c6 blocks remain the candidate binding motifs for cooperatively acting proteins at some other time point. It also seems plausible to speculate that two similar terminal blocks c1 and c7 that bind proteins in uninfected cell, as well as in early infection [9], represent such recognition motifs for cooperatively acting proteins, and the strictly predetermined distance between them appears to be the prerequisite of their cooperation and, thus defines the fixed CNE length.

Further experimental research is necessary to infer the essence of extreme conservation of the alphabaculovirus CNE, so the functional and structural studies of this element are under way.

REFERENCES

1. *Elgar G.* Pan-vertebrate conserved non-coding sequences associated with developmental regulation. *Brief Funct Genomic Proteomic.* 2009;**8**(4):256–65.

2. *Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G.* Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 2005;**3**(1):e7.

3. *Harmston N, Baresic A, Lenhard B.* The mystery of extreme non-coding conservation. *Philos Trans R Soc Lond B Biol Sci.* 2013;**368**(1632):20130021.

4. *Dubchak I, Brudno M, Loots GG, Pachter L, Mayor C, Rubin EM, Frazer KA.* Active conservation of noncoding sequences revealed by three-way species comparisons. *Genome Res.* 2000;**10**(9):1304–6.

5. *Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D.* Ultraconserved elements in the human genome. *Science.* 2004;**304**(5675):1321–5.

6. *Clarke SL, VanderMeer JE, Wenger AM, Schaar BT, Ahituv N, Bejerano G.* Human developmental enhancers conserved between deuterostomes and protostomes. *PLoS Genet.* 2012;**8**(8):e1002852.

7. *Kikhno I.* Identification of a conserved non-protein-coding genomic element that plays an essential role in Alphabaculovirus pathogenesis. *PLoS One.* 2014;**9**(4):e95322.

8. *Chen YR, Zhong S, Fei Z, Hashimoto Y, Xiang JZ, Zhang S, Blissard GW.* The transcriptome of the baculovirus Autographa californica multiple nucleopolyhedrovirus in Trichoplusia ni cells. *J Virol.* 2013;**87**(11):6391–405.

9. *Krappa R, Behn-Krappa A, Jahnel F, Doerfler W, Knebel-Mörsdorf D.* Differential factor binding at the promoter of early baculovirus gene PE38 during viral infection: GATA motif is recognized by an insect protein. *J Virol.* 1992;**66**(6):3494–503.

10. *Carson DD, Summers MD, Guarino LA.* Molecular analysis of a baculovirus regulatory gene. *Virology.* 1991;**182**(1):279–86.

11. *Thompson JD, Higgins DG, Gibson TJ.* CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 1994;**22**(22):4673–80.

12. *Herniou EA, Jehle JA.* Baculovirus phylogeny and evolution. *Curr Drug Targets.* 2007;**8**(10):1043–50.

13. *Chen X, Zhang WJ, Wong J, Chun G, Lu A, McCutchen BF, Presnail JK, Herrmann R, Dolan M, Tingey S, Hu ZH, Vlak JM.* Comparative analysis of the complete genome sequences of Helicoverpa zea and Helicoverpa armigera single-nucleocapsid nucleopolyhedroviruses. *J Gen Virol.* 2002;**83**(Pt 3):673–84.

14. *Tang P, Zhang H, Li Y, Han B, Wang G, Qin Q, Zhang Z.* Genomic sequencing and analyses of HearMNPV--a new Multinucleocapsid nucleopolyhedrovirus isolated from Helicoverpa armigera. *Virol J.* 2012;**9**:168.

15. *Gomi S, Majima K, Maeda S.* Sequence analysis of the genome of Bombyx mori nucleopolyhedrovirus. *J Gen Virol.* 1999;**80**(Pt 5):1323–37.

16. *Harrison RL, Bonning BC.* Comparative analysis of the genomes of Rachiplusia ou and Autographa californica multi-

ple nucleopolyhedroviruses. *J Gen Virol.* 2003;**84**(Pt 7): 1827–42.

17. *Lauzon HA, Jamieson PB, Krell PJ, Arif BM.* Gene organization and sequencing of the Choristoneura fumiferana defective nucleopolyhedrovirus genome. *J Gen Virol.* 2005;**86**(Pt 4):945–61.

18. *Nai YS, Wu CY, Wang TC, Chen YR, Lau WH, Lo CF, Tsai MF, Wang CH.* Genomic sequencing and analyses of Lymantria xylina multiple nucleopolyhedrovirus. *BMC Genomics.* 2010;**11**:116.

19. *Ahrens CH, Russell RL, Funk CJ, Evans JT, Harwood SH, Rohrmann GF.* The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear polyhedrosis virus genome. *Virology.* 1997;**229**(2):381–99.

20. *Kuzio J, Pearson MN, Harwood SH, Funk CJ, Evans JT, Slavicek JM, Rohrmann GF.* Sequence and analysis of the genome of a baculovirus pathogenic for Lymantria dispar. *Virology.* 1999;**253**(1):17–34.

21. *Thumbi DK, Eveleigh RJ, Lucarotti CJ, Lapointe R, Graham RI, Pavlik L, Lauzon HA, Arif BM.* Complete sequence, analysis and organization of the Orgyia leucostigma nucleopolyhedrovirus genome. *Viruses.* 2011 Nov;**3**(11):2301–27.

22. *Chen X, IJkel WF, Tarchini R, Sun X, Sandbrink H, Wang H, Peters S, Zuidema D, Lankhorst RK, Vlak JM, Hu Z.* The sequence of the Helicoverpa armigera single nucleocapsid nucleopolyhedrovirus genome. *J Gen Virol.* 2001;**82**(Pt 1):241–57.

23. *Rost B.* Twilight zone of protein sequence alignments. *Protein Eng.* 1999;**12**(2):85–94.

24. *Pearson WR.* An introduction to sequence similarity ("homology") searching. *Curr Protoc Bioinformatics.* 2013;Chapter 3:Unit3.1.

25. *Miele SA, Garavaglia MJ, Belaich MN, Ghiringhelli PD.* Baculovirus: molecular insights on their diversity and conservation. *Int J Evol Biol.* 2011;**2011**:379424.

26. *Hayakawa T, Rohrmann GF, Hashimoto Y.* Patterns of genome organization and content in lepidopteran baculoviruses. *Virology.* 2000;**278**(1):1–12.

27. *Singh J, Singh CP, Bhavani A, Nagaraju J.* Discovering microRNAs from Bombyx mori nucleopolyhedrosis virus. *Virology.* 2010;**407**(1):120–8.

28. *Huang J, Levin DB.* Identification and functional analysis of a putative non-hr origin of DNA replication from the Spodoptera littoralis type B multinucleocapsid nucleopolyhedrovirus. *J Gen Virol.* 1999;**80** (Pt 8):2263–74.

29. *Rodems SM, Friesen PD.* Transcriptional enhancer activity of hr5 requires dual-palindrome half sites that mediate binding of a dimeric form of the baculovirus transregulator IE1. *J Virol.* 1995;**69**(9):5368–75.

30. *Andreassen R, Worren MM, Høyheim B.* Discovery and characterization of miRNA genes in Atlantic salmon (Salmo salar) by use of a deep sequencing approach. *BMC Genomics.* 2013;**14**:482.

31. *Morris TD, Miller LK.* Mutational analysis of a baculovirus major late promoter. *Gene.* 1994;**140**(2):147–53.

32. *Weyer U, Possee RD.* Analysis of the promoter of the Autographa californica nuclear polyhedrosis virus p10 gene. *J Gen Virol.* 1989;**70**(Pt 1):203–8.

33. *Krejmer M, Skrzecz I, Wasag B, Szewczyk B, Rabalski L.* The genome of Dasychira pudibunda nucleopolyhedrovirus (DapuNPV) reveals novel genetic connection between baculoviruses infecting moths of the Lymantriidae family. *BMC Genomics.* 2015;**16**(1):759.

34. *Brázda V, Laister RC, Jagelská EB, Arrowsmith C.* Cruciform structures are a common DNA feature important for regulating biological processes. *BMC Mol Biol.* 2011;**12**:33.

35. *Guturu H, Doxey AC, Wenger AM, Bejerano G.* Structure-aided prediction of mammalian transcription factor complexes in conserved non-coding elements. *Philos Trans R Soc Lond B Biol Sci.* 2013;**368**(1632):20130029.

36. *Merika M, Thanos D.* Enhanceosomes. *Curr Opin Genet Dev.* 2001;**11**(2):205–8.

37. *Panne D, Maniatis T, Harrison SC.* An atomic model of the interferon-beta enhanceosome. *Cell.* 2007;**129**(6):1111–23.

Надзвичайна еволюційна стабільність консервативного некодуючого елементу (CNE) геному бакуловірусів

В. Е. Макаренко, І. В. Кіхно, В. І. Кашуба

Мета. Визначити рівень еволюційної стабільності CNE (conserved non protein-coding element) – некодуючого функціонального елементу бакуловірусного геному, що грає виключну роль у патогенезі бакуловірусів. **Методи.** Для ідентифікації ортологічних послідовностей в геномах 50 альфабакуловірусів застосовували NCBI-BLAST. Ортологічні послідовності вирівнювали за допомогою алгоритму вирівнювання ClustalW. Отримували матрицю значень відсотків ідентичності (percent identity, %ID) для попарних порівнянь за допомогою програми Alistat. **Результати.** Середній рівень ID (73 %), розрахований для 1225-членної вибірки, виявився порівнюваним з такими показниками кодуючих послідовностей найконсервативніших бакуловірусних генів *polh* та *pif2*, а також з показниками промоторних ділянок *polh* та *p18*, що є найконсервативнішими представниками невеликої групи еволюційно стабільних промоторних ділянок пізніх генів. **Висновки.** CNE є одним з найконсервативніших елементів геномів альфабакуловірусів.

К л ю ч о в і   с л о в а: бакуловіруси, консервативний некодуючий елемент.

Чрезвычайная эволюционная стабильность консервативного некодирующего элемента (CNE) геномов бакуловирусов

В. Э. Макаренко, И. В. Кихно, В. И. Кашуба

Цель. Определить уровень эволюционной стабильности CNE (conserved non protein-coding element) – некодирующего функ-

ционального элемента бакуловирусного генома, который играет исключительную роль в патогенезе бакуловирусов. **Методы.** Для идентификации ортологичных последовательностей в геномах 50 альфабакуловирусов применяли NCBI-BLAST. Ортологичные последовательности выравнивали при помощи алгоритма выравнивания ClustalW. Получали матрицу значений процентов идентичности (percent identity, %ID) для попарных сравнений при помощи программы Alistat. **Результаты.** Средний уровень ID (73 %), рассчитанный для 1225-членной выборки, является сравнимым с таковыми показателями кодирующих последовательностей наиболее консервативных баку-

ловирусных генов *polh* и *pif2*, а также с показателями промоторных участков *polh* и *p18,* которые являются самыми консервативними представителями небольшой группы эволюционно стабильных промоторных участков поздних генов. **Выводы.** CNE является одним из наиболее консервативных элементов геномов альфабакуловирусов.

К л ю ч е в ы е   с л о в а: бакуловирусы, консервативный некодирующий элемент