Doklady Biochemistry and Biophysics, Vol. 401, 2005, pp. 104–107. Translated from Doklady Akademii Nauk, Vol. 401, No. 1, 2005, pp. 112–116.

BIOCHEMISTRY, BIOPHYSICS, AND MOLECULAR BIOLOGY

Discovery of Retrotransposons of the Cockroach *Blattella germanica*

A. G. Chumachenko*, C. Schal, and D. V. Mukha***

Presented by Academician Yu.P. Altukhov June 8, 2004

Received July 14, 2004

Transposable elements are present in practically all eukaryotes studied and account for tens of percent of genomic DNA. Although numerous studies were performed and important results were obtained, the question on the biological role of transposable elements and their presumable relationship with viruses remains open and requires further investigation. From our standpoint, molecular-genetic studies of novel species that have not been studied earlier (and which, therefore, are not model) are principally important for a complex understanding the entire diversity of genetic processes underlying the functional activity and evolutionary variability of transposable elements.

This was the first work to reveal, clone, and sequence extended DNA fragments of the cockroach *Blattella germanica*, which share a high extent of similarity with the highly conserved reverse-transcriptase domains of insect retrotransposons. The discovery of fragments of transposable elements makes it possible to clone the full-length copies and to perform further study of the structural and functional organization and biological role of these genomic elements of *B. germanica.*

Within the framework of studying the structural and functional organization of *Bg*DNV—a densovirus of *B. germanica* that has been discovered recently [1], we amplified and cloned DNA fragments corresponding to open reading frames (ORFs) of the viral genome. This study is a necessary stage of creation of recombinant molecules and their subsequent expression in *E. coli* cells in order to accumulate viral polypeptides and obtain virus-specific antibodies. The scheme of relative position of the viral ORFs and location of primers used for amplification of the viral genome regions specified is shown in Fig. 1a. For amplification we used the following pairs of primers: 1a–1b, 2a–2b, 3a/4a–3b, 3a/4a–4b, and 5a–5b. Amplification was conducted using the GenePakCore kit according to the recommendations of the manufacturer (Laboratoriya Izogen). Total DNA extracted from cockroaches that were infected with the virus was used as a template. Amplification included the following stages: denaturation at 95°C for 5 min; 30 cycles (denaturation at 94°C for 1 min, annealing of primers at 53°C for 2 min, and elongation at 72°C for 3 min); and final elongation for 7 min.

As a result, the viral genome fragments located between the corresponding pair of primers were amplified and cloned (data not shown).

it is known that polymerase chain reaction (PCR) is an effective highly sensitive procedure that allows genomic regions specified to be amplified. However, when using this method, it is necessary to use the maximal number of controls confirming the specificity of the reaction product. At the stage of selecting the optimal amplification conditions, especially in those cases when DNA used as a template contains subrepeats and can form complex secondary structures (an example of such template is DNA of the virus studied), it is very important to use negative controls (in particular, primers and their different combinations), which, is the case of specific PCR, should not amplify the template DNA. For example, in the case of ORF4 amplification, in addition to the pair of primers 3a/4a (5'-ctttggagtcactatgggagatg-3')–4b (5'-ttattcctgtagctcctccg-3'), which ensure PCR amplification of the specified region of viral genome under optimal conditions, we also used the pair of primers 3a/4a' (5'-gaaacctcagtgataccctctac-3') and 4b (5'-aataaggacatcgaggaggc-3'), where the primers 3a/4a' and 4b' are DNA sequences complementary to primers 3a/4a and 4b, respectively. Apparently, when performing PCR with the use of primers 3a/4a' and 4b', the product of viral genome amplification should not be detected.

Surprisingly, when the pair of primers 3a/4a' and 4b' was used within a broad range of temperature and Mg^{2+} concentration, PCR of total DNA of virus-infected cockroaches resulted in stable amplification of three DNA fragments of 2.1, 0.9, and 0.73 kb in length (Fig. 1b). PCR conditions used in this experiments do not allow us to precisely determine the proportion of

* *Vavilov Institute of General Genetics, Russian Academy of Sciences, ul. Gubkina 3, Moscow, 119991 Russia*

** *Northern Caroline State University, Northern Caroline, Rali, United States*

1607-6729/05/0304-0104 © 2005 Pleiades Publishing, Inc.

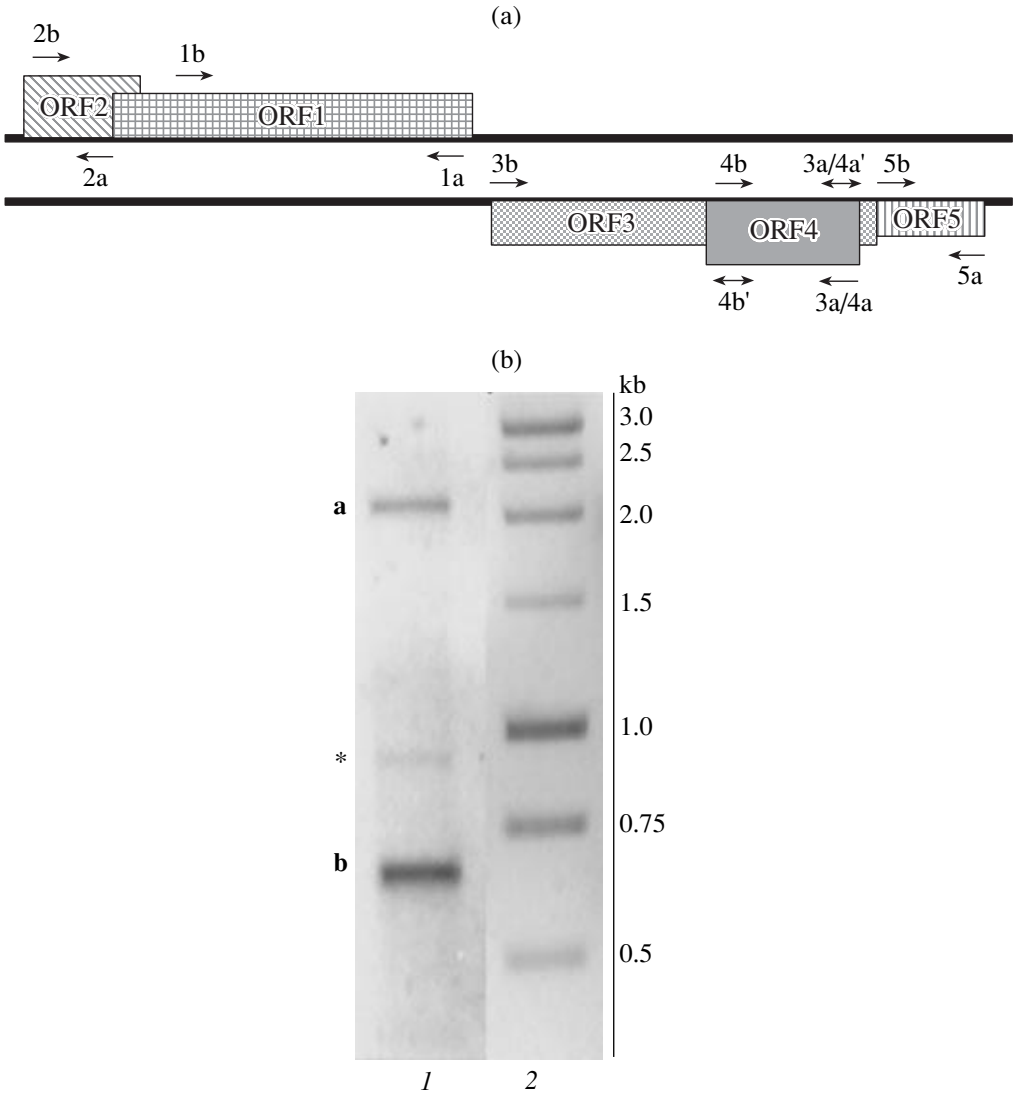

Fig. 1. Polymerase chain reaction (PCR) of total DNA of densovirus-infected *B. germanica* using the primers corresponding to the boundaries of the open reading frames (ORFs) of the viral genome. (a) Scheme of the structural organization of the *Bg*DNV densovirus. Designations: ORF1, ORF2, ORF3, ORF4, and ORF5, the corresponding ORFs; 1a–1b, 2a–2b, 3a/4a–3b, 3a/4a–4b, and 5a–5b, pairs of primers flanking the corresponding ORFs (the directin of areas show the orientation of primers); 3a/4a' and 4b', primes complementary to the primers 3a/4a and 4b, respectively (for explanation see text). (b) Results of PCR with the primers 3a/4a' and 4b'. Lane *1* shows the major amplified fragments (**a** and **b**) and the minor fragment (shown with asterisk). Lane *2* shows the DNA marker.

the fragments amplified. However, it should be noted that the 0.9-kb fragment, unlike the other two fragments, was always (in more than ten experiments) represented by minor DNA quantities. The DNA fragments of 2.1 and 0.73 kb in length (designated on Fig. 1b as **a** and **b**, respectively) have been cloned and sequenced.

A comparison of the nucleotide sequences of the fragments **a** and **b** with the nucleotide sequences contained in the GenBank database, as well as with one another using the Blast (blastn)software, did not reveal a significant degree of similarity. However, when comparing the amino acid sequences of all possible ORFs of the fragments **a** and **b** with the amino acid sequences of proteins deposited in the Swissprot database using the BLAST (blastx) software, significant degrees of similarity with the amino acid sequences corresponding to the highly conserved amino acid domain of revertase (reverse transcriptase) belonging to different retrotransposons of insects have been revealed (Fig. 2). Both types of comparisons (blastn and blastx) were performed using the corresponding software available at the site http://www.ncbi.nlm.nih.gov/BLAST/.

As seen in Fig. 2a, translated DNA of the fragment **a** shared the maximal degree of similarity with the amino acid sequence of reverse transcriptase of the butterfly

(a)

Amino acids: identical, 46% (32/69); positive, 63% (44/69)

```
a       224 VPSNWNTAVLHMIWKGKGRRDEIDKYRPIALTSIFRKVLEKTMIAQLQEY-EDR--LDVA 394
            VPS+W TA +H I K KG R +   YRPIA+TS+  KV+E+ +  QL +Y EDR  +
Rt-Bm   512 VPSSWKTAHVHPIPK-KGDRSDPSSYRPIAITSLLSKVMERIINIQLLKYLEDRQLISDR 570

a       395 QGGFIKGKA 421
            Q GF  G++
Rt-Bm   571 QYGFRHGRS 579
```

(b)

Amino acids: identical, 31% (40/128); positive, 53% (68/128)

```
b       375 KGGVAGSLLSPILFNIFIDSLPRTQRTKHSSFLLGSHRINSLLYADDIVLVSSTRSGLQS 554
            +G   G  LSP+LFN+ ID L R+   K     +G+ ++N+  +ADD++L +ST  GLQ
R2-Dm   528 RGVKQGDPLSPVLFNLIIDRLLRSL-PKDIGVHVGNAKVNACAFADDLMLFASTPKGLQE 586

b       555 MLDTCERHSISHGYVFAPSKCEVIAPQG----------KKTPCVTMYGEKVRKTPSFKYL 704
            +L+T  +   S G      KC  I+ +G          ++T C+     +++++  +KYL
R2-Dm   587 LLNTTVKFLSSVGLTLNADKCFTISIKGQPKQKVTVVEQRTFCIGRARVQLKRSEEWKYL 646

b       705 GVPFNDKG 724
            G+ F   G
R2-Dm   647 GIHFTADG 654
```

(c)

Amino acids: identical, 31% (44/140); positive, 53% (75/140)

```
b       375 KGGVAGSLLSPILFNIFIDSLPRTQRTKHSSFLLGSHRINSLLYADDIVLVSSTRSGLQS 554
            +G   G  LSP+LFN  +D++ R +  +++ FL+G+ +I +L++ADD+VL++ TR GLQ+
R2-Nv   511 RGVRQGDPLSPLLFNCVMDAVLR-RLPENTGFLMGAEKIGALVFADDLVLLAETREGLQA 569

b       555 MLDTCERHSISHGYVFAPSKCEVIA--PQGKKTPC-------VTMYGEKVRK---TPSFK 698
             L   E      G    P KC  +A  P GK+           T+  +++ +      +K
R2-Nv   570 SLSRIEAGLQEQGLEMMPRKCHTLALVPSGKEKKIKVETHKPFTVGNQEITQLGHADQWK 629

b       699 YLGVPFNDKGVDMAELCVEG 758
            YLGV +N  G    ++ + G
R2-Nv   630 YLGVVYNSYGPIQVKINIAG 649
```

Fig. 2. Comparative analysis of the amino acid sequences of translated DNA of (a) fragment **a** and (b, c) fragment **b** using the BLAST (blastx) software. Designations: Rt-Bm, R2-Dm, and R2-Nv are highly conserved domains of reverse transcriptase of retrotransposons of *B. mori, D. mercatorum*, and *N. vitripennis*, respectively.

Bombyx mori retrotransposon containing no long terminal repeats (non-LTR; accession no. AY359886 at GenBank). Translated DNA of the fragment **b** also shared the maximal degree of similarity with this type of transposable elements—specifically, with the amino acid sequence of reverse transcriptase of type R2 retrotransposons of the fly *Drosophila mercatorum* (accession no. AF015685 at GenBank) [2] (Fig. 2b) and the ichneumon wasp *Nasonia vitripennis* (accession no. L00950 at GenBank) [3] (Fig. 2c).

The non-LTR retrotransposons are one of numerous and diverse classes of transposable elements of eukary-

otic genome. Today, based on the structural characteristics and similarity of nucleotide composition, eleven groups (specifically, CRE, R1, R2, R4, L1, RTE, Tad1, LOA, Jockey, CR1, and I) of non-LTR retrotransposons are distinguished. Transposable elements of one group of represent part of genomes of evolutionarily distant taxons of eukaryotes. For example, group L1 transposable elements were found in the genome of human, animals, plants, and protozoans [4]. It was shown that non-LTR retrotransposons usually contain two ORFs coding for reverse transcriptase and endonuclease, which ensure the transposition and increase in the number of copies of the transposable element [4].

The biological role of transposable elements remains largely obscure. From our standpoint, the key approach to solving this problem may include studies of characteristics of the structural and functional organization and the consistent patterns of biological activity of non-LTR retrotransposons that have specific integration sites in eukaryotic genomes. The R1 and R2 retrotransposons are examples of such transposable elements.

The clusters of ribosomal genes of almost all arthropods (in particular, insects) that have been described to date, contain the R1 and R2 transposable elements [5]. Both transposable elements have specific integration sites in the 28S rRNA gene sequence [6]. It was shown that the 28S rRNA genes may contain either one transposable element (R1 or R2) or both retrotransposons [7, 8]. In *D. melanogaster*, depending on the strain, 17 to 67% of the 28S genes contain the R1 transposable elements, and only 2–28% contain the R2 elements. The common percentage of the 28S genes containing either R1 or R2 elements varies from 32 to 77 [8]. Insertions of both R1 and R2 elements result in inactivation of the 28S genes [9, 10].

The phylogenetic analysis of the R1 and R2 retrotransposons of the genus *Drosophila* showed that vertical transmission is characteristic of transposable elements of this type [11, 12]. Furthermore, the phylogenetic analysis of the R1 and R2 retrotransposons of evolutionarily distant taxons of arthropods allows the existence of horizontal transmission to be assumed and the approximate time of origination of the R1 and R2 transposable elements to be determined. Currently, it is believed that these elements could appear approximately 600 billion years ago [4, 5].

Currently, the transposable elements revealed in the *B. germanica* genome are actively studied in our laboratory. We plan to create a cosmid library of *B. germanica* genes and to perform screening of the library using probes containing cloned retrotransposon fragments. This approach will allow us to clone the full-length copies of these transposable elements and to begin studying the biological role of the latter. In addition, the evolutionary succession between the discovered transposable elements of *B. germanica* and the densovirus (the primers complementary to this virus allowed the transposable elements to be amplified) remains to be understood.

ACKNOWLEDGMENTS

This study was supported by the Russian Foundation for Basic Research, project no. 04-04-49297a; and the program on basic research from the Presidium of the Russian Academy of Sciences “The Dynamics of Gene Pools of Plants, Animals, and Human,” project no. 24-IOG-09-2004.

REFERENCES

1. Mukha, D.V. and Shal, K., *Mol. Biol.* (Moscow), 2003, vol. 37, pp. 607–618.
2. Malik, H.S., Burke, W.D., and Eickbush, T.H., *Genetics*, 1999, vol. 151, pp. 653–665.
3. Burke, W.D., Eickbush, D.G., Xiong, Y., *et al., Mol. Biol. Evol.,* 1993, vol. 10, pp. 163–185.
4. Malik, H.S., Burke, W.D., and Eickbush, T.H., *Mol. Biol. Evol.*, 1999, vol. 16, pp. 793–805.
5. Burke, W.D., Malik, H.S., Lathe, W.C., and Eickbush, T.H., *Nature*, 1998, vol. 392, pp. 141–142.
6. Perez-Gonzalez, C.E. and Eickbush, T.H., *Genetics*, 2001, vol. 158, pp. 1557–1567.
7. Jakubczak, J., Xiong, Y., and Eickbush, T.H., *J. Mol. Biol.*, 1990, vol. 212, pp. 37–52.
8. Jakubczak, J., Zenni, M.K., Woodruff, R.C., and Eickbush, T.N., *Genetics*, 1992, vol. 131, pp. 129–142.
9. Long, E.O. and Dawid, I.B., *Cell*, 1979, vol. 18, pp. 1185–1196.
10. Kidd, S.J. and Glover, D.M., *J. Mol. Biol.*, 1981, vol. 151, pp. 645–662.
11. Lathe, W.C. and Eickbush, T.H., *Mol. Biol. Evol.*, 1997, vol. 14, pp. 1232–1241.
12. Gentile, K., Burke, W.D., and Eickbush, T.H., *Mol. Biol. Evol.*, 2001, vol. 18, pp. 235–245.