INTRODUCTION

The R1 retroposons are site-specific mobile elements (ME) of eukaryotes, whose transposition occurs within the ribosomal gene cluster. These ME contain two open reading frames (ORFs). The protein corresponding to ORF1 is homologous to GAG protein of retroviruses, while the protein corresponding to ORF2 plays a key role in the process of transposition [1, 2]. It was demonstrated that integration of R1 retroposons occurs into certain target site of the 28S rRNA gene, which results in inactivation of this gene copy [3–5].

In the genomes of a number of organisms, several retroposon subfamilies with similar target sites, but displaying different degrees of differences in nucleotide composition, were described. For instance, in the genome of fruit fly Drosophila melanogaster has two retroposon R1 subfamilies [6], while in the genome of jewel wasp Nasonia vitripennis, at least four retroposon R2 subfamilies are identified [7].

The appearance of several ME subfamilies in one genome can be caused by horizontal transfer [8]. However, the existence of closely relative subfamilies of R1 retroposons with similar structure and nucleotide sequences clearly points to the possibility of intragenomic divergence of this type of mobile elements. Because of this, these mobile elements can serve as convenient model to study the patterns of molecular evolution of eukaryotic genome as a whole.

In our earlier studies, cloning and amplification of different variants of 5' truncated copies of R1 retroposons from the genomic DNA of German cockroach Blattella germanica was performed. Sequence comparison of the clones obtained provided identification of two subfamilies of this retroposon in German cockroach [9].

In the present study, a library of German cockroach genes constructed with the help of cosmid vector, was screened with the probes generated using cloned fragments of 5'-truncated copies of two subfamilies of R1 retroposons. Full-length copies of retroposons of interest were identified and partly sequenced. It was demonstrated that in German cockroach, retroposons R1, belonging to different subfamilies, differed from one another in domain structure of the C-terminal region of the second reading frame. For the first time, in the sequence of retroposon R1 transmembrane domains were detected.

MATERIALS AND METHODS

Cloning and sequencing. A genomic library of German cockroach was generated using the SuperCos I (Stratagene) cosmid vector, according to the recommendations of the manufacturer. Clones containing recombinant DNA were screened with radioactively labeled probes, represented by the fragments (about 1000 bp) of 5'-truncated copies of the two subfamilies of retroposons R1, described earlier [9]. The assignment cloned mobile elements to one or another subfamily types, as well as their full-length was endorsed by sequencing of the ME flanking regions. Integrated copies of retroposons were attributed to native (full-length) copies in case of identification of extended 5' untranslated regions. Determination of extended 5' and 3' terminal sequences of mobile elements examined was performed using the method of stepwise sequencing.
Recombinant cosmid DNA was extracted using the Wizard Plus Minipreps DNA Purification System (Promega) kit. Full-length inserts of the mobile elements of interest in cosmid vectors were amplified using the method of polymerase chain reaction (PCR) with primers flanking German cockroach retroposon R1 target sites [9]. Amplification was performed using PCR, optimized for amplification of extended DNA fragments. The reaction was carried out in a final volume of 50 μl and using the Long-PCR Enzyme Mix (Fermentas) reagent kit according to the recommendations of the manufacturer. The conditions of Long-PCR included initial denaturing for 30 s at 94°C, followed by 10 cycles of 95°C for 30 s; 55°C for 30 s; 68°C for 7 min; 20 cycles of 95°C for 30 s, 55°C 30 s; 68°C for 7 min (with 10 s increase in each cycle); and final elongation at 68°C for 7 min. PCR was carried out in automated regime using the Primus (MWG-Biotech) thermal cycler. The amplificates obtained were sequenced using the method of Sanger, the ABI Wizard Plus Minipreps DNA Purification System and using the Long-PCR Enzyme Mix (Fermentas) reagent kit according to the recommendations of the manufacturer. The conditions of Long-PCR included initial denaturing for 30 s at 94°C, followed by 10 cycles of 95°C for 30 s; 55°C for 30 s; 68°C for 7 min; 20 cycles of 95°C for 30 s, 55°C 30 s; 68°C for 7 min (with 10 s increase in each cycle); and final elongation at 68°C for 7 min. PCR was carried out in automated regime using the Primus (MWG-Biotech) thermal cycler. The amplificates obtained were sequenced using the method of Sanger, the ABI Wizard Plus Minipreps DNA Purification System and using the Long-PCR Enzyme Mix (Fermentas) reagent kit according to the recommendations of the manufacturer. Fluorescent chromatograms of the sequenced fragments were treated using the Chromas-Pro software program.

Bioinformatic analysis. The computer databases and Internet-recourses used are available at the NCBI server (http://www.ncbi.nlm.nih.gov/BLAST/). Multiple alignments of amino acid sequences were carried out using the ClustalW2 and MAFFT software programs (http://www.ebi.ac.uk/Tools/science.html).

Domain structure of extended amino acid sequences was examined using the SMART software (http://smart.embl-heidelberg.de/). Three-dimensional protein structures were predicted with the help of the CPHmodels 3.0 program (http://www.cbs.dtu.dk/services/CPHmodels/). Comparison of the predicted three-dimensional protein structures with the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) was performed using the DaliLite v. 3 program (http://ekhidna.biocenter.helsinki.fi/dali_server/). To obtain and further analyze hydrophobic profiles of amino acid sequences, the Topology prediction of membrane proteins (TopPred) program was used (http://mobyle.pasteur.fr/cgi-bin/portal.py?form=toppred).

The phylogenetic tree was constructed using the Molecular Evolutionary Genetics Analysis (MEGA 4) software (http://www.megasoftware.net/) with neighbor joining algorithm [10].

For comparative analysis, the ORF2 amino acid sequences of the following mobile elements (the GenBank database accession numbers are shown in brackets) were used: TRAS1 (#D38414), SART1 (#D85594), Bombyx mori; R1 (#X51968), Waldo-B (#AJ278684) Drosophila melanaster; R6Ag2 (#AB090818), R7Ag1 (#AB090820), RT1 (#M93690), MinoAg1 (#AB090816) Anopheles gambiae.

RESULTS AND DISCUSSION

From the genomic library of German cockroach, generated based on the cosmid vector, four cosmids were isolated. Two of these cosmids contained full-length copies of retroposon R1 from one of the subfamilies (R1a), while another two cosmids contained full-length retroposon copies belonging to another subfamily (R1b).

In two retroposons belonging to different subfamilies, the extended sequences of 3' terminal regions, R1a (2197 bp) and R1b (2438 bp) were determined (GenBank accession numbers GenBank GU205093 and GU205094, respectively). The degree of similarity of these sequences constituted 51.7%. It was demonstrated that these sequences contained the regions corresponding to the second open reading frame (ORF2), 1395 bp of retroposon R1a, and 16989 bp of retroposon R1b. The segments of 3' untranslated regions of retroposons R1a and R1b were 801 and 784 bp in size, respectively. The 3' ends of the two ME types described contained “polyamide tails” of 20 bp in size. These findings were consistent with our previous data obtained in the course of the analysis of 5' truncated copies of retroposon R1 [9].

The amino acid sequences corresponding to the ORF2 C-terminal regions of retroposons R1a and R1b with the sizes of 465 and 566 amino acid residues, respectively, are demonstrated in Fig. 1a. The degree of similarity of these sequences constituted 43.4% of amino acid identity. Using the SMART program [11], it was demonstrated that the ORF2 regions of both retroposons examined contained the fragments of reverse transcriptase domain. This domain is an essential structural functional element of all autonomous retroposons [1, 2]. Furthermore, it was shown that the ORF2 segment of retroposon R1b examined, unlike the same fragment of retroposon R1a, contained two transmembrane domains (Figs. 1b, 1c). Transmembrane domains TM1 and TM2, detected by the SMART program, were represented by relatively short sequences enriched in hydrophobic amino acids, 9 out of 20 and 10 out of 23 amino acids, respectively (in Fig. 1a, hydrophobic amino acids are shown in italics).

Analysis of amino acid sequences with the methods of bioinformatic analysis requires the use of a number of algorithms to confirm statistical significance of the data obtained. For this purpose, in this study the following approach was used. The sequence of 379 amino acids, corresponding to the ORF2 C-terminal region without the reverse transcriptase domain of retroposon R1b (designated by the letters without background in the Fig. 1a) was examined with the help of the CPHmodels 3.0 program. The essence of the algorithm of this program is the prediction of tree-dimensional structure of the peptide of interest based on the search of similarity with the proteins represented in the Protein Data Bank, i.e., with the proteins, for which three-dimensional structure was determined using...