Mol Gen Genomics (2006) 275: 288–296
DOI 10.1007/s00438-005-0079-3

ORIGINAL PAPER

Sergei Glushkov · Olga Novikova · Alexander Blinov
Victor Fet

Divergent non-LTR retrotransposon lineages from the genomes of scorpions (Arachnida: Scorpiones)

Received: 27 June 2005 / Accepted: 12 November 2005 / Published online: 3 December 2005

Abstract We screened across the taxonomic diversity of order Scorpiones (22 species belonging to 21 genera and 10 families) for the presence of seven different clades of non-LTR retrotransposons in their genomes using PCR with newly designed clade-specific consensus-degenerate hybrid oligonucleotide primers. Scorpion genomes were found to contain four known non-LTR retrotransposon clades: R1, I, Jockey, and CR1. In total, 35 fragments of reverse transcriptase genes of new elements from 22 scorpion species were obtained and analyzed for three clades, Jockey, I, and CR1. Phylogenies of different clades of elements were built using amino acid sequences inferred from 33 non-LTR retrotransposon clones. Distinct evolutionary lineages, with several major groups of the non-LTR retroelements were identified, showing significant variation. Four lineages were revealed in Jockey clade. The phylogeny of I clade showed strong support for the monophyletic origin of such group of elements in scorpions. Three separate lineages can be distingiushed in the phylogenetic tree of CR1 clade. The large fraction of the isolated elements appeared to be defective.

Keywords Retrotransposons · Non-LTR retrotransposons · Scorpions · Degenerate PCR primers

Communicated by G. Reuter

S. Glushkov · O. Novikova · A. Blinov (✉)
Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences, Prospekt Lavrentyeva 10, 630090, Novosibirsk, Russia
E-mail: blinov@bionet.nsc.ru

V. Fet
Department of Biological Sciences, Marshall University, Huntington, WV 25755-2510, USA
E-mail: fet@marshall.edu

Introduction

Non-LTR retrotransposons are mobile genetic elements that propagate themselves by reverse transcription of an RNA intermediate. The elements of this class have no long terminal repeats and utilize a simpler target-primed reverse transcription (TPRT) mechanism for their retrotransposition (Luan et al. 1993). Non-LTR retrotransposons presumably evolved from group II introns (Malik et al. 1999) that also make use of TPRT in their mobilization (Zimmerly et al. 1995).

Non-LTR retrotransposons have been found in many eukaryotes investigated up to date. The copy number of these elements may vary from just several copies per genome (some elements in *Drosophila melanogaster*; Berezikov et al. 2000) to over 800,000 copies (~20% of the genome) for L1 elements in human (International Human Genome Sequencing Consortium 2001). Several enzymatic activities can be distinguished in proteins encoded by non-LTR retrotransposons (Malik et al. 1999). The key component is reverse transcriptase that is present in all non-LTR elements. The second component is endonuclease, which is provided by restriction-enzyme-like endonuclease (REL-endo) domain in some elements and by apurinic/apyrimidinic (AP) endonuclease in other elements. The first ORF, if present, encodes a *gag*-like protein with the function of a nucleic acid chaperone (Martin and Bushman 2001). Finally, some elements also contain ribonuclease H (RNase H) domains.

Phylogenetic analysis of non-LTR retrotransposons based on the reverse transcriptase domains allowed distinguishing 15 phylogenetic clades (Malik et al. 1999; Malik and Eickbush 2000; Volff et al. 2000; Lovsin et al. 2001; Arkhipova and Morrison 2001; Burke et al. 2002). Based on structural and phylogenetic features of different elements, Malik et al. (1999) developed a scenario for the evolution of non-LTR retrotransposons and demonstrated that non-LTR elements are inherited strictly

by vertical transmission. According to this scenario, the most ancient clades of non-LTR retrotransposons (GENIE, CRE, R2, NeSL-1, and R4) contain only one ORF and show site-specific distribution in the genomes (Malik et al. 1999; Malik and Eickbush 2000), which is provided by REL-endo encoded by these elements. During further evolution of mobile elements, the REL-endo domain was substituted with an apurinic/apyrimidinic (AP) endonuclease domain acquired from the host cells. All younger clades (L1, RTE, Tad, R1, LOA, I, Jockey, CR1, Rex1, and L2) possess this AP endonuclease domain. The acquisition of the AP endonuclease resulted in losing target site specificity for all the elements (except for the R1 clade and some elements from the L1 clade), and coincided with the origin of a second ORF in front of the RT-encoding ORF. Finally, elements of some clades obtained one more enzymatic domain in the second ORF, the RNase H domain.

Mobile elements, particularly non-LTR retrotransposons, are the powerful tools for phylogenetic analysis. The integration of a non-LTR retrotransposon to a new place is an irreversible event. Non-LTR retrotransposons, once inserted in chromosomal DNA, appear to be fixed. On the other hand, insertions are insignificantly trifling into the same independent locus in unrelated lineages. One more phenomenon, horizontal transfer of non-LTR retrotransposons, is believed to be very rare; therefore, the distribution of non-LTR retrotransposons among species generally reflects their phylogenetic relationships. The most popular transposon-based marker method is the sequence-specific amplification polymorphism approach (S-SAP), also called "transposon display" (Casa et al. 2000). The S-SAP markers were developed for a wide range of taxa, in particular in plants (Casa et al. 2000; Kentner et al. 2003; Vershinin et al. 2003), insects (Zampicinini et al. 2004), and fungi (Taylor et al. 2004; Keiper et al. 2003).

Scorpions (Arachnida: Scorpiones) have been a subject of intensive, systematic, and evolutionary investigation in recent decades (Kjellesvig-Waering 1986; Polis 1990; Gantenbein et al. 1999; Fet et al. 2000; Brownell and Polis 2001; Fet and Selden 2001; Froy and Gurevitz 2003; Soleglad and Fet 2003; Coddington et al. 2004). This fact is due to many unique features of this ancient arachnid group as well as high toxicity for mammals known in some of its taxa, namely in the family Buthidae. The first comprehensive morphology-based analysis of extant scorpion relationships at the high systematic level has been published recently (Soleglad and Fet 2003). A number of molecular markers applied to scorpion phylogeny included mostly mitochondrial genes at species and genus levels (Soleglad and Fet 2003). One of our objectives was to select a new set of molecular markers for the ongoing phylogenetic analysis of various scorpion groups.

No transposable elements from scorpions have been previously identified. The main objective of this study was to screen a number of scorpion species across the taxonomic diversity of this arachnid order for the presence of non-LTR retrotransposon families, which can be used in the clarification of phylogenetic relationships. Distinct evolutionary lineages of the non-LTR retroelement sequence were identified, showing significant variation. Some of the identified groups can be useful for resolving the relationships not only among families or genera but also at species level.

Materials and methods

Species collecting and total DNA isolation

Table 1 lists all the taxa of scorpions used in the present study and their geographic origin. Taxonomy is given after Soleglad and Fet (2003), with further amendments from Soleglad et al. (2005). Live scorpions were collected in nature and preserved in 96% ethyl alcohol. Detailed label data are available from the authors. Total DNA extraction and PCR amplification have been done according to the standard techniques as described previously (Gantenbein et al. 1999; Guryev et al. 2001; Soleglad and Fet 2003).

PCR amplification

Degenerate oligonucleotides for seven selected clades of non-LTR retrotransposons were newly designed by inspection of conserved amino acid sequences in the RT domains of different published non-LTR retroelements. In total, four sense and six antisense degenerate primers were selected. Seven combinations of these primers were unique to seven selected clades of non-LTR retrotransposons: R2 clade (R2-S = TATCTTCTTCTCCnggnccng aygg and R2-A = CAATAGGCGATAAnggrtcncc ytg); R4 clade (R4-S = GTTTACATAATTGgaarkcnccngg and R4-A = TAAGCGGCGACAAnswrtcnccytg); Jockey clade (Jockey-S = AGCTCAAGCCAAAramrkcncc ngg and Jockey-A = CAAAAACACTGCCytgnggnacn cc); CR1 clade (R2-S and Jockey-A); R1 clade (R2-S and R1-A = CAGAGATCGATCCytgngkrcncc); LOA clade (LOA-S = CACTTAAAGGTTCngcnccnggnyt and LOA-A = AGGGAGATAAAGGnswnccytrngg AAGG CGATAAAACncynccytgngg); and I clade (R2-S and I-A = AAGGCGATAAAACncynccytgngg), where Y = C + T, R = A + G, K = G + T, W = A + T, S = G + C, and N = A + G + C + T.

PCR amplification was performed using 0.1 μg of genomic DNA in 10-μl volume of 10 mM Tris–HCl (pH 8.9), 1 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 200 μM each of four dNTPs, 0.5 μM primers, and 2.5 units of *Taq* polymerase. After an initial denaturation step for 3 min at 94°C, the PCR reactions were subjected to 30 cycles of amplification consisting of 30 s denaturation at 94°C, 42 s annealing at 52°C, and 1 min extension at 72°C. PCR results were assayed by agarose gel electrophoresis. PCR fragments of expected size were cloned into pBlueScript (KS +) vector using standard procedures.

Table 1 Scorpion species used in this study, their taxonomy and geographic origin, and isolated clones

Parvorder (2)	Superfamily (4)	Family (10)	Genus (21) and species (22)	Geographic locality	Jockey clones	I clones	CR1 clones
Buthida	Buthoidea	Buthidae	*Centruroides exilicauda* (Wood, 1863)	Arizona, USA		CexI [AY894771]	
			Grosphus madagascariensis (Gervais, 1843)	Madagascar			
			Liobuthus kessleri Birula, 1900	Turkmenistan	LkeJockey [AY894766]		
			Lychas mucronatus (Fabricius, 1798)	Vietnam		LmuI* [AY894770]	
			Mesobuthus gibbosus (Brullé, 1832)	Thessaly, Greece		MgiI-1* [AY894769]	
						MgiI-2 [AY894774]	
						MgiI-3 [AY894775]	
						MgiI-4 [AY894776]	
Iurida	Iuroidea	Caraboctonidae	*Hadruroides charcasus* (Karsch, 1879)	Ecuador			HchCR1 [AY894787]
			Hadrurus hirsutus (Wood, 1863)	Baja California Sur, Mexico			HhiCR1-1* [AY894779]
							HhiCR1-2 [AY894786]
		Iuridae	*Iurus dufoureius* (Brullé, 1832)	Crete, Greece			IduCR1-1[AY894783]
							IduCR1-2* [AY894784]
	Chactoidea	Chactidae	*Anuroctonus pococki* Soleglad & Fet, 2004	California, USA	ApoJockey [AY894767]		
			Brotheas granulatus Simon, 1877	Brazil	BgrJockey [AY894764]		
			Chactas reticulatus Kraepelin, 1912	Colombia	CreJockey* [AY894765]		CreCR1-1* [AY894758]
							CreCR1-2 [AY894759]
							CreCR1-3* [AY894760]
			Nullibrotheas allenii (Wood, 1863)	Baja California Sur, Mexico			NalCR1 [DQ084212]
		Euscorpiidae	*Euscorpius carpathicus* (Linnaeus, 1767)	Romania			EcaCR1 [AY894761]
		Superstitioniidae	*Superstitionia donensis* Stahnke, 1940	California, USA			SdoCR1*[AY894757]
		Vaejovidae	*Paruroctonus stahnkei* (Gertsch & Soleglad, 1966)	Arizona, USA			PstCR1 [DQ084213]
			Pseudouroctonus apacheanus (Gertsch & Soleglad, 1972)	Arizona, USA	PapJockey* [AY894763]		PapCR1* [AY894780]
			Smeringurus mesaensis (Stahnke, 1957)	California, USA			SmeCR1-1* [AY894778]
							SmeCR1-2 [AY894762]
			Vaejovis spinigerus (Wood, 1863)	Arizona, USA			VspCR1* [AY894777]
			Vaejovis puritanus Gertsch, 1958	California, USA			VpuCR1 [AY894788]
	Scorpionidea	Bothriuridae	*Bothriurus flavidus* Kraepelin, 1911	Argentina		BflI-1 [AY894772]	
						BflI-2 [AY894773]	
		Hemiscorpiidae	*Hadogenes bicolor* Purcell, 1899	South Africa	HbiJockey [AY894768]		HbiCR1-1* [AY894781]
							HbiCR1-2* [AY894782]
							HbiCR1-3 [AY894789]
		Scorpionidae	*Heterometrus spinifer* (Ehrenberg, 1828)	Southeast Asia			HspCR1 [AY894785]

The GenBank accession numbers are given in brackets and the interrupted sequences are marked with an asterisk

DNA sequencing

Clones were amplified by PCR with M13 primers, and 40 ng of the product was used in a 10 µl cycle sequencing reaction with the ABI BigDye Terminator Kit on an ABI 377 DNA sequencer. Sequences were deposited to GenBank under accession numbers AY894757–AY894789 and DQ084212, DQ084213 (Table 1).

Phylogenetic analysis

The amino acid sequences of newly identified RT and RT from GenBank database were aligned using ClustalW software (Thompson et al. 1994) and edited manually. Phylogenetic trees were generated by Neighbor-Joining (NJ) method using MEGA2 software package (Kumar et al. 2001). RT domains of *R2* and *CRE* elements were employed for rooting the trees. Statistical support for the trees was evaluated by bootstrapping (100 replications) (Felsenstein 1985).

Results and discussion

Detection of retroelement-like sequences

Our aim was to screen a large set (22 species) of scorpions from various systematic groups in order to identify the non-LTR retrotransposons. The known distribution and phylogenetic analysis of non-LTR retrotransposons based on the reverse transcriptase domains allowed expecting the presence of at least seven clades of non-LTR retrotransposons in scorpions: R2, R1, R4, LOA, Jockey, CR1, and I. All these clades contain elements known from the arthropods. However, three of those clades consist of site-specific elements, which are not useful as molecular markers in sequence-specific amplification polymorphism analysis. Elements from R2 and R1 clades are site-specific for the rRNA genes, while members of the R4 clades are site-specific for either the rRNA genes or simple repeats (Jakubczak et al. 1991; Burke et al. 1995). In addition, elements from CR1 clade are known from a wide range of animal taxa in comparison with R2, R1, LOA, Jockey, and I clades, members of which were detected mainly in insects. Thus, the possibility to discover CR1-like elements in scorpions was higher than for other groups. The degenerate oligonucleotide primers for all seven clades were designed and several species were screened. In total, we designed four sense and six antisense primers based on the sequences of known elements. Seven combinations of these primers were unique to seven selected groups of non-LTR retrotransposons from arthropods (see Materials and methods).

Twenty-two scorpion species were screened by PCR. Consistent with the spacing of the amino acid sequence domains, the primers amplified PCR product of ca. 500 bp. None of the investigated species showed the positive PCR results with primer combinations for R2, R4, and LOA groups of elements (Fig. 1). It is possible that these groups can be altogether absent from scorpion genomes, or they can be highly divergent from known elements so that our primers were not appropriate for detection.

The products with expected size were obtained for the R1, Jockey, CR1, and I groups (Fig. 1). In total, 155 clones were isolated for Jockey, CR1, and I groups; 75 of them were obtained for CR1 group, 51 for I, and 29 for Jockey. More than half of the clones were homologous to the RT. Some products were found to have the same primer at both ends but showed the presence of RT domain: they can represent rearranged copies or fragments inserted into each other.

Jockey group of elements

The primers designed for Jockey group isolation proved effective at amplifying fragments of expected size in 19 out of 22 species investigated. We selected eight species belonging to the two different parvorders and five families (Soleglad and Fet 2003): parvorder Buthida: *Liobuthus kessleri* and *Mesobuthus gibbosus* (Buthidae); parvorder Iurida: *Hadrurus hirsutus* (Caraboctonidae), *Anuroctonus pococki*, *Chactas reticulatus*, and *Brotheas granulatus* (Chactidae), *Pseudouroctonus apacheanus* (Vaejovidae), and *Hadogenes bicolor* (Hemiscorpiidae). The amplicons obtained using Jockey primers were cloned and sequenced for these species. In total, 29 nucleotide sequences were isolated, which generally included two to three clones for the each eight species investigated. After preliminary identification of reverse transcriptase fragments by comparison with sequences in the GenBank databases, it was found that two and three clones which were isolated from *M. gibbosus* and *H. hirsutus* displayed no presence of reverse transcriptase. The majority of other obtained Jockey clones contained an interrupted reverse transcriptase. However, four clones isolated from four different species held intact reverse transcriptase (Table 1). Additionally, translated products of two clones were interrupted by single stop codons and no frameshifts were detected. The BLAST search demonstrated a strong similarity of all these six sequences to the known elements from Jockey clade. Therefore, the designed Jockey primers proved to be suitable for isolation of the appropriate RT domain from scorpions.

Pairwise comparisons of the 24 Jockey-like retrotransposon fragments showed nucleotide similarity of 88.1–93.9% within species and a wide range of 50.3–72.7% between species. Pairwise comparisons of the amino acid sequences were implemented only for six Jockey-like retroelements, which were either intact or contained single stop codon as in the case of *PapJockey* from *P. apacheanus* and *CreJockey* from *C. reticulatus*.

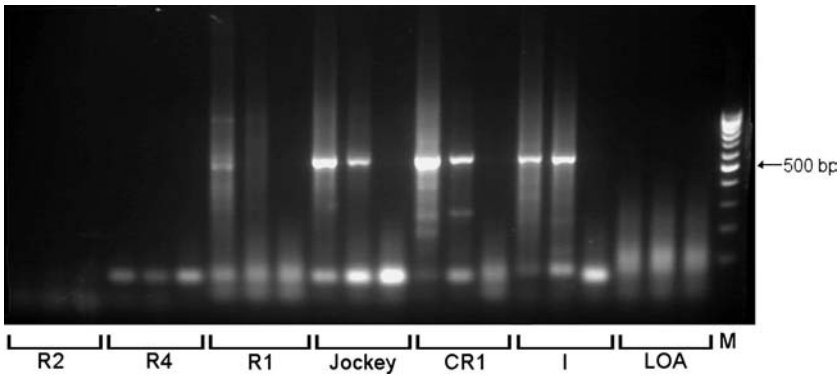

Fig. 1 Electrophoretic analysis of PCR product of seven non-LTR retrotransposons clades from two scorpion species—*Hadogenes bicolor* and *Grosphus madagascariensis*. The *third line* in each set is the negative control. Size marker (100 bp DNA Ladder, Medigen) is indicated in the *rightmost lane* of gel. The *arrowhead* indicates the position of band 500 bp in length

Amino acid similarity ranged from 83.1% between *BgrJockey* and *CreJockey* to 39.4% between *HbiJockey* and *LkeJockey*.

More detailed analysis of the elements was performed by NJ analysis of six new amino acid Jockey-like sequences (see Table 1) and the non-scorpion sequences from the GenBank database (accession numbers shown on the tree) as shown in Fig. 2. Elements *R2* from *D. melanogaster* and *CRE1* from *Crithidia fasciculata* were used as the outgroup. NJ phylogeny demonstrated 59% bootstrap support for the monophyletic origin of the Jockey-like elements from the arthropods including newly isolated elements but showed high diversity among the scorpion elements. Two major clusters of Jockey-like scorpion elements can be distinguished with relatively high bootstrap supports at 85 and 69%.

One well-supported cluster consists of three elements with more than 80% similarity isolated from scorpion species belonging to the same family, Chactidae (*ApoJockey* from *A. pococki*, *CreJockey* from *C. reticulatus*, and *BgrJockey* from *B. granulatus*) and one element from *Bombyx mori*, which displayed more than 70% similarity with *ApoJockey*, *CreJockey*, and *BgrJockey* elements (71.5, 73.6, and 76.5%, respectively).

The second group included the remaining three elements from scorpions (*LkeJockey*, *HbiJockey*, and *PapJockey*) and known elements from invertebrates. It was not expected that elements from different scorpion species would show higher similarity with known elements from different insects rather than with other scorpion elements. However, the similarity between *LkeJockey* and retroelement from *B. mori* was 62.6% while its similarities with *HbiJockey* and *PapJockey* were only 39.4 and 42.3%, respectively. The same situation was observed with *HbiJockey* and *PapJockey*: the similarity between *HbiJockey* and element from the tick *Amblyomma americanum* is 52%, and between *PapJockey* and element from the mosquito *Anopheles arabiensis* is 54.5%, whereas similarity between *HbiJockey* and *PapJockey* was only 43.4%. The relationships among these three different lineages of elements remain unresolved.

Thus, we for the first time identified diverse Jockey-like elements from various groups of scorpions. The

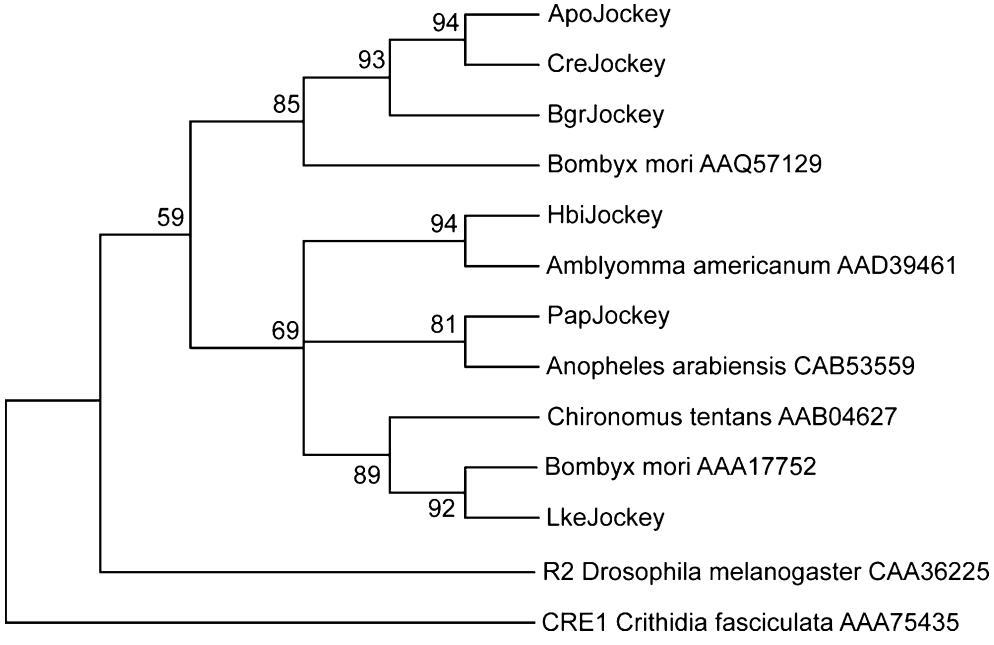

Fig. 2 Phylogenetic tree based on the RT domain of Jockey-like retrotransposons showing the position of the newly identified elements from scorpions: *ApoJockey*; *CreJockey*; *BgrJockey*; *HbiJockey*; *PapJockey*; and *LkeJockey*. Percentages of bootstrap support, from 100 replications, are indicated for branches with $>50\%$ support. The accession numbers of RT protein sequences of known elements are given following the species names

similarity and phylogenetic analyses revealed a high divergence in the Jockey-like retroelements among scorpion families, with at least four divergent lineages. Scorpions are the most ancient arthropod lineage in which the representatives of Jockey-like elements are found. In addition, the Jockey clade includes subclades of elements common for both scorpions and insects rather than being specific for these arthropod groups. We could assume therefore that the origin not only of Jockey clade but also of its subclades predates the evolution of insects, and was present already in the common arthropod ancestor of insects and chelicerates. Earlier we studied Jockey clade elements in detail in Chironomidae (Insecta: Diptera). These data, in addition to the results of the present study, indicate that the Jockey clade has a relatively recent origin compared to other retroposons since it contains not only a large number of groups but also a large number of intact elements.

I group of elements

Following the PCR analysis, in which the product of expected 500 bp in length was obtained for 20 species, amplicons were cloned and sequenced from 9 scorpion species belonging to the 2 parvorders and 7 families: parvorder Buthida: *Centruroides exilicauda, Lychas mucronatus*, and *Mesobuthus gibbosus* (Buthidae); parvorder Iurida: *Iurus dufoureius* (Iuridae), *Brotheas granulatus* (Chactidae), *Euscorpius carpathicus* (Euscorpiidae), *Heterometrus spinifer* (Scorpionidae), *Paruroctonus stahnkei* (Vaejovidae), and *Bothriurus flavidus* (Bothriuridae). Altogether, 51 clones were isolated as potentially containing reverse transcriptase gene sequence, from 3 to 7 clones for the each of 9 species analyzed. After sequencing and comparison with GenBank database, only 31 of the sequences were found to be homologous to the RT domain of non-LTR retroelements. None from three clones isolated from *I. dufoureius*, five clones from *E. carpathicus*, and four clones from *H. spinifer* contained reverse transcriptase. Among the clones obtained from *B. granulatus* and *P. stahnkei*, some did not carry any sequences similar to RT (four clones from each species), and some showed clear similarity to RT (two and three clones, correspondingly). Unfortunately, the translation products of the RT-inclusive clones from *B. granulatus* and *P. stahnkei* contained frameshifts and many stop mutations. They were obtained from defective elements and could not be included into the phylogenetic analysis. Most of RT-containing clones from the remaining 4 species were also interrupted, in total 19 clones. However, several isolated clones were intact: three clones *MgiI-2*, *MgiI-3*, and *MgiI-4* from *M. gibbosus*, two clones *BflI-1* and *BflI-2* from *B. flavidus*, and *CexI* from *C. exilicauda* showed non-defective translation product (Table 1). Additionally, translation of clones *LmuI* from *L. mucronatus* and *MgiI-1* each carried only one stop codon and no frameshifts.

For the scorpion I-like elements analyzed, intraspecific nucleotide similarity ranged from 42.5 to 95.4% in 26 sequences, while interspecific similarity ranged from 34.2 to 51.6%. Pairwise comparisons of the amino acid sequences within species showed an average similarity of 84.3% among *MgiI-1*, *MgiI-2*, *MgiI-3*, and *MgiI-4* elements, and 76% between elements *BflI-1* and *BflI-2*. Interspecific similarity ranged from 44.8% between *MgiI-2* and *LmuI* to 63.1% between *MgiI-2* and *BflI-2*.

The phylogenetic tree resulting from the NJ analysis of I-like elements sequences from scorpions and known elements from GenBank database showed strong support for a monophyletic origin of lineage I-like elements from scorpions, with bootstrap value 84% (Fig. 3). Two clusters could be recognized within isolated elements. The first included the elements from scorpions which belong to the parvorder Buthida: *MgiI-2, MgiI-3*, and *MgiI-4* elements isolated from *M. gibbosus*, and *CexI* elements from *C. exilicauda*. The bootstrap support for

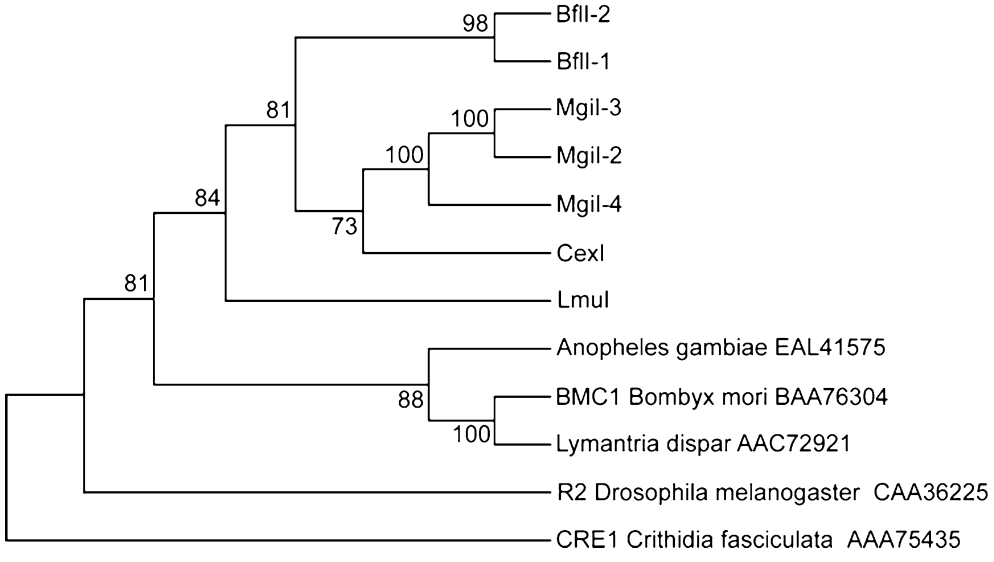

Fig. 3 Phylogenetic tree based on the RT domain of I-like retrotransposons showing the position of the newly identified elements from scorpions: *BflI-1, 2*; *MgiI-2, 3, 4*; *CexI*; and *LmuI*. Percentages of bootstrap support, from 100 replications, are indicated for branches with $>50\%$ support. The accession numbers of RT protein sequences of known elements are given following the species names

separation of this cluster was 73%. The second cluster consisted of elements *BflI-1* and *BflI-2* from *B. flavidus* which belongs to another parvorder, Iurida. This expected result reflects the basic dichotomy between these two parvorders and the relatively ancient (possibly Permian-Triassic) event of clear separation between these two major scorpion parvorders. Independent evolution of Buthida and Iurida agrees firmly with all earlier morphological and molecular studies (Soleglad and Fet 2003). Finally, *LmuI* element was placed by the NJ analysis not into the first or second cluster but represented a separate clade. This element could belong to a different lineage of I-like retrotransposons from scorpions. On the other hand, the divergence of *LmuI* element most likely is a result of the inactivation and degradation of this element. As opposed to Jockey clade, I-like elements from scorpions did group along the phylogeny of their host organisms, so we can suggest that only one group of elements exists within this clade. This evolutionary lineage of retrotransposons most likely is nearing the end of its existence, which is confirmed by the absence of intact elements in it, and by our unsuccessful attempts to find these elements in many tested scorpion species.

CR1 group of elements

Originally, we expected the presence of the CR1 group in scorpions more than for other groups of elements. Therefore, we implemented the broadest analysis for this group.

The band of expected size was obtained for all species investigated during the PCR analysis with primers designed for CR1 group. Products were cloned and sequenced for 22 species. The total number of potentially sequenced RT-carrying clones was 75, from 2 to 5 clones for each species. In total, 35 of the isolated clones showed clear similarity with RT sequences from GenBank. As for the Jockey- and I-like RT-containing clones, the majority of CR1-like isolated elements revealed defective translated products (24 clones). Nevertheless, several intact RT sequences were obtained and analyzed from 11 species. Moreover, some of the RT sequences were found to contain only one stop codon and no frameshifts. Such sequences were included in further analysis together with intact RTs (Table 1).

The pairwise comparative analysis of nucleotide sequences showed both high heterogeneity within and between species, the average interspecific similarity of 37.4%, and the similarity within species ranging from 92.7 to 42.6%.

In phylogenetic analysis, we used 16 amino acid CR1-like RT sequences isolated from scorpions and known elements from GenBank. The preliminary NJ analysis showed strong support of monophyletic origin of the scorpion CR1-like elements (bootstrap 93%) and the presence of three separate clusters (data not shown). We added known CR1 clade members to the phylogenetic analysis. In the resulting tree (Fig. 4), two out of three clusters of scorpion CR1-like elements grouped with the elements from other organisms.

It was unexpected to detect a common group of CR1 retrotransposons from scorpions and vertebrates (*Gallus gallus* and *Platemys spixii*; Haas et al. 1997; Kajikawa et al. 1997). Eight scorpion elements from four families of Iurida belonged to this first cluster: *NalCR1*, *VpuCR1*, *IduCR1-1* and *IduCR1-2*, *SmeCR1, HhiCR1-1* and *HhiCR1-2*, and *PapCR1* (Table 1; Fig. 3b). The relationships among these elements remain unresolved because of low bootstrap values for nodes within cluster I. It is interesting that three scorpion species (*V. puritanus*, *S. mesaensis*, and *P. apacheanus*) belonged to the same family, Vaejovidae; however, the divergence among the elements from these three species was higher than that between *V. puritanus* and *I. dufoureius* (Iuridae). The similarity among elements from Vaejovidae ranged from 61.2 to 65.7%, while similarity between *VpuCR1* and *IduCR1-2* was 68.4%. The retroelements from cluster I could not be subdivided into the subgroups or families because of very high degree of divergence.

The more expected was the cluster III of newly identified scorpion elements *CreCR1-1, CreCR1-2, CreCR1-3*, *PstCR1*, *EcaCR1*, and *SdoCR1*. Among the known retrotransposons, the closest elements were *T1* and *Q* from *Anopheles gambiae* and *Sam3* from *Caenorhabditis elegans* (Besansky 1990; Besansky et al. 1994; Marin et al. 1998). The elements from cluster III showed a high level of divergence as did the elements from cluster I. At the same time, we can separate two subgroups: *CreCR1-1*, *CreCR1-2*, *CreCR1-3*, and *PstCR1* (bootstrap support 98%; similarity 63.1%); *EcaCR1* and *SdoCR1* (bootstrap support 92%; similarity 69.2%). Four species, from which retroelements of the cluster III were isolated, are members of different families of scorpions: Chactidae, Euscorpiidae, Vaejovidae, and Superstitioniidae, representing all four families of superfamily Chactoidea (Soleglad and Fet 2003).

An additional cluster of isolated CR1-like elements consisted of four sequences from three species: *VspCR1*, *HspCR1*, *HbiCR1-1*, and *HbiCR1-2* (cluster II on the phylogenetic tree, Fig. 4). NJ analysis puts cluster II together with the cluster III with a weak bootstrap support (only 52%). Nevertheless, we can suggest that cluster II of elements is more closely related to the cluster III than to the cluster I.

The elements isolated with the CR1 clade primers demonstrated a high degree of divergence; it is likely that several different lineages of CR1-like retroelements are present in the scorpion genomes. At least three major groups of elements can be distinguished by the NJ analysis. The elements from the scorpions which belonged to the same family or genus could be found in different clusters. For example, elements *VpuCR1* and *VspCR1* from two species belonging to the genus *Vaejovis* were found in the clusters I and II (Fig. 4). It is interesting to note that none of the RT-like sequences

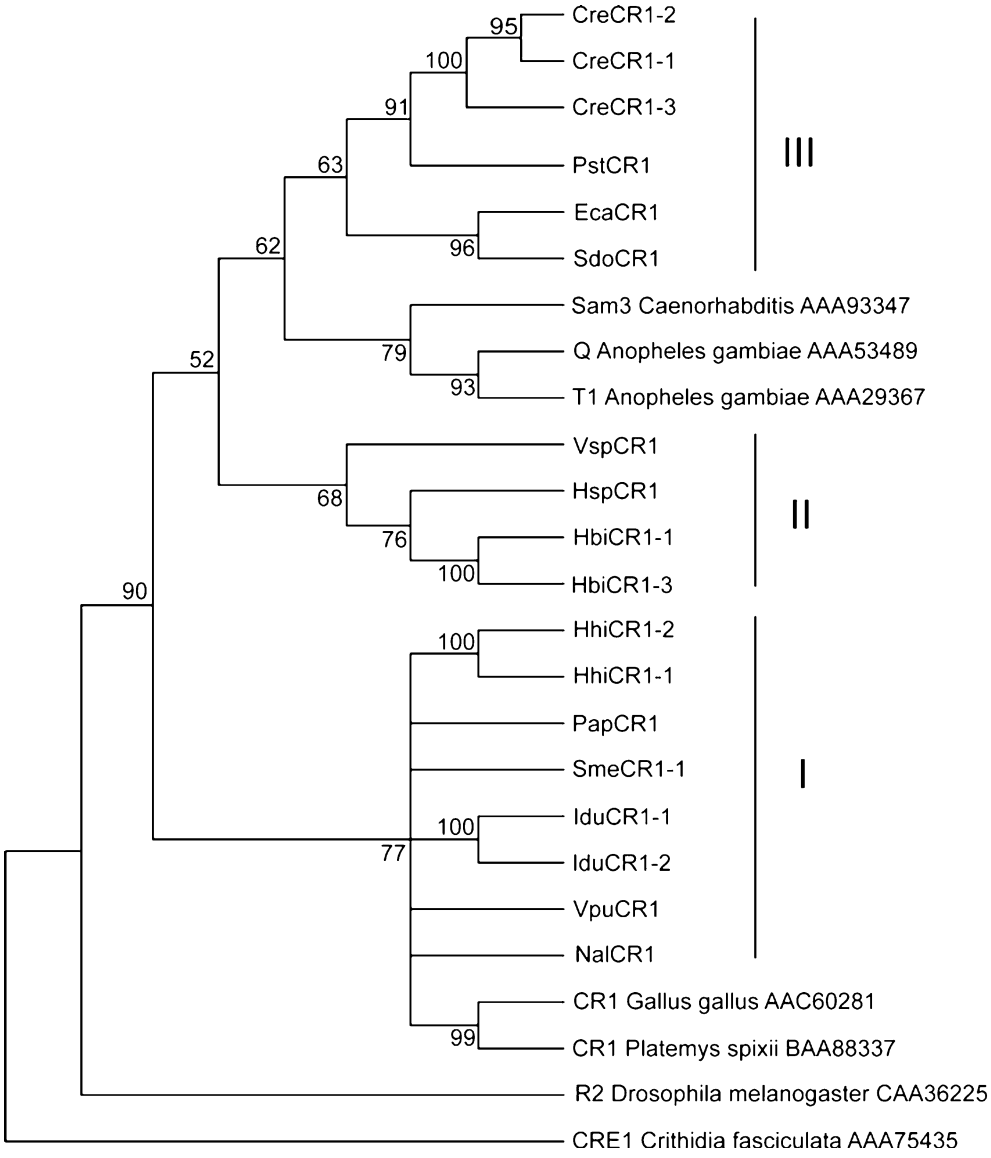

Fig. 4 Phylogenetic tree based on the RT domain of CR1-like retrotransposons showing the position of the newly identified elements from scorpions. Percentages of bootstrap support, from 100 replications, are indicated for branches with $>50\%$ support. The accession numbers of RT protein sequences of known elements are given following the species names

from CR1 clade were isolated from scorpions belonging to the parvorder Buthida. It is possible therefore that these elements were independently lost in Buthida after its divergence from the common scorpion ancestor.

Conclusions

We analyzed three divergent clades of non-LTR retrotransposon elements isolated from scorpions. The screening procedure was efficient and more than 50% of the cloned sequences included retrotransposons. The large fraction of the elements appeared to be defective (68 clones) which is in accordance with known correlation of active and inactive elements from other organisms (Penzkofer et al. 2005; Sassaman et al. 1997). In total, we isolated 35 relatively intact RT sequences.

The similarity and phylogenetic analyses demonstrated the presence of highly divergent lineages of retrotransposons in scorpions. In each examined clade, several lineages could be distinguished with the exception of I clade in which all isolated elements are monophyletic and compose a single group. It is intriguing that one group of isolated elements from scorpions turned out to be closer to the elements from vertebrates than to the elements from arthropods (cluster I of CR1-like elements; Fig. 4). The non-LTR retrotransposon elements should be further investigated as a potential rich system of phylogenetic markers.

Acknowledgments This work was supported in part by the state contract 10002-251/П-25/155-270/200404-082 and Siberian Branch of the Russian Academy of Sciences (project No. 10.4). The work has been conducted in compliance with the current laws governing genetic experimentation in the USA and Russia. For their help in obtaining the scorpion specimens for this study we thank Luis Acosta, Hoang Ngoc Anh, Joe Bigelow, Daniil Blinov, Galina Fet, Benjamin Gantenbein, Alexander Gromov, Wilson Lourenço, Graeme Lowe, Valentin Popa, Svetlana Shcherbik, Michael E. Soleglad, and Iasmi Stathi.

References

Arkhipova IR, Morrison HG (2001) Three retrotransposon families in the genome of *Giardia lamblia*: two telomeric, one dead. Proc Natl Acad Sci USA 98:14497-14502

Berezikov E, Bucheton A, Busseau I (2000) A search for reverse transcriptase-coding sequences reveals new non-LTR retrotransposons in the genome of Drosophila melanogaster. Genome Biol 1:RESEARCH0012

Besansky NJ (1990) Evolution of the T1 retroposon family in the *Anopheles gambiae* complex. Mol Biol Evol 7:229–246

Besansky NJ, Bedell JA, Mukabayire O (1994) Q: a new retrotransposon from the mosquito *Anopheles gambiae*. Insect Mol Biol 3:49–56

Brownell PH, Polis GA (2001) Scorpion biology and research. Oxford University Press, New York

Burke WD, Malik HS, Rich SM, Eickbush TH (2002) Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. Mol Biol Evol 19:619–630

Burke WD, Muller F, Eickbush TH (1995) R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. Nucleic Acids Res 23:4628–4634

Casa AM, Brouwer C, Nagel A, Wang L, Zhang Q, Kresovich S, Wessler SR (2000) Inaugural article: the MITE family heartbreaker (Hbr): molecular markers in maize. Proc Natl Acad Sci USA 97:10083–10089

Coddington JA, Giribet G, Harvey MS, Prendini L, Walter DE (2004) Arachnida. In: Cracraft J, Donoghue MJ (eds) Assembling the tree of life. Oxford University Press, New York, pp 296–318

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fet V, Selden PA (eds) (2001) Scorpions 2001. In Memoriam Gary A. Polis. British Arachnological Society, Burnham Beeches, Bucks

Fet V, Sissom WD, Lowe G, Braunwalder ME (2000) Catalog of the scorpions of the world (1758–1998). New York Entomological Society, New York

Froy O, Gurevitz M (2003) New insight on scorpion divergence inferred from comparative analysis of toxin structure, pharmacology and distribution. Toxicon 42:549–555

Gantenbein B, Fet V, Largiadèr C, Scholl A (1999) First DNA phylogeny of the genus *Euscorpius* Thorell 1876 (Scorpiones, Euscorpiidae) and its bearing on the taxonomy and biogeography of this genus. Biogeographica (Paris) 75:59–72

Guryev V, Makarevitch I, Blinov A, Martin J (2001) Phylogeny of the genus *Chironomus* (Diptera) inferred from DNA sequences of mitochondrial cytochrome b and cytochrome oxidase I. Mol Phylogenet Evol 19:9–21

Haas NB, Grabowski JM, Sivitz AB, Burch JB (1997) Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. Gene 197:305–309

International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Jakubczak JL, Burke WD, Eickbush TH (1991) Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc Natl Acad Sci USA 88:3295–3299

Kajikawa M, Ohshima K, Okada N (1997) Determination of the entire sequence of turtle CR1: the first open reading frame of the turtle CR1 element encodes a protein with a novel zinc finger motif. Mol Biol Evol 14:1206–1217

Keiper FJ, Hayden MJ, Park RF, Wellings CR (2003) Molecular genetic variability of Australian isolates of five cereal rust pathogens. Mycol Res 107:545–556

Kentner EK, Arnold ML, Wessler SR (2003) Characterization of high-copy-number retrotransposons from the large genomes of the louisiana iris species and their use as molecular markers. Genetics 164:685–697

Kjellesvig-Waering EN (1986) A restudy of the fossil scorpionida of the world. Palaeontographica Americana, 55. Paleontological Research Institution, Ithaca

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1255

Lovsin N, Gubensek F, Kordis D (2001) Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. Mol Biol Evol 18:2213–2224

Luan DD, Korman MH, Jakubczak JL, Eickbush TH (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell 72:595–605

Malik HS, Burke WD, Eickbush TH (1999) The age and evolution of non-LTR retrotransposable elements. Mol Biol Evol 16:793–805

Malik HS, Eickbush TH (2000) NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans*. Genetics 154:193–203

Marin I, Plata-Rengifo P, Labrador M, Fontdevila A (1998) Evolutionary relationships among the members of an ancient class of non-LTR retrotransposons found in the nematode *Caenorhabditis elegans*. Mol Biol Evol 15:1390–1402

Martin SL, Bushman FD (2001) Nucleic acid chaperone activity of the ORF1 protein from the mouse LINE-1 retrotransposon. Mol Cell Biol 21:467–475

Penzkofer T, Dandekar T, Zemojtel T (2005) L1Base: from functional annotation to prediction of active LINE-1 elements. Nucleic Acids Res 33:D498–500

Polis GA (ed) (1990) Biology of scorpions. Stanford University Press, Stanford

Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas TP, DeBerardinis RJ, Gabriel A, Swergold GD, Kazazian HH Jr (1997) Many human L1 elements are capable of retrotransposition. Nat Genet 16:37–43

Soleglad ME, Fet V (2003) High-level systematics and phylogeny of the extant scorpions (Scorpiones: Orthosterni). Euscorpius 11:1–175

Soleglad ME, Fet V, Kovařík F (2005) The systematic position of the scorpion genera *Heteroscorpion* Birula, 1903 and *Urodacus* Peters, 1861 (Scorpiones: Scorpionoidea). Euscorpius 20:1–30

Soleglad ME, Sissom WD (2001) Phylogeny of the family Euscorpiidae Laurie, 1896: a major revision. In: Fet V, Selden PA (eds) Scorpions 2001. In memoriam Gary. A. Polis Burnham Beeches, Bucks: British Arachnological Society, pp 25–111

Taylor EJ, Konstantinova P, Leigh F, Bates JA, Lee D (2004) Gypsy-like retrotransposons in Pyrenophora: an abundant and informative class of molecular markers. Genome 47:519–525

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Vershinin AV, Allnutt TR, Knox MR, Ambrose MJ, Ellis TH (2003) Transposable elements reveal the impact of introgression, rather than transposition, in *Pisum* diversity, evolution, and domestication. Mol Biol Evol 20:2067–2075

Volff J-N, Korting C, Schartl M (2000) Multiple lineages of the non-LTR retrotransposon Rex1 with varying success in invading fish genomes. Mol Biol Evol 17:1673–1684

Zampicinini G, Blinov A, Cervella P, Guryev V, Sella G (2004) Insertional polymorphism of a non-LTR mobile element (NLRCth1) in European populations of *Chironomus riparius* (Diptera, Chironomidae) as detected by transposon insertion display. Genome 47:1154–1163

Zimmerly S, Guo H, Perlman PS, Lambowitz AM (1995) Group II intron mobility occurs by target DNA-primed reverse transcription. Cell 82:545–554