CMLS, Cell. Mol. Life Sci. 60 (2003) 803–809
1420-682X/03/040803-07
DOI 10.1007/s00018-003-2329-z

CMLS Cellular and Molecular Life Sciences

Research Article

The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1-like elements

R. Albalat [a], J. Permanyer [a], C. Cañestro [a,b], A. Martínez-Mir [a,c], O. Gonzàlez-Angulo [a] and R. Gonzàlez-Duarte [a,*]

[a] Departament de Genètica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal, 645, 08028 Barcelona (Spain), Fax + 34 934110969, e-mail: rgonzalez@ub.edu
[b] Present address: Institute of Neuroscience, 1254 University of Oregon, 1425 E. 13th Avenue, Eugene, Oregon 97403 (USA)
[c] Present address: Department of Dermatology, Columbia University, College of Physicians & Surgeons, 630 West 168th Street VC 15-204, New York, New York 10032 (USA)

Received 25 November 2002; received after revision 14 February 2003; accepted 5 March 2003

Abstract. BfCR1 is the first non-long terminal repeat retrotransposon to be characterised in the amphioxus genome. Sequence alignment of the predicted translation product reveals that BfCR1 belongs to the CR1-like retroposon class, a family widely distributed in vertebrate and invertebrate lineages. Structural analysis shows conservation of the specific motifs of the ORF2-CR1 elements: the N-terminal endonuclease, the reverse transcriptase and the C-terminal domains. The BfCR1 element possesses an atypical 3′ terminus consisting of the tandem repeat $(AAG)_6$. We gathered evidence supporting the mobility of this element and report an estimated 15 copies of BfCR1, mostly truncated, per haploid genome, a remarkably low number when compared to that of vertebrates. Phylogenetic analysis, including the amphioxus element, seems to indicate that (i) CR1-like retroposons cluster in a monophyletic group and (ii) the CR1-like family was already present in the chordate ancestor. Our data provide further support for the horizontal transmission of CR1-like elements during early vertebrate evolution.

Key words. Chordate; CR1-like retrotransposon; genome evolution; horizontal transmission; mobile element.

Introduction

The genome organisation of the cephalochordate amphioxus remains largely uncharacterised. Little is, therefore, known of the type and abundance of transposable elements, gene density and architecture [1–3] and base composition [1], all of which are key features in understanding the evolution of chordates and explaining the transition to vertebrate genomes.

One of the most abundant classes of transposable elements are the non-long terminal repeat (non-LTR) retrotransposons (also termed LINE-like elements or retroposons). They are widely distributed in eukaryotes, although the current collection of non-LTR elements is clearly biased towards insects [4]. Evolutionary relationships based on the alignment of the reverse transcriptase domain, and the study of sequence organisation and the transposition mechanism have suggested that these non-LTR retroposons are closely related to bacterial and organellar group II introns [5]. Eleven clades of non-LTR retrotransposons have been described, all of which ap-

* Corresponding author.

peared before the divergence of the major animal phyla. Of these, L1, CR1, and RTE retrotransposons constitute the three major lineages present in vertebrates [4]. Full-length CR1-type elements contain two protein-coding regions, ORF1 and ORF2, which encode a nucleic acid-binding protein and an endonuclease and reverse transcriptase (RT), respectively [6–8]. Members of the CR1-like family have been identified in the vertebrates chicken (CR1) [6, 7], turtle (PsCR1) [8], fugu (Maui) [9], and shark (HER1) [10], and in the invertebrates *Anopheles* (T1 and Q) [11, 12], *Caenorhabditis* (Sam6 and Sam3) [13, 14] and *Schistosoma* (SR1) [15]. The copy number of CR1-type elements per genome varies greatly depending on the species, and on the segment used for estimation, since only a very small percentage are full length. Most elements are truncated in their 5′ regions, as the level of processivity of the RT determines the proportion copied from the RNA template at the target integration site.

Very little is known about repetitive DNA sequences in the amphioxus genome [16]. Indeed, not a single member of a transposon family has been described in cephalochordates, the sister group of vertebrates. This contrasts with the extensive knowledge that has been built up of many individual amphioxus genes, each of which is known to correspond to several vertebrate counterparts in a number of gene families, and upon which the idea that vertebrate genomes evolved through two rounds of genome duplication is based [17]. The lack of information on transposable elements in these animals has restrained the reconstruction of the invertebrate/vertebrate genome transition. Hence, the fact that a 3.2-kb DNA sequence with high homology to the CR1-type elements was found in the *Branchiostoma floridae* presenilin gene [1] became an issue of interest. We characterised this sequence, which we have named BfCR1, estimated the copy number per haploid genome, performed a multiple sequence alignment of the deduced translated ORF, and carried out a phylogenetic analysis with its vertebrate and invertebrate counterparts.

Material and methods

Sequence analysis and construction of phylogenetic trees

For sequence comparisons against public databases, WU-BLAST-X 2.0+BEAUTY, TBLASTN 2.1.2 and BLOCKS search software were used. Sequence data were obtained from GenBank. The accession numbers of the sequences used in the present analysis were as follows: *Anopheles gambiae* T1 (M93689), *A. gambiae* Q (U03849), *B. floridae* BfCR1 (AF369890), *Caenorhabditis elegans* Sam3 (U46668), *C. elegans* Sam6 (Z82275), *Drosophila melanogaster* Doc (X17551), *Fugu rubripes* Maui (AF086712), *Gallus gallus* CR1 (U88211), *Platemys spixii* PsCR1 (AB005891), *Schistosoma mansoni* SR1 (U66331), *Scyliorhinus torazame* HER1(AB027737).

Multiple sequence alignment and neighbour-joining (NJ) tree construction were performed with Clustal X [18] and drawn with the TreeViewPPC program [19]. Confidence in each node was assessed by 1000 bootstrap replicates. Maximum likelihood (ML) analysis was performed using the quartet sampling and the NJ parameter estimation procedure of Tree Puzzle, with the JTT model of amino acid substitution [20], estimating the amino acid frequencies from the data set, and taking into account the among-site rate heterogeneity with eight gamma-distributed categories.

Genomic library screening, Southern and slot blot analyses

A *B. floridae* FIX II genomic library [21] was screened with a 1445-bp DNA fragment of BfCR1 (from Leu539 to Glu1020 of the deduced protein). The probe was labelled with [α-^{32}P]dCTP by random-hexamer priming and hybridised to phage DNA transferred on Amersham nylon (Hybon-N) filters in duplicate. Approximately 210,000 phage clones were screened. Hybridisations were carried out in phosphate-SDS solution [22] at 60 °C overnight. 4 × 15 min washes were performed at 60 °C in 2 × SSC, 0.1 % SDS. Hybridisation signals were detected by autoradiography. Only the signals present in the original and duplicated filters were considered.

Total genomic DNA from single *B. floridae* animals was isolated using the guanidine isothiocyanate method [23] with minor modifications [24]. For Southern studies, 10 µg of genomic DNA was digested with *Pst*I, resolved in 0.8% agarose gels, transferred to an Amersham nylon (Hybon-N) filter, and UV fixed. Membranes were hybridised with the same DNA probe as that used for library screening. Labelling, hybridisation and washing conditions were the same as those described for genomic library screening. The Southern blot was rehybridised with a 5′ cDNA fragment (exons 1 and 2) of *BfPS* as described elsewhere [1].

For slot blot analysis, 5 µg, 500 ng, 50 ng and 5 ng of total genomic *B. floridae* DNA (≈ 8×10^6, 8×10^5, 8×10^4 and 8×10^3 haploid genomes, respectively) *Xba*I restricted, and 100 ng, 10 ng, 1 ng, 100 pg, 10 pg and 1 pg of phage DNA containing the BfCR1 element (from 2×10^9 to 2×10^4 copies of BfCR1, respectively) were denatured with 0.4 M NaOH and 25 mM EDTA in a final volume of 200 µl. The samples were applied to an Amersham nylon (Hybon-N) membrane using a slot blot device. Prior to sample loading, the membrane was soaked in H_2O, and after loading, neutralized with 2 M sodium acetate, pH 5.4 and fixed with UV light. Membranes were hybridised with the same 1445-bp DNA fragment as used for library screening. Additionally, a second probe of

461 bp encompassing Pro43 to Leu145 of the amphioxus ORF2 predicted protein was used. Labelling, hybridisation and washing conditions were also the same. Quantification of the slot blot signal was performed with the GS525 Molecular Imager System from Bio-Rad.

Results and discussion

Similarities between *BfCR*1 and *CR*1-like elements

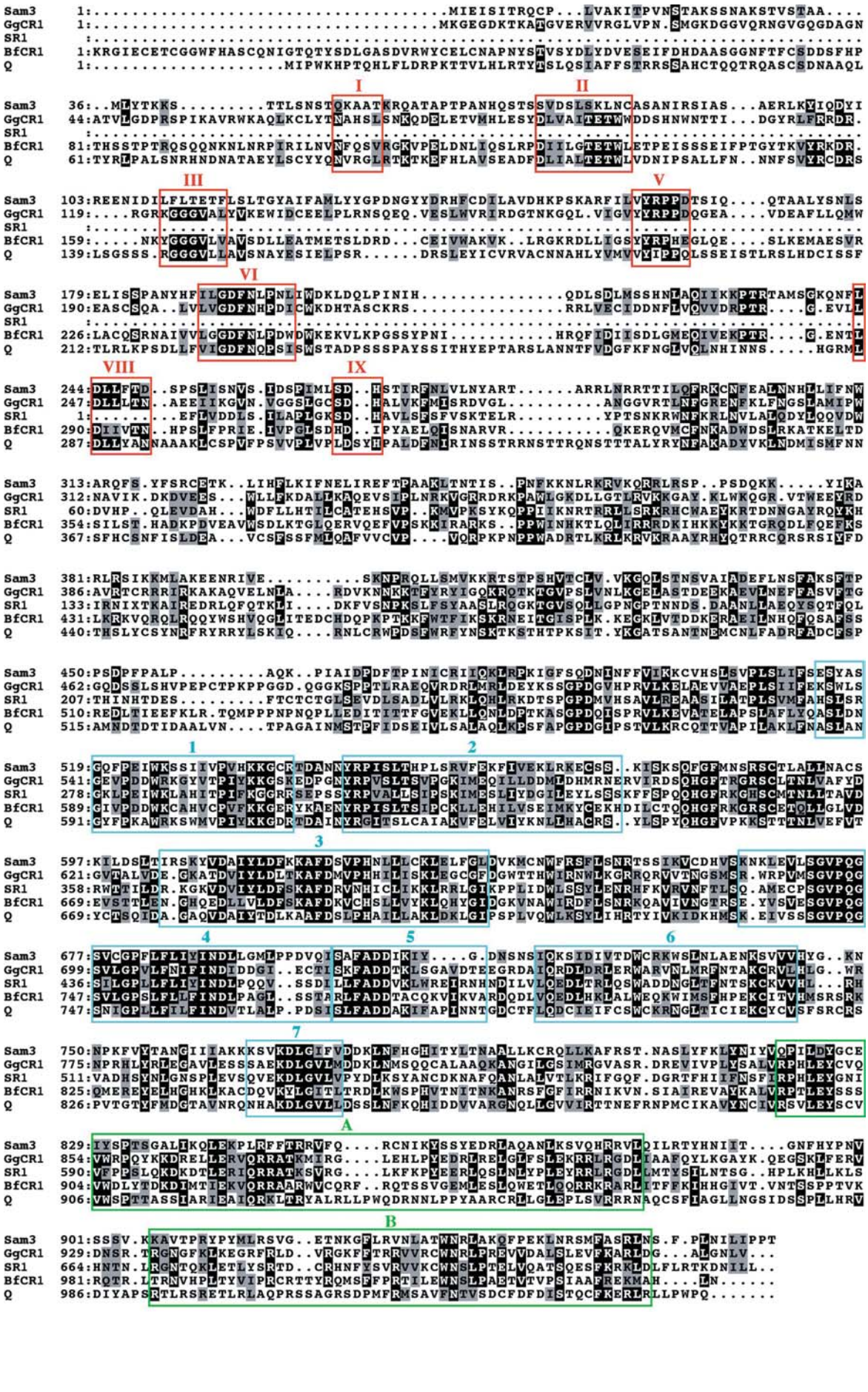
A large fraction of repetitive sequences are associated with mobile or transposable elements, which due to their ability to move within genomes can generate mutations and influence genomic organisation and gene expression. Here, we describe the first non-LTR retrotransposon, BfCR1, in the cephalochordate subphylum. The 3275-bp sequence covered almost the full-length ORF2 of non-LTR retrotransposons (fig. 1) and encoded an EN/RT protein (BLOCK search program gave an E value: $4.9e^{-18}$ for RT domains). Comparison of the predicted amino acid sequence of ORF2-BfCR1 with the translated nucleotide databases showed that the greatest similarity was to be found with the retrotransposon family CR1 (E value: e^{-121} for *P. spixii* and e^{-111} for *G. gallus* CR1 elements). Further analysis was performed on the specific motifs of the ORF2-CR1 elements, the N-terminal endonuclease domain (amino acids 106–313 of BfCR1), the RT domain (amino acids 584–850) and the C-terminal domain (amino acids 895–1037) of unknown function. The endonuclease is thought to be responsible for nicking the target DNA for retrotransposition. Alignment of the BfCR1 N-terminal endonuclease revealed homology to the reported I, II, III, V, VI, VIII and IX conserved domains [25], which suggested conservation of the endonucleolytic activity. In the RT-BfCR1 domain, the seven conserved blocks reported by Xiong and Eickbush [26] were clearly identified. This region was further used for drawing the phylogenetic relationships among CR1-like elements. Concerning the A and B segments of the C-terminal domain, BfCR1 showed conservation of region A (40% identity over 67 amino acids when compared with chicken CR1) but a weak degree of preservation of region B (26% identity over 49 amino acids). These figures were within the range for non-vertebrate versus vertebrate comparisons, but differed from the high sequence similarity that both segments show in vertebrate versus *S. mansoni* CR1 comparisons [15]. Finally, another feature of CR1-like members is an 8-bp tandem direct repeat in the 3′ untranslated region (UTR). Instead, in our case, the 3′ UTR of BfCR1 showed an AAG triplet repeated six times. Therefore, the 3′ end of BfCR1 differed in sequence and in length from other members of the CR1 clade.

Evidence supporting the mobility of BfCR1 can be derived from the fact that this element was not always present in the presenilin gene. First, some of the *BfPS* bands clearly did not match with any of the BfCR1 bands after the Southern analysis (compare the high-molecular-weight bands in lanes 1, 3 and 6 in the right panel with those in the left panel of fig. 2C). Second, PCR amplification of genomic DNA produced a fragment containing *BfPS* intron 1 without the BfCR1 element (data not shown). Sequence comparison of the PCR product with the previously characterized *BfPS* intron 1 [1] enabled us to ascertain the limits of the characterised element.

Copy number of *BfCR1* in the amphioxus genome

The BfCR1 copy number per haploid genome was estimated using three independent approaches. First, quantitative slot blot analysis was used to compare the intensity of slot blot hybridisation for BfCR1 in genomic DNA samples with those of serial dilutions of a single-copy BfCR1-containing phage. This indicated that there were approximately 11 copies of BfCR1 per haploid genome (fig. 2A). To verify BfCR1 representation further, we performed a library screening on seven 30,000-phage plates and obtained an average of 15 positives per plate (fig. 2B). Considering an average insert size of 15 kb, and an amphioxus genome length of 5.8×10^8 bp, the figure obtained was 20 BfCR1 elements. Finally, the discrete banding pattern (less than 20 bands) revealed by Southern analysis with DNA samples from single animals agreed with the above frequencies (fig. 2C). Considering the three approaches together, an average estimate of partial plus full-length BfCR1 elements is 15 copies per haploid genome. However, when a probe corresponding to a 5′ segment of the ORF2-CR1 was used, the number of elements containing this region decreased to three to four per haploid genome. Thus, the genome of amphioxus has 15 copies of the BfCR1 element which extend 1.5 kb from the 3′ terminus but only 3–4 copies which extend as far as 3.1 kb.

This value is quite low if compared with other CR1-like elements: for example, 30,000 truncated and 30 full-length in chicken [6], 10,000 and 400 in turtle [8], 3000 and 240 in fugu [9], and 200 and 50 in *S. mansoni* [15], but within the range of other non-LTR retrotransposons described in the tunicate *Ciona intestinalis* (50 Cili-1 and Cili-2) [27]. The low abundance of retroposons in tunicates and cephalochordates could arguably be related to their relatively small genome, and the vertebrate ancestor may also have shared this trait. However, vertebrate genomes not only differ in size, they also show other specific features: large heterogeneity in base composition, great variations in gene density, higher gene number, and complex exon-intron architecture. And among these features, they might harbour the defence mechanisms (i.e. methylation) to control the damaging effects associated with these elements. However, the analysis of a large segment of the *C. intestinalis* genome showed that methylation is not necessarily targeted to transposable elements [28].


Sam3
GgCR1
SR1
BfCR1
Q
I
II
III
V
VI
VIII
IX
1
2
3
4
5
6
7
A
B

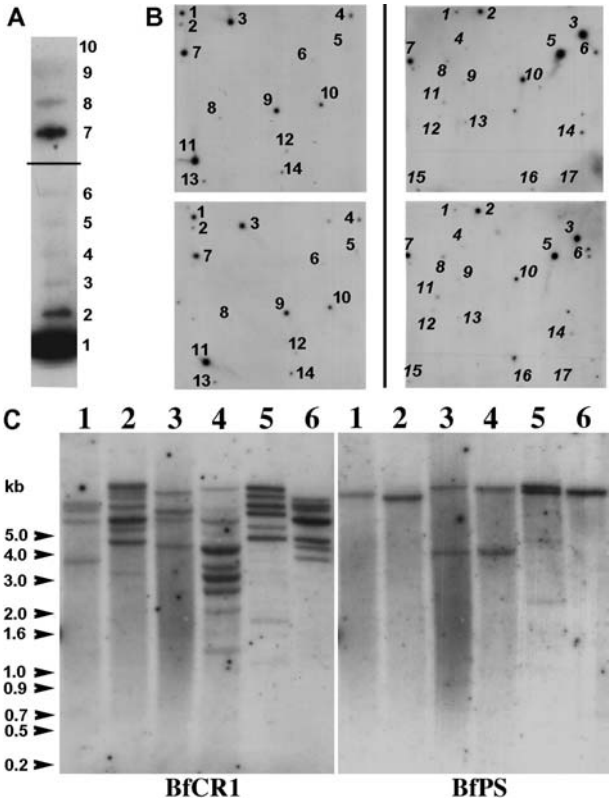


Figure 2. (*A*) Slot blot analysis of BfCR1 number in the amphioxus genome. Slots 1–6: decreasing quantities of a single phage containing the BfCR1 element. Slots 7–10: 5 µg, 500 ng, 50 ng and 5 ng of total genomic *B. floridae* DNA digested with *Xba*I. (*B*) Two representative examples (left and right panels) of the hybridisation of a *B. floridae* FIX II genomic library screened with the BfCR1 element. Positive signals have been numbered in the two original filters (top) and their duplicates (bottom). (*C*) Southern analysis of genomic DNA from single *B. floridae* animals, *Pst*I restricted and probed with BfCR1 (left) and *BfPS* (right).

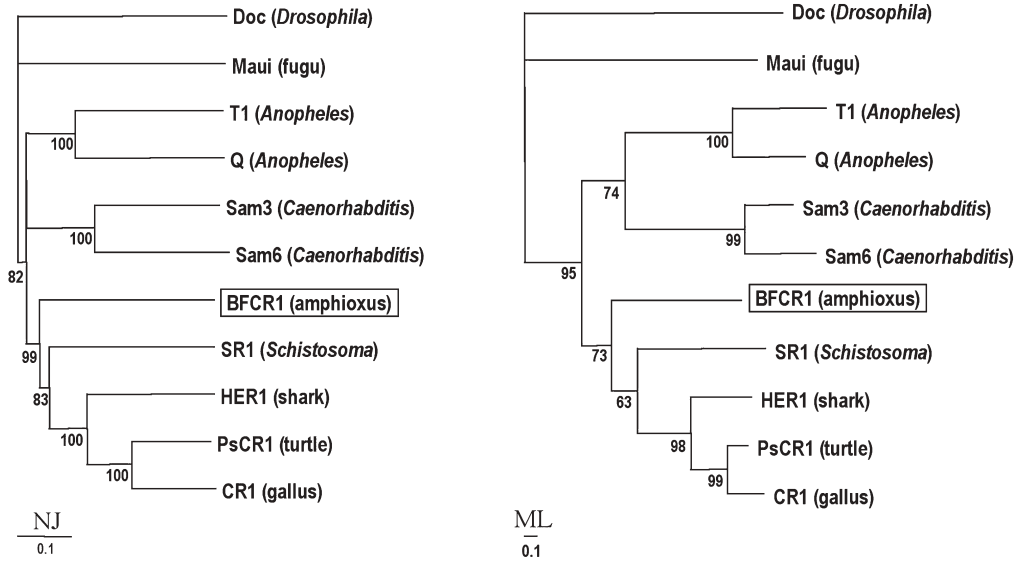


Figure 3. Neighbour-joining and maximum-likelihood phylogenetic trees from the alignment of the RT and C-terminal domains of deduced proteins of the CR1-like retroposon family. Figures at nodes are scores from 1000 bootstrap resamplings of the data (NJ) or quartet puzzling support values (ML). Both methods produced the same overall tree topology.

Figure 1. Alignment of deduced protein sequences of members of the CR1-like retroposon family. N-terminal endonuclease domains I, II, III, V, VI, VIII and IX (boxed in red), 1–7 conserved blocks found in all RTs (boxed in blue) and A and B segments in the C-terminal region (boxed in green) are indicated. Amino acid conservation and similarity are shown in black and grey backgrounds, respectively. Sam3 (Caenorhabditis), GgCR1 (Gallus), SR1 (schistosoma), BfCR1 (amphioxus, Q (Anopheles)

Phylogenetic relationships

Sequence analysis of the 3275-bp of BfCR1 revealed that the closest match was to be found with the CR1 family, showing 43% similarity to and 31% identity with the ORF2 deduced amino acid sequence of chicken CR1. These values increased when the segment being compared was restricted to the RT or the C-terminal conserved domains. Phylogenetic analyses based on the predicted protein of the RT-encoding domain and the C-terminal region of ORF2 were performed using the NJ and ML methods (fig. 3). The trees were rooted using *D. melanogaster* Doc as the outgroup, since the insect element belongs to the Jockey family, which is considered the sister group of CR1. The fact that BfCR1 did not cluster within the vertebrate CR1-like clade, as revealed by high bootstrap values (100 and 98 NJ and ML, respectively), suggests that the CR1-like elements were already present in the genome of the ancestral chordate, and strengthens the hypothesis that non-LTR elements originated in the Precambrian era [4, 27]. Unexpectedly, our analysis indicated that the Maui retrotransposon from the fugu fish failed to cluster with either the vertebrate or with the invertebrate CR1-like elements. Recently, the Maui element, initially described as a member of the CR1 family [9], was included in the newly defined LINE2 clade [29].

The phylogenetic position of the non-LTR retroelement of *S. mansoni* (SR1) within the deuterostome CR1 cluster (bootstrap values of 99 and 73 for NJ and ML, respectively) was surprising. Furthermore, SR1 appeared to be more closely related to vertebrate CR1-like elements than was BfCR1 (bootstrap values of 83 and 63 for NJ and ML, respectively). This close phylogenetic relationship was further supported by the high degree of conservation of the A and B domains of the SR1 C-terminal region, defined as highly preserved segments in vertebrate (but not in invertebrate) CR1-like retroposons. Although horizontal transfer has been considered for the SR1 mobile element [15], no reliable evidence was obtained after a comprehensive phylogenetic analysis of non-LTR retrotransposons [4]. However, when BfCR1 was incorporated into this analysis, the SR1 position with respect to the vertebrate elements supported the proposed horizontal transmission, facilitated by the fact that schistosomes are vertebrate blood vessel parasites. The SR1 branch at the base of the vertebrate cluster, the variability reported among SR1 elements [15] and our phylogenetic analysis suggest that the transfer was an early event which probably occurred after the cephalochordate-vertebrate split.

Acknowledgements. We wish to thank J. Garcia-Fernàndez and C. Minguillon for providing the *B. floridae* genomic library, G. Marfany for helpful discussions and R. Rycroft for revising the English text. This study was supported by grants from DGICYT (Ministerio de Educación y Cultura, Spain, BMC2000-0536), and FPI fellowships from the CIRIT (Generalitat de Catalunya) and from Universitat de Barcelona to C. C. and J. P., respectively.

1 Martínez-Mir A., Cañestro C., Gonzàlez-Duarte R. and Albalat R. (2001) Characterization of the amphioxus presenilin gene in a high gene-density genomic region illustrates duplication during the vertebrate lineage. Gene **279:** 157–164
2 Abi-Rached L., Gilles A., Shiina T., Pontarotti P. and Inoko H. (2002) Evidence of en bloc duplication in vertebrate genomes. Nat. Genet. **31:** 100–105
3 Cañestro C., Albalat R., Hjelmqvist L., Godoy L., Jörnvall H. and Gonzàlez-Duarte R. (2002) Ascidian and amphioxus *Adh* genes correlate functional and molecular features of the ADH family expansion during vertebrate evolution. J. Mol. Evol. **54:** 81–89
4 Malik H. S., Burke W. D. and Eickbush T. H. (1999) The age and evolution of non-LTR retrotransposable elements. Mol. Biol. Evol. **16:** 793–805
5 Curcio M. J. and Belfort M. (1996) Retrohoming: cDNA-mediated mobility of group II introns requires a catalytic RNA. Cell **84:** 9–12
6 Burch J. B., Davis D. L. and Haas N. B. (1993) Chicken repeat 1 elements contain a pol-like open reading frame and belong to the non-long terminal repeat class of retrotransposons. Proc. Natl. Acad. Sci. USA **90:** 8199–8203
7 Haas N. B., Grabowski J. M., Sivitz A. B. and Burch J. B. (1997) Chicken repeat 1 (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames. Gene **19:** 305–309
8 Kajikawa M., Ohshima K. and 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
9 Poulter R., Butler M. and Ormandy J. (1999) A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. Gene **227:** 169–179
10 Ogiwara I., Miya M., Ohshima K. and Okada N. (1999) Retropositional parasitism of SINEs on LINEs: identification of SINEs and LINEs in elasmobranchs. Mol. Biol. Evol. **16:** 1238–1250
11 Besansky N. J. (1990) A retrotransposable element from the mosquito *Anopheles gambiae*. Mol. Cell. Biol. **10:** 863–871
12 Besansky N. J., Bedell J. A. and Mukabayire O. (1994) Q: a new retrotransposon from the mosquito *Anopheles gambiae*. Insect Mol. Biol. **3:** 49–56
13 The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science **282:** 2012-2018
14 Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Bonfield J. et al. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans.* Nature **368:** 32–38
15 Drew A.C. and Brindley P.J. (1997) A retrotransposon of the non-long terminal repeat class from the human blood fluke *Schistosoma mansoni*: Similarities to the chicken-repeat-1-like elements of vertebrates. Mol. Biol. Evol. **14:** 602–610
16 Cañestro C., Gonzàlez-Duarte R. and Albalat R. (2002) Minisatellite instability at the *Adh* locus reveals somatic polymorphism in amphioxus. Nucleic Acids Res. **30:** 2871–2876
17 Holland P.W. (1996) Molecular biology of lancelets: insights into development and evolution. Israel J. Zool. **42:** 247–272
18 Thompson J. D., Gibson T. J., Plewniak F., Jeanmougin F. and Higgins D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25:** 4876–4882
19 Page R. D. M. (1996) An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. **12:** 357–358
20 Jones D. T., Taylor W. R. and Thornton J. M. (1992) The rapid generation of mutation data matrices from protein sequences. CABIOS **8:** 275–282

21 Ferrier D. E. K., Minguillón C., Holland P. W. H. and Garcia-Fernàndez J. (2000) The amphioxus Hox cluster: deuterostome flexibility and Hox14. Evol. Dev. **2:** 284–293
22 Church G. M. and Gilbert W. (1984) Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:** 1991–1995
23 Chirgwin J. M., Przybyla A. E., MacDonald R. J. and Rutter W. J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18:** 5294–5299
24 Garcia-Fernàndez J., Baguñà J. and Saló E. (1993) Genomic organization and expression of the planarian homeobox genes *Dth-1* and *Dth-2*. Development **118:** 241–253
25 Martín F., Olivares M. and López M.C. (1996) Do non-long terminal repeat retrotransposons have nuclease activity? Trends Biochem. Sci. **21:** 283–285
26 Xiong Y. and Eickbush T. H. (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:** 3353–3362
27 Simmen M. W. and Bird A. (2000) Sequence analysis of transposable elements in the sea squirt, *Ciona intestinalis*. Mol. Biol. Evol. **17:** 1685–1694
28 Simmen M. W., Leitgeb S., Charlton J., Jones S. J. M., Harris B. H., Clark V. H. et al. (1999) Nonmethylated transposable elements and methylated genes in a chordate genome. Science **283:** 1164–1167
29 Lovsin N., Gubensek F. and Kordi D. (2001) Evolutionary dynamics in a novel L2 clade of non-LTR retrotransposons in Deuterostomia. Mol. Biol. Evol. **18:** 2213–2224