

Research paper

A SINE in the genome of the cephalochordate amphioxus is an Alu element**Linda Z. Holland**

Marine Biology Research Division, Scripps Institution of Oceanography, University of California

Corresponding address: Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA USA 92093-0202. Tel: 01 858-534-5607; Fax 01 858-534-7313. E-mail: lzholland@ucsd.edu

Received: 2006.02.14; Accepted: 2006.02.21; Published: 2006.04.10

Transposable elements of about 300 bp, termed "short interspersed nucleotide elements or SINEs are common in eukaryotes. However, Alu elements, SINEs containing restriction sites for the AluI enzyme, have been known only from primates. Here I report the first SINE found in the genome of the cephalochordate, amphioxus. It is an Alu element of 375 bp that does not share substantial identity with any genomic sequences in vertebrates. It was identified because it was located in the FoxD regulatory region in a cosmid derived from one individual, but absent from the two FoxD alleles of BACs from a second individual. However, searches of sequences of BACs and genomic traces from this second individual gave an estimate of 50-100 copies in the amphioxus genome. The finding of an Alu element in amphioxus raises the question of whether Alu elements in amphioxus and primates arose by convergent evolution or by inheritance from a common ancestor. Genome-wide analyses of transposable elements in amphioxus and other chordates such as tunicates, agnathans and cartilaginous fishes could well provide the answer.

Key words: Branchiostoma, transposon, cis-regulation, SINE, Alu, chordate evolution

1. Introduction

Eukaryotic genomes contain two main classes of transposable elements. The Class I elements include the long terminal repeat (LTR) retrotransposons, the non-LTR retrotransposons and the short interspersed nucleotide elements (SINEs) [1, 2]. The Class II elements have terminal inverted repeats with transposase-binding sites, and include the *P* elements found in *Drosophila*. As implied by the name, SINEs are the shortest elements. They are about 300 bp long and usually contain an RNA polymerase III promoter. The 3' end typically has a polyA stretch of varying length. The best-known SINEs are the Alu elements of primates, which are thought to have a common origin with the B1 elements of rodents [1, 3, 4]. Alu elements are the most abundant transposable elements in the human genome, being present in about 500,000 copies. These retroposons consist of two related monomers in tandem and, as the name implies, contain a restriction site for the enzyme AluI. They have no open reading frames, and the factors necessary for their amplification are thought to come from long interspersed elements (LINEs) which code for functional reverse transcriptases with an endonuclease domain [2, 4].

The present work describes the first SINE detected in the genome of the cephalochordate amphioxus. Previous studies of transposable elements in amphioxus are limited to a Class I non-LTR retrotransposon (BfCR1) [5] and a class II, non autonomous transposable element (ATE-1) [6]. Amphioxus (*Branchiostoma*) is an aquatic invertebrate with a much simpler body plan than vertebrates (e.g.

no paired eyes, ears or limbs). Moreover, the amphioxus genome appears to have a similar organization as that of vertebrates but lacks the extensive gene duplications that occurred in the vertebrate lineage [7]. Traditional phylogenetic analyses have placed amphioxus as the sister group of vertebrates. However, recent analyses done with large gene sets have placed amphioxus basal in the chordates with the rapidly evolving tunicates as the sister group of vertebrates [8, 9]. Amphioxus is, therefore, the most appropriate organism for comparison with vertebrates to understand how the vertebrates evolved from invertebrate chordate ancestors. The surprising finding of the present work is not that amphioxus has transposable elements or that this element is a SINE, which are common in eukaryotic genomes, but that it contains three AluI sites, two of which are in tandem. Alu elements have previously been found only in primates [4]. Thus, the finding of an Alu element in amphioxus raises the question whether amphioxus and primates are an example of convergent evolution or whether they both arose from an Alu element in the ancestral chordate.

2. Methods

The transposable element was located by blastn searches of the amphioxus (*Branchiostoma floridae*) genome sequences in the Trace Archives of GenBank with portions of the upstream regulatory region of the FoxD gene (AF512537), which we had sequenced from cosmid MPMGc117O0129 available from the RZPD (<http://www.rzpd.de>). Sequences that were present in this cosmid, but absent from at least one of the

FoxD alleles in the Trace Archives, were suspected transposable elements. Additional copies of this transposable element were located by blastn searches of amphioxus BAC clones deposited in GenBank. BAC clones are available from the BACPAC resources at <http://bacpac.chori.org/>.

3. Results and discussion

We have previously identified a notochord-specific enhancer in the FoxD gene in amphioxus (*Branchiostoma floridae*) [10]. This enhancer encompassed 4.7 kb upstream of the ATG start codon and has several binding sites for the notochord marker *brachyury* as well as for Fox D itself, suggesting autoregulation. Subsequently, we determined that the 5' portion of this enhancer, which includes these binding sites, is essential for directing expression to the notochord [J.-K. Yu, L.Z. Holland, unpubl.] However, just downstream of the FoxD binding sites are two sites for FoxA2 (HNF3 β), which, like *brachyury*, is expressed in the notochord (Figure 1a). Therefore, it seemed likely that this region with the FoxA2 sites is also involved in directing notochordal expression. Surprisingly, blast searches of the amphioxus genome sequences in the Trace Archives of GenBank, which represent the two alleles of a single individual, revealed that both alleles lacked the region with the FoxA2 binding sites that corresponds to -1168 to -1543 upstream of the FoxD ATG start codon in the clone we previously sequenced, which was from a different individual (Figure 2). Moreover, we found this 375 bp region in numerous traces with flanking regions that did not correspond to the FoxD gene. An estimate would be 50-100 copies in the genome of the individual sequenced. Thus, this 375 bp sequence is clearly a SINE. Figure 1b compares this SINE with eight representative homologous SINES found in the Trace Archives and in cosmid sequences, all of which are from the two alleles of the same individual. The first two of these eight SINES are located in cosmids containing respectively, the Tbx15/18/21 gene and the Pax1/9 gene. In the former, the SINE is 4200 bp upstream of the ATG start codon of Tbx15/18/21 and in the latter it is 40 kb upstream of the 5' end of Pax 1/9 and located within an intron of a hypothetical gene coding for a cyclic nucleotide-gated cation channel. The remaining SINES in Figure 1 were contained in relatively short BAC end sequences and genomic traces, and therefore, their location relative to particular genes cannot be determined until the full genome sequence is available.

The amphioxus SINE has several characteristic features of SINES in general and of primate Alu elements in particular, and I, therefore, term it the Amphi-Alu element. SINES generally are about 300 bp long with a polyA tail of varying length and have similar, but not identical, left and right halves [11-13]. The insertion site is usually flanked by short direct repeats. SINES typically have a region derived from tRNA that contains a consensus RNA polymerase promoter [14]. The amphioxus element is 375 bp long,

the polyA tail ranges from 1-13 bases (Figure 1b), and the right and left halves are 52% identical (Clustal W alignment). Most of the copies of this element have from one to three AluI sites, although they can be mutated (Figure 1b). Moreover, alignments with tRNA sequences from sea urchin available at <http://lowelab.ucsc.edu/GtRNAdb/> reveal a high percentage of identities with a sea urchin tRNA for glutamine with an imperfect match for the RNA polymerase promoter "A" site and a nearly perfect one for the "B" site (Figure 3). As is typical for SINES, deletions of portions of either the left arm or the right arm are common. One example of each is shown in Figure 1b. At the 5' end of the Alu element is a conserved repeat sequence GAATGAAT. Comparison with the alleles of FoxD lacking the Alu element (Figure 2) show that these repeats are within the Alu element and not in the flanking region. Such repeats are lacking at the 3' end of the Alu element and in the adjacent 3' flanking sequence. Figure 2 shows the insertion site in the FoxD gene for an allele lacking Amphi-Alu. The potential insertion site is between 2 "T" residues and is flanked by direct repeats of CTTTGT. In the clones containing the Amphi-Alu element, there are direct repeats in the flanking regions of some clones (e.g. GTTTTATTG/GTTTCTTG for FoxD). However, direct repeats are not always present on either side of Amphi-Alu, raising the possibility, which can be tested when the full genome assembly is available, that some Amphi-Alu SINES may be longer than the one in the AmphiFoxD allele shown in Figure 1 [10].

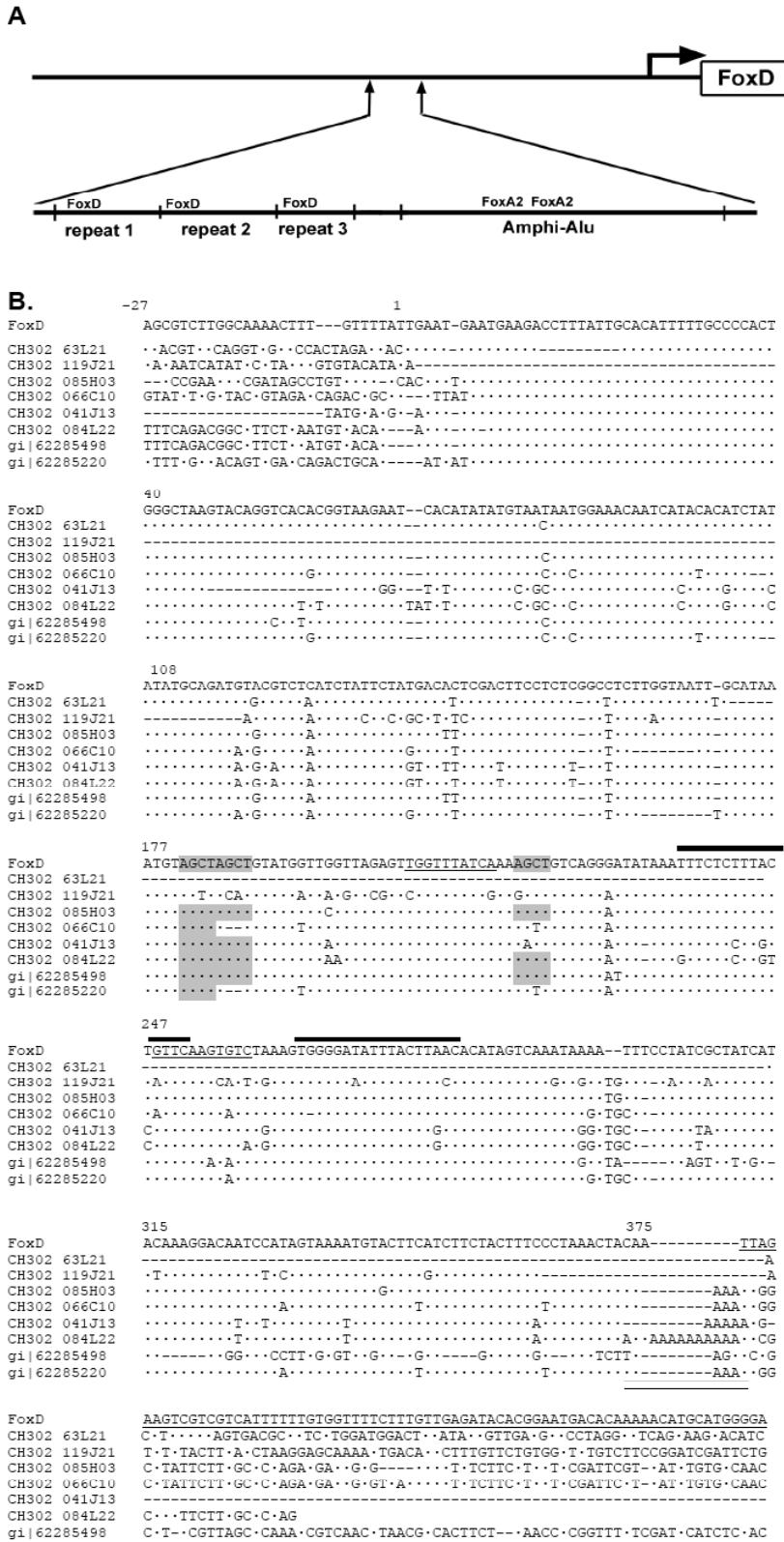
One unusual feature of Amphi-Alu is that the region with the highest identity to a tRNA is from base 199 to base 183, rather than at the 5' end [13] as is typical for SINES. Although it is possible that Figure 1 depicts Amphi-Alu in the reverse orientation, this is not likely since a varying number of "A" residues were present in all clones at the end designated as the 3' end in Figure 1b. Moreover, comparisons of the reverse strand of Amphi-Alu with tRNAs from a range of species, including sea urchin, had few matches. This raises the possibility that the Amphi-Alu SINE is a chimeric element, but it is puzzling that in either orientation, the putative RNA polymerase III sites are not near the ends of the element.

Figure 1. The Amphi-Alu SINE from the FoxD gene in an individual of *Branchiostoma floridae*. A. Schematic diagram of the upstream regulatory region showing the FoxD repeat region in cosmid MPMGc117O0129, which we have shown is essential for directing expression of FoxD to the notochord just upstream of the Amphi-Alu element, which is inserted into the FoxD gene at base -1168 upstream of the ATG start codon. This element contains two binding sites for FoxA2 which is also expressed in the notochord, but as it is not present in two other FoxD alleles, it cannot be essential for notochord expression. B. Representative sequence variations of the Amphi-Alu SINE from the FoxD gene in cosmid MPMGc117O0129 and from six BAC clones and two genomic traces from the trace archive database (<http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?>). All

eight sequences are from the same individual. The Amphi-Alu SINE in BAC CH302 63L21 is located in the presumed upstream regulatory region of the Tbx15/18/21 gene and that from CH302 119J21 within an intron of a hypothetical gene coding for a cyclic nucleotide gated cation channel. Since the other sequences are either from BAC end sequencing or shotgun sequencing, it is not possible as yet to determine if they are within regulatory regions of genes. The Amphi-Alu sequence is shown in the reverse orientation of its insertion into the FoxD locus to orient the polyA stretches as polyA tails (double-underlining). The 5' end of the Alu element is shown as 1, the Alu sites are shaded in grey, two regions corresponding to the binding sites for RNA polymerase on the tRNA-like region are underlined and the binding sites for FoxA2 are shown by thick lines over the sequence

Alu elements have previously been described only in primates [15], raising the question of whether Amphi-Alu and primate Alu elements are descended from a common ancestral Alu, or whether the presence of Alu sites in these two SINEs from primates and amphioxus represent convergent evolution. It has been argued that the FLA (free left arm) Alu family arose at the origin of mammals about 112 mya [15]. However, alignments of human Alu elements and Amphi-Alu do not reveal a high level of identity; in general, Amphi-Alu is not as GC rich as human Alu elements, and the most AT-rich region is not central as in human Alu elements but starts at base 290. Moreover, human Alus and the related B1 elements of rodents have a 7SL RNA-related region [3, 13, 16], while Amphi-Alu is related to a tRNA for asparagine. Thus, since amphioxus and vertebrate lineages are estimated to have split about 520 mya, it seems more likely that the Alu sites in the amphioxus and mammalian Alu elements have arisen through convergent evolution. However, there are few descriptions transposable elements in basal vertebrates such as agnathans [17] or in tunicates [18], which recent phylogenetic analyses place as the sister group to vertebrates, amphioxus being basal in the chordates [8, 9]. Alignments of a tunicate SINE (Cics-1), which lacks AluI sites, with Amphi-Alu

reveal only a few short regions of identity [18]. A



thorough study of transposable elements in amphioxus and these animals is critical for understanding the evolutionary history of Alu elements.

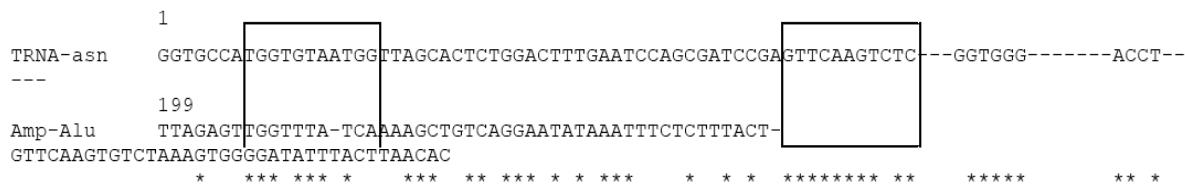
In the three instances where complete cosmid or BAC sequences were available, the Amphi-Alu elements were located in likely regulatory DNA either upstream of the ATG start codon (*FoxD* and clone CH302 63L21) or within an intron (CH302 119J21). Human Alu elements within introns have been implicated in alternative splicing and exonization whereby part of the SINE is not spliced out of the mRNA and can cause frame shifts and gene inactivation [2, 19]. Moreover, whether in the upstream regulatory region or in an intron, SINEs may provide binding sites for transcription factors and thus influence gene regulation. The Amphi-Alu

element has probable binding sites for FoxA2, which is normally expressed in the endoderm and notochord of amphioxus [20]. Its location in the *FoxD* gene just downstream of an important enhancer (Figure 1a), suggests that it might be contributing to regulation of the *FoxD* gene in the individuals that have it in the *FoxD* regulatory region. Alu elements have been implicated in regulation of many genes [21, 22]. To what extent such transposable elements might mediate evolution of cis-regulation in animals in nature is not known. However, it is certainly a possible mechanism for effecting large changes in cis-regulation.

Figure 2. The AmphiFoxD genomic region lacking the Amphi-Alu SINE from a sequence in the Trace Archives of GenBank (gnl|ti|545126576 name:AFSA504495.g2). The site corresponding to the insertion site of the Amphi-Alu SINE in the *FoxD* gene in cosmid MPMGc117O0129 is shown by the arrowhead. Repeated sequenced flanking the insertion site are underlined.

```
GGAGCCCATGGGTCGCCCTCCTCCACCCAAACACATCAGTAAACATGCATAAAAGATGGGACACTGAAATATGAA
CCCTAACAGTCATCTCAGGGAAAAAAACGGGACAAATCGGAGCACATGCGGAGTTAAGTAAACTGGAAAGAGGGGAG
GGTCCTTAAGAAAACAAGCAGCACAAGAAACCTCCTCTGTGAAGCAAAGACAAGGTGGTGTATGGTGAAGTGTG
TAAGCACACGGGAAACAGCTTAGCTTGCCGTGCCAGATCACGTCTGCCGAACCCGACCCGGGGTGGAGCAGGG
GGAGGTACGGACAGCTGCCGCCCCCTCCCTCTCCTCACGTCCCAGACGCGCAGCGGAAGGTGGTGGCTTGAAG
AAATAGGTTGTTACTTCCTCCCATCTCAACAAACAACAATCGATCATCGTAAACATTTGTTCTGGTTGA
GATACTGTTGATTAGCCGGTTGGCAAACACCCCGTGTGGTACACACACACACACATACACACAGCGTCTTGGC
AAAACTTGTTTATTTAGAAGTCGTCGTCATTTGTTGGTTCTTGTGAGATAACCGAATGACACAAAACA
TGCACGGGGAGAGGTTGAAATCGTATAACACAAAAAGAAAGTGAATTTACACAAAGTCAATGGAAAGTAACAAGTAA
AAATAACTAAACATGCCATCTGATGCCATCGATGCAACG
```

Figure 3. Alignment of the sea urchin (*Strongylocentrotus purpuratus*) tRNA for asparagine and the corresponding region of the Amphi-Alu element. The regions with high identity to the A and B binding sites for RNA polymerase III are boxed.



In summary, the finding of an Alu element in the amphioxus genome that is present in the regulatory DNA of some, but not all, alleles of the *FoxD* gene and in the regulatory regions of other genes raises questions both about the evolution of SINEs in chordates and suggests that the insertion of such SINES into regulatory DNA could mediate evolution of cis-regulatory regions. There is some evidence for positive selection for such transposon-mediated changes in regulatory DNA [21, 22], and it may become evident with the sequencing of the genomes of more organisms like amphioxus with relatively short life cycles are sequenced to what extent such changes in regulatory DNA mediate large evolutionary changes.

Acknowledgements

This research was supported by grants to L.Z. H. from the National Science Foundation of the USA

numbers IOB 0416292, IOB 0236171 and IOB 1111/0208138.

Conflict of interests

The author has declared that no conflict of interest exists.

References

- Kazazian HH. Mobile elements: drivers of genome evolution. *Science* 2004; 303:1626-1632.
- Dewannieux M, Heidmann T. LINEs, SINEs and processed pseudogenes: parasitic strategies for genome remodeling. *Cytogenetic Genome Res* 2005; 110(1-4):35-48.
- Quentin Y. A master sequence related to a free left Alu monomer (FLAM) at the origin of the B1 family in rodent genomes. *Nucleic Acids Res* 1994; 22:2222-2227.
- Jurka J. Evolutionary impact of human Alu repetitive elements. *Curr Opin Genet Dev* 2004; 14(6):603-608.
- Albalat R, Permanyer J, Cañestro C et al. The first non-LTR retrotransposon characterised in the cephalochordate amphioxus, BfCR1, shows similarities to CR1-like elements. *Cell Mol Life Sci* 2003; 60:803-809.

6. Cañestro C, Albalat R, González-Duarte R. Isolation and characterization of the first non-autonomous transposable element in amphioxus, ATE-1. *Gene* 2003; 318:69-73.
7. Minguillon C, Gardenyes J, Serra E et al. No more than 14: the end of the amphioxus Hox cluster. *Int J Biol Sci* 2005; 1:19-23.
8. Blair JE, Hedges SB. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol* 2005; 22(11):2275-2284.
9. Philippe H, Lartillot N, Brinkmann H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* 2005; 22(5):1246-1253.
10. Yu J-K, Holland ND, Holland LZ. Tissue-specific expression of *FoxD* reporter constructs in amphioxus embryos. *Dev Biol* 2004; 274(2):452-461.
11. Kazazian HH. Mobile elements: Drivers of genome evolution. *Science* 2004; 303:1626-1632.
12. Jurka J, Zuckerkandl E. Free left arms as precursor molecules in the evolution of Alu sequences. *J Mol Evol* 1991; 33(1):49-56.
13. Batzer MA, Deininger PL. Alu repeats and human genomic diversity. *Nat Rev Genet* 2002; 3(5):370-379.
14. Smit AFA, Riggs AD. MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res* 1995; 23:98-102.
15. Kapitonov V, Jurkal J. The age of Alu subfamilies. *J Mol Evol* 1996; 42(1):59-65.
16. Zhou Y-H, Zheng JB, Gu X, et al. A novel Pax-6 binding site in rodent B1 repetitive elements: coevolution between developmental regulation and repeated elements? *Gene* 2000; 245(2):319-328.
17. Takahashi A, Nakata O, Kasahara M et al. Structures for the proopiomelanocortin family genes proopiocortin and proopiomelanotropin in the sea lamprey *Petromyzon marinus*. *Gen Comp Endocrinol* 2005; 144(2):174-181.
18. Simmen MW, Bird A. Sequence analysis of transposable elements in the sea squirt, *Ciona intestinalis*. *Mol Biol Evol* 2000; 17(11):1685-1694.
19. Lev-Maor G, Sorek R, Shomron N et al. The birth of an alternatively spliced exon: 3' splice-site selection in Alu exons. *Science* 2003; 300(5623):1288-1291.
20. Shimeld SM. Characterisation of amphioxus *HNF-3* genes: Conserved expression in the notochord and floor plate. *Dev Biol* 1997; 183(1):74-85.
21. Girard L, Freeling M. Regulatory changes as a consequence of transposon insertion. *Dev Genet* 1999; 25(4):291-296.
22. Britten RJ. DNA sequence insertion and evolutionary variation in gene regulation. *PNAS* 1996; 93(18):9374-9377.

Author biography

Linda Z. Holland, B.A., M.A. (Stanford University, Palo Alto, CA, USA), Ph.D. (University of California San Diego, San Diego, CA, USA), is Research Professor in the Marine Biology Research Division, Scripps Institution of Oceanography, University of California San Diego. After receiving her M.A., she worked for fifteen years as a laboratory technician before embarking on her own research on chordate evolution, which led to a Ph.D. and to her current position. She pioneered research on developmental genetics of amphioxus and has been instrumental in the effort to sequence the amphioxus genome. She currently directs a laboratory focused on understanding mechanisms of embryonic patterning in amphioxus as key to understanding how vertebrates evolved from their invertebrate chordate ancestors.