Two Novel Families of Short Interspersed Repetitive Elements from the Mud Loach (*Misgurnus mizolepis*)

Hak-Seob Lim[1,2], Moo-Sang Kim[1,3], Ok-Soon Kim[1], Ji-Yeon Kim[1], Young-Mi Choi[1], Sang Jung Ahn[1] and Hyung-Ho Lee[1]*

[1]Department of Biotechnology and Bioengineering, College of Fisheries Sciences, Pukyong National University, 599-1, Daeyean 3-dong, Nam-gu, Busan, 608-737, Korea
[2]Institute of Bio-Millennium Promise Co., LTD, Jeonggwan-myeon Gijang-gun, Busan 619-961, Korea
[3]Department of Aquatic Life Medicine, Pukyong National University, Busan 608-737, Korea

Abstract Short interspersed repetitive elements (SINEs) are dispersed throughout eukaryotic genomes. These SINEs have been shown to be excellent phylogenetic markers for the closed related species. In this report, we isolated two novel families of SINEs from the mud loach. The two SINE families, mlSINE-L and mlSINE-S, have genomic lengths of about 410bp and 270bp, respectively. 5' and 3' ends of the SINE families are well conserved and highly homologous to each of corresponding ends of RSg-1 and *Sma*I SINEs. Phylogenetic analysis shows that mlSINEs are unique to the mud loach. A dot blot hybridization experiment shows that mlSINE-L has an estimated copy number of 1×10^3 per 2×10^9bp (2.8 pg) and is more frequently distributed at nuclear matrix attachment regions (MARs) than loop DNAs. The result suggests that mlSINEs may preferentially integrate in or near MARs.

Key words : SINEs, matrix attachment regions (MARs), mud loach

Introduction

At least 30% of human chromosomal DNA is composed of short and long interspersed repetitive elements (SINEs and LINEs) [6]. Similarly, more than 50% of the some higher plant genomes, such as those of maize and *Arabidopsis*, also consist of repetitive sequences [2]. To date, almost all SINEs have been derived from tRNAs [15-18], with the exception of the primate Alu and the rodent B1 families, which are derived from 7SL RNA [26,27]. The tRNA-derived SINEs are not simple pseudogenes for tRNAs but have a composite structure including a region homologous to a tRNA, a middle tRNA-unrelated region, and a terminal AT-rich region [15-18]. SINEs are distinguished from LINEs based on their high copy number, relatively short length, and inability to encode for enzyme, such as retrotranscriptase. They typically range from 70 to 500 bp in length and may present in over 10^4 total copies in the eukaryotic genome [16]. SINEs are dispersed in the genomes of various multicellualr organisms [16,17,21]. Certain families of SINEs found specifically within members of the specific clade. It is thought that SINEs are not excised precisely, and have not been inserted independently at orthologous loci within different evolutionary lineages. Thus, SINEs have been shown to be excellent phylogenetic markers for the closely related species [4,12,13,16,22].

It is generally accepted that the integration of SINE at a new locus is an irreversible event [16] and that the sites of such integration are chosen almost at random [8]. However, human Alus may preferentially integrate in locally AT-rich or regions of R-bands of chromosomes [9,11,20]. Tikhonov *et al.* [25] also reported that S1 SINE elements in the genome of *Brassica* showed an interesting preference for matrix attachment region (MARs). DNA recombination is occurred on the nuclear matrix [1]. These reports suggested that MARs

* Corresponding author
Phone: +82-51-620-6184, Fax: +82-51-620-6180
E-mail: hyunghl@pknu.ac.kr

might be one of the strong candidates related to the integration of SINEs.

In previous study, we cloned an ARS (autonomously replicating sequence), ARS101, from matrix attachment regions of mud loach [10]. The ARS contained a partial Tc1-like element [5] and a new SINE-like element. We called the SINE-like sequence as mlSINE1. In this report, we used the conserved regions of mlSINE1 as primers for isolating other related mlSINEs and screened the genomic libraries of the mud loach (*Misgurnus mizolepis*) and the genomic DNA by PCR. As a result, we isolated 28 clones and sequence analysis shows they are classified into two families, mlSINE-L and mlSINE-S. We also analyzed the distribution of mlSINEs in the genome, especially in MARs.

Materials and Methods

Screening of mlSINE from genomic library

Total genomic DNA from the mud loach was digested with Sau3AI. The DNA fragments were ligated with λGem-11 *Bam*HI arm vector (Promega) and were then packaged in vitro. Screening was performed with a DIG-labeled oligonucleotide, designed from mlSINE1, by PCR. Hybridization was performed according to instructions supplied with the DNA Detection Kit (Roche). Positive phage clones were isolated and their inserts were amplified by PCR with the primers (forward, 5'-CGGAATTCTCTGCWGCYAATGATCGGA-3'; reverse, 5'-CGGAATTCTTTATTTAACCAGGTA-3') designed from mlSINE1. The products of PCR were isolated and inserted into pURY19 vector [10] and then the sequences were determined.

Polymerase chain reaction

The reaction mixture for amplification by PCR contained Taq buffer (50 mM KCl, 10 mM Tris-HCl, pH9.0, 0.1% Triton X-100, and 1.5 mM $MgCl_2$), 0.25 mM dNTP, 10 pmole of each primer (forward, 5'-CGGAATTCTCTGCWGCYAATGATCGGA-3'; 5'-reverse, 5'-CGGAATTCTTTATTTAACCAGGTA-3'), 1 μg of genomic DNA, and 1 unit of *Taq* DNA ploymerase (Promega) in a final volume of 50 μl. The thermal cycling involved 30 repeats of denaturation at 94℃ for 1 min, annealing at 52℃ for 30 sec, and expansion at 72℃ for 30 sec. The reaction mixture then was analyzed by electrophoresis in a 1% of agarose gel. The bands of products were cut from the gel, and the DNA fragment were purified and digested with *Eco*RI. The *Eco*RI-digested DNAs were ligated to pURY19 vector digested by *Eco*RI [10].

Dot blot hybridization

Dot blot hybridization was performed to estimate the number of copies of SINE sequences in the genome of mud loach. Increasing amounts of genomic DNA (0.62, 1.25, 2.5, 5, 10, and 20 μg) or mlSINE1 clones in pURY19 vector (0.16, 0.13, 0.63, 1.25, 2.50, and

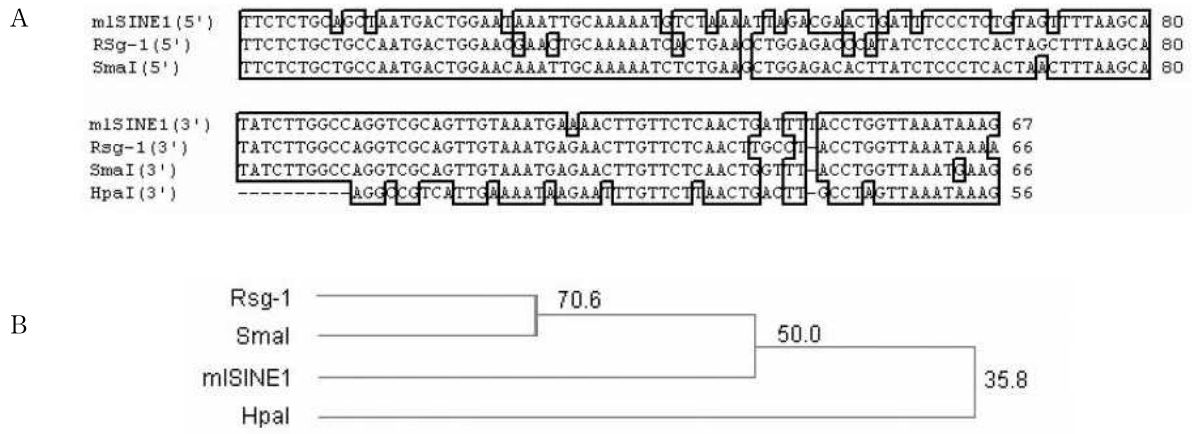


Fig. 1. Alignment and phylogenetic tree of RSg-1, *Sma*I, *Hpa*I-MS type I, and mud loach mlSINE1. Sequence alignment and phylogenetic tree construction were carried out using DNAsis software (Hitachi). (a) Sequence comparison of mlSINE1 with RSg-1, *Sma*I, and *Hpa*I-MS type I at their 5' (upper alignment) and 3' (low alignment) ends. Dots indicate gaps inserted to improve the alignment. (b) Phylogenetic relationship of RSg-1, *Sma*I, mlSINE1 and *Hpa*I-MS type I based on their 3' sequence. Numbers above branches mean percentage sequence similarity.

5×10^{10}) were dissolved in 400 μl of 0.4 M NaOH. After denaturation in 0.4 M NaOH, each DNA sample was blotted onto a GeneScreen Plus membrane (DuPont NEN products, USA) with a dot-blotting apparatus (model DP-96; Advantec, JAPAN). The membrane was neutralized in a solution of 0.5 M Tris-HCl (pH 7.0) and 1 M NaCl and then dried. Hybridization was performed by a DIG-labeled oligonucleotide designed from mlSINE1. Intensities of the spots were measured by densitometry via image analysis system (Bio Profil®, FRANCE). The procedure to obtain nuclear matrix associated DNAs (MARs) and released DNAs (loop DNAs) followed Lim *et al* [10].

Results

From the clone, ARS101, we isolated from the mud loach [10], and a new SINE-like element (mlSINE1) is found. We compared the mlSINE1 sequence with *Hpa*I [3], RSg-1 [29] and *Sma*I [24] SINEs, the most related sequences resulted from BLAST search. The analysis showed that the 5' and 3' ends (80 bp and 67bp, respectively) of them were highly conserved (Fig. 1A) even if *Hpa*I SINE showed low conservation at

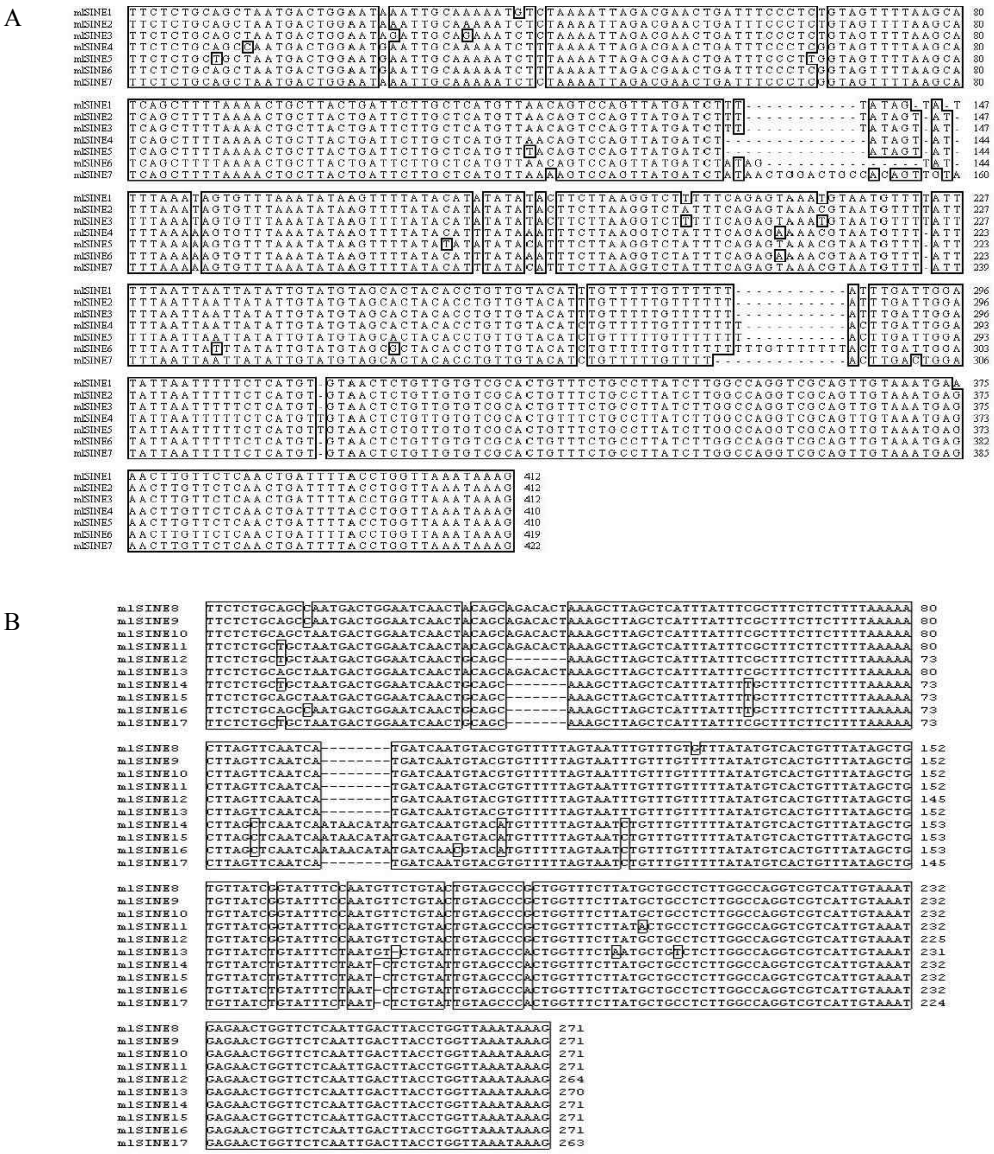


Fig. 2. Sequence alignment of 17 members of the mlSINEs from the mud loach. The mlSINEs are parted into two groups, with 410 bp (a) and 270 bp (b) in length, respectively. Large boxes indicate identical nucleotides. Dashes indicate gaps inserted to improve the alignment.

the 5'. However, the phylogenetic relationship of mlSINE1, RSg-1, *Sma*I and HapI elements indicated that mlSINE1 is distinguished from the related sequences (Fig. 1B). This suggests that mlSINE1 is a new SINE sequence.

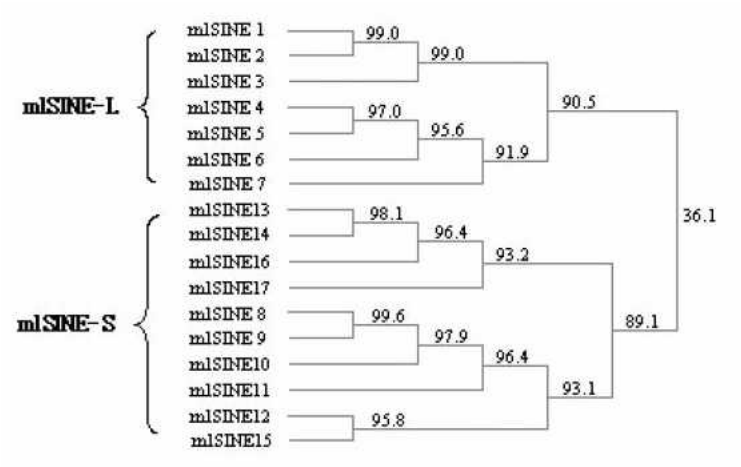
Isolation and characterization of mlSINEs families

To obtain more information about the mlSINEs, we isolated additional 28 clones by screening from genomic libraries and PCR from genomic DNA of the mud loach and then determined the sequences. The sequence analysis shows that clones are clearly assigned to 17 members of two types of mlSINEs, with about 410 and 270 bp in length, respectively (Figs. 2 and 3A). Sequences of some members, AGACACT (35-41) in mlSINE12, 14, 15, 16, 17 and ATAACATA (94-101) in mlSINE8, 9, 10, 11, 12, 13, 17, AACTGGACTGCC (140-151) in mlSINE7 and TTTGTTTTTTT (282-292) in mlSINE6 show deletions and/or insertions. Phylogenetic analysis suggests that two families of the mlSINEs, mlSINE-L and mlSINE-S are distinctly separated (Fig. 3A). As shown in Figure 3b, the 5' and 3' regions of the consensus sequences of mlSINE-L and mlSINE-S were relatively conserved and specially the 3' region of the sequences were more conserved than the 5' region.

Distribution of mlSINEs in mud loach

Eukaryotic genomes contain a large number of mo-

A



B

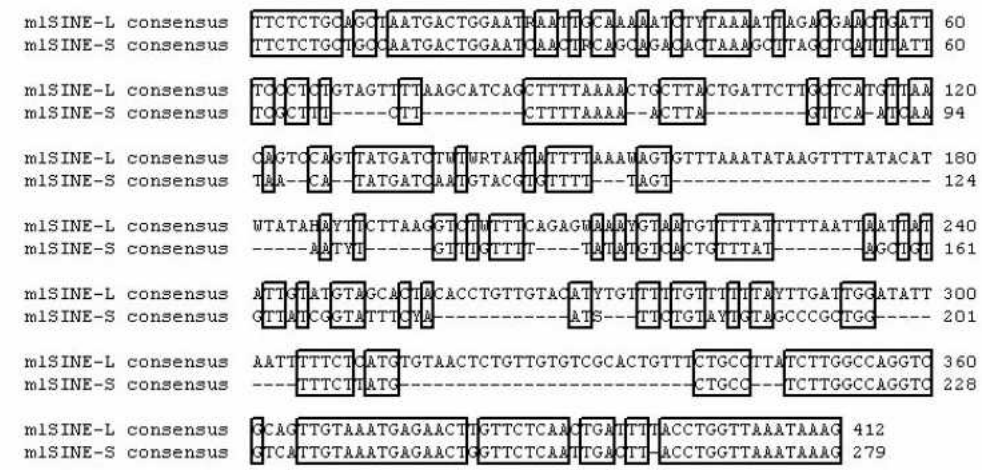


Fig. 3. Clustering and the consensus sequences of two families of mlSINEs from the mud loach. (A) Clustering of mlSINE-L and mlSINE-S was derived from alignments of 17 members of mlSINEs in Figure 2a and b. Numbers above branches mean percentage sequence similarity. (B) The consensus sequences of mlSINE-L and mlSINE-S were derived from the data in Figure 2a and b. Boxes indicate identical nucleotides.

bile elements that have proliferated by retroposition to produce numerous copies of interspersed repeats [19,23,28]. To estimate the number of copies of the mlSINEs in genomes of mud loach, we performed the dot hybridization with mlSINE1 as the probe. As shown in Figure 4a, 5 μg of DNA from the mud loach genome gave same intensity as spot of the 0.63×10^{10} copies of the plasmid containing mlSINE1 based on the densitometry measurement. Based on the genome size of mud loach (2.8 pg=2×10^9 bp) [14], we can infer that mud loach has 1×10^3 copies of the mlSINEs, judging from the intensities of spots on the membrane by densitometry via image analysis. In addition, Figure 4B shows that the intensity of matrix attachment region (MAR) was more than two times higher than that of loop DNA. The result indicates the preference of mlSINEs in the integration for MARs.

Discussion

We isolated two new SINE families, mlSINE-L and mlSINE-S, from the mud loach and compared the sequences with related SINEs. The mlSINE-L consists of about 410-422bp, while mlSINE-S consists of about 264-271 bp (Fig. 3). The 5' and 3' ends of mlSINEs are significantly homologous to the corresponding ends of *Hpa*I, RSg-1, and *Sma*I SINEs, except the 5' end of mlSINE is less conserved with 5' end of *Hpa*I (Fig. 1 A). However, phylogenetic analysis indicates that the mlSINEs is a new SINE families (Fig. 1B). The RSg-1 family, whose members have well-defined 3' ends that contain poly A segments and heterogeneous 5' ends, was reported from the rainbow trout [29]. The salmonid *Hpa*I SINE was first described as one of three families of SINEs, each of which was distributed in a different lineage during the evolution of salmonids [7]. The other two families of SINEs are the salmon SINE, which is restricted to chum and pink salmon, and the charr FokI SINE, which is only found in species in the genus *Salvelinus chauus* [7]. The salmon *Sma*I family is a family of SINEs, whose members are restricted to the genomes of chum and pink salmon in the genus *Oncohynchus* [7]. Kido *et al*. [7,8] suggested that 5' and 3' ends of *Ava*II, *Fork*I, *Hpa*I, and *Sma*I from salmonid fishes are well conserved. These may suggest that mlSINEs are uniquely evolved in cobitid fishes and phylogenetic markers for the family. To clarify that point the further study is needed.

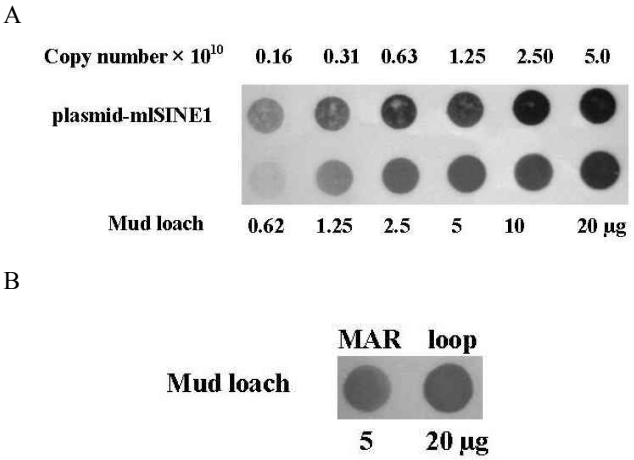


Fig. 4. The copy numbers and distribution of the mud loach mlSINEs in the genome by a dot-blot hybridization. (a) The copy numbers of mlSINEs. The copy number of cloned element (mlSINE1) (0.16-5.0 x 10^{10}) and amount of genomic DNA (0.62-20 μg) blotted on the membrane are on the top and bottom, respectively. (b) Distribution of mlSINEs in the mud loach genome. MAR and loop indicate nuclear matrix-bound and -unbound DNA, respectively. Two different amounts of genomic DNA (5 and 20 μg) used.

A dot-hybridization analysis shows that mlSINEs are present at about 10^3 copies per 2×10^9 bp and more frequently distributed in matrix attachment regions (MARs) than loop DNAs (Fig. 4). It is generally accepted that the integration sites of SINE are chosen almost at random [8]. In recent study, however, S1 SINEs of *Brassica* are more frequently integrated into MAR than loop DNA [25]. Our result is also similar to that of Tikhonov *et al* [25]. It is known that DNA recombination is occurred on the nuclear matrix [1]. Thus, matrix attachment region (MAR) might be one of the strong candidate related to integration.

요 약

짧은 집단 반복 요소 (Short Interspersed Repetitive Elements, SINE) 는 수백개 정도의 염기로 구성된 반복염기서열로서 LINE (Long Interspersed Nucleotide Elements)와 함께 바이러스와는 구별되는 레트로트랜스포존 (Retrotransposon)의 하나로 알려져 있다. 이들의 생체 내 역할은 정확하게 밝혀진 것은 없지만 게놈 내에서 반복염기서열의 재배열을 통해 완전히 새로운 유전자를 창조하거나 기존의 유전자를 변형시킴으로써 유전물질의 운반수단 및 진화적 변화에 있어서 중요한 역할을 할 것이라 예상되며, 질병의 원인이 된다고도 밝혀져 있다. 본 연구에서는 미

꾸라지로부터 SINE의 새로운 두 그룹을 분리하였다. 두 SINE 그룹, mlSINE-L과 mlSINE-S는 각각 약 410bp와 270bp의 염기로 구성되어 있다. 두 SINE 그룹의5'과 3'말단의 서열은 RSg-1와 *Sma*I SINE 의 그것과 높은 유사도를 보였다. 계통발생분석결과, mlSINE들은 미꾸라지에서 유일하였으며, dot blot hybridization의 결과는 mlSINE-L이 미꾸라지 게놈 2×10^9bp (2.8 pg)당 1×10^3 copy를 가지는 것으로 추정되며, loop DNA보다 핵기질부착부위 (nuclear matrix attachment regions, MARs)에서 그 분포도가 높았다. 이런 결과는 미꾸라지의 새로운 SINE 들이 핵기질 부착부위 내에서나 혹은 가까운 주변에 우선적으로 삽입될 수 있음을 나타낸다.

Acknowledgements

This research was supported by a grant (B-2005-08) from Marine Bioprocess Research Center of the Marine Bio 21 Center funded by the Ministry of Maritime Affairs & Fisheries, Republic of Korea.

References

1. Dave, V. P., Modak, M. J. and Pandey, V.N. 1991. Nuclear matrix bound V(D)J recombination activity in rat thymus nuclei: an in vitro system. *Biochemistry* **30**, 4763-4767.
2. Grandbastien, M. A. 1992. Retroelements in higher plants. *Trends Genet.* **8**, 103-108.
3. Hamada, M., Kido, Y., Himberg, M., Reist, J. D., Ying, C., Hasegawa, M. and Okada, N. 1997. A newly isolated family of short interspersed repetitive elements (SINEs) in coregonid fishes (whitefish) with sequences that are almost identical to those of the *Sma*I family of repeats: Possible evidences for the horizontal transfer of SINEs. *Genetics* **146**, 355-367.
4. Hamada, M., Takasaki, N., Reist, J. D., DeCicco, A. L., Goto, A. and Okada, N. 1998. Detection of the ongoing sorting of ancestrally polymorphic SINEs toward fixation or loss in populations of two species of charr during speciation. *Genetics* **150**, 301-311.
5. Izsvák, Z., Ivics, Z. and Hackett, P. B. 1995. Characterization of a *Tc1*-like transposable element in zebrafish (*Danio rerio*). *Mol. Gen. Genet.* **247**, 312-322.
6. Kazazian, H. H. Jr and Moran, J. V. 1998. The impact of L1 retrotransposons on the human genome. *Nat. Genet.* **19**, 19-24.
7. Kido, Y., Aono, M., Yamaki, T., Matsumoto, K., Murata, S., Saneyoshi, M. and Okada, N. 1991. Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. *Proc. Natl. Acad. Sci. USA.* **88**, 2326-2330.
8. Kido, Y., Himberg, M., Takasaki, N. and Okada, N. 1994. Amplification of distinct subfamilies of short interspersed elements during evolution of the Salmonidae. *J. Mol. Biol.* **241**, 633-644.
9. Korenberg, J. R. and Rykowski, M. C. 1988. Human genome organization: Alu, LINE and the molecular structure of metaphase chromosome bands. *Cell* **53**, 391-400.
10. Lim, H. S., Kim, M. S., Park, J. Y., Choi, K. E., Hwang, J. Y., Kim, D. S. and Lee, H. H. 2002. Molecular cloning and characterization of ARS elements from the mud loach (*Misgurnus mizolepis*). *Mol. Cells* **13**, 185-193.
11. Matera A. G., Hellman, U. and Schmid, C. W. 1990. A transpositionally and transcriptionally competent Alu subfamily. *Mol. Cell. Biol.* **10**, 5424-5432.
12. Murata, S., Takasaki, N., Saitoh, M. and Okada, N. 1993. Determination of the phylogenetic relationships among Pacific salmonids by using short interspersed elements (SINEs) as temporal landmarks of evolution. *Proc. Natl. Acad. Sci. USA.* **90**, 6995-6999.
13. Murata, S., Takasaki, N., Saitoh, M., Tachida, H. and Okada, N. 1996. Details of retropositional genome dynamics that provide a rationale for a generic division: the distinct branching of all the Pacific salmon and trout (*Oncorhynchus*) from the Altantic salmon and trout (*Salmo*). *Genetics* **142**, 915-926.
14. Nam, Y. K., Cho, Y. S., Cho, H. J. and Kim, D. S. 2002. Accelerated growth performance and stable germ-line transmission in androgenetically derived homozygous transgenic mud loach, *Misgurnus mizolepis*. *Aquaculture* **209**, 257-270.
15. Ohshima, K., Koishi, R., Matsuo, M. and Okada, N. 1993. Several short interspersed repetitive elements (SINEs) in distant species may have originated from a common ancestral retrovirus: characterization of a squid SINE and a possible mechanism for generation of tRNA-derived retroposons. *Proc. Natl. Acad. Sci. USA.* **90**, 6260-6264.
16. Okada, N. 1991. SINEs. *Curr. Opin. Genet. Dev.* **1**, 498-504.
17. Okada, N. 1991. SINEs: short interspersed repeated elements of the eukaryotic genome. *Trends Ecol. Evol.* **6**, 358-361.
18. Okada, N. and Ohshima, K. 1995. Evolution of tRNA-derived SINEs. In: Maraia, R. J. editors.The impact of short interspersed elements (SINEs) on the Host genome. R.G. Landes Company: Austin, Texas, pp 61 – 79.
19. Rogers, J. H. 1985. The orogin and evolution of retrpposons. *Int. Rev. Cytol.* **93**, 187-279.
20. Schmid, C. W. 1996. Alu: structure, origin, evolution, significance and function of one-tenth of human DNA. *Prog. Nucl. Acid Res. Mol. Biol.* **53**, 283-319.
21. Schmid, C. W. and Maraia, M. 1992. Transcriptional regulation and transpositional selection of active SINE sequence. *Curr. Opin. Genet. Dev.* **2**, 874-882.
22. Shimamura, M., Yasus, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I. and Okada, N. 1997. Molecular evidence from retroposons that whales from a clade within even-toed ungulates. *Nature* **388**, 666-670.

23. Smit, A. F. 1996. The origin of interspersed repeats in the human genome. *Curr. Opin. Genet. Dev.* **6**, 743-748.
24. Takasaki, N., Ymaki, T., Hamada, M., Park, L. and Okada, N. 1997. The salmon *Sma*I family of short interspersed repetitive elements (SINEs): interspecific and intraspecific variation of the insertion of SINEs in the genomes of chum and pink salmon. *Genetics* **146**, 369-380.
25. Tikhonov, A. P., Lavie, L., Tatout, C., Bennetzen, J. L., Avramova, Z. and Deragon, J. M. 2001. Target sites for SINE integration in Brassica genomes display nuclear matrix binding activity. *Chromosome Res.* **9**, 325-37.
26. Ullu, E. and Tschudi, C. 1984. *Alu* sequences are processed 7SL RNA genes. *Nature* **312**, 171-172.
27. Weiner, A. M. 1980. An abundant cytoplasmic 7S RNA is complementary to the dominant interspersed middle repetitive DNA sequence family in the human genome. *Cell* **22**, 209-218.
28. Weiner, A. M., Deininger, P. L. and Efstratiadis, A. 1986. Nonviral retroposon: genes pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55**, 631-661.
29. Winkfein, R. J., Moir, R.D., Krawetz, S.A., Blanco, J., States, J.C. and Dixon, G.H. 1988. A new family of repetitive, retroposon-like sequences in the genome of the rainbow trout. *Eur. J. Biochem.* **176**, 255-264.