BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **220,** 226–232 (1996)
ARTICLE NO. 0385

Mermaid: A Family of Short Interspersed Repetitive Elements Widespread in Vertebrates

Nobuyoshi Shimoda,* Mario Chevrette,† Marc Ekker,‡ Yutaka Kikuchi,*,§ Yoshiki Hotta,*,¶ and Hitoshi Okamoto*,§,[1]

**Division of Cellular Communication, National Institute for Basic Biology, Nishigonaka 38, Myodaiji-cho, Okazaki, Aichi 444, Japan; †Department of Surgery, Urology Division, McGill University and Montreal General Hospital Research Institute, 1650 Avenue Cedar, Montreal, Quebec, Canada, H3G 1A4; ‡Loeb Institute for Medical Research, Ottawa Civic Hospital, University of Ottawa, 725 Parkdale Avenue, Ottawa, Ontario, Canada, K1Y 4E9; ¶Department of Physics, Graduate School of Science, University of Tokyo, Hongo 7-3-1, Bunkyo, Tokyo 113, Japan; and §Department of Physiology, Keio University, School of Medicine, 35 Shinanomachi, Shinjuku, Tokyo 160, Japan*

Received February 1, 1996

We have discovered a family of short interspersed repetitive elements (SINEs) that are present in the genomes of fish, amphibian and primates. The family of the SINEs, designated *mermaid,* is distinctive in each species except for a conserved region of approximately 80 bp. Some members of the *mermaid* family were found in transposon-like repetitive elements, including Tc1-like elements which are also distributed in the genomes of fish and amphibian. This raises the possibility of horizontal transfer of the *mermaid* family between vertebrates via transposons.

Repetitive sequences constitute a substantial portion of the genomes of eukaryotes (1). They are categorized into two classes: tandemly repeated sequence and transposable elements. Eukaryotic transposable elements are further divided into two main classes, according to their mode of transposition: elements that move directly through DNA copies and elements that transpose by reverse transcription of an RNA intermediate. The latter elements are often referred to as retroposon. Short interspersed repetitive elements (SINEs) are supposed to be members of retroposons, although they do not encode proteins needed for reverse transcription. It has been proposed that SINEs have originated from tRNAs because most of them contain sequences homologous to tRNA molecules at their 5′ end (2). Other structural characteristics of SINEs are tandem repeats of short oligonucleotide present at the 3′ end and terminal direct repeats.

Although all of the SINEs previously characterized were considered to be specific to a few species, a genus or a family, we have found a novel family of SINEs, designated *mermaid,* that are distributed in the genomes of fish, amphibian and primates. Mechanisms that could be involved in the wide distribution of *mermaid* in the animal kingdom will be discussed.

MATERIALS AND METHODS

Computer sequence analysis. Identification of the *mermaid* related sequences in GenBank was carried out using the software Blastn (3). Sequences were aligned using GeneWorks (Intelligenetics) followed by manual optimization.

DNA samples. Genomic DNA from zebrafish and medaka fish were prepared from the whole bodies of adult fishes by standard procedures (4). Other genomic DNAs were purchased from BIOS Laboratories. For the cloning of *mermaid* sequences from the zebrafish, a zebrafish genomic library was screened using the [α-^{32}P]-labeled medaka DNA fragment that contains the medaka *mermaid* sequence. Three independent positive clones were obtained and were designated Z. mer1, Z. mer2, and Z. mer3.

Mermaid PCR. mermaid PCR was carried out in a total volume of 50 μ1 with 50 ng of genomic DNA, primers at 1 μM, in 50 mM KCl, 10 mM TRIS-HCl pH 8.0, 1.5 mM $MgCl_2$, 0.01% gelatin, 200 μM dNTPs, and 2.5 units of Taq polymerase (Promega) for 30 cycles of 94°C denaturation (1 min), 63°C annealing (45 sec), and 72°C extension (1 min) in an automated thermal cycler (ASTEC, model PC-700). The sequences of the degenerated nucleotides, oligoA and oligoB, are 5′-

[1] To whom correspondence should be addressed. Fax: 81-3-3357-5445. e-mail: QZB00077@nifty serve.or.jp.

0006-291X/96 $18.00

AGAA(C/T)(A/G)TGCAAACTCCACACAGA-3′ and 5′-CCTGGAG(G/A)AAACCCAC(G/A)CA(G/A)ACA-3′, respectively.

Zoo blot analysis. The oligoA and oligoB were 5′ end labeled with [γ-^{32}P] ATP (Amersham, 5000 Ci/mmol) and Megalabeling kit (Takara). DNA samples were denatured in alkaline buffer (0.4 M NaOH, 10 mM EDTA) and transferred to nylon membranes (Hybond-N plus, Amersham) as described previously (4). Hybridization and posthybridization washes were performed as described previously (4,5). The hybridization was carried out in 5×SSC, 0.1% SDS (w/v) and 5×Denhardt at 48°C for 12 h, followed by two washes at room temperature for 15 min, two washes at 48°C for 20 min and one wash at 60°C for 20 min, in 5×SSC, 0.1% SDS (w/v).

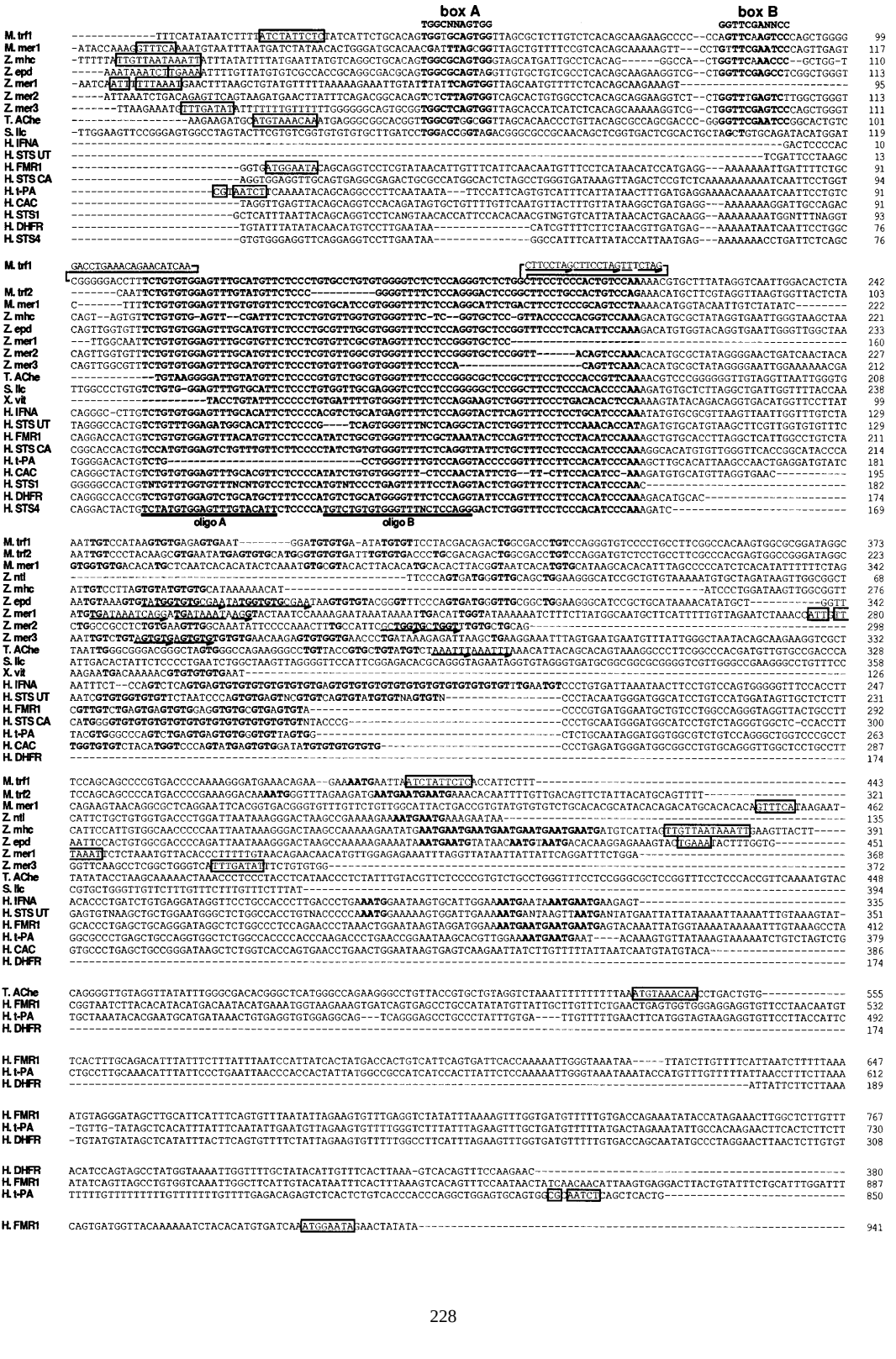
RESULTS

Characterization of Sequences of the Mermaid Family

In the course of our efforts to isolate Tc1-like transposable element in the Japanese medaka fish, we found a sequence similar to the C terminal region of the putative transposase of Tes1, a Tc1-like repetitive element found in a hagfish (6, data not shown). However, sequences corresponding to the N terminal region of the putative transposase were not observed within 1 kbp upstream from the segment. A search of the GenBank data base revealed that the upstream region contained 77 nucleotides that showed significant homology to sixteen sequences that are present in the genomes of medaka fish and other vertebrates, such as zebrafish (*Danio rerio*), ray (*Torpedo marmorata*), frog (*Xenopus lavies*), and human (*Homo Sapiens*). Of seventeen sequences, eight are present in introns and intergenic regions, five are found in human sequence tagged sites, one is detected in a human cDNA and three are associated with transposable elements. Of the latter three sequences, one is present in the insertion element causing the zebrafish *no tail* mutation (7), another is located in a zebrafish Tc1 -like repetitive element, named Tdr1 (8), and the last one is adjacent to a transposon, termed Xori, found in the *Xenopus* genome (9). These sequences are shown in Fig. 1 with three additional sequences (designated Z. mer1, Z. mer2, and Z. mer3) that were cloned from the zebrafish, based on their similarities to the 77 base pair consensus. An alignment of these sequences showed that they are distinct in each species outside the conserved region of 77 base pairs. We designated these sequences as members of the *mermaid* family, which can be divided into subfamilies according to their hosts. The *mermaid* family of fish represented a typical SINE structure (2): they had split promoters of RNA polymerase III (10), called boxA and boxB at the 5′ end, a simple repeated unit of (AATG) (in medaka and zebrafish), (T) (ray), (GTTTCTTT) (sandbar shark) at the 3′ ends, and were flanked by direct repeats (medaka, zebrafish, and ray). In addition, internal direct repeats as common in other SINEs (11,12) were present in the *mermaid* sequences of the zebrafish and the ray. The consensus sequence of RNA polymerase III promoter regions in zebrafish *mermaid* sequences showed significant similarity to some tRNAs (Fig. 2). In contrast to the fish sequences, the human *mermaid* sequences contained no sequences resembling the consensus sequence of an RNA polymerase III promoter. However, two of the longest human *mermaid* sequences, designated H. FMR1 and H. t-PA, had terminal direct repeats and the H. t-PA sequence ended in simple repeat of $(T)_s$. The (AATG) motifs were also detected in the middle of the human *mermaid* sequences. It should be noted that the *mermaid* sequences of zebrafish and human were variable in length probably due to deletions. Conservation of the *mermaid* sequences within each species was low compared to other SINEs characterized to date (12,13).

Distribution of the Mermaid Family in Vertebrates

To examine the distribution of the *mermaid* family in vertebrate genomes, we performed zoo blot experiments using DNA from eight different vertebrates. Two degenerate oligonucleotides, termed oligoA and oligoB, corresponding to conserved *mermaid* sequences were used as probes (see Fig. 1). As expected from the results of the data base search, both probes hybridized to the genomes of medaka fish, zebrafish, and human, and the oligoB hybridized to frog DNA (Fig. 3A and B). Moreover, both probes hybridized to the genomes of chimpanzee, gorilla, and orangutan. This


box A
TGGCNNAGTGG
box B
GGTTCGANNCC
oligo A
oligo B

result indicated that *mermaid*-related sequences also exist in nonhuman primates. Neither of these probes, however, hybridized to DNAs from mouse, chicken, or snake. When we screened a zebrafish genomic library, whose average insert size is 14 kbp, with a DNA fragment that contains the Z. mer2 clone, approximately 10% of the plaques gave positive signals of varying intensity, presumably due to the different degrees of homology of the genomic sequences to the probe. Alternatively, these differences could be due to the presence of more than one *mermaid* like-sequences/clone. This percentage of positive clones corresponds to about 12000 copies of the *mermaid* sequence per haploid genome of the zebrafish whose estimated size is 1.7×10^9 bp (14), giving an average distance between copies of approximately 140 kb. The intensity of signals of both zoo blots showed that the copy numbers of the *mermaid* family of primates were approximately 125 times smaller than that of the zebrafish. Assuming that all primates have a genome size of 3×10^9 bp, the copy number of the *mermaid* family in primates was thus estimated to be 170 copies per haploid genome under our experimental conditions.

Although the presumed average distance between the zebrafish *mermaid* sequences is far beyond the maximum length that PCR can amplify, we performed PCR with one of the two oligonucleotides described above as primer to examine whether we could obtain DNA fragments between two close *mermaid* sequences from each vertebrate genome (hereafter, we call this experiment as *mermaid* PCR). Unexpectedly, amplification of DNAs from medaka fish, zebrafish, and primates with oligoA or B gave many distinct or smearing bands, while oligoB primer also gave bands with frog DNA (Fig. 4A and B). Neither primers gave discernible bands with chicken, snake, or mouse DNA. The presence or absence of PCR products were not only consistent with the presence or absence of hybridizing signals in the zoo blots, but the number of PCR fragments also appeared to be correlated with the intensity of the signals on the zoo blots. To characterize the structure of the *mermaid* PCR products, we determined the sequences of six of them derived from the zebrafish with oligoB primer, and found that they all contained the primer sequence at both ends. However, sequences related to the zebrafish *mermaid* were observed only at one end of these PCR products (data not shown). Similarly the shortest *mermaid* PCR products (≈ 300 bp) amplified from the chimpanzee and orangutan using oligoA contained oligoA sequences at both ends, and human *mermaid* related sequences at one end. These results indicated that both specific and non-specific binding of the primer to genomic DNA were involved in the generation of these products. The amplified fragments were, therefore, not genuine inter-*mermaid* PCR products and should be called *mermaid*-tagged, or *mermaid*-anchored PCR products.

FIG. 1. Sequences of members of the *mermaid* family. M. mer1 is the nucleotide sequence of the *mermaid* family member found in the upstream region of the Tc1-like element of the medaka fish (D78164; unpublished results). Z. mer1, Z. mer2, and Z. mer3 are the nucleotide sequences of members of the *mermaid* family from clones Z. mer1, Z. mer2, and Z. mer3, respectively (D78161, D78162, D78163). M. trf1, M. trf2, Z. mhc, Z. epd, Z. ntl, T. AChe, X. vit, H. FMR1, H. t-PA, H. CAC and H. DHFR are the nucleotide sequences of members of the *mermaid* family associated with the genes indicated: M. trf1 and M. trf2, the transferrin gene in the medaka fish (D64033); Z. mhc, the gene for major histocompatibility class II protein in the zebrafish (U08874); Z. epd, the gene for ependymin beta and gamma chains in the zebrafish, whose upstream region contains Tdr1 (M89643); Z. ntl, the insertion sequence, caused the zebrafish *no tail* mutation (X71596); T. AChe, the gene for acetylcholinesterase in *Torpedo californica* (X56517); X. vit, the vitellogenin A2 gene in *Xenopus laevis*, whose upstream region contains Xori (Y00354); H. FMR1, the gene for fragile X mental retardation protein in human (L29074); H. t-PA, the human gene for tissue plasminogen activator (K03021); H. CAC, the cDNA of RNA from human lymphocytes (U00954); H. DHFR, the gene for human dihydrofolate reductase (X00856). H. IFNA, H. STS UT, H. STS CA, H. STS1, and H. STS 4 show the nucleotide sequences of members of the *mermaid* family associated with the sequence tagged sites of human indicated: H. IFNA, human interferon alpha gene related dinucleotide repeat (M98545); H. STS UT, Human STS UT7566 (L30507); H. STS CA, AFM238yf8 (Z17075); H. STS1, human STS STS1-cSRL-30b4-uA/cSRL-30b4-uZ (G02317); H. STS4, STS4-408 (L00843). Numbers in parentheses indicate GenBank accession numbers. Terminal direct repeats flanking the *mermaid* family are boxed. The two internal promoters for the RNA polymerase III are shown as boxA and boxB. The conserved region of the *mermaid* family is given in boldface in the second block. GT dinucleotide rich regions and AATG motif are given in boldface in the third block and the fourth block, respectively. Horizontal arrows present internal direct repeats.

```
Z.mer cons  GCGCAGTGGCTCAGTGGTTAGCACTGTCGCCTCACAGCAAGAAGGTCGCTCGTTCGAGTCTCGGCTGGGTCAGTT
tRNA-Val    GTTTCCGTAGTGTAGTGGTTATCACGTTCGCCTTACACGCGAAAGGTCCCCGGTTCGAAACCGGGCGGAAACACCA
tRNA-Trp    GACCTCGTGGCGCAATGG-TAGCGCGTCTGACTCCAGATCAGAAGGTTCCGTGTTCGAATCACGTCGGGGTCACCA
tRNA-Met    GCCTCGTTAGCGCAGTAGGTAGCGCGTCAGTCTCATAATCTGAAGGTCGTGAGTTCGATCCTCACACGGGGCACCA
tRNA-Gly    GCATTGGTGGTTCAGTGG-TAGAATTCTCGCCTGCCACGCGGCAGG-CCCGGGTTCGATTCCCGGCCAATGCACCA
```

FIG. 2. Sequence comparisons between the RNA polymerase III promoter related region in the consensus sequence of the zebrafish *mermaid* family and tRNAs from bovine and human (20).

DISCUSSION

We characterized a family of SINEs of vertebrates, which we named "*mermaid*". The *mermaid* family is the most widespread SINE currently known in the animal kingdom, with members present in the genomes of fish, frog, and primates, but not in the genomes of mouse, snake, or bird. This family thus do not follow phylogenetic lines, indicating that the broad distribution of this family is not due to evolutionary conservation. In addition, the structure of this family is considerably diverged between species. Of particular note is the observation that the 5′ end of the human *mermaid* sequences exhibit no homology to any tRNAs. These findings would imply that the *mermaid* family had arisen independently in each species, although we cannot exclude the possibility that the family had followed phylogenetic lines, but for some reason has been only conserved in a subset of lineages. We found that *mermaid* PCR with primate DNA produced some similar bands (Fig. 4A) and confirmed by sequencing that the *mermaid*-anchored PCR products of ≈ 300 bp amplified from the chimpanzee and orangutan using oligoA showed significant homology even outside the *mermaid* related sequences (data not shown). These facts suggest that the distribution of human *mermaid* sequences occurred before the primate radiation, and at least some *mermaid* sequences in the human genome are fixed at the equivalent chromosomal position in nonhuman primate genomes following an expansion of this family in the ancestor of primates.

Another possible mechanism for the generation of the *mermaid* family involves horizontal transmission across species borders. We found that the three fish and one frog *mermaid* sequences are located adjacent to or inside of transposon-like repetitive elements including Tc1-like elements. Tc1 was first found in *C. elegans* and is believed to transpose from DNA to DNA via excision and insertion (15). In recent years, a number of Tc1-like elements have been detected in fish and frog

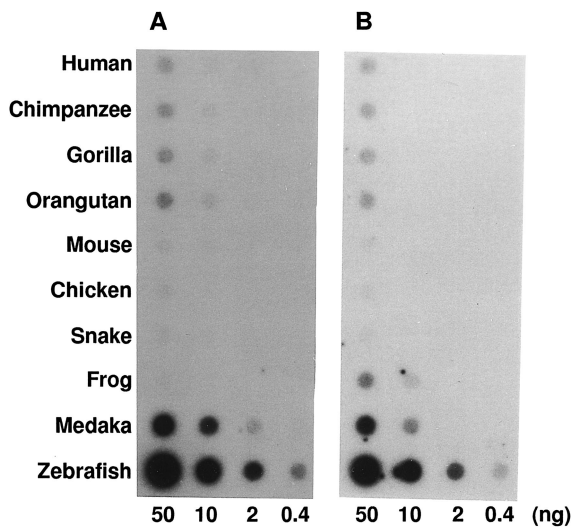


FIG. 3. Distribution of the *mermaid*-related sequences in vertebrates. The genomic DNAs from vertebrates were blotted onto the nylon membrane and hybridized with oligoA (A) and oligoB (B) probes directed to the conserved region of the *mermaid* family. Numbers indicate the quantity of DNA blotted.

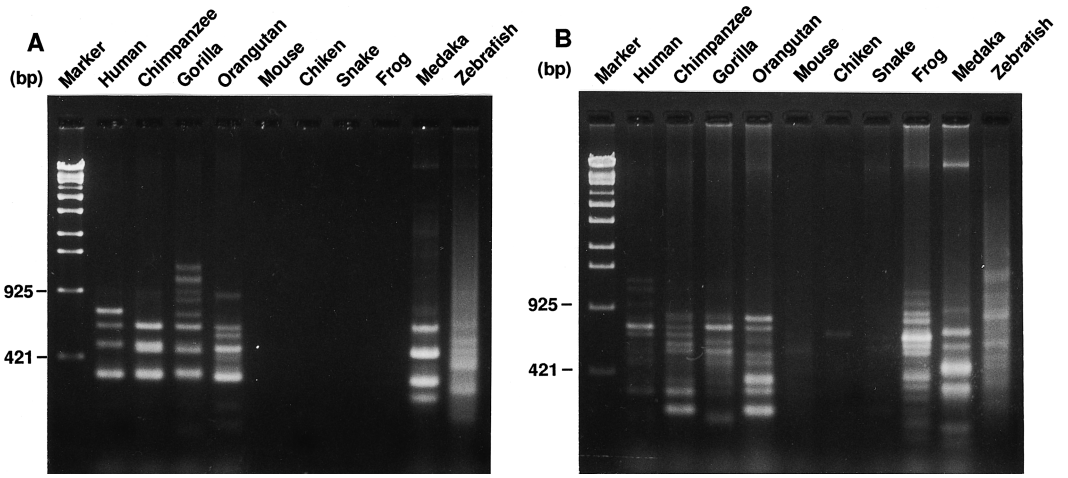


FIG. 4. Amplification of genomic DNAs from ten vertebrate species with oligoA primer (A) and oligoB primers (B). PCR-amplified products were fractionated on a 1% agarose gel by using *Sty*I-digested lambda DNA as a size standard.

(8,16,17,18). However, phylogeny deduced from structures of Tc1-like elements in fish is not consistent with the established phylogeny of teleost fish that carry them (8,17, unpublished results). Based on these findings, horizontal transmission of Tc1-like elements has been suggested (17,18,19). If this is the case, transposons such as a Tc1-like element might be a carrier of the *mermaid* family and allow horizontal transmission of this family through a DNA form. Horizontally transmitted *mermaid* sequences might have been shaped in each lineage to constitute distinct subfamilies during evolution. Another SINE, termed *Nhe*I repeat, is also found within a Tc1-like element isolated from Atlantic salmon (11). Tc1-like elements may thus form a structure that has susceptibility to insertion of SINEs in a stably integrated state in the genome or during the process of transposition.

ACKNOWLEDGMENTS

We are grateful to S. Yasumasu for providing the medaka genomic library, and S. Matsubayashi and K. Takamatsu for their assistance in maintaining the fish colony. We thank Cristofre Martin for critical reading of the manuscript. This work was supported by a Grant-in-Aid (07262217, 07279107, 07np0701) from Ministry of Education, Science and Culture of Japan, Special Coordination Funds for Promoting Science and Technology from Science and Technology Agency of Japan, and Uehara Memorial Foundation. N.S. and Y.K. are supported by National Institute for Basic Biology and Y.K. is supported by a fellowship from the Japan Society for the Promotion of Science for Japanese Junior Scientists.

REFERENCES

1. Charlesworth, B., Sniegowski, P., and Stephan, W. (1994) *Nature* **371,** 215–220.
2. Deininger, P. L., and Batzer, M. A. (1993) *in* Evolutionary Biology (Hect, M. K., MacIntyre, R. J., and Clegg, M. T., Eds.), Vol. 27, pp. 157–196, Plenum, New York.
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
4. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K., Albright, L. M., Coen, D. M., Varki, A., and Janssen, K. (1994) Current Protocols in Molecular Biology, John Wiley & Sons, New York.
5. Batzer, M. A., Schmid, C. W., and Deininger, P. L. (1993) *Methods. Enzymol.* **224,** 213–232.
6. Heierhorst, J., Lederis, K., and Richter, D. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 6798–6802.
7. Schulte-Merker, S., van Eeden, F. J. M., Halpern, M. E., Kimmel, C. B., and Nüsslein-Volhard, C. (1994) *Development* **120,** 1009–1015.
8. Izsvák, Z., Ivics, Z., and Hackett, P. B. (1995) *Mol. Gen. Genet.* **247,** 312–322.
9. Riggs, C. D., and Taylor, J. H. (1987) *Nucleic Acids Res.* **15,** 9551–9565.
10. Galli, G., Hofstetter, H., and Birnstiel, M. L. (1981) *Nature* **294,** 626–631.
11. Goodier, J. L., and Davidson, W. S. (1994) *Genome* **37,** 639–645.
12. Ohshima, K., Koishi, R., Matsuo, M., and Okada, N. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 6260–6264.

13. Alexander, L. J., Rohrer, G. A., Stone, R. T., and Beattie, C. W. (1995) *Mammal. Genome* **6,** 464–468.
14. Postlethwait, J. H., Johnson, S. L., Midson, C. N., Talbot, W. S., Gates, M., Ballinger, E. W., Africa, D., Andrews, R., Carl, T., Eisen, J. S., Horne, S., Kimmel, C. B., Hutchinson, M., Johnson, M., and Rodriguez, A. (1994) *Science* **264,** 699–703.
15. Moerman, D. G., and Waterston, R. H. (1989) *in* Mobile DNA (Howe, M. and Berg, D., Eds.), pp. 537–556, Amer. Soc. Microbiol., Washington, DC.
16. Radice, A. D., Bugaj, B., Fitch, D. H. A., and Emmons, S. W. (1994) *Mol. Gen. Genet.* **244,** 606–612.
17. Goodier, J. L., and Davidson, W. S. (1994) *J. Mol. Biol.* **241,** 26–34.
18. Doak, T. G., Doerder, F. P., Jahn, C. L., and Herrick, G. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 942–946.
19. McDonald, J. F. (1993) *Curr. Opin. Genet. Dev.* **3,** 855–864.
20. Sprinzl, M., Hartmann, T., Weber, J., Blank, J., and Zeidler, R. *Nucleic Acids Res.* Suppl. 17, r1–r172.