A Newly Isolated Family of Short Interspersed Repetitive Elements (SINEs) in Coregonid Fishes (Whitefish) With Sequences That Are Almost Identical To Those of the Smal Family of Repeats: Possible Evidence for the Horizontal Transfer of SINEs

Mitsuhiro Hamada,* Yukiharu Kido,*1 Mikael Himberg,1 James D. Reist,2 Cao Ying,*3 Masami Hasegawa4 and Norihiro Okada*

*Faculty of Bioscience and Biotechnology, Tokyo Institute of Technology, Midori-ku, Yokohama 226, Japan, 1State Educational Institute of Fisheries and Aquaculture, 21 610 Kirjala, Finland, 2Freshwater Institute, Winnipeg, Manitoba, R3T2N6, Canada and 4The Institute of Statistical Mathematics, Minato-ku, Tokyo 106, Japan

Manuscript received August 16, 1996
Accepted for publication February 10, 1997

ABSTRACT

The Smal family of repeats is present only in the chum salmon and the pink salmon, and it is not present in five other species in the same genus or in other species in closely related genera. In the present study, we showed that another short interspersed repetitive elements (SINES) family, which is almost identical to the Smal family, is present in all fishes in the subfamily Coregoninae, being regarded as the most primitive salmonids. This new family of SINES was designated the Smal-cor family (Smal family of repeats in coregonids). The consensus sequence of the Smal-cor family was found to be 98.6% homologous to that of the Smal family. Accordingly, it is difficult to explain the high degree of homology between these two families of SINES by any mechanism other than the horizontal transfer of SINES. The estimates of the rate of neutral mutation of nuclear genes, comparing chum salmon and European whitefish, confirmed this possibility. Our results strongly suggest that a member(s) of the Smal-cor family might have been transferred horizontally from one coregonid species to a common ancestor of chum and pink salmon or to these two species independently, to allow subsequent amplification of the Smal family in their respective genomes.

TRANSPLOSABLE elements (TEs) have been found in almost all species surveyed (Berg and Howe 1989). TEs can be classified into two groups, namely, those derived from RNAs that are transposed to DNA via cDNA intermediates and those that are transposed directly from DNA to DNA. The former TEs are called retroposons, and the mechanism by which they are transposed is called retroposition (Rogers 1985; Weiner et al. 1986). The latter TEs are called transposons, and the mechanism by which they are transposed is called transposition (Singer and Berg 1991). Analysis of several TEs has shown that their distribution among species has sometimes been discontinuous in phylogeny. Moreover, the phylogenetic relationships deduced from TEs are not always the same as the phylogenetic relationships among species that have been obtained by analysis of a variety of characteristics, such as morphological traits and molecular data. Such discrepancies have been explained by the postulated horizontal or lateral transfer of TEs between species (Kidwell 1993; Syvanen 1994; Capy et al. 1994 and references therein). TEs that encode a transposase or a reverse transcriptase (RTase) have the potential capacity for horizontal transfer because they have their own machinery for transposition encoded within them. Many examples of the horizontal transfer of TEs, such as the P element (Daniels et al. 1990; Clark et al. 1994), mariner elements (Maruyama and Hartl 1991; Lawrence et al. 1992; Robertson 1993; Lohe et al. 1995; Garcia-Fernández et al. 1995; Oosumi et al. 1995), 1 factor (Bucheton et al. 1986) and jockey (Mizoguchi and Mazo 1990), have been reported. Host species involved include not only Drosophila but also many insects (Houck et al. 1991; Robertson 1993), the sea urchin (Springer et al. 1995), bacteria (Lawrence et al. 1992) and plants (Smith et al. 1989; Flavell et al. 1992).

Short interspersed repetitive elements (SINES; Singer 1982) are one group of retroposons, and they are often present at more than 105 copies per genome. With the exception of the rodent type 1 and primate Alu families, both of which are derived from 7SL RNA (Weiner 1980; Ulku and Tschudi 1984), all of the SINE families examined to date have been shown to be derived from tRNAs (Ohshima et al. 1993; Okada 1991a,b; Okada and Ohshima 1995 and references therein). Each SINE has internal promoters for RNA polymerase III, but none encodes any protein that can
catalyze retroposition, such as a RTase. We previously characterized three families of rRNA-derived SINEs in salmonid genomes (Kido et al. 1991). The SmaI family of repeats is restricted to the genomes of chum salmon (Oncorhyncus keta) and pink salmon (O. gorbuscha). The char: FokI family of repeats is present only in species that belong to the genus Salvelinus. The salmonid HpaI family of repeats is present in all species in the family Salmonidae but not in other species (Matsumoto et al. 1986; Kido et al. 1991; Koishi and Okada 1991; Murata et al. 1993, 1996; Kido et al. 1995). It was proposed that, among these three kinds of SINE family, the SmaI family was amplified relatively recently because of its restricted distribution and the limited sequence divergence of the members of this family (Kido et al. 1991; see also the accompanying article by Takasaki et al. 1997).

By contrast to hypotheses related to transposable elements that encode a transposase or an RTase, such as those discussed above, it has not previously been seriously proposed that SINEs can be transferred horizontally from one species to another during evolution. Recently, however, we discussed this possibility on the basis of our detailed analysis of the distribution of the HpaI SINEs in salmonid species (Takasaki et al. 1994, 1996). We have also discussed this possibility from a theoretical perspective (Oshshima et al. 1993) in the absence of concrete evidence. In the present report, we describe evidence that is suggestive of the horizontal transfer of SINEs between distantly related salmonid species.

**MATERIALS AND METHODS**

DNA samples: Total genomic DNA of each fish was extracted by the method of Blin and Stafford (1976). The species analyzed and their geographic sources are shown in Table 1.

Construction and screening of genomic libraries, subcloning and sequencing: Total genomic DNA from Coregonus lavaretus, C. albula, C. artedi, SternoGLOSSUS leucichthys and Prosopium splionotus was digested with EcoRI to construct a genomic library for each of these species. Digests were size fractionated by centrifugation on sucrose gradients (10 to 40%, w/v). DNA fragments of 2 to 4 kilobases (kb) were ligated with Agt10 arms (Stratagene, La Jolla, CA) and then packed in vitro. Screening was performed with an end-labeled oligonucleotide, designated Smol0 (see positions 97 to 117 in Figure 4) as the probe. Hybridization was allowed to proceed at 42° overnight in a solution of 6X SSC (SSC is 0.15 m NaCl, 0.015 m trisodium citrate, pH 7), 1% (w/v) sodium dodecyl sulfate (SDS), 5X Denhardt’s reagent [1X Denhardt’s reagent is 0.02% (w/v) Ficoll 400, 0.02% (w/v) polyvinylpyrrolidone and 0.02% (w/v) bovine serum albumin] and 100 μg.ml−1 herring DNA. Washing was performed in 2X SSC plus 1% SDS at 50° for 30 min. Positive phage clones were isolated and their inserts were subcloned into phUCl8 or phUC19. The inserts then were sequenced with primers that corresponded to or were complementary to the consensus sequence for the SmaI or Smol-cor family.

Dot blotting and hybridization: For dot-blot hybridization, an aliquot of genomic DNA (100 ng) or linearized plasmid DNA (from 2 to 100 ng) was adjusted to 1 μg of DNA by addition of calf thymus genomic DNA as a carrier. After de-

The presence in 0.25 M NaOH, each DNA sample was blotted onto a GeneScreen Plus membrane (Du Pont NEN Products, Boston) with a dot-blotting apparatus (model DP-96; Advan-tec, Tokyo). 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 with the SmaI oligonucleotide as the probe, as described above and under the same conditions as those used for screening. Washing was performed as described above (Mura ta 1993).

PCR for the detection of orthologous loci in the genomes of salmonid fish and cloning of products amplified by PCR: The reaction mixture for amplification by PCR contained 1th buffer (Toyobo, Tokyo), 0.2 mM dNTPs, 100 ng of each primer, 1 μg of genomic DNA, and 2 units of Tth DNA polymerase (Toyobo) in a final volume of 100 μl. The thermal cycling involved 30 repeats of denaturation at 93° for 1 min, annealing at 55° for 1 min, and extension at 72° for 1 min. The reaction mixture then was analyzed by electrophoresis in a gel that contained 2.2% (w/v) NuSieve GTG and 0.8% (w/v) SeaKem GTG agarose (FMC BioProducts, Rockland, ME). The bands of products were cut from the gel, and the DNA fragments were purified with Ultrafree-C3HV (Nihon Millipore, Tokyo). The DNA fragments were blunt ended, phosphorylated and ligated to pUC or M13 vectors that had been digested by SmaI.

Preparation of a probe and Southern hybridization: Products of PCR were transferred from gels to GeneScreen Plus membranes in 0.4 M NaOH and 0.6 M NaCl. Membranes were neutralized in 0.5 M Tris-HCl (pH 7.0) and 1 M NaCl and then dried. As the probe for detection of each SmaI-related sequence, we used the insert of the SmaI (OK)-2 clone (Kido et al. 1991), which had been amplified by PCR with primers that corresponded to the termini of the consensus sequence of the SmaI family of repeats in the presence of [α-32P]dCTP (NEN Research Products). Hybridization was performed at 42° overnight in a solution of 50% (v/v) formamide, 1 M NaCl, 1% SDS, 2X Denhardt’s solution and 100 μg.ml−1 herring DNA.

Sequence analysis: The alignments of sequences of the SmaI-cor family of repeats and those of the orthologous loci of IPM-4, respectively, were performed initially with CLUSTAL V (Higgins et al. 1992). The alignment then was inspected by eye and gaps were introduced to improve the alignment. The numbers of synonymous and nonsynonymous substitutions per site in the gene for growth hormone of Coregoninae and Salmoninae were estimated by the procedure of Ina (Ina 1995) based on Kimura’s two-parameter model (Kimura 1980). The numbers of substitutions per site in the SmaI family of repeats, the SmaI-cor family of repeats and the orthologous loci of IPM-4 were estimated by the NucML program in MOLPHY (Adachi and Hasegawa 1996) based on the same model. The sequences of the SmaI-cor family reported in the present article (see Figure 3) and those of the SmaI family of repeats reported in the accompanying article (Takasaki et al. 1997) were used to estimate their divergence.

The nucleotide sequence data reported in this article will appear in the DDBJ, EMBL, and GenBank databases with the following accession numbers: AB001839-AB001865.

**RESULTS**

The presence of a new family of SINEs in coregonids: The family Salmonidae consists of three subfamilies, namely, Salmoninae (salmon, trout, char and huchen), Thymallinae (grayling) and Coregoninae (whitefish). To examine the presence of SmaI-related sequences in the genomes of salmonids, we isolated the genomic DNA from 17 species in these three subfamilies, as well as from
Horizontal Transfer of SINEs?

### TABLE 1

The species of fish analyzed in this study and their geographic sources

<table>
<thead>
<tr>
<th>Family</th>
<th>Subfamily</th>
<th>Genus</th>
<th>Species</th>
<th>Common name</th>
<th>Geographic source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salmonidae</td>
<td>Salmoninae</td>
<td>Oncorhynchus</td>
<td>keta</td>
<td>Chum salmon</td>
<td>Okhotsk Sea</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gorbuscha</td>
<td></td>
<td>Pink salmon</td>
<td>Japan Sea</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nerka adonis</td>
<td></td>
<td>Kokanee</td>
<td>Lake Shikotsu, Japan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tshawytscha</td>
<td></td>
<td>Chinook salmon</td>
<td>Willapa River, United States</td>
</tr>
<tr>
<td></td>
<td></td>
<td>kisutch</td>
<td></td>
<td>Coho salmon</td>
<td>Lake Eva Creek, United States</td>
</tr>
<tr>
<td></td>
<td></td>
<td>masou</td>
<td></td>
<td>Cherry salmon</td>
<td>Japan Sea</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mykiss</td>
<td></td>
<td>Steelhead trout</td>
<td>University of Washington</td>
</tr>
<tr>
<td></td>
<td></td>
<td>trutta</td>
<td></td>
<td>Brown trout</td>
<td>Nikko Branch, National Research Institute of Aquaculture, Japan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plecoglossidae</td>
<td></td>
<td>Salmo</td>
<td>keta</td>
<td>Atlantic salmon</td>
<td>Atlantic Ocean</td>
</tr>
<tr>
<td></td>
<td></td>
<td>salar</td>
<td></td>
<td>Dolly Varden</td>
<td>Montana Creek, United States</td>
</tr>
<tr>
<td></td>
<td></td>
<td>malma</td>
<td></td>
<td>White-spotted charr</td>
<td>Shokotan River, Japan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>leucomaenis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>leucomaenis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>leucomaenis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>leucomaenis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pluvius</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>namaycush</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Osmeridae</td>
<td></td>
<td>Salvelinus</td>
<td>salar</td>
<td>Japanese common charr</td>
<td>Miyako River System, Japan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>maltus</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>leucomaenis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>leucomaenis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>leucomaenis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>pluvius</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>namaycush</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coregoninae</td>
<td></td>
<td>Hucho</td>
<td>perryi</td>
<td>Japanese huchen</td>
<td>Sorachi River system, Japan</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lenok</td>
<td></td>
<td>Lenok</td>
<td>Amur River, Russia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>thymallus</td>
<td></td>
<td>Grayling</td>
<td>Muhlhos Hatchery, Finland</td>
</tr>
<tr>
<td></td>
<td></td>
<td>artedi</td>
<td></td>
<td>Lake cisco</td>
<td>Canada</td>
</tr>
<tr>
<td></td>
<td></td>
<td>autunnalis</td>
<td></td>
<td>Arctic cisco</td>
<td>Canada</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nasus</td>
<td></td>
<td>Broad whitefish</td>
<td>Canada</td>
</tr>
<tr>
<td></td>
<td></td>
<td>pedel</td>
<td></td>
<td>Peled</td>
<td>Canada</td>
</tr>
<tr>
<td></td>
<td></td>
<td>alburna</td>
<td></td>
<td>Vendace</td>
<td>Ob River, Russia</td>
</tr>
<tr>
<td></td>
<td></td>
<td>lavaretus</td>
<td></td>
<td>European whitefish</td>
<td>Lake Pyhaarvi, Finland</td>
</tr>
<tr>
<td></td>
<td></td>
<td>widgremii</td>
<td></td>
<td>Tugun</td>
<td>Knopkagla, Finland</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tugun</td>
<td></td>
<td></td>
<td>Ob River, Russia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coregoninae</td>
<td></td>
<td>Stenodus</td>
<td>leucichthys</td>
<td>Inconnu</td>
<td>Woody Bay, Canada</td>
</tr>
<tr>
<td></td>
<td></td>
<td>cylindraceum</td>
<td></td>
<td>Round whitefish</td>
<td>The Great Lakes, United States</td>
</tr>
<tr>
<td></td>
<td></td>
<td>spilonotus</td>
<td></td>
<td>Bonneville whitefish</td>
<td>Bear Lake, United States</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ayu fish</td>
<td>Lake Biwa, Japan</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pond smelt</td>
<td>Lake Kasumigaura, Japan</td>
</tr>
</tbody>
</table>

2 species in the families Plecoglossidae and Osmeridae as controls. The samples of genomic DNA and a DNA fragment corresponding to a cloned Smal sequence were blotted onto a nylon membrane, and then they were allowed to hybridize with an oligonucleotide probe that contained part of the Smal sequence. Under high-stringency washing conditions, the DNA from chum and pink salmon and that from all the fish in Coregoninae gave similar strong signals, but the DNA from the other fish in Salmoninae and that from fish in Thymallinae, Plecoglossidae and Osmeridae gave only weak signals or no signal at all, as shown in Figure 1A. One microgram of DNA from C. albula gave a signal of the same intensity as 50 ng of DNA of the cloned plasmid. If the genome of salmonids is $2 \times 10^9$ base pairs in length (OHNO et al. 1968), we can infer that the coregonid species have $3.5 \times 10^4$ copies of the Smal-related sequence, judging from the intensities of spots of the hybridized genomic DNAs and the cloned DNA. To confirm the presence of the Smal-related sequences in other coregonid species, we blotted genomic DNA that we had isolated from seven coregonid species, which included species in three genera in the subfamily Coregoninae, namely Prosopium, Stenodus and Coregonus, and we allowed these samples of DNA to hybridize with the same oligonucleotide. As shown in Figure 1B, the DNA from all the coregonid species gave strong hybridization signals, confirming the presence of Smal-related sequences in all these coregonids.

Cloning and sequencing of the Smal-cor family of repeats: Using the oligonucleotide that contained part of the Smal sequence as a probe, we isolated several dozen phage clones from genomic DNA libraries of *P. spilonotus*, *S. leucichthys*, *C. lavaretus*, *C. albula* and *C. artedi*, respectively (Figure 2). We randomly selected 20 clones and determined their sequences. Figure 3 shows an alignment of 20 such sequences with the consensus sequence of the Smal family of repeats. As expected, the sequences were highly homologous to that of the Smal family of repeats, which is restricted to chum and pink salmon. This newly isolated family of SINEs was designated the Smal-cor family of SINEs.
in coregonids), whereas the Smal family of repeats present in the genomes of chum and pink retains its original designation for convenience. Members of the Smal-cor family of repeats could be divided into at least two subfamilies, namely, type I and type II, on the basis of diagnostic nucleotide substitutions. The consensus sequences of the two types of members of the Smal-cor family of repeats and that of the Smal family are compared in Figure 4. The Smal-cor type I subfamily was found to have two nucleotide substitutions, one nucleotide deletion and one nucleotide insertion when it was compared with the Smal family. The G residues at positions 18 and 84 in the Smal family were both replaced by A residues in the Smal-cor type I subfamily. Furthermore, in this subfamily of repeats, one A residue was inserted between positions 86 and 87 in the Smal family, and the T residue at position 6 in the Smal family was absent. On the other hand, when compared with the Smal family, the Smal-cor type II subfamily was found to have six nucleotide substitutions, one to seven nucleotide insertions and one nucleotide deletion. The T residues at positions 19 and 30 in the Smal family were replaced by C residues; similarly, G at position 47 was replaced by A, GT at positions 81 and 82 was replaced by AC, G at position 84 was replaced by A and A at position 96 was replaced by G. A single nucleotide deletion at position 6 was common to the Smal-cor type I and type II subfamilies of repeats. The differences between the consensus sequences of the Smal-cor type I subfamily and the Smal family and that between the Smal-cor type II subfamily and the Smal family were 1.4 and 4.9%, respectively.

The average sequence divergences among members of the two subfamilies of the Smal-cor family were calculated and compared with those of members of the Smal family of repeats. For Smal-cor type I, the value was found to be 2.5% and for Smal-cor type II it was 1.2%, both values being higher than that for the Smal family.
of repeats (~0.5%; the accompanying article by TAKASAKI et al. 1997). The consensus sequence of a subfamily must correspond to the source (master) sequence that gave rise to the members of the subfamily via retroposition. After dispersion into genomes, individual members of a subfamily accumulate mutations randomly in the subfamily consensus sequence to an extent that is proportional to the time since the original dispersion. Thus, the average sequence divergence of members of a subfamily from the consensus sequence of the subfamily should roughly reflect the age of the subfamily. Therefore, the Smal-cor type I subfamily seems to be older than the type II subfamily, and the Smal family appears to be the youngest among these three (sub)families of SINEs.

Comparison between the Smal family and the Smal-cor family: The differences between the consensus sequences of the Smal-cor type I subfamily and the Smal family and that between the Smal-cor type II subfamily and the Smal family are 1.4 and 4.9%, respectively (Figure 4). These values do not directly reflect the genetic distance between these individual (sub)families of SINEs but they do reflect the genetic distance between the source sequences of these (sub)families of SINEs. To estimate the genetic distance between the Smal family and Smal-cor family, we compared sequences of all possible pairs of the SINEs under consideration and computed the pairwise average sequence divergences per site. The pairwise average sequence divergence between the Smal family and Smal-cor type I was calculated to be 0.058 ± 0.021, and the pairwise average sequence divergence between the Smal family and Smal-cor type II was 0.086 ± 0.026 (Table 2; mean ± SE).

Highly diverged Smal-like sequences are present in all salmonid species: When we performed Southern hybridization analysis of salmonid DNA using the Smal
are indicated above the sequence.

sequence of the SmuI family derived from 25 members of the SmuI family from the chum and pink salmon

\[ \text{SmaI-cor(Cla)331} \]
\[ \text{SmaI-cor(Cla)304} \]
\[ \text{SmaI-cor(Cla)361} \]
\[ \text{SmaI-cor(Cla)505} \]

exposed the autoradiogram for a long time, we detected fishes. These results indicated the existence of small

sequence of the SmuI family derived from 25 members of the SmuI family from the chum and pink salmon (Kido et al. 1991; Takasaki et al. 1997, accompanying article). The second set of lines shows members of the SmuI-cor type I subfamily, and the third set of lines shows members of the SmuI-cor type II subfamily. Y indicates C or T. Dots indicate nucleotides identical to those in the consensus sequence. Dashes indicate gaps inserted to improve the alignment. Nucleotide positions in the consensus sequence are indicated above the sequence. Underlining shows flanking direct repeats. Psp, P. spironotus; Slc, S. leucichthys; Cla, C. lavaretus; Gab, C. albula; and Car, C. artedi.

**Figure 3.** Compilation of sequences of 20 members of the SmuI-cor family from Coregoninae. The top line is the consensus sequence of the SmuI family derived from 25 members of the SmuI family from the chum and pink salmon (Kido et al. 1991; Takasaki et al. 1997, accompanying article). The second set of lines shows members of the SmuI-cor type I subfamily, and the third set of lines shows members of the SmuI-cor type II subfamily. Y indicates C or T. Dots indicate nucleotides identical to those in the consensus sequence. Dashes indicate gaps inserted to improve the alignment. Nucleotide positions in the consensus sequence are indicated above the sequence. Underlining shows flanking direct repeats. Psp, P. spironotus; Slc, S. leucichthys; Cla, C. lavaretus; Gab, C. albula; and Car, C. artedi.

**Figure 4.** Compilation of consensus sequences of SmuI, SmuI-cor type I and SmuI-cor type II. The consensus sequences of SmuI-cor type I and of SmuI-cor type II were derived from the data in Figure 3. Dots indicate nucleotides identical to those in the consensus sequences. Dashes indicate gaps inserted to improve the alignment. Nucleotide positions in the consensus sequence are indicated above the sequence.
nones. To examine the presence or absence of the Smal-like sequences in the genomes of the fishes that apparently did not include the Smal or Smal-cor families of repeats, we screened a genomic library of kokanee (*O. nerka*) using the Smal sequence as a probe, and we sequenced several cloned DNAs. Figure 5 shows an alignment of four such sequences, namely, IPM-4, R-403, U-12 and RU-8, together with the Smal consensus sequence and the Smal-cor consensus sequences. The extent of sequence divergence between the Smal-like sequence and the Smal consensus sequence was calculated. For IPM-4, R-403, U-12 and RU-8, the values were 11.3, 10.6, 9.1 and 8.5%, respectively, indicating that these sequences are too divergent to be regarded as members of the Smal-cor or Smal families of SINEs.

To confirm the presence of the highly diverged Smal-like sequences in salmonid fishes, which apparently did not have members of the Smal or Smal-cor families of repeats in their genomes, we designed a pair of primers that should have been able to amplify almost the entire unit of the Smal sequence by PCR. We amplified such sequences from the respective genomes and cloned the products of PCR. We determined sequences for five clones from kokanee, eight clones from cherry salmon and five clones from grayling (*Thymallus thymallus*). All of the sequences were very different from those of the Smal family of repeats (data not shown). However, they did exhibit homology to those of the Smal and Smal-cor families to a similar extent to the clones discussed above, reflecting the weak signals obtained in the Southern analysis described above. The numbers of copies of these Smal-like sequences were estimated to be ~100 per genome by dot-blot analysis (see the accompanying article by Takasaki *et al.* 1997).

To estimate the time of insertion of the Smal-like sequences, we performed an analysis by PCR. Pairs of primers that flanked these sequences were designed for specific amplification by PCR of the orthologous loci from the genomes of salmonid fishes, including the coregonid species. In the case of IPM-4 (Figure 6A), the Smal-like sequence was found to have been inserted before the divergence of the subfamily Coregoninae, probably into a common ancestor of the family Salmonidae, since products of PCR were amplified from all the salmonid fishes. Hybridization experiments, for which results are shown in Figure 6B, confirmed that all the fish contained the Smal-like sequence at orthologous loci (the signal for *S. namaycush* was very weak). In the case of U-12, a similar result was obtained, indicating that the Smal-like unit in U-12 was inserted before the divergence of the subfamily Coregoninae (data not shown). These Smal-like sequences, which were highly diverged and appeared to have been amplified in a common ancestor of the family Salmonidae, were tentatively named the Smal-div family (highly diverged Smal family), although at present we do not know whether these sequences constitute a single family.

**Figure 5.**—Sequences of Smal-div sequences obtained from the genome of kokanee. Dots indicate nucleotides identical to those in the consensus sequences. Dashes indicate gaps inserted to improve the alignment.
quences of 630 nucleotides in the coding region of each gene for growth hormone were compared (Table 3). The number of substitutions in the nuclear genes of these two species was estimated by Ina's procedure (INA 1995). The two genes for growth hormone are presumed to have been generated via a process of tetraploidization of the ancestral salmonid genome between 25 and 100 mya (OHNO et al. 1968; ALLENDORF and THORGAARD 1984). Therefore, they might also be expected to be present in the genome of European whitefish, and one of them was isolated and sequenced in the present study (M. HAMADA and N. OKADA, unpublished results). In general, to estimate the genetic distance between two species, we must compare sequences at orthologous loci. Since we could not determine which gene for growth hormone in chum salmon was orthologous to the gene from the European whitefish that we had sequenced, we compared both possible pairs of sequences. The number of synonymous substitutions per site is almost the same as the rate of mutation in the case of pseudogenes, and it is considered to be equal to the rate of neutral mutation (NEI 1987). SHEN et al. (1991) reported previously that the rate of mutation of the Ah master gene(s) is close to the rate of unselected (neutral) mutation. If such is also the case for salmon SINES, the rate of mutation of the master sequences of the Smal and Smal-cor families of repeats can be presumed to be close to the rate of neutral mutation. As shown in Table 3, we found that the minimum value for the rate of synonymous substitution between genes for growth hormone in the two species was much higher than the divergence between Smal and Smal-cor type I and a little higher than that between Smal and Smal-cor type II (Table 2). The partial sequence of a cDNA for the gene for growth hormone of the arctic cisco (C. autumnalis) has been reported (GenBank accession no. x77245). This sequence yielded almost the same conclusion when we compared it with the sequences from chum salmon.

TABLE 3

<table>
<thead>
<tr>
<th>Genes compared</th>
<th>Synonymous substitution</th>
<th>Amino acid replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>OKGH I and OKGH II</td>
<td>0.101 ± 0.025</td>
<td>0.030 ± 0.008</td>
</tr>
<tr>
<td>OKGH I and CLGH</td>
<td>0.099 ± 0.024</td>
<td>0.036 ± 0.009</td>
</tr>
<tr>
<td>OKGH II and CLGH</td>
<td>0.106 ± 0.024</td>
<td>0.030 ± 0.008</td>
</tr>
</tbody>
</table>

Comparison of three genes for growth hormone (GH). OKGH indicates the gene for GH from chum salmon (O. keta) and CLGH indicates the gene for GH from European whitefish (C. lavaretus). Boldface indicates the rates adopted for estimates of the divergence between O. keta and C. lavaretus.
we compared the orthologous loci of the SmuI-div sequence of salmon (designated ON for *O. nerka*) and Bonneville whitefish (designated PS for *P. spilonotus*) by sequencing the respective products of PCR. The three sequences are aligned in Figure 7. The sequence divergence was estimated by calculating the numbers of substitutions per site between the two loci of IPM-4. The calculated sequence divergence between chum salmon and kokanee was 0.050 ± 0.012, whereas that between chum salmon and Bonneville whitefish was 0.090 ± 0.021 and that between kokanee and Bonneville whitefish was 0.095 ± 0.021. As in the cases of growth hormone and the sequence divergence of the IPM-4 loci from chum salmon and European whitefish (0.090 ± 0.021) was higher than that between the SmuI and the SmuI-cor type I (0.058 ± 0.021) and almost identical to that between the SmuI and the SmuI-cor type II (0.086 ± 0.026). From our results, we propose that the SmuI family of repeats might have been generated by horizontal gene transfer of a member(s) of the SmuI-cor family of repeats during evolution (see DISCUSSION).

**FIGURE 7.**—Sequences of orthologous loci of the SmuI-div sequence IPM-4. OK, O. *keta*; ON, O. *nerka*; and PS, *P. spilonotus*. The sequence in bold type represents a unit of integrated SmuI-like sequence. Primer sequences are underlined. Identical nucleotides are indicated by dots and deletions by dashes.

**Horizontal Transfer of SINEs:**

SINE. Moreover, this similarity is much higher than that expected from the genetic distance between the species in these two distantly related groups. (3) The sequence divergence of members of the SmuI family of repeats is low, indicating that this family of SINEs is relatively young in evolutionary terms. By contrast, the SmuI-cor family of repeats is much older than the SmuI family of repeats and members of the SmuI-cor family are distributed in all coregonid species.

There are two possible explanations for the discontinuous distribution of these two families of SINEs. One explanation is that the horizontal transfer of SINEs occurred between distantly related salmonid species. In this case, as judged from the salmonid phylogeny and the ages of the two families of SINEs, the direction of transfer would have been from one species of Coregonidae to a common ancestor of the chum and pink salmon or, alternatively, to each of these two species independently. It is unknown at present whether the amplification of the SmuI family of repeats occurred in a common ancestor of the chum and pink salmon or whether it occurred in the lineage of the pink salmon with subsequent transfer of the active source gene(s) to the chum salmon (see the accompanying article by TAKASAKI et al. 1997). Since the SmuI-cor type I subfamily is more similar to the SmuI family than is the SmuI-cor type II subfamily, a SINE(s) of the SmuI-cor type I might have been transferred horizontally during evolution to become a master (source) gene of the SmuI family of repeats. The divergence between the SmuI family of repeats and SmuI-cor type I is 0.058 ± 0.021, whereas the number of synonymous substitutions per site in genes for growth hormone between chum salmon and European whitefish is 0.099 ± 0.024. In addition, the sequence divergence of the IPM-4 loci between chum salmon and Bonneville whitefish (0.090 ± 0.021) is higher than that between SmuI and SmuI-cor type I (0.058 ± 0.021). The differences in these values (even though the level of statistical significance is only 10%) support our hypothesis that horizontal transfer of SINEs occurred during the evolution of salmonids.
We can consider introgression as another possibility. However, introgression seems less likely than the horizontal transfer of SINEs since introgression requires hybridization, and the occurrence of hybridization between such distantly related species seems very unlikely. We previously discussed the possibility of horizontal transfer of SINEs (Oshima et al. 1993; Takasaki et al. 1994, 1996). Recently, Kordis and Gubensk (1995) proposed that a member(s) of the group of art-2 retroposons, which is present specifically in the suborder Ruminantia (Duncan 1987; Lenstra et al. 1993), might have been transferred horizontally to the genome of snake, claiming this to be the first example of the horizontal transfer of a SINE(s). After publication of their article, however, it was demonstrated that the art-2 family belongs to a long interspersed element (LINE; Singer 1982) family (Szmraj et al. 1995; Okada and Hamada 1997). Therefore, the report by Kordis and Gubensk (1995) provides an interesting example of the horizontal transfer of a LINE(s) but not of a SINE(s).

Another possibility: a maintained master gene: Studies of diagnostic mutations in various subfamilies of the primate Alu family provided two possible models for the amplification of SINEs. Shen et al. (1991) and Deininger et al. (1992) proposed that all the subfamilies of Alu could have been formed by sequential mutations that occurred at a single (or a very few) master gene locus (the “master gene” hypothesis; for review, see Tachida 1996). In fact, in the case of the ID (identifier) SINE family in rodents, it has been demonstrated that the members were generated from a single master gene, namely, a BCI RNA gene (Kim et al. 1994; Kass et al. 1996). Deininger et al. (1992) pointed out that a certain function of the master gene must be essential for the host’s survival since the master gene is maintained in the genome for long periods on the evolutionary time scale (Deininger et al. 1992; Sinnett et al. 1992). The sequence of the BCI RNA gene has been well conserved for a long period of time, and it has been suggested that this tRNA-derived RNA sequence has now been “exapted” (to use the nomenclature proposed by Brosius and Gould 1992) to a new function that results in selective pressure for maintenance of the gene sequence.

In an alternative model, namely, the “multiple source genes” model (for review, see Schmid and Marais 1992), simultaneous amplification from a number of source genes (Matera et al. 1990) or a relay of active genes (Britten et al. 1988) is proposed to be responsible for the amplification of SINES. Since recent reports (Batzer et al. 1995; Shaiikh and Deininger 1996) revealed an activity responsible for the simultaneous amplification of multiple dispersed Alu source genes, it is now recognized that the “multiple source genes” model is applicable to the amplification of the primate Alu family. Our recent detailed analysis of the salmonid HpaI family of repeats also supports this type of model (Kido et al. 1994; Takasaki et al. 1994).

Another possible explanation for the generation of the SmaI family and the SmaI-cor family is that a master (source) gene for these two families of SINEs was first generated in a common ancestor of the family Salmonidae, and then this gene was vertically inherited during evolution of the salmonid lineage. In this model, members of the SmaI family of repeats and the SmaI-cor family of repeats are direct progeny of this master gene. However, there appears to be no reason to believe that the “master gene” model was operative, in particular in the case of these families of repeats, since the primate Alu and other major SINEs, such as the salmonid HpaI family of repeats, seem to support the “multiple source genes” model. Moreover, in the present case, after divergence of the subfamily Coregoninae, the master gene was not subject to amplification in the lineage of the Salmoninae and the Thymallinae but must have been maintained in their genomes during this period. After divergence of kokanee, the retained master gene acquired the capacity for amplification to become a source gene(s) for the SmaI family of repeats. Such discontinuous amplification of SINEs from a single master gene appears unlikely. Moreover, if it were to have occurred, the sequence that has been fully conserved might be expected to be present in all salmonid genomes. Using PCR, we failed, however, to find an example of the consensus sequence of the SmaI family in the genome of kokanee or in the genomes of species of Salmoninae other than chum and pink salmon (data not shown). Thus, the “master gene” model for the generation of the SmaI-cor and SmaI families of SINEs appears unlikely, even though we cannot yet completely rule out this model.

Highly diverged SmaI-div sequences: We demonstrated that highly diverged SmaI-like sequences are present in all salmonid genomes and named them members of the SmaI-div family as a whole. They were probably generated in a common ancestor of the family Salmonidae, with each accumulating neutral mutations during salmonid evolution. The relationship between the SmaI-div sequences and the SmaI-cor family of repeats can be explained by the “multiple source genes” model. In this model, progeny genes, generated from a source gene, become potential source genes when several conditions are fulfilled, for example, conditions related to local chromatin structure, hypomethylation and the acquisition of 5’ promoters and 3’ terminators that might also affect the folding of RNA, self-priming and so on (for review, see Schmid and Marais 1992; Matera et al. 1990; Schmid 1991; Liu and Schmid 1993). We suggested previously that the local environment of the genome might be important for the generation of a dominant source gene of the salmonid HpaI family of repeats at a newly retroposed locus, and we indicated that our data supported the “multiple source genes” model for this family of SINEs (Kido et al. 1994;
and pink salmon, as we discussed above. After divergence was integrated within a Tc-l-like transposon in the lineage of the subfamily Coregoninae. In the lineage of the subfamilies Thymalinae and Salmoninae, such a lineage might have been transferred horizontally to chum salmon, as we discussed above.

**The vector for horizontal transfer:** In a previous study, we found that a unit of the salmonid Hpa I SINE was integrated within a Tc-l-like transposon in the cloned DNA of Atlantic salmon (Takasaki et al. 1996). Goodier and Davidson (1994) suggested that Tc-l-like transposons might have been transferred horizontally among multicellular organisms. Thus, a Tc-l-like transposon with a SINE element might have acted as a vector for the horizontal transfer of SINES. In the Smal-cor family of repeats, we found examples similar to that described above. We found two cases in which a unit of the Smal-cor family of repeats was integrated into the Rsg-1 LINE, which was first characterized in the genome of rainbow trout (Winkfein et al. 1988) and was shown to be widely distributed in salmonid species (Y. Terai and N. Okada, unpublished results). Figure 8 shows the alignments of the sequences of Smal-cor-331 (Figure 8A) and Smal-cor-361 (Figure 8B), indicating their flanking sequences, with the consensus sequence of a unit of the integrated Smal-cor SINE.

**Figure 8.—Sequence of the Rsg-1 consensus sequences and sequences of Smal-cor-331 (A) and Smal-cor-361 (B).** Rsg-1 indicates the consensus sequence of the Rsg-1 LINE, as reported by Winkfein et al. (1988). N indicates a position of which the consensus nucleotide is unknown. Identical nucleotides are indicated by dots and deletions by dashes. Bold type indicates the sequence of a unit of the integrated Smal-cor SINE.

**Takasaki et al. 1994, 1996.** Studies of pol III-initiated B1 transcripts (Maraia 1991), as well as other lineages, also suggested that some regulatory mechanisms might preferentially "choose" the transcripts of master sequences at the RNA level during retroposition. In the case of the Smal-cor family of repeats, one of the Smal-div sequences, which included the diagnostic nucleotides of the present Smal-cor family of repeats, was chosen in such a way, and it became the source gene of the Smal-cor family of repeats in the lineage of the subfamily Coregoninae. In the lineage of the subfamilies Thymalinae and Salmoninae, such a choice did not occur during evolution. After divergence of kokanee, one member of the Smal-cor type I subfamily might have been transferred horizontally to chum and pink salmon, as we discussed above.

The vector for horizontal transfer: In a previous study, we found that a unit of the salmonid Hpa I SINE was integrated within a Tc-l-like transposon in the cloned DNA of Atlantic salmon (Takasaki et al. 1996). Goodier and Davidson (1994) suggested that Tc-l-like transposons might have been transferred horizontally among multicellular organisms. Thus, a Tc-l-like transposon with a SINE element might have acted as a vector for the horizontal transfer of SINES. In the Smal-cor family of repeats, we found examples similar to that described above. We found two cases in which a unit of the Smal-cor family of repeats was integrated into the Rsg-1 LINE, which was first characterized in the genome of rainbow trout (Winkfein et al. 1988) and was shown to be widely distributed in salmonid species (Y. Terai and N. Okada, unpublished results). Figure 8 shows the alignments of the sequences of Smal-cor-331 (Figure 8A) and Smal-cor-361 (Figure 8B), indicating their flanking sequences, with the consensus sequence of Rsg-1. In the case of Smal-cor-331, a unit of the Smal-cor family of repeats is integrated at position 212 of the reported Rsg-1 consensus sequence in the
same orientation, and, in the case of Smal-cor-361, a unit is integrated at position 485 in the reverse orientation. The Rsg-1 LINEs at these two loci might not have the ability to retropose because these sequences are truncated at the 5′ ends of the elements. If an active Rsg-1 LINE, including a SINE sequence, were to have been present in the genomes of the subfamily Coregoninae and if it were to have been transferred horizontally as are other LINEs, it is possible that a SINE might have been transferred horizontally, with Rsg-1 as the vector, from a coregonid fish to chum or pink salmon.

Recently, on the basis of observations that the 3′ ends of several families of tRNA-derived SINEs are derived from the 3′ ends of respective LINEs, our group proposed the possibility that a RTase encoded by the respective LINE family might recognize the 3′ end of a cognate SINE family, which might provide a general scheme for the mechanism by which SINEs could acquire retropositional activity (Ohshima et al. 1996; for review, see Smit 1996). Moreover, we suggested the presence of a new LINE family in the eel genome, whose 3′-end sequence is identical to the 3′-end sequence of the Smal family (Otshima et al. 1996). Therefore, this new LINE family could have been responsible for retroposition of members of the Smal family that were horizontally transferred in pink and chum salmon, as well as for that of members of the Smal-cor family in coregonid fishes.

The authors are grateful to Dr. R. A. Bodaly for gifts of coregonid samples and to Dr. T. Yamai for help in working with experiments. This work was supported by a Grant-in-Aid for Specially Promoted Research from the Ministry of Education, Science and Culture of Japan.

LITERATURE CITED


Berg, D. E., and M. M. Howe, (Editors), 1989 Mobile DNA, American Society for Microbiology, Washington, DC.


Capp, P., D. Annelabere and T. Langin, 1994 The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genet. 10: 7–12.


