

Gamera, a family of LINE-like repetitive sequences widely distributed in medaka and related fishes

A Koga¹, H Hori¹ and Y Ishikawa²

¹Division of Biological Science, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan; ²National Institute of Radiological Sciences, Chiba 263-8555, Japan

A family of repetitive sequences, designated Gamera, has been identified in the genome of the Hainan medaka fish *Oryzias latipes*, a closely related species to the common medaka fish *O. latipes*. Sequencing and Southern blot analyses of this family revealed: (1) amino acid sequence similarity to reverse transcriptase domains of long interspersed nuclear elements (LINEs); (2) 5' truncation of dispersed cop-

ies; and (3) the disruption of another genetic element, indicating a past transposition event. These results suggest that Gamera belongs to the LINE superfamily. Gamera is widely distributed in the genus *Oryzias*, and the phylogenetic relationship might indicate its presence in the common ancestor of the genus.

Heredity (2002) **89**, 446–452. doi:10.1038/sj.hdy.6800162

Keywords: transposon; LINE; medaka; fish; distribution; phylogeny

Introduction

The medaka fish *Oryzias latipes* is a small freshwater teleost native to eastern Asia, including China, Korea and Japan. It is, together with the zebrafish *Danio rerio*, a useful model animal for studies of vertebrate genetics and developmental biology, mainly because of its small body size, high reproductive rate, and large and transparent eggs allowing easy manipulation (Iwamatsu, 1997; Ishikawa, 2000; Packer, 2001; Wittbrodt *et al.*, 2001). Evolution, especially on the time scale of species divergence, is also a research field in which the medaka fish has advantages. This is because several closely related species have been identified and their detailed evolutionary relationship is known, as in the case of the *melanogaster* species subgroup of *Drosophila*. Phylogenetic links have been studied not only in terms of nucleotide sequence data (Naruse *et al.*, 1993; Naruse, 1996; Koga *et al.*, 2000), but also with regard to karyotypes (Magtoon and Uwa, 1985; Uwa, 1991) and reproductive isolation based on meiotic segregation in interspecific hybrids (Uwa, 1991; Sakaizumi *et al.*, 1992). Their habitats cover a wide range of Asia, stretching from Japan to India.

Transposable elements are thought to be factors contributing to genome evolution because of their transposition activity causing mutations and their repetitive nature giving rise to chromosomal rearrangements (Moran *et al.*, 1999; Kidwell and Lisch, 2000). Representatives of each of the two major classes of transposable elements, DNA-based elements and RNA-mediated elements, have been identified in the medaka fish. *Tol1* (Koga *et al.*, 1995) and *Tol2* (Koga *et al.*, 1996) are elements of the former

class. Examples of the latter include OLR1 (Naruse *et al.*, 1992), *mermaid* (Shimoda *et al.*, 1996), *Swimmer 1* (Duvernell and Turner, 1998), *Rex1* (Volff *et al.*, 2000), *Rex3* (Volff *et al.*, 1999), *Rex6* (Volff *et al.*, 2001b) and *Poseidon* (Volff *et al.*, 2001a). Some of these have been examined for their distribution among species and an impact on genome evolution has been suggested.

Long interspersed nuclear elements (LINEs) comprise a major type of RNA-mediated element (Hutchison *et al.*, 1989; Smit, 1999). They lack the long terminal repeats (LTRs) found in retrovirus-like elements, hence they are also called non-LTR retrotransposons. We have recently identified a family of LINE-like repetitive sequences residing in the medaka genome. This family, which we have named Gamera, does not exhibit a strong nucleotide sequence similarity to the LINEs so far found in the medaka fish, but amino acid sequence similarities are observed with LINEs from various organisms. The present paper describes these common features. We have also investigated the element's distribution in the genus *Oryzias*, and the results suggest Gamera is an ancient resident of the genus.

Materials and methods

Medaka fish and other species

A laboratory stock of the medaka fish *O. latipes*, originally collected in the Nagoya area, was used as the sample for this species in the present study. Three more laboratory stocks were also employed for the examination of the intraspecific distribution of Gamera, as explained in the results and discussion section.

Eight species of the genus *Oryzias* were obtained from the World Medaka Aquarium of the Nagoya City Higashiyama Zoological Garden. These species, and their original collection sites are: *O. celebensis*, Ujung Pandang,

Correspondence: A Koga, Division of Biological Science, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan.

E-mail: koga@bio.nagoya-u.ac.jp

Received 20 March 2002; accepted 16 July 2002

Indonesia; *O. curvinotus*, Hong Kong, China; *O. javanicus*, Singapore; *O. luzonensis*, Ilocos Norte, Philippines; *O. mekongensis*, Kara Sin, Thailand; *O. melastigma*, Chidambaram, India; *O. minutillus*, Bangkok, Thailand; *O. nigrimas*, Lake Poso, Indonesia.

In addition, swordtail fish (*Xiphophorus helleri*) and zebrafish (*Danio rerio*) were purchased from a pet shop in Nagoya. Commercially available genomic DNAs were used for comparison with the chicken and human.

Analysis of genomic DNA

Southern and dot blotting and subsequent hybridization experiments were performed as described by Koga and Hori (1999), except that the AlkPhos Direct System (Amersham Pharmacia Biotech Ltd), instead of radioisotopes, was used for probe labelling and signal detection. For the nine species of the genus *Oryzias*, a 4 µg aliquot of genomic DNA was used for each gel slot. For species outside the genus, 20 µg, five times this amount, was applied. This was to retain the detectability of the hybridization analysis, which is expected to decrease as the genome size of the target species increases. The size of the haploid genome of the medaka fish has been estimated to be 0.68–0.85 × 10⁹ bp (Tanaka, 1995). Its lower limit is two-fifths of one estimate for zebrafish (1.7 × 10⁹ bp, Postlewait et al, 1994) and slightly more than one fifth of that for human beings (3.0 × 10⁹ bp, Gardiner, 1995). Thus, at least for zebrafish and the human, the genome size should not be a significant factor lowering the detectability when hybridization signals are weak or absent.

Hybridization probes were prepared by polymerase chain reaction (PCR) amplification followed by purification with QIAquick columns (QIAGEN GmbH). The regions used for the probes are explained in each case.

Other molecular techniques

PCR, cloning and sequencing were conducted as previously described (Koga and Hori, 1999, 2000).

Results and discussion

Encounter with Gamera

Tol1 is a DNA-based transposable element found in the medaka fish *O. latipes*, and also present in the Hainan medaka fish *O. curvinotus* (Koga et al, 1995). It is known to be an active element because we have recently observed its *de novo* transposition (unpublished results). However, an autonomous copy, which is expected to carry a gene for its transposase, has hitherto not been identified. For the purpose of finding such a copy, we examined length variation of *Tol1* copies randomly collected from genomic libraries of *O. latipes* and *O. curvinotus*. This is because internal deletions, giving rise to shorter elements, are a common phenomenon in DNA-based elements and, therefore, a longer *Tol1* copy, if found, could be taken as a candidate for an autonomous copy. Among 42 *Tol1* copies examined, the majority of which were 1.9 kb, one copy was encountered exhibiting a length of 6.0 kb. The genomic DNA clone containing this 'long *Tol1* copy' along with its flanking chromosomal regions, which originated from *O. curvinotus*, was designated curLT and sequenced. Comparing the sequence data with the first found, 1.9-kb-long, *Tol1* copy (EMBL accession number D42062; designated as *Tol1-tyr* because

found in the *tyrosinase* gene) revealed that, in the curLT clone, a *Tol1* copy is disrupted by an extra DNA fragment of about 4.5 kb (Figure 1). This extra fragment is a repetitive sequence because, as shown below (Figure 3), multiple hybridization bands appeared on genomic Southern blots using part of it as a probe. This repetitive sequence family is different from *Tol1* because it did not hybridize to any of the 42 *Tol1*-containing clones, except for the curLT clone, in dot blot assays (data not shown). The family consisting of the newly found repetitive sequences was named Gamera, and the particular Gamera copy found in the curLT clone was designated Gamera-cur1.

To define the extent of Gamera-cur1, we compared the sequence of curLT with, in addition to *Tol1-tyr*, four other *Tol1* copies randomly chosen from the 42 *Tol1*-containing clones. Nucleotides matching at least one of the five *Tol1* copies were regarded as part of *Tol1*, and the remaining sequence of 4439 bp was taken to be Gamera-cur1. This sequence has been deposited in the EMBL database with the accession number AB081572. Its terminal regions may involve nucleotides that are not part of Gamera but of *Tol1*, because their boundaries could not be defined precisely due to sequence variation among the *Tol1* copies. Cloning and sequencing of other Gamera copies may help to determine the boundaries.

Gamera has amino acid sequence similarity to LINE reverse transcriptases

As shown in Figure 1, Gamera-cur1 contains three 'open reading frames' (defined as described in the legend to Figure 1) of more than 150 amino acids. The longest 'open reading frame' is for 388 amino acids and a BLAST search with its sequence as the query resulted in a list of LINEs from various organisms. Amino acid sequence alignments among elements with relatively high scores are shown in Figure 2. The aligned regions were their reverse transcriptase domains. The highest score was obtained with the SjR2 element of the human blood fluke *Schistosoma japonicum*. However, this was the case when the 388 amino acid sequence was used as the query sequence, and the score order changed when other parts were used. In addition, the 388 amino acid region itself may be part of the entire reverse transcriptase domain of a more complete copy of Gamera which may be present somewhere else in the genome. For these reasons, it is not clear, from the present data, which LINE family Gamera is most closely related to.

Gamera copies are truncated at various 5' sites

It is common among LINE families for them to consist of full-length copies and shorter copies of various lengths, with the latter lacking the 5' regions (Eickbush, 1994). This phenomenon is thought to be caused by incomplete reverse transcription, a step included in the proliferation cycle of LINEs. We therefore examined if this phenomenon of 5' truncation also occurs in Gamera. Four parts of Gamera-cur1, designated a to d in the direction of 5' to 3' (see Figure 1), were amplified by PCR, and Southern blots against medaka fish genomic DNA were performed with these probes separately (Figure 3). The result was fewer hybridization bands observed with probe a than with b, b than c, and c than d, indicating that Gamera involves copies that are truncated at various 5' sites, as is the case for most LINEs reported to date.

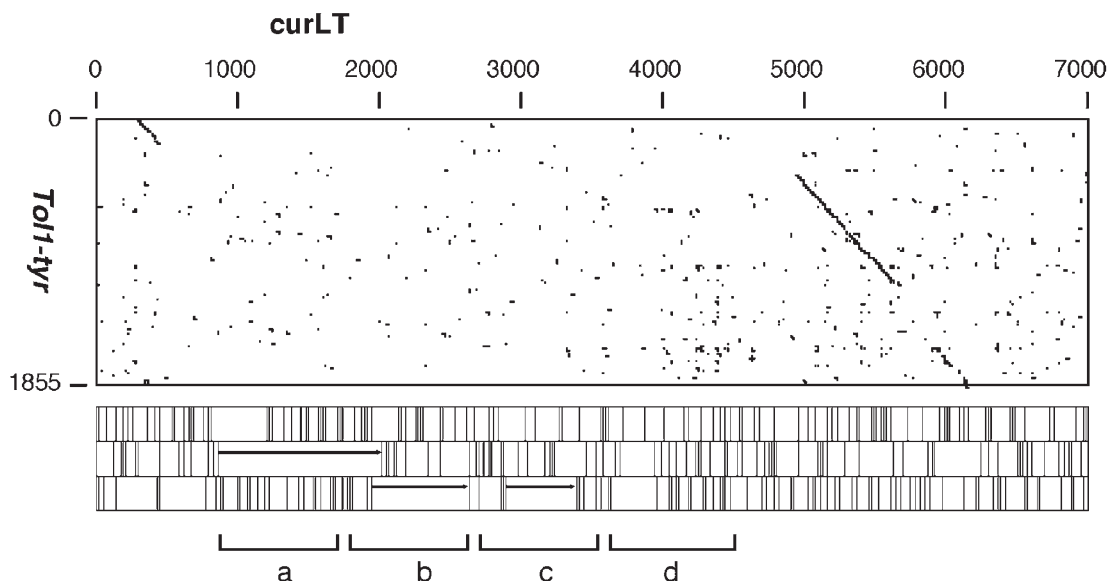


Figure 1 Dot matrix analysis of the curLT and *Tol1-tyr* nucleotide sequences. curLT (abscissa) is a genomic DNA fragment of *Oryzias curvinotus* containing a 6.0-kb-long *Tol1* copy. *Tol1-tyr* (ordinate) is a 1.9-kb-long *Tol1* copy first found in *O. latipes*. The criterion for matching was a 70% match over a window of 20 nucleotides. The dot matrix shows that, in curLT, *Tol1* is disrupted by an extra DNA fragment of 4.5 kb. Positions of possible stop codons are shown in the boxes under the dot matrix. The arrows in the boxes are 'open reading frames' of more than 150 amino acids included in the extra fragment. An 'open reading frame' was defined here as a stretch of amino acids not necessarily starting with the ATG codon, and is thus more loosely defined than this term's usual meaning. The bars with letters a to d are regions used for probes in Southern hybridization experiments. Their locations on the EMBL AB081572 sequence are: a, 453-1252; b, 1353-2152; c, 2253-3052; d, 3153-4052.

The tail region of Gamera exhibits a high copy number. To estimate the copy number of Gamera in the medaka fish genome, we conducted a dot blot assay (Figure 4). Using the already known copy number of the medaka fish *Tol2* transposable element as the standard, we estimated the Gamera copy number per diploid genome to be as few as 40 for probe a and more than 2500 for probe d. Such a drastic variation in the copy number along the element is often observed for LINES (cf. Hutchison *et al*, 1989).

Gamera-cur1 appears to be an incomplete copy

Full-length copies of human L1 contain two open reading frames (starting with the ATG codon as generally defined), designated ORF1 and ORF2. ORF1 encodes a DNA-binding protein and ORF2 includes endonuclease and reverse transcriptase domains. In case of the 6.0-kb-long L1 with EMBL accession number U93569, ORF1 encodes 338 amino acids and ORF2 is for 1275 amino acids. Equivalent structures of similar sizes are seen with full-length copies of other LINES such as *Swimmer 1* of the medaka fish (Duvernell and Turner, 1998) and Maui of fugu (Poulter *et al*, 1999). The longest 'open reading frame' in Gamera-cur1 is for 388 amino acids, and this region corresponds to only a central portion of the L1 ORF2. In addition, it is located near to the 5' terminus of Gamera-cur1 (see Figure 1). A probable explanation for these findings is that Gamera-cur1 was first generated as a truncated copy and mutational changes have since accumulated.

An adenine-rich tail region and target site duplication are also features of LINES. Possible nucleotide segments are seen in the curLT clone but convincing evidence is lacking at present.

Gamera is widespread in the *Oryzias* genus

Southern blot analysis for the presence/absence of Gamera-hybridizing sequences was performed against genomic DNAs of nine species of the genus *Oryzias* and a number of other species outside the genus (Figure 5). All the nine species in the genus exhibited hybridization bands, the intensity differing among species. There appears to be a negative correlation between the signal intensity and the genetic distance from *O. curvinotus*, from which the probe DNA originated. This result suggests that Gamera was present in the common ancestor of these species and has accumulated mutational changes in each lineage. Proliferation of different variants in different lineages is another possible explanation, but this is less likely than mutation accumulation because it does not necessarily require a negative correlation between the hybridization signal intensity and the genetic distance.

Hybridization signals were not observed for swordtail fish, zebrafish, chicken and human. Gamera thus seems to be absent in these species or else too divergent to be detected with Southern hybridization.

We also examined the distribution of Gamera within the species *O. latipes*, which demonstrates considerable geographical variation and is composed of four major regional populations (Sakaizumi, 1986; Sakaizumi *et al*, 1987). Southern blot analysis of four stock fishes representing the four regional populations revealed Gamera to be present in all the fishes examined, at similar copy numbers (data not shown).

Gamera phylogeny is in line with that of the hosts

To confirm, by nucleotide sequence analysis, the inference that Gamera is an ancient resident of the genus, part of Gamera was amplified by PCR and then sequenced.

a

```

Gamera  INTDPPTKEEIVSVIKSLKSNKAPGHDNLTAELFKDTTETASKILOPFKTVWQSASITPE
SjR2    INIGPPSLSEVVKATGNLKRGRAGAPDGLTPEIFKEGGPVLAARITELLARIWELDVIPS
Swimmer KLLTAITEETIQFAIRKLKSGKMAGADGFSPEWYKTMETHLIPTLLKTENWVMEKKTTP
L1Hs    SLNREITGSEIVAIINSLPTKKSFGPDGFTAIFYQRYKEELVPFLKLFQSIKEGILEN
PsCR1   GEXPSVLKEQVQDYLERIDXHKSMDXDXHPRVLRRELADVXEPLAIFEXSIRLGEVDP

Gamera  EWTGKVIKVPKKG-TLSECNWNGITLLSIPSKIMAKIINRLSEAVN--TTLRKEQAG
SjR2    DWSRTLTPVFKKG-QKSSCDNYRGISLTNIVSKILASILLRRLTKAHE--EQTRNQGG
Swimmer SNNKAITSIIPKEGKDRLDCANYRPVSVINIDYKLFSTIISRLETILP--MLIHKDOTG
L1Hs    SFYEASITLIPKPGRDTHKKENFRPTSLMNTDAKILNKILANRIQQHIK--KLIIHDOVG
PsCR1   DTKKANIVPIFKKG-KKEDPGNYRPVSLTSXPGKIMEQVLKESILRHLERKVIKXSQHG

Gamera  FRKGRRCCTEQIFALRNIEQSAEWQQQ-LFFNFIDFEKAFDSVHRDSLWRILRAYGVPTH
SjR2    FRPGRGCIDQIETLROVLEHRHTERRP-TIAVELDLKAAFDSDIKVMWQCLSLKGVPEK
Swimmer FIKRQQTQDSIRKVLHITHQVQKQKE-TLVISLDAEKAFDSVRWTFIYKVLGKFGCKS
L1Hs    FIPGMQGFNIRKSIINVLQHINRAKDKNHVSIISDAEKAFDKIQPPMLKTLNKLIDGM
PsCR1   FTKGKSCLTNLIIFYEEVTGSDVVGKA-VDVIYLDSEKAFDTVSHSITISKLKKVGLDEW

Gamera  MIKLIRSFYNN--CRCCVGGSDIWFEEVKTGVRQGCVMSSALLEFNVVIDWVMRHTTEAG-E
SjR2    YNLIQALYSNTTCRVRAYGRLSSELTSSGVRQACPLSPFLNFIIIDILLELTLSSDF
Swimmer ITETISGLYNKPTARIKINGDFTETITLERGTROGCNMSALLEALYIPLGQWIRQRADI
L1Hs    YLKIIRAIYDKPTANIILNGQKLEAFPLKTGTROGCPLSPFLNFIVLEVLARAIQKEKI
PsCR1   TKIWIENWLXHQARVINGSMSSVQPPVTSQVPGSGVLGPVLENIIFINDLEDGVDCTLSK

Gamera  TGIWTLFSTLEDLDFADDLALISHNYRHMQDKTDRLSRFGSQVGLKINQSK
SjR2    PGVDLFPQDKLTDLEYADDIVLSEADAKMODFLTTLNMNVSMGMRFSPSK
Swimmer KGVVSGKEQKLSLFADDLLLTISOPTKTLPIIMDSKDECTLSGKINNVK
L1Hs    KGTQLGKEEVKLSLFADDIMIVLENPIVSAQNLLKLISNFSKVSGYKINVOK
PsCR1   FADDTKLGVDVDTLEGRDRIQKDLNKLEDWAKRNLMRFNKDKCRVLHLGWKN

```

b

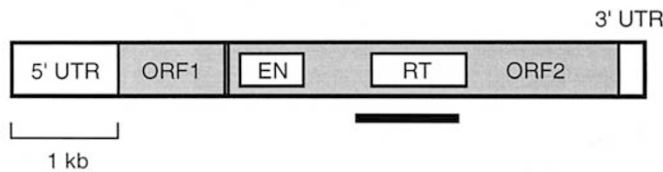


Figure 2 Amino acid sequence alignments for Gamera and LINEs. (a) Alignments were made using the CLUSTAL W program (Thompson et al, 1994). The Gamera sequence in the top line is heavily shaded, along with identical amino acid residues in the other four sequences. Lightly shaded residues are chemically similar amino acids. Grouping was: nonpolar, AFILMPVW; polar, CGNQSTY; basic, DE; acidic, HKR. Designations of sequences and their EMBL accession numbers are: Gamera, Gamera-cur1 of *Oryzias curvinotus*, AB081572; SjR2, the SjR2 element of the human blood fluke *Schistosoma japonicum*, AY027869; Swimmer, the Swimmer 1 element of the medaka fish, AF055641; L1Hs, the L1 element of human, U93569; PsCR1, the PsCR1 element of the turtle *Platemys spixii*, AB005891. (b) The aligned regions in panel (a) were reverse transcriptase domains. The range of the region in the case of human L1 is indicated by a bar under the overall structure of L1. It corresponds to amino acid positions 448-737 in ORF2 consisting of 1275 amino acids. The scheme was redrawn from Sassaman et al, (1997). Designations are: UTR, untranslated regions; ORF, open reading frame; EN, endonuclease domain; RT, reverse transcriptase domain.

The PCR primers (nucleotides 688-715 and 1538-1511 of AB081572) were designed to represent two 28-bp regions of Gamera-cur1 where amino acid sequences are relatively highly conserved among Gamera and human L1, and, in addition, to end at the first or second nucleotides of codons. Amplification was successful with five of the nine species shown in Figure 5, which are relatively closely related to *O. curvinotus*. Phylogenetic analysis of these five sequences (Figure 6) provided results in line with those for the host species.

Gamera is a family different from known medaka LINEs. Five LINE or LINE-like families, to our knowledge, have been reported in the medaka fish. They are *Swimmer 1* (EMBL AF055641, Duvernell and Turner, 1998), *Rex1* (AJ288486, Volff et al, 2000), *Rex3* (AJ400430, Volff et al, 1999), *Rex6* (AJ293518, Volff et al, 2001b) and *Poseidon* (AJ293655, Volff et al, 2001a). Dot matrix analyses of

nucleotide sequences were conducted between Gamera and these elements, and also Gamera and human L1 (U93569, Sassaman et al, 1997). There was no indication of any difference in similarity among the five combinations (Gamera vs *Swimmer 1*, Gamera vs *Rex1*, etc.; data not shown).

Of the above five medaka fish elements, a complete copy has been identified only for *Swimmer 1*. This element has all the components that characterize LINEs: a 5' untranslated region (UTR), an ORF1, an ORF2, a 3' UTR and an adenine-rich tail. In other words, its structure is similar to those of most known LINEs. However, *Swimmer 1* has specific features. First, its copy number per genome is as low as 10-20, in contrast to the 10^4 or more for many LINEs. Second, 5' truncation is rare, although truncated copies are much more frequent than complete copies with many LINEs. In these respects, Gamera does not resemble *Swimmer 1*, exhibiting features of 'typical'

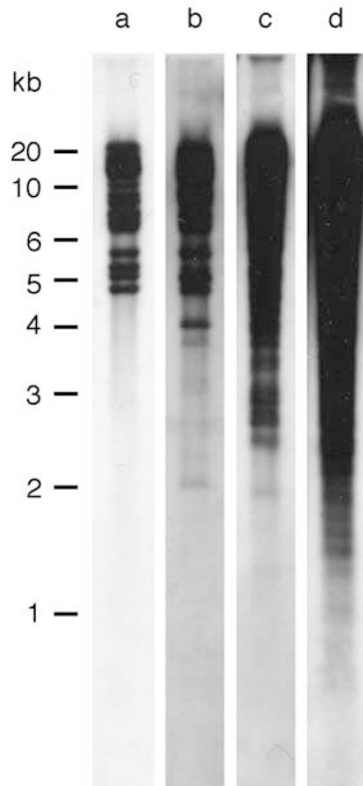


Figure 3 Genomic Southern blot analysis with different parts of Gamera-cur1. Medaka fish genomic DNA, 16 μ g from a single fish, was digested with *Sac*I, divided into 4 aliquots, each containing 4 μ g, and electrophoresed on a single 0.8% agarose gel. The nylon membrane blotted with the DNA was cut into four parts and hybridized with probes a to d (see Figure 1) separately. The conditions for hybridization and signal detection were the same for all four sets. The probes used are shown over the autodiagrams. The sizes and mobilities of the size marker DNA fragments are indicated along the left margin.

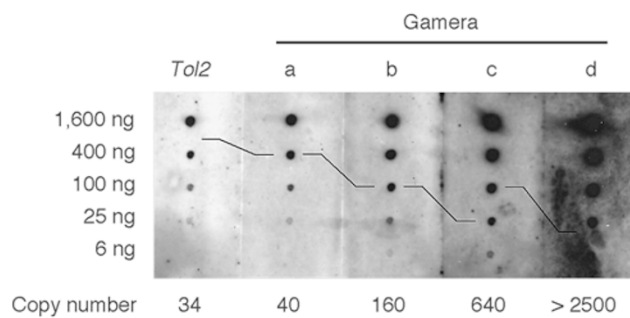


Figure 4 Dot blot assay to estimate the copy number of Gamera. Five strips of nylon membranes were blotted with medaka fish genomic DNA of different concentrations, indicated in the left margin. The strips were, separately but under the same conditions, hybridized with the probes indicated at the top. *Tol2* is a multi-copy DNA-based transposable element and its copy number per diploid genome of the fish used had been estimated to be exactly 34 by the method described in Koga *et al* (2000). Dots exhibiting nearly the same intensities were selected as indicated by the lines, and the copy numbers for probes a to d were estimated by using the copy number of *Tol2* as the standard. The estimated copy numbers per diploid genome are shown at the bottom. All the five probes were 800–900 bp in length. The region used for the *Tol2* probe is nucleotides 575–1392 of EMBL D84375. For Gamera probes a to d, see legend to Figure 1.

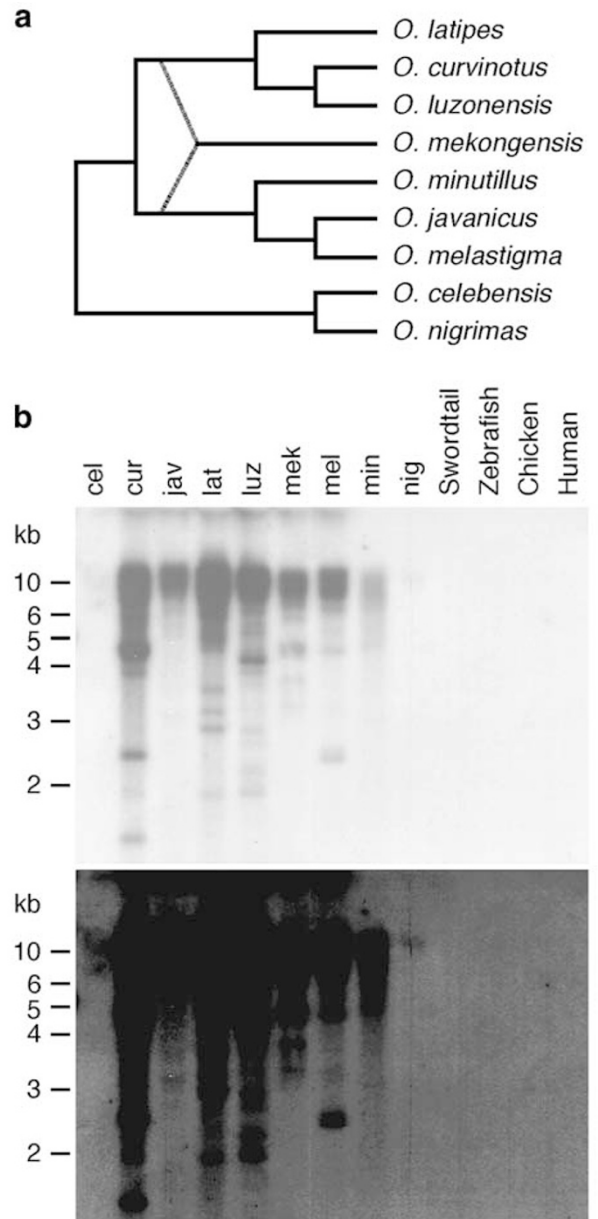
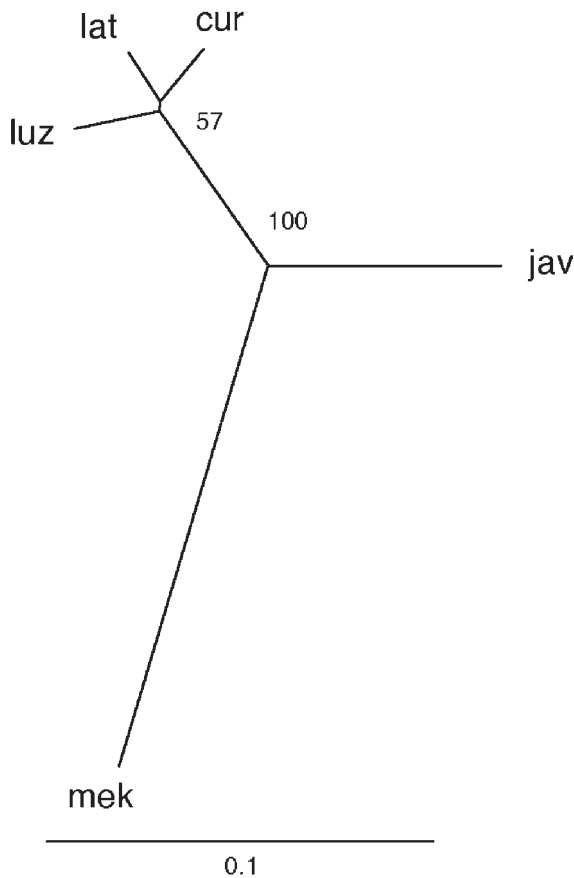


Figure 5 Distribution of Gamera among species. (a) Phylogenetic tree of the nine species of the genus *Oryzias* used in the present study. The scheme was redrawn from references (Naruse *et al*, 1993; Naruse, 1996) and personal communication (JS Albert and K Naruse). Note that the phylogenetic position of *O. mekongensis* has not been well established. (b) Southern blot analysis to examine the distribution of Gamera among species. The names of species in the genus *Oryzias* are abbreviated to their first three letters. Genomic DNAs (see text for amounts) were digested with *Bam*HI and electrophoresed on a 0.8% agarose gel. The nylon membrane blotted with the DNAs was hybridized with probe b (see Figure 1). The sizes and mobilities of the size marker DNA fragments are indicated along the left margin. The two autodiagrams were produced from the same hybridization membrane, with different durations of exposure to film. The upper is after 2-h exposure, and the lower after 18-h exposure. Hybridization signals appear in the lower autodiagram for *O. celebensis* and *O. nigrimas*. Further longer exposure did not produce band-like signals for the four species outside the genus *Oryzias* (not shown).



References

- Duvernell DD, Turner BJ (1998). *Swimmer 1*, a new low-copy-number LINE family in teleost genomes with sequence similarity to mammalian L1. *Mol Biol Evol* **15**: 1791–1793.
- Eickbush TH (1994). Origin and evolutionary relationships of retroelements. In: Morse (ed) *The Evolutionary Biology of Viruses*, Raven Press: New York. pp 121–157.
- Gardiner K (1995). Human genome organization. *Curr Opin Genet Dev* **5**: 315–322.
- Hutchison CA, Hardies SC, Loeb DD, Shehee WR, Edgell MH (1989). LINEs and related retrotransposons: long interspersed repeated sequences in the eukaryotic genome. In: Berg DE, Howe MM (eds) *Mobile DNA*, American Society for Microbiology: Washington, DC. pp 593–617.
- Ishikawa Y (2000). Medakafish as a model system for vertebrate developmental genetics. *Bioessays* **22**: 487–495.
- Iwamatsu T (1997). *The Integrated Book for the Biology of the Medaka* (in Japanese), Daigaku Kyoiku Publisher: Okayama, Japan.
- Kidwell MG, Lisch DR (2000). Transposable elements and host genome evolution. *Trends Ecol Evol* **15**: 95–99.
- Koga A, Hori H (1999). Homogeneity in the structure of the medaka fish transposable element *Tol2*. *Genet Res Camb* **73**: 7–14.
- Koga A, Hori H (2000). Detection of *de novo* insertion of the medaka fish transposable element *Tol2*. *Genetics* **156**: 1243–1247.
- Koga A, Inagaki H, Bessho Y, Hori H (1995). Insertion of a novel transposable element in the tyrosinase gene is responsible for an albino mutation in the medaka fish, *Oryzias latipes*. *Mol Gen Genet* **249**: 400–405.
- Koga A, Shimada A, Shima A, Sakaizumi M, Tachida, Hori H (2000). Evidence for recent invasion of the medaka fish genome by the *Tol2* transposable element. *Genetics* **155**: 273–281.
- Koga A, Suzuki M, Inagaki H, Bessho Y, Hori H (1996). Transposable element in fish. *Nature* **383**: 30.
- Magtoon W, Uwa H (1985). Karyotype evolution and relationship of small ricefish, *Oryzias minutillus*, from Thailand. *Proc Jpn Acad Ser B* **61**: 157–160.
- Moran JV, DeBerardinis RJ, Kazazian HH Jr. (1999). Exon shuffling by L1 retrotransposition. *Science* **283**: 1530–1534.
- Naruse K (1996). Classification and phylogeny of fishes of the genus *Oryzias*. *Fish Biol J Medaka* **8**: 1–10.
- Naruse K, Mitani H, Shima A (1992). Highly repetitive interspersed sequence isolated from genomic DNA of the Medaka, *Oryzias latipes*, is conserved in three other related species within the genus *Oryzias*. *J Exp Zool* **262**: 81–86.
- Naruse K, Shima A, Matsuda M, Sakaizumi M, Iwamatsu T, Soroto B et al (1993). Distribution and phylogeny of rice fish and their relatives belonging to the suborder Arianichthyoidei in Sulawesi, Indonesia. *Fish Biol J Medaka* **5**: 11–15.
- Packer A (2001). Medaka on the move. *Nat Genet* **28**: 302.
- Postlewait JH, Johnson SL, Midson CN, Talbot WS, Gates M, Ballinger EW et al (1994). A genetic linkage map of the zebrafish. *Science* **264**: 699–703.
- Poulter R, Butler M, Ormandy J (1999). A LINE element from the pufferfish (*fugu*) *Fugu rubripes* which shows similarity to the CR1 family of non-LTR retrotransposons. *Gene* **227**: 169–179.
- Sakaizumi M (1986). Genetic divergence in wild populations of Medaka, *Oryzias latipes* (Pisces: Oryziatidae) from Japan and China. *Genetics* **69**: 119–125.
- Sakaizumi M, Uwa H, Jeon SR (1987). Genetic diversity of the East Asian populations of the freshwater fish, *Oryzias*. *Zool Sci* **4**: 1003.
- Sakaizumi M, Shimizu Y, Hamaguchi S (1992). Electrophoretic studies of meiotic segregation in inter- and intraspecific hybrids among East Asian species of the genus *Oryzias* (Pisces: Oryziatidae). *J Exp Zool* **264**: 85–92.
- Sassaman DM, Dombroski BA, Moran JV, Kimberland ML, Naas

Figure 6 Phylogenetic tree constructed with Gamera sequences from five *Oryzias* species. PCR amplification was successful for five of the nine species (see text), and the PCR products were cloned into a plasmid vector. For each species, a single clone was randomly chosen and sequenced, and the data have been deposited in the EMBL database. The five species designated by their first three letters, the lengths of the sequences and the EMBL accession numbers are: cur, 795 bp, AB081573; jav, 795 bp, AB081574; lat, 795 bp, AB081575; luz, 789 bp, AB081576; mek, 782 bp, AB081577. The tree was constructed using the neighbour-joining method. Gaps in the luz and mek data were excluded from the calculation. The bootstrap percentages obtained with 1000 replicates are shown at the nodes.

LINEs, as also suggested to be the case for *Rex6* and *Poseidon*. The medaka fish genome appears to contain multiple, and possibly many, lineages of high-copy-number LINEs, similarly to mammalian genomes.

Acknowledgements

The authors thank H Hashikawa, M Sato and S Susaki (World Medaka Aquarium of the Nagoya City Higashi-yama Zoological) for providing fish material. We are also grateful to DL Hartl (Harvard University), H Tachida (Kyushu University) and S Hamada (Nagoya University) for helpful discussion. This work was supported by grant no. 13216205 to AK from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

- TP, DeBerardinis RJ et al (1997). Many human L1 elements are capable of retrotransposition. *Nat Genet* **16**: 37–43.
- Shimoda N, Chevrette M, Ekker M, Kikuchi Y, Hotta Y, Okamoto H (1996). *Mermaid*: a family of short interspersed repetitive elements widespread in vertebrates. *Biochem Biophys Res Commun* **220**: 226–232.
- Smit AF (1999). Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* **9**: 657–663.
- Tanaka M (1995). Characteristics of medaka genes and their promoter regions. *Fish Biol J Medaka* **7**: 11–14.
- Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Uwa H (1991). Cytosystematic study of the Hainan medaka, *Oryzias curvinotus*, from Hong Kong (Teleostei: Oryziidae). *Ichthyol Explo Freshwaters* **1**: 361–367.
- Volff JN, Hornung U, Scharf M (2001a). Fish retrotransposons related to the *Penelope* element of *Drosophila virilis* define a new group of retrotransposable elements. *Mol Genet Genomics* **265**: 711–720.
- Volff JN, Korting C, Froschauer A, Sweeney K, Scharf M (2001b). Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J Mol Evol* **52**: 351–360.
- Volff JN, Korting C, Scharf M (2000). Multiple lineages of the non-LTR retrotransposon *Rex1* with varying success in invading fish genomes. *Mol Biol Evol* **17**: 1673–1684.
- Volff JN, Korting C, Sweeney K, Scharf M (1999). The non-LTR retrotransposon *Rex3* from the fish *Xiphophorus* is widespread among teleosts. *Mol Biol Evol* **16**: 1427–1438.
- Wittbrodt J, Shima A, Scharf M (2001). Medaka, a model organism from the Far East. *Nat Rev Genet* **3**: 53–64.