SIZE VARIATION AND STRUCTURAL CONSERVATION OF VERTEBRATE TELOMERASE RNA*

Mingyi Xie[‡], Axel Mosig[‖], Xiaodong Qi[‡], Yang Li[‡], Peter F. Stadler[¶,,#], Julian J.-L. Chen[‡,§,1]**

From the [‡]Department of Chemistry & Biochemistry and [§]School of Life Sciences, Arizona State University, Tempe, AZ 85287, USA

[‖]Department of Combinatorics and Geometry, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences Campus, Shanghai, China

[¶]Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, D-04107 Leipzig, Germany

[**]Department of Theoretical Chemistry, University of Vienna, A-1090 Wien, Austria

[#]Santa Fe Institute, Santa Fe, NM 87501, USA

Running head: Structure and function of vertebrate telomerase RNA

[1]Address correspondence to: Julian J-L Chen, Department of Chemistry & Biochemistry, Arizona State University, Tempe, AZ 85287-1604, Tel: 480-965-3650; Fax: 480-965-2747; E-mail: JLChen@asu.edu

Telomerase extends chromosome ends by copying a short template sequence within its intrinsic RNA component. Telomerase RNA (TR) from different groups of species varies dramatically in sequence and size. We report here the bioinformatic identification, secondary structure comparison and functional analysis of the smallest known vertebrate TRs from five teleost fishes. The teleost TRs (312-348 nt) are significantly smaller than the large cartilaginous fish TRs (478-559 nt) and tetrapod TRs. This remarkable length reduction of teleost fish TRs correlates positively with their genome size, reflecting an unusual structural plasticity of TR during genome evolution. The teleost TR structure, lacking most of the variable sequences, defines the minimum consensus of the vertebrate TR structure. Despite their small sizes, the medaka and fugu TRs, when assembled with their catalytic reverse transcriptase protein counterparts, reconstituted active and processive enzymes. The CR4-CR5 domain of teleost fish TR is relatively small and yet displays a higher efficiency than the pseudoknot domain in reconstituting telomerase activity. With the identification and characterization of teleost fish TR, we now have a complete view of the evolutionary divergence of vertebrate TR.

Telomeres are specialized DNA protein complexes that cap chromosome ends and are important for genome stability and cellular proliferation (1). Telomeres consist of repetitive DNA sequences and a variety of telomere-associated proteins. The length of telomeric DNA in most eukaryotes is maintained by telomerase, a specialized ribonucleoprotein (RNP) enzyme. Telomerase adds telomeric DNA repeats to chromosome ends to counterbalance the natural shortening that occurs during DNA replication. The telomerase RNP enzyme consists of at least two essential core components, the catalytic protein component telomerase reverse transcriptase (TERT), and the telomerase RNA (TR) that provides a template for telomeric DNA synthesis.

TR is remarkably variable in size, sequence and even secondary structure between different groups of eukaryotes. To date, TR sequences have been identified in 28 ciliates, 14 yeasts and 38 vertebrates. Due to the lack of sequence similarity between groups of species, the TR secondary structures have been determined independently for each of these three groups (2). The vertebrate TR secondary structure is composed of three highly conserved structural domains: the pseudoknot/template domain, the CR4-CR5 domain and the scaRNA domain (3-5). The pseudoknot/template domain contains a template region for telomeric DNA synthesis, and a conserved pseudoknot structure essential for telomerase activity. The CR4-CR5 domain together with the pseudoknot/template domain are both required for reconstituting active telomerase (6). However, their mechanistic roles are unclear. The scaRNA domain is crucial for the 3'-end processing of TR and telomerase RNP biogenesis

in vivo (3,7). While this three-domain TR structure is conserved in tetrapods and cartilaginous fish (4), TR has not yet been identified or studied in teleost fish which comprises almost half of the extant vertebrate species.

Teleost fish is the most diverse vertebrate group (8), and is distinct from the cartilaginous fish. The teleost and tetrapods (including amphibian, reptile, birds and mammals) diverged from each other around 450 million years ago. Since then, teleost fish have undergone genome duplication and rediploidization, resulting in the amazing level of genomic diversity. The relatively faster evolution rate and the consequent diversity in teleost fish offer an attractive model for evolutionary studies. However, identification of TR from teleost fish using degenerate PCR or BLAST search has not been successful due to a high degree of sequence variation in TR.

Here we report the identification of TRs from five teleost fishes, *Danio rerio, Oryzias latipes, Gasterosteus aculeatus, Takifugu rubripes* and *Tetraodon nigroviridis*, using a novel bioinformatics method. To functionally analyze the teleost TR structure, we have cloned TERT protein genes from medaka, fugufish, and zebrafish, and reconstituted telomerase activity for medaka and fugufish. The structural and functional analyses of teleost fish telomerase enzyme provide important new insights into the evolution of vertebrate telomerase RNP.

Experimental Procedures

Bioinformatics search of teleost fish TR sequences. A sequence search was performed using fragrep2. The input pattern, shown in supplemental Fig. S1, consists of eight position-specific weight matrices (PWMs). The quality of match between a PWM and a DNA sequence is measured as a fraction of similarity above an unavoidable background (9). The computational approach and implementation details of fragrep2 are described in detail in Mosig et al, 2007. Our search pattern was generated by annotating the eight conserved regions in the TR alignment published in Chen et al., 2000, and converted to a fragrep2 search pattern using the aln2pattern tool. Both fragrep2 and aln2pattern are available for download from http://www.bioinf.uni-leipzig.de/Software/. The initial search pattern (Fig. S1) resulted in a single plausible hit in the medaka genome (assembly MEDAKA1). A BLAST search using medaka sequence as query against other teleost fish genomes revealed homologs in the stickleback (assembly BROAD S1), fugu (assembly FUGU 4.0) and tetraodon (assembly TETRAODON 7). Based on the four teleost TR sequences, a modified and less stringent search pattern was generated, with which we found 79 candidate sequences in the zebrafish genome (assembly Zv6). These were screened using INFERNAL (10) and the secondary structure annotated TR alignment from the Rfam database (11), resulting in a single sequence which fit well with both other teleost candidates and with the previously known vertebrate TR sequences. The alignment of all 40 known vertebrate TR sequences can be obtained from the Telomerase Database at http://telomerase.asu.edu/.

Genomic DNA and total RNA isolation. For isolation of genomic DNA and total RNA, medaka fish (*Oryzias latipes*) were purchased from Aquatic Eco-Systems (Apopka, FL), and Zebrafish (*Danio rerio*) were obtained from Dr. Yung Chang (Arizona State University, AZ) or purchased from Aquatical Tropicals, Inc (Plant City, FL). Green spotted pufferfish (*Tetraodon nigroviridis*) were purchased from AquariumFish.net. Liver tissue of fugu (*Takifugu rubripes*) fish was obtained from Dr. Shugo Watabe (University of Tokyo, Japan).

Genomic DNA was isolated from 50-100 mg of fish tissue using the DNAzol reagent (Invitrogen) following manufacture's instruction. Stickleback (*Gasterosteus aculeatus*) genomic DNA was a generous gift from Dr. David Kingsley (Stanford University, CA). Total RNA was isolated from 100-200 mg gill or liver tissues using 1 ml Trizol reagent (Life Technologies, USA) following manufacture's instructions. Concentrations of DNA and RNA samples were determined by OD_{260} measurement using the Nanodrop ND-1000 spectrophotometer (NanoDrop Technologies).

Sequencing and cloning of TR genes. To verify the sequences, teleost fish TR genes were PCR amplified from genomic DNA and the PCR products were sequenced directly. The verified sequences of five teleost fish TR genes were deposited into GenBank with the following

accession numbers: EF569636 (*Danio rerio*), EF569637 (*Oryzias latipes*), EF569638 (*Takifugu rubripes*), EF680233 (*Tetraodon nigroviridis*) and EF680234 (*Gasterosteus aculeatus*).

For medaka, zebrafish and fugu, the PCR products of TR genes were cloned into the EcoR-V site of the pZero vector (Invitrogen) to generate pMedaka-TR, pZebrafish-TR and pFugu-TR. Plasmids were sequenced to confirm sequence accuracy of the cloned TR genes.

Identification and cloning of teleost fish TERT genes. To reconstitute telomerase activity, we cloned telomerase reverse transcriptase (TERT) genes from medaka, zebrafish and fugu. The fugu TERT (AY861384) and medaka TERT (DQ248968) gene sequences have been identified and were available from GenBank (12). The zebrafish TERT gene was identified in this study via a BLAST search of the zebrafish genome database using the fugu TERT protein sequence as query. The exact 5' and 3'-ends of the full-length zebrafish TERT cDNA sequence were determined by the 5'- and 3'-rapid amplification of cDNA ends (RACE) using a SMART-RACE cDNA Amplification Kit (Clontech, USA). The cDNA sequence was determined by direct sequencing of the RT-PCR products. The sequence of zebrafish TERT gene has been deposited into GenBank with the accession number EF202140.

To clone the TERT genes, the coding sequences of medaka and zebrafish TERT genes were PCR amplified from the cDNA samples that were reverse transcribed from total RNA samples using Thermoscript reverse transcriptase (Invitrogen) and an oligo-dT18 reverse primer. The fugu TERT cDNA was PCR amplified from a cDNA library obtained from Dr. Byrappa Venkatesh (Institute of Molecular and Cell Biology, Singapore). The PCR products of the medaka, zebrafish and fugu TERT cDNAs were cloned into the pCITE vector for *in vitro* synthesis of the recombinant TERT proteins.

In vitro transcription of TR. RNA was prepared by T7 in vitro transcription using PCR DNA fragments as template as described previously (13,14).

Northern blotting analysis. Twenty micrograms of total RNA were resolved on a 4% polyacrylamide/8 M urea denaturing gel and electrotransferred to Hybond-XL membrane (Amersham) at 0.5 A for 1h. The membrane was UV cross-linked and prehybridized at 65°C for 30 min in 20 ml of UltraHyb hybridization buffer (Ambion). Riboprobes with sequences complementary to the target RNA were generated by in vitro transcription from a PCR DNA template that contained the T7 promoter. Riboprobes were synthesized and labeled with [α-^{32}P] UTP using a MaxiScript kit (Ambion). After incubation at 37°C for 1h, one microliter of RNase-free DNase I (2 U/μl) was added followed by a 20 min incubation at 37°C to remove the DNA template. Riboprobes were then purified using microspin G-25 columns (GE Healthcare). The membrane was hybridized at 65°C overnight in 20 ml of prehybridization buffer (see above) with the riboprobe added at 1X10^6 c.p.m./ml. The hybridized membrane was washed twice in 20 ml of 1X SSC (3.0 M NaCl and 0.3 M sodium citrate, pH 7.0)/ 0.2% SDS for 10 min and twice in 20 ml of 0.2X SSC/ 0.1% SDS for 30 min at 65°C. The blot was analyzed using a phosphorimager, Bio-Rad FX Pro.

In vitro reconstitution of telomerase. Human, medaka, fugu and zebrafish telomerases were reconstituted using the TnT (Transcription and translation) quick coupled rabbit reticulocyte lysate system (Promega). Briefly, recombinant TERT protein was synthesized in 10 μl of rabbit reticulocyte lysate at 30ºC for 60 min following the manufacturer's instructions. To assemble the telomerase complex, in vitro synthesized TR was added to a final concentration of 1 μM in the TnT reaction of TERT synthesis, and incubated at 30°C for 30 min. For human and medaka RNA titration experiments, the pseudoknot/template or CR4-CR5 RNA fragment was saturated at 3 μM, while the other RNA fragment was added at various concentrations.

Conventional telomerase activity assay. Enzymatic activity of in vitro reconstituted telomerase was determined by a direct primer extension assay. A 10 μl reaction was carried out with 3 μl of in vitro reconstituted telomerase sample in the presence of 1x PE buffer (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 2 mM DTT, 3 mM $MgCl_2$ and 1 mM spermidine), 1 mM dATP, 1

mM dGTP, 1 mM dTTP, 2 pmole 5'-^{32}P end labeled $(TTAGGG)_3$ at 30°C for 2 hours. The products were subjected to phenol/chloroform extraction and ethanol precipitation, followed by 10% denaturing PAGE. Gels were dried, and products were detected and analyzed using a Bio-Rad FX Pro Imager. For each reaction, activity was determined by measuring the total intensity of extended substrate primer, correcting for background, and normalizing against unextended primer (loading control). Relative activities were obtained by dividing the activity of each reaction by that of the reaction with saturated concentration of RNA fragments. For the titration assay, the relative activities were plotted against concentrations of RNA fragment and the nonlinear regression curve fitting was carried out using one site binding (hyperbola) equation, Y=Bmax*X/(Kd+X) (Prism 5, Graphpad software, Inc.).

Results

A novel bioinformatics approach to identify TR sequences. Despite significant efforts to clone TRs from a diverse array of vertebrate species, TR sequences have not been identified from teleost fish (4). Similarly, computational searches using the Basic Local Alignment Search Tool (BLAST) of the available fish genomes have been unsuccessful (data not shown). The inability to identify TR sequences in teleost fish using PCR or BLAST presumably stems from the fact that vertebrate TRs are conserved only in eight relatively short regions (called Conserved Region 1-8, or CR1-CR8) that are interrupted by highly variable sequences with a large number of indels (4). BLAST as well as specialized tools for searching RNA genes were inadequate for identifying reasonable TR candidates in the teleost fish genomes.

To identify TR sequences, we employed an improved homology search tool, fragrep2, to search teleost fish genomes. The original version of fragrep program implements a specialized algorithm for homology search that considers gap-free sequence patterns separated by variable-length regions of non-aligned sequence (15). This approach has been demonstrated to work well for genome-wide searches of non-coding RNAs (ncRNAs) (15,16). However, it had not been successful in finding teleost fish TRs. This is because even the relatively well-conserved blocks, i.e. CR1-CR8, contained many variations to be well represented by a single consensus sequence. To circumvent this, in fragrep2, we have replaced consensus sequences by position-specific weight matrices (PWMs) to search for matched DNA sequences (17). As shown in supplemental Fig. S1, the initial search pattern contains a collection of PWMs as well as minimal and maximal distances between these PWM blocks.

Using this new approach, we successfully found a TR candidate in the medaka genome. Homologs of this medaka sequence could then be readily found by means of BLAST in stickleback, fugu and tetraodon genomes. All four sequences are flanked upstream by an ADP-ribosylation factor and downstream by homologs of human LASP1 and/or PLXDC2 (Table S1). Based on the alignment of the four teleost fish sequences, we modified the search pattern and were able to retrieve a single convincing candidate from the zebrafish genome. Surprisingly, the genomic location of the zebrafish TR candidate is neither syntenic with that of the other teleost sequences nor with the human locus (Table S1). All five teleost TR genes were PCR amplified from genomic DNA samples and the PCR DNA products were sequenced directly to verify the sequences identified from the genome databases (see Experimental Procedures).

Unique transcription elements of fish TR genes. Analysis of genomic sequences upstream of the fish TR-coding sequences revealed transcriptional elements typical of an RNA polymerase II promoter: a conserved TATA box-like and a CCAAT box element (Fig. S2). This suggests that, like other vertebrate TRs, teleost TRs are products of RNA polymerase II. Interestingly, a putative CRE-BP1/c-Jun binding element, located between the TATA and CCAAT boxes, is conserved in all fishes and some amphibians (bullfrog and horned frog), but not in other vertebrates (Fig. S2). This data suggest an evolutionary change in transcriptional regulation of TR gene along the tetrapod lineage.

The compact size of teleost fish TR. To confirm the presence of the identified teleost TR

transcripts in cells, we performed northern blotting analysis to detect the endogenous TRs. The medaka and zebrafish TRs were each detected as a single band on the northern blot (Fig. 1A, lane 1). Based on the northern result, the size of the endogenous medaka and zebrafish TRs are slightly smaller than the in vitro transcribed RNA markers that are 317 and 322 nt, respectively (Fig. 1A, compare lane 1 and 2).

To determine the actual size of the endogenous TR, we mapped the 5'-ends of medaka and zebrafish TRs by 5'-RACE. The results showed that the 5'-ends of both medaka and zebrafish TRs lie 14 nucleotides upstream of the template sequence. Assuming that the 3'-end of the fish TR is located, like other vertebrate TRs, 3 residues downstream of the box ACA motif, the medaka and zebrafish TRs are predicted to be 312 and 317 nt long, respectively, consistent with the sizes observed from the northern analysis. Based on sequence alignment, the other three teleost TR homologs are predicted to be 348 nt (stickleback), 325 nt (fugu) and 328 nt (Tetraodon). This makes teleost TRs the smallest among all known vertebrates, as the size of previously known vertebrate TRs ranges from 382 to 559 nt (4).

Teleost fishes have notably small genomes, while the cartilaginous fishes have relatively large genomes (18). Intriguingly, teleost fishes with smaller genomes have the smallest TRs, while cartilaginous fishes with larger genomes have the largest TRs (from 478 to 559 nt) among vertebrates. By plotting the TR size over the genome size, we found that the dramatic size variation of TR correlates positively with their genome size with an R^2 value of 0.5007 and a p value < 0.0001 (Fig. 1B). This strong correlation suggests that the size variation of fish TR resulted from evolution of the entire genome.

Secondary structure of teleost fish TR. To determine if these small teleost TRs share a similar secondary structure with other vertebrate TRs, we constructed secondary structure models for teleost fish TRs using phylogenetic comparative analysis. The primary sequences of the five teleost TRs identified were aligned manually as described previously (4). The eight conserved regions CR1-CR8 found previously in 35 vertebrate TRs, are largely conserved in the teleost TRs (Fig. 2). Because of their small size and the presence of the CR sequences, teleost fish TR sequences can be readily aligned without much ambiguity. The aligned sequences were analyzed for covariations to derive a conserved secondary structural model for the teleost TR (Fig. 3A). Homologous to the structures of other vertebrate TRs, the proposed teleost structure contains eleven helices (P1, P2a, P2b, P3, P4, P5, P6, P6.1, P7a, P7b and P8) grouped into three separate structural domains: the pseudoknot/template domain, the CR4-CR5 domain and the snoRNA domain (Fig. 3A). All helices, except for the P6.1 and P7a, were supported with at least one co-variation per helix. All five teleost TRs share a similar secondary structure with variation mostly in the hypervariable region between the P4 and P5 helices (Fig. 3A and supplemental Fig. S3).

Being the smallest, the teleost TR resembles the essential core of vertebrate TR (Fig. 3B). It contains shorter linker sequences between the three conserved domains. The commonalities and differences of the vertebrate TR structures are discussed in detail below.

Pseudoknot/template domain. The pseudoknot/ template domain consists of a highly conserved pseudoknot structure, the template sequence, and the P1 helix that defines the boundary of the RNA template. The pseudoknot structure consists of the P2a-P2b and the P3 helices that are universally present in vertebrate TRs (Fig. 3B). The mammalian pseudoknot however contains an additional helix P2a.1 that extends the P2a helix (Fig. 3B, human TR). This mammal-specific P2a.1 helix is essential for human telomerase activity and is possibly involved in binding to the TERT protein (19). In teleost TR, the P2a and P2b helices are separated by a conserved asymmetric (0/6) internal loop (Fig. 3A), whereas, in other groups of vertebrates, this internal loop contains a slightly variable number of residues. The P3 helix, in tetrapods, is conserved as a 9 base-pair helix with a single nucleotide bulge (Fig. 4A, tetrapods). The shark and ray P3 helix has the same length but with a two-nucleotide bulge at a different position (Fig. 4A, sharks and rays). Medaka TR interestingly lacks any bulge in its P3 helix, while other teleost TRs have a one-nucleotide bulge at the position identical to the shark's. Notably, the lack of bulge

in the medaka P3 helix seems to be compensated by extensions of the P3 helix and J2b/3 loop (Fig. 4A, medaka). The variation of the length and position of the bulge in the P3 helix suggests that it might not be a critical element for the function or structure of the pseudoknot structure. Deletion of the bulge in the human P3 helix results in a minor reduction of telomerase activity (20,21). The real role of the P3 bulge however remains to be revealed. Based on an NMR solution structure, the pseudoknot of human TR forms a triple helix that involves 5 base triples and a base pair at the junction of P2b and P3 helices (21). The sequences that form the triple helix are absolutely conserved even in teleost TR, confirming its critical role in telomerase function (Fig. 4A). In contrast, the distal portion of the P3 helix and the J2b/3 are less conserved, and are slightly variable in length and sequence (Fig. 4A, teleost panel).

In all five teleost and most vertebrates, except for some rodents, TRs possess a long-range interacting P1 helix upstream of the template region (Fig. 3). In human TR, the P1 helix consists of two individual helices, P1a and P1b, separated by an internal loop. The teleost P1 helix is substantially shorter, containing only the P1b equivalent portion while lacking the P1a portion. The integrity of P1b helix and its distance from the template defines the boundary of the RNA template (22). In human telomerase, disruption of the P1b helix alters the template boundary, resulting in template usage outside of the normal template. Likewise, disruption of the P1 helix in medaka TR also altered the template boundary (data not shown). This supports the notion that the P1 helix is also the element for template boundary definition in teleost telomerase.

CR4-CR5 domain. The CR4-CR5 domain, in addition to the pseudoknot/template domain, is a structural element essential for *in vitro* telomerase activity. The P6 and P6.1 helices in this domain are universally present in all know vertebrate TRs (Fig. 3B). Remarkably, the sequence (5'-**A<u>AGAG</u>N<u></u>UNGN<u>CUCU</u>G**-3') of the P6.1 stem-loop is highly conserved even in the teleost fish. It was previously thought that the invariant sequence of the P6.1 helix loop was due to a biased sequence collection that resulted from the PCR amplification strategy used for cloning most of the vertebrate TRs (4). This PCR strategy presumably amplified only the TR sequences with conserved sequence in the P6.1 stem-loop, part of the annealing site of the PCR reverse primers. However, all five teleost fish TRs were identified through bioinformatic approach, instead of PCR. The structure, not the sequence, of the P6.1 helix is known to be important for telomerase activity *in vitro* as compensatory mutations that maintain the helical structures of P6.1 do not reduce activity of reconstituted telomerase (14). Surprisingly, similar compensatory mutations of P6.1 helix resulted in reduced telomerase activity reconstituted *in vivo* (23). The absolute sequence conservation in the P6.1 helix suggests that, in addition to its based-paired structure, the sequence of this helix might be important for the telomerase function *in vivo*.

The teleost TR, lacking the distal stem-loop P6b, consists only of the P6 (i.e. homologous to the P6a in human TR), P.6.1 and P5 helices in the CR4-CR5 domain (Fig. 4B). While the P6b helix is dispensable in teleost fish and some tetrapods such as turtle and frog, the proximal part of P6b stem-loop is required, as a species-specific element, for human telomerase activity (24). The single-stranded regions, J5/6 and J6.1/5, at the three-way junction between P5, P6 and P6.1 helices are relatively more variable in teleost than in other vertebrates. Although its essential role in telomerase function is evident, the mechanistic function of the CR4-CR5 domain remains to be uncovered.

SnoRNA/scaRNA domain. The 3'-portion of vertebrate TR contains a unique secondary structure (hairpin-hinge-hairpin-tail) and sequence motifs (box H and ACA) that are critical for TR biogenesis and shared by the box H/ACA snoRNAs (7). Most vertebrate TRs contain an additional motif called CAB box that is shared by the small Cajal body RNAs (scaRNAs) (3). While the box H and ACA are important for RNA localization to nucleoli, the CAB box is important for localization to the Cajal body where RNP complex assembly is thought to take place (25). Interestingly, teleost TR lacks an obvious CAB box (UGAG) in the CR7 region (Fig. 4C). The lack of CAB box implies that teleost TR might not localize to the Cajal body. Since the Cajal body has been suggested to play a role in telomerase regulation and telomere recruitment (26), it would be interesting to understand TR localization in

teleost and its correlation with the regulation of telomerase function.

Medaka and fugu telomerases are processive. Telomerase activity *in vitro* requires both the TR component and the catalytic TERT protein. To functionally characterize the structural elements of teleost TR, we reconstituted telomerase complex from *in vitro* synthesized TERT protein and TR. We cloned the fugu and madaka TERT genes and synthesized the recombinant TERT proteins in vitro in rabbit reticulocyte lysate followed by assembling with *in vitro* transcribed TR fragments (see Experimental Procedures). Active telomerases were successfully reconstituted for medaka and fugu, confirming the authenticity of the teleost telomerase components cloned (Fig. 5).

Conventional primer-extension assay using the reconstituted enzymes allowed us to examine processivity of various teleost telomerases. One of the determinants for telomerase processivity is the length of the alignment region in the RNA template (13). As predicted from the presence of the 4-nucleotide alignment sequence in their RNA templates, the reconstituted medaka and fugu telomerases are processive, generating a typical 6-nucleotide ladder pattern of the elongated products (Fig. 5). However, the pseudoknot fragment of zebrafish TR failed to generate telomerase activity when assembled with medaka and fugu TERT proteins (Fig. 5, lanes 7-9 and 16-18), suggesting a cross-species incompatibility of zebrafish pseudoknot with TERT protein.

To analyze activity of zebrafish telomerase, we thus identified and cloned zebrafish TERT cDNA (see Experimental Procedures). Unexpectedly, the *in vitro* synthesized zebrafish TERT protein failed to reconstitute a detectable activity when assembled with zebrafish, medaka or fugu TRs (data not shown). Based on the alignment of TERT amino acid sequences, the cloned zebrafish TERT protein was unlikely to be a product of alternative splice variant, as it contained all essential motifs. The possibility of mutations in the cloned zebrafish TERT gene was ruled out as identical sequences were found from two individual zebrafish obtained from different sources. While gene duplication is relatively common in teleost, more rigorous BLAST searches of the zebrafish genome did not reveal any other candidate sequences for the TERT gene. We speculate that the *in vitro* synthesized zebrafish TERT protein, unlike the medaka and fugu TERT proteins, might not fold correctly as the recombinant zebrafish TERT protein migrated faster than expected on SDS-PAGE (data not shown).

The CR4-CR5 domain is the main determinant in TR for functional binding to medaka TERT. Vertebrate TERT protein possesses two RNA-binding sites that bind independently to the CR4-CR5 and pseudoknot domains of the TR. As shown previously, human TERT is functionally compatible with the mouse CR4-CR5 domain but not the mouse pseudoknot domain (13). In this study, we also showed that the fugu TERT protein reconstituted telomerase activity with CR4-CR5 RNA fragments, but not the pseudoknot domain, from other teleost fish species (Fig. 5, lanes 10-15) or even distantly related vertebrates such as human, quoll, chicken, turtle, frog and shark (data not shown). This difference in cross-species compatibility indicated that the CR4-CR5 domain is functionally more conserved across a wide variety of species than the pseudoknot domain. Unlike the fugu TERT, the medaka TERT assembled with the fugu pseudoknot RNA to reconstitute telomerase activity with a low processivity (Fig. 5, lanes 4-6), suggesting a more relaxed RNA binding specificity of medaka TERT protein.

While reconstituting teleost fish telomerase, we observed a significantly lower activity of the reconstituted enzyme using two RNA fragments, than that of the enzyme reconstituted using the full-length RNA (data not shown). To determine which RNA fragment was responsible for the lower activity reconstituted, we carried out the *in vitro* reconstitution with titrations of each of the two RNA fragments as well as the full-length RNA. We define the median effective concentration (or EC_{50}) as the RNA concentration required to generate 50% of the saturated activity of reconstituted telomerase. It is noteworthy that this EC_{50} value measured in this assay is related only to the functional binding (or assembly) of the RNA fragment to the TERT protein, excluding non-specific or non-functional bindings. A lower EC_{50} value of the RNA indicates a higher assembly efficiently with the TERT protein to

generate active telomerase enzyme. Remarkably, the CR4-CR5 fragments and the full-length TR gave rise comparable EC50 values. The medaka CR4-CR5 and full-length RNAs had EC_{50} values of 87.4 and 85.9 nM, respectively, while the human CR4-CR5 and full-length RNAs had EC_{50} values of 203.9 and 241.6 nM, respectively (Fig. 6). In comparison, the medaka and human pseudoknot RNA fragments had high EC_{50} values of 506.2 and 523.5 nM, respectively (Fig. 6). The reduction of reconstituted activity at high concentrations of the full-length TR might be due to the multimerization or aggregation of TR as previously reported (27). Our result indicates that the CR4-CR5 domain is the main determinant for efficient binding and assembly of TR to the TERT protein.

Discussion

Unlike to the TERT, TR is prominently divergent in size, sequence and even structure. The lack of sequence similarity between groups of species has made it difficult to identify TR genes through sequence homology searches. In this study, by using a novel bioinformatics approach, we have successfully identified TR sequences from five teleost genomes. Unexpectedly, the teleost fish TRs are the smallest, to date, among known vertebrate TRs, and remarkably resemble the minimal consensus of vertebrate TRs. The structural and functional analyses of teleost fish telomerase provide important insights into the structural evolution of vertebrate TR as well as the co-evolution of the TR and TERT protein.

Fast evolution of TR structure and size. Owing to the various numbers of species-specific structural elements, the size of TR is remarkably variable, up to one order of magnitude, from 150 nt in ciliates to 1500 nt in yeasts. Even within vertebrates, the size of TR varies from 312 nt in medaka fish to 559 nt in shark. Since all vertebrate TRs share the same essential structural domains, the size variation is mainly due to deletions or insertions in the linker regions between conserved domains.

From the evolutionary point of view, the emergence or disappearance of structural elements or linker sequences in TR over a short evolutionary time scale is rather intriguing. The unusual plasticity of TR structure was likely facilitated by the non-lethal and progressive nature of the consequences of TR mutations. In organisms with long telomeres, the impact of telomerase mutations is delayed for a number of generations (28). Such delay could allow an accumulation of secondary mutations, some of which might compensate for the initial deleterious mutation, eventually leading to emergence of novel structural elements in TRs.

A possible scenario for the emergence of new structural elements is the insertion of a transposable element into the TR gene during evolution. For example, the scaRNA or snoRNA domain in the vertebrate TR is absent in both the ciliate and yeast TRs, and has been acquired during evolution along the vertebrate lineage. Since some snoRNA and scaRNA have recently been suggested to contain characteristics of retrotransposons (29), it is possible that a transposition event may occurred and fused a mobile scaRNA gene with an ancestral TR gene. Identification of TRs from early branching chordates such as sea squirt will provide crucial clues on the origin of the vertebrate-specific structural domains.

Because most vertebrates, including the early-branched cartilaginous fish, contain the scaRNA-specific motif (CAB box), we propose that it was a scaRNA, rather than a snoRNA, being inserted into the vertebrate TR gene. Teleost fish and some bird TRs that lack an obvious CAB box, might have subsequently evolved to function without a CAB box motif. Notably, other scaRNAs, e.g. U100, from teleost fish contain a conserved CAB box sequence (30).

Based on the phylogenetic tree derived from the aligned TR sequences, tetrapods, teleost fishes and cartilaginous fishes are grouped into three monophyletic clades (Fig. 7), representing three separated evolutionary lineages that lead to three distinct size groups of TR molecules. Cartilaginous and teleost fish TRs evolved in opposite directions toward size expansion and reduction, respectively, corresponding to their genome size evolution. It is generally believed that the small sizes of teleost genomes are mainly due to the low abundance of transposable elements and the significant reduction in intron size (31). Our data suggest that genome compression affected not

only the intergenic or intronic DNA sequences but also the RNA genes. Similarly, teleost RNase P RNA is about 50 nt shorter than the 350 nt long human RNase P RNA.

Interestingly, teleost fish TR appears to be more divergent than cartilaginous fish TR from tetrapod TR (Fig. 7). This is consistent with a recent comparative genomic study that showed a higher degree of sequence conservation between the human and elephant shark genomes than that of human and teleost fish genomes (32,33). It is generally believed that the teleost fish has experienced a genome duplication after diverging from tetrapod lineage and before the fish radiation (34). However, no extra TR gene or pseudoknot gene was found in the 5 teleost fish species, suggesting either the teleost TR gene was not duplicated or the duplicated TR copy has been lost from the common ancestor of teleost fish.

Co-evolution of the TR and TERT protein. During structural diversification, the function of the telomerase RNP has to be conserved through co-evolution between the RNA and protein components, which can be reflected by the interspecies compatibility of the components. For example, the CR4-CR5 domain of TR appears to be highly conserved and exchangeable between species to reconstitute telomerase activity with medaka TERT (Fig. 5 and data not shown). In contrast, the pseudoknot/template RNA domain appears to be incompatible even between the closely related species (e.g. between medaka and fugu, or between human and mouse), suggesting a fast co-evolution rate between the pseudoknot RNA domain and the TERT protein.

The triple helix within the pseudoknot domain and the P6.1 helix in the CR4-CR5 domain, both containing invariant sequences, are the two most conserved structural elements in vertebrate TRs, (Fig. 4A). However, the detailed functional roles of these two structural elements have yet to be uncovered. Interestingly, the triple helix seems to be an ancient feature conserved in many species (2,21,35). It is, thus, unlikely to be responsible for the interspecies incompatibility of the pseudoknot domains. The distal helix of P3 stem and J2b/3 loop, on the other hand, demonstrate some extent of variation among vertebrate species (Fig. 4A). Swapping the whole pseudoknot structure (P3, P2b and J2b/3) between medaka and fugu TRs did not improve their inter-species compatibility (data not shown).

The teleost CR4-CR5 domain is considerably smaller than other vertebrates as it lacks the distal P6b helix. Nonetheless, the smaller medaka CR4-CR5 RNA fragment (50 nt) exceeds its human counterpart (89 nt) in effectiveness of reconstituting telomerase activity *in vitro* (Fig.. 6). The higher assembly efficiency is likely due to a higher binding affinity between the medaka TERT protein and the CR4-CR5 RNA fragment, which will require co-evolution between the medaka TERT protein and the TR. Since the P6b helix in the CR4-CR5 domain of human TR is essential for binding to the human TERT protein (24), the human TERT might have evolved with an additional binding pocket for the P6b helix.

While we were able to reconstitute activity from medaka and fugu telomerases, it is unclear why the zebrafish TERT failed to reconstitute detectable telomerase activity. Among the five teleost species studied, zebrafish branches out early and is more divergent than the other four teleost fishes (36). While the zebrafish TR is considerably conserved compared with the other four fishes, we speculate that the zebrafish TERT protein might require additional factors, such as specific chaperone proteins, for correct folding and proper assembly with the TR *in vitro*.

In summary, the identification of teleost TR and characterization of its structure and function now provide a complete picture of the unusual divergence of vertebrate TR. The novel bioinformatic tool fragrep2 is an effective approach to find notoriously divergent TR sequences in eukaryotic genomes. The small teleost fish TR and the large cartilaginous fish TR reflect the unusual plasticity of TR structure during evolution. Teleost fish telomerase is very processive and contains a functional P1 helix that defines the template boundary. The conservation of the structure and function of teleost fish telomerase supports the use of teleost fish as a model organism for the study of telomerase biology.

Acknowledgments - We thank Tracy Niday and Drs Mary Armanios, Jim Allen and Juli Feigon for critical reading of the manuscript. We

are grateful to Drs. Byrappa Venkatesh, Shugo Watabe, David Kingsley and Yung Chang for providing fugu ovary cDNA, fugu liver tissue, stickleback genomic DNA and zebrafish, respectively.

References

1. Ferreira, M. G., Miller, K. M., and Cooper, J. P. (2004) *Molecular cell* **13**(1), 7-18
2. Chen, J.-L., and Greider, C. W. (2004) *Proc Natl Acad Sci U S A* **101**(41), 14683-14684
3. Jady, B. E., Bertrand, E., and Kiss, T. (2004) *J Cell Biol* **164**(5), 647-652
4. Chen, J.-L., Blasco, M. A., and Greider, C. W. (2000) *Cell* **100**(5), 503-514.
5. Chen, J.-L., and Greider, C. W. (2004) *Trends Biochem Sci* **29**(4), 183-192
6. Tesmer, V. M., Ford, L. P., Holt, S. E., Frank, B. C., Yi, X., Aisner, D. L., Ouellette, M., Shay, J. W., and Wright, W. E. (1999) *Mol. Cell. Biol.* **19**(9), 6207-6216
7. Mitchell, J. R., Cheng, J., and Collins, K. (1999) *Molecular and cellular biology* **19**(1), 567-576
8. Nelson, J. S. (2006) *Fishes of the world*, 4th Ed., Wiley, New York
9. Kel, A. E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O. V., and Wingender, E. (2003) *Nucleic Acids Res* **31**(13), 3576-3579
10. Eddy, S. R. (2006) *Cold Spring Harbor symposia on quantitative biology* **71**, 117-128
11. Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S. R. (2003) *Nucleic Acids Res* **31**(1), 439-441
12. Yap, W. H., Yeoh, E., Brenner, S., and Venkatesh, B. (2005) *Gene* **353**(2), 207-217
13. Chen, J.-L., and Greider, C. W. (2003) *EMBO J* **22**(2), 304-314
14. Chen, J.-L., Opperman, K. K., and Greider, C. W. (2002) *Nucleic Acids Res* **30**(2), 592-597
15. Mosig, A., Sameith, K., and Stadler, P. F. (2006) *Genomics Proteomics Bioinformatics* **4**(1), 56-60
16. Backofen, R., Bernhart, S. H., Flamm, C., Fried, C.*, et al.* (2007) *Journal of experimental zoology. Part B* **308**(1), 1-25
17. Mosig, A., Chen, J.-L., and Stadler, P. F. (2007) *Lect. Notes Comput. Sc.* **4645**, 335-345
18. Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B.*, et al.* (2007) *Nucleic Acids Res* **35**(Database issue), D332-338
19. Ly, H., Blackburn, E. H., and Parslow, T. G. (2003) *Molecular and cellular biology* **23**(19), 6849-6856
20. Comolli, L. R., Smirnov, I., Xu, L., Blackburn, E. H., and James, T. L. (2002) *Proc Natl Acad Sci U S A* **99**(26), 16998-17003
21. Theimer, C. A., Blois, C. A., and Feigon, J. (2005) *Molecular cell* **17**(5), 671-682
22. Chen, J.-L., and Greider, C. W. (2003) *Genes Dev* **17**(22), 2747-2752
23. Ly, H., Calado, R. T., Allard, P., Baerlocher, G. M., Lansdorp, P. M., Young, N. S., and Parslow, T. G. (2005) *Blood* **105**(6), 2332-2339
24. Mitchell, J. R., and Collins, K. (2000) *Molecular cell* **6**(2), 361-371.
25. Richard, P., Darzacq, X., Bertrand, E., Jady, B. E., Verheggen, C., and Kiss, T. (2003) *EMBO J* **22**(16), 4283-4293
26. Tomlinson, R. L., Ziegler, T. D., Supakorndej, T., Terns, R. M., and Terns, M. P. (2006) *Molecular biology of the cell* **17**(2), 955-965
27. Ren, X., Gavory, G., Li, H., Ying, L., Klenerman, D., and Balasubramanian, S. (2003) *Nucleic Acids Res* **31**(22), 6509-6515
28. Blasco, M. A., Lee, H. W., Hande, M. P., Samper, E., Lansdorp, P. M., DePinho, R. A., and Greider, C. W. (1997) *Cell* **91**(1), 25-34.
29. Weber, M. J. (2006) *PLoS genetics* **2**(12), e205
30. Vitali, P., Royo, H., Seitz, H., Bachellerie, J. P., Huttenhofer, A., and Cavaille, J. (2003) *Nucleic Acids Res* **31**(22), 6543-6551

31. Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., *et al.* (2004) *Nature* **431**(7011), 946-957
32. Venkatesh, B., Kirkness, E. F., Loh, Y. H., Halpern, A. L., *et al.* (2006) *Science* **314**(5807), 1892
33. Venkatesh, B., Kirkness, E. F., Loh, Y. H., Halpern, A. L., *et al.* (2007) *PLoS biology* **5**(4), e101
34. Meyer, A., and Van de Peer, Y. (2005) *Bioessays* **27**(9), 937-945
35. Shefer, K., Brown, Y., Gorkovoy, V., Nussbaum, T., Ulyanov, N. B., and Tzfati, Y. (2007) *Molecular and cellular biology* **27**(6), 2130-2143
36. Benton, M. J., and Donoghue, P. C. (2007) *Molecular biology and evolution* **24**(1), 26-53
37. Kumar, S., Tamura, K., and Nei, M. (2004) *Brief Bioinform.* **5**(2), 150-163

FOOTNOTES

* This work was supported by NSF CAREER Award (MCB0642857) to J.L.C.
[2]The abbreviations used are: TR, telomerase RNA; TERT, telomerase reverse transcriptase, RNP, ribonucleoprotein; CR, conserved region; BLAST, basic local alignment search tool; PWM, position-specific weight matrices; snoRNA, small nucleolar RNA; scaRNA, small Cajal body RNA; CAB box, Cajal body specific box.

FIG. LEGENDS

Fig. 1. **(A)** Northern blotting analysis of medaka and zebrafish TRs. Twenty micrograms of total RNA (lane 1), 50 pg (Lane 2) or 500 pg (Lane 3) of *in vitro* transcribed medaka or zebrafish TRs were electrophoresed on 4% denaturing polyacrylamide gels. Blots were each hybridized with riboprobes specific to each TR. Endogenous TR bands are indicated by solid triangles. The in vitro transcribed medaka TR (317 nt) and zebrafish TR (322 nt) serve as markers for size estimation and mass quantitation. The levels of endogenous TR in liver cells were quantitated to be 508 pg and 110 pg per 20 µg of total RNA for medaka and zebrafish, respectively. **(B)** Positive correlation between the TR size and the genome size. The genome sizes (Mbp) were derived from C-values (pg) obtained from The Animal Genome Size Database (http://www.genomesize.com/). The sizes of TRs are based on data from Chen et al, 2000 and this study. Five teleost and four cartilaginous (sharks and rays) fishes are clustered into two separated groups at the lower-left and higher-right ends of the graph, respectively. The 95% confidence band (dashed) of the linear regression line (solid) is shown. The p value is <0.0001.

Fig. 2. Sequence alignment of teleost fish TR. The alignment includes TR sequences from zebrafish (*Danio rerio*), medaka (*Oryzias latipes*), stickleback (*Gasterosteus aculeatus*), fugu (*Takifugu rubripes*) and tetraodon (*Tetraodon nigroviridis*). Residues that are 100% (in red) or 80% (in blue) conserved in non-teleost vertebrates TRs (Chen et al., 2000) are shown below the alignment. The eight conserved regions (CRs) are indicated with red brackets. Black lines above the alignment indicate helices (P1-P8) in the secondary structures. Conserved motifs, i.e. the template, box H and box ACA, are indicated with red lines above the aligned sequences. Residues shaded in blue indicate conserved nucleotides that form Watson-Crick base pairings, while the ones shaded in green indicate nucleotides that co-vary and maintain base-pairing. The residues shaded in yellow are located in the single-stranded regions and universally conserved among the five teleost fishes. Dashes (-) denote alignment gaps. Every tenth nucleotide of the zebrafish sequence is marked with dots above the alignment. The size of each RNA is indicated at the end of the respective sequence. Asterisks indicate organisms for which the 5'-end of the RNA was determined by 5'-RACE.

Fig. 3. Vertebrate TRs share a conserved secondary structure. **(A)** Secondary structures of medaka and fugu TRs. Residues conserved in all five teleost TRs are shown in red. Three structural domains (pseudoknot/template, CR4-CR5 and snoRNA) are outlined and labeled. On the medaka TR structure, eleven helices (P1, P2a, P2b, P3, P4, P5, P6, P6.1, P7a, P7b and P8) and every tenth nucleotide of the sequence are labeled. The template region, box H and ACA motifs are indicated by black boxes. **(B)**

Comparison of secondary structures of medaka, human and shark TRs. The pseudoknot and CR4-CR5 domains are shown in green, while the scaRNA domain (or snoRNA domain in the teleost TR) is shown in cyan. The structural determinants (the P1 helix) for template boundary definition are shown in magenta. In the human TR structure, the mammal-specific structural elements required for activity are shown in brown.

Fig. 4. Structural comparison of the pseudoknot and CR4-CR5 domains, and sequence alignment of the CR7 domains. **(A)** Comparison of the triple helix region within the pseudoknot domain. The schematic of the triple helix region from human (tetrapods), sharpnose shark (cartilaginous) and five teleost are shown, based on an NMR structure reported previously (21). For human, structural elements, etc. P2b, P3, J2a/3, J2b/3, are labeled. The triple helix forming sequence (red) conserved in all species, the bulge on P3 helix (purple) and the conserved G-C base pair close to the triple helix (cyan) are highlighted. The green bars indicate the hoogsteen base pair. The size of the bulge, p3 stem and J2b/3 loop are indicated to the right of the schematics. The dashed line in J2a/3 represents omitted sequences. **(B)** Comparison of medaka and human CR4-CR5 secondary structure. Helixes P5, P6a, P6b and P6.1 are labeled. Residues in red indicate conserved nucleotides in all vertebrates. Nucleotides in green indicate conservation in 5 teleost. While nucleotides in blue indicate conservation in other vertebrates excluding teleost. **(C)** Teleost TR lacks an obvious CAB box motif (UGAG). The five teleost TR sequences shaded in gray are aligned manually with the alignment of 35 non-teleost TR sequences derived from Chen et al., 2000. Conserved CAB box is indicated with red lines above the aligned sequences. Residues identical to human sequence are shaded in blue (helix P8) or in yellow (loop L8). Dashes (-) denote alignment gaps.

Fig. 5. Activity assay of *in vitro* reconstituted teleost telomerase. Medaka and fugu TERT proteins synthesized *in vitro* were assembled with *in vitro* transcribed pseudoknot/template and CR4-CR5 RNA fragments of medaka (md), fugu (f) or zebrafish (z). The RNA fragments, medaka pseudoknot (1-150), medaka CR4-CR5 (154-241), fugu pseudoknot (1-139), fugu CR4-CR5 (143-253), zebrafish pseudoknot (1-134) and zebrafish CR4-CR5 (137-242), were assembled in different combinations with either medaka or fugu recombinant TERT protein as indicated above the gel. The assembled telomerases were analyzed for activity using a conventional telomerase assay (see Experimental Procedures). The numbers on the left (+4, +10, +16, +22, +28, +34 etc.) indicate the number of nucleotides added to the primer for each major band seen.

Fig. 6. Effective concentrations of the pseudoknot and CR4-CR5 domains to assemble active telomerase *in vitro*. Titration experiments were performed with pseudoknot and CR4-CR5 RNA fragments or full-length TR alone for reconstituting medaka (upper panel) and human (lower panel) telomerase enzymes. Various concentrations of pseudoknot or CR4-CR5 RNA fragments were assembled with the other RNA fragment at a saturated 3 μM and the *in vitro* synthesized TERT protein, followed by the conventional telomerase assay. The pseudoknot (medaka: nt 1-150 and human: nt 32-195) and CR4-CR5 (medaka: nt 170-220 and human: nt 241-328) RNA fragments were titrated as indicated. The relative activity represents the ratio of total activity of each reaction over the total activity of the reaction with saturated concentrations of both RNA fragments. The median effective concentration (EC_{50}) values of each RNA fragment are indicated.

Fig. 7. The neighbor-joining tree inferred from the vertebrate TR sequences. The tree was derived using the neighbor-joining method from the aligned TR sequences of 14 vertebrates including 5 tetrapods (human, mouse, macaw, turtle and frog), 5 teleost fishes (fugu, tetraodon, stickleback, medaka and zebrafish) and 4 cartilaginous fishes (stingray, cownose ray, sharpnose shark and dogfish shark). The phylogenetic tree was constructed using the program MEGA3.1 (37). The number next to each node indicates a value as a percentage of 1000 bootstrap replicates. Branch lengths are proportional to the number of residue changes. Scale bar indicates an evolutionary distance of 0.05 nucleotide substitution per position in the sequence.

Fig. 1

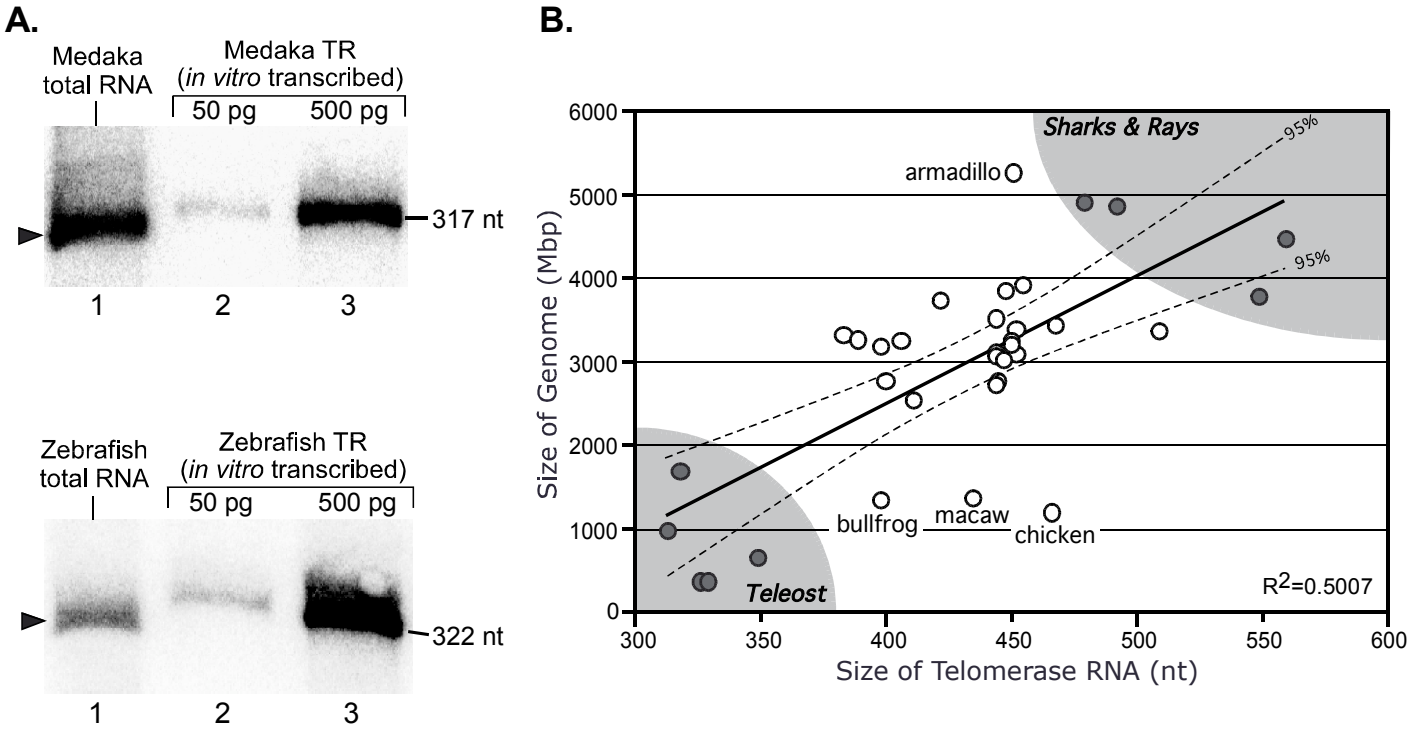

A.
Medaka total RNA
Medaka TR (in vitro transcribed)
50 pg
500 pg
317 nt
1
2
3
Zebrafish total RNA
Zebrafish TR (in vitro transcribed)
50 pg
500 pg
322 nt
1
2
3
B.
Sharks & Rays
95%
95%
armadillo
bullfrog
macaw
chicken
Teleost
R2=0.5007
Size of Genome (Mbp)
Size of Telomerase RNA (nt)
0
1000
2000
3000
4000
5000
6000
300
350
400
450
500
550
600

Fig. 2

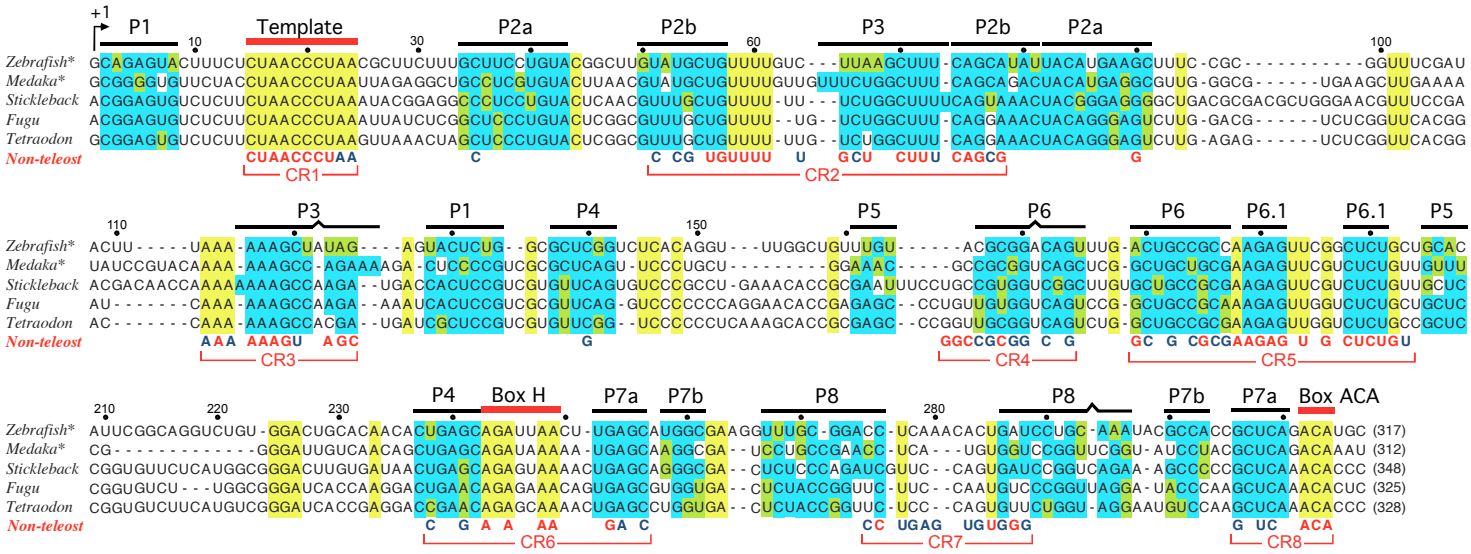
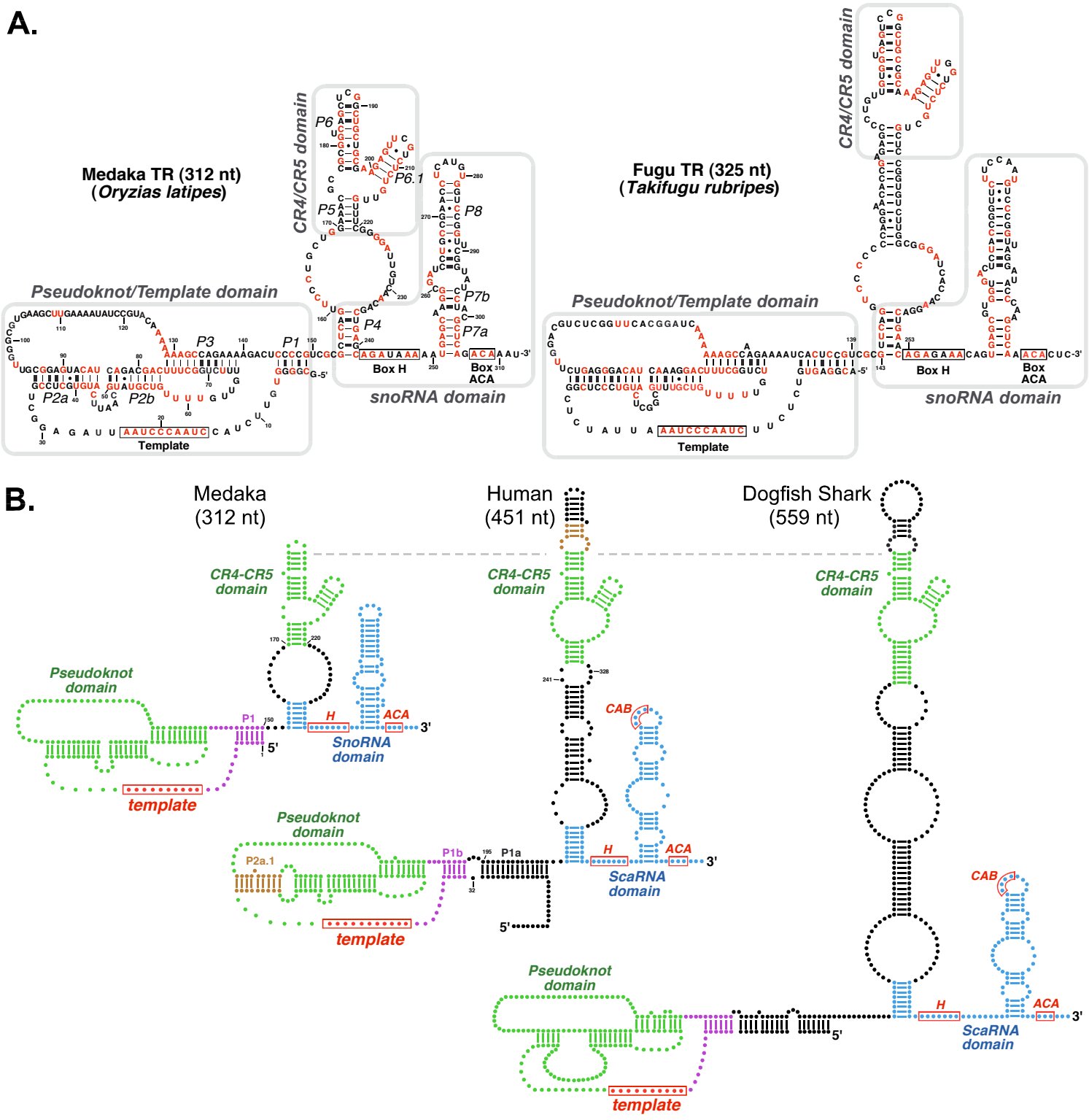

+1
P1
Template
P2a
P2b
P3
P2b
P2a
P3
P1
P4
P5
P6
P6
P6.1
P6.1
P5
P4
Box H
P7a
P7b
P8
P8
P7b
P7a
Box ACA
CR1
CR2
CR3
CR4
CR5
CR6
CR7
CR8
Zebrafish*
Medaka*
Stickleback
Fugu
Tetraodon
Non-teleost

Fig. 3


A.
Medaka TR (312 nt)
(Oryzias latipes)
Pseudoknot/Template domain
CR4/CR5 domain
P6
P6.1
P5
P8
P7b
P7a
P4
P3
P1
P2a
P2b
Box H
Box ACA
snoRNA domain
Template
Fugu TR (325 nt)
(Takifugu rubripes)
Pseudoknot/Template domain
CR4/CR5 domain
Box H
Box ACA
snoRNA domain
Template
B.
Medaka
(312 nt)
Human
(451 nt)
Dogfish Shark
(559 nt)
CR4-CR5 domain
Pseudoknot domain
P1
H
ACA
SnoRNA domain
template
P2a.1
P1b
P1a
CAB
ScaRNA domain
5'
3'

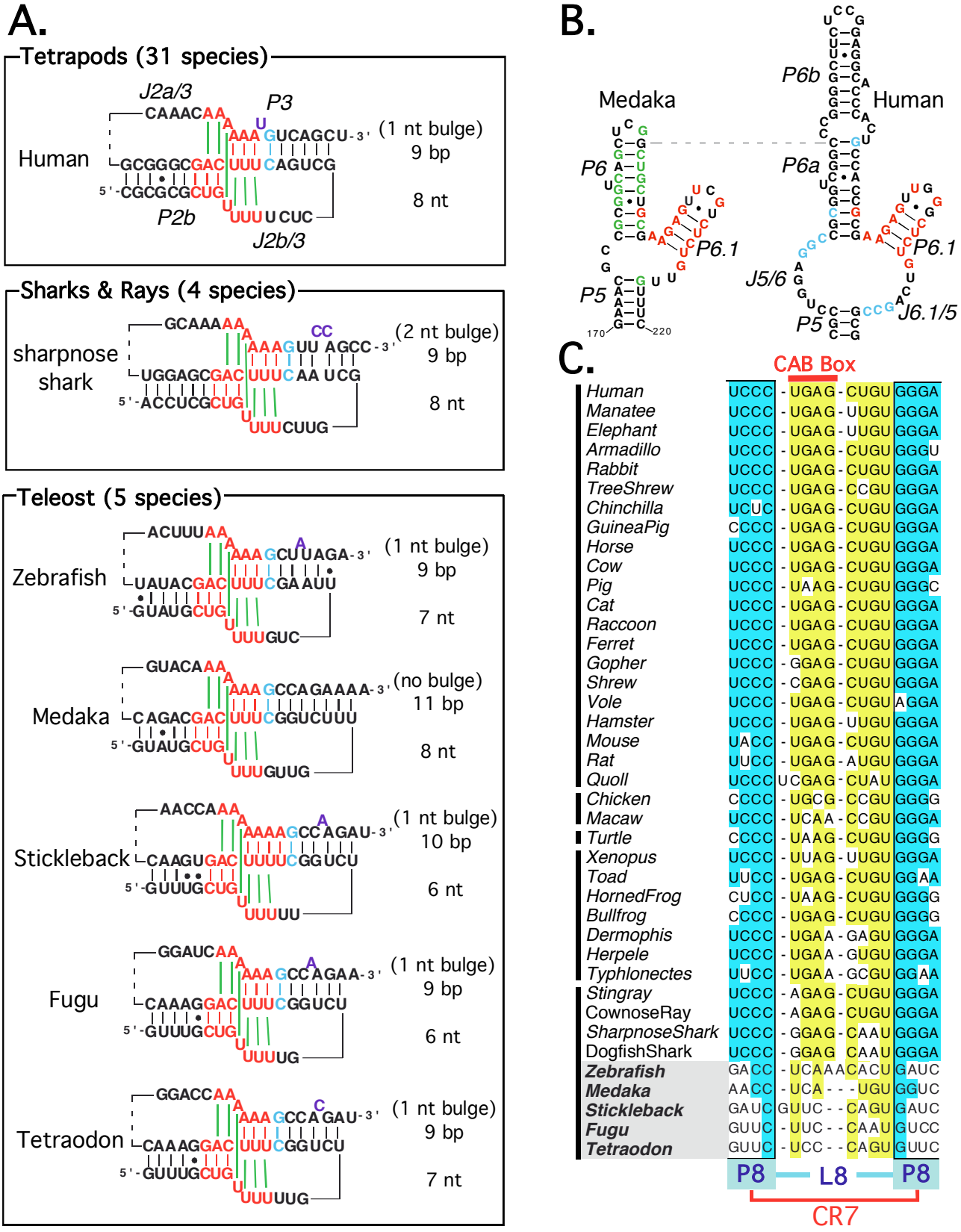

A.
Tetrapods (31 species)
J2a/3
P3
Human
(1 nt bulge)
9 bp
8 nt
P2b
J2b/3
Sharks & Rays (4 species)
sharpnose shark
(2 nt bulge)
9 bp
8 nt
Teleost (5 species)
Zebrafish
(1 nt bulge)
9 bp
7 nt
Medaka
(no bulge)
11 bp
8 nt
Stickleback
(1 nt bulge)
10 bp
6 nt
Fugu
(1 nt bulge)
9 bp
6 nt
Tetraodon
(1 nt bulge)
9 bp
7 nt
B.
Medaka
Human
P6
P6.1
P5
P6b
P6a
J5/6
J6.1/5
C.
CAB Box
Human
Manatee
Elephant
Armadillo
Rabbit
TreeShrew
Chinchilla
GuineaPig
Horse
Cow
Pig
Cat
Raccoon
Ferret
Gopher
Shrew
Vole
Hamster
Mouse
Rat
Quoll
Chicken
Macaw
Turtle
Xenopus
Toad
HornedFrog
Bullfrog
Dermophis
Herpele
Typhlonectes
Stingray
CownoseRay
SharpnoseShark
DogfishShark
Zebrafish
Medaka
Stickleback
Fugu
Tetraodon
P8
L8
P8
CR7

Fig. 5

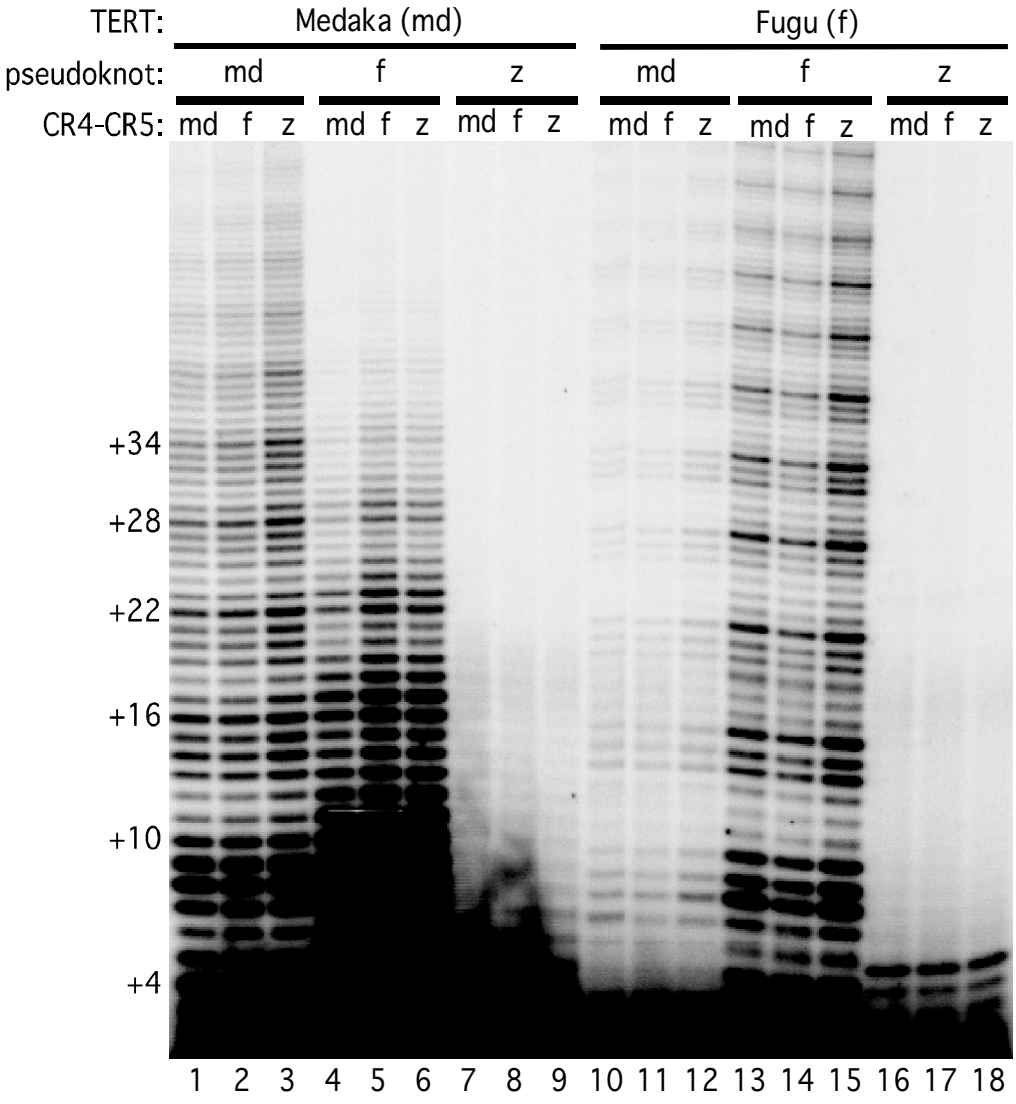

TERT:
Medaka (md)
Fugu (f)
pseudoknot:
md
f
z
md
f
z
CR4-CR5:
md f z md f z md f z md f z md f z md f z
+34
+28
+22
+16
+10
+4
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18

Fig. 6

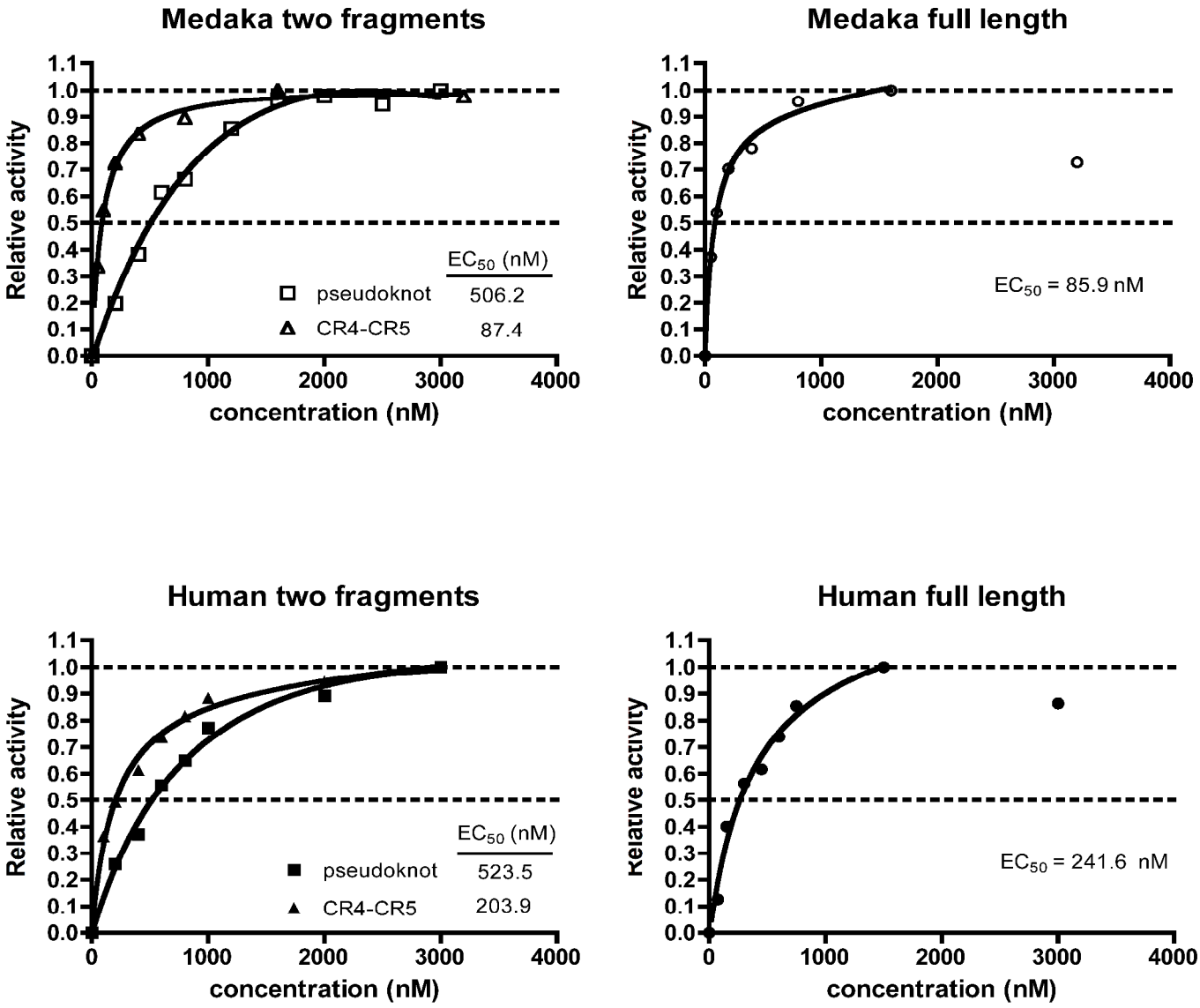
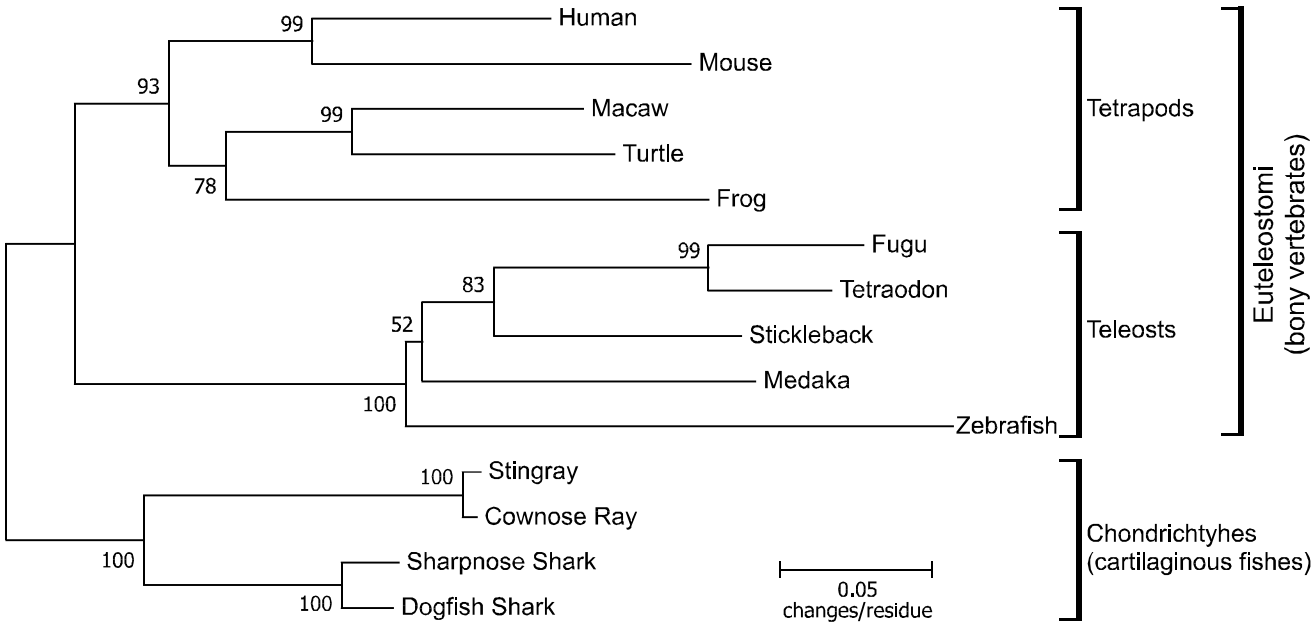

Medaka two fragments
Relative activity
concentration (nM)
EC50 (nM)
pseudoknot 506.2
CR4-CR5 87.4
Medaka full length
EC50 = 85.9 nM
Human two fragments
pseudoknot 523.5
CR4-CR5 203.9
Human full length
EC50 = 241.6 nM

Fig. 7


Human
Mouse
Macaw
Turtle
Frog
Fugu
Tetraodon
Stickleback
Medaka
Zebrafish
Stingray
Cownose Ray
Sharpnose Shark
Dogfish Shark
99
93
99
78
99
83
52
100
100
100
100
0.05
changes/residue
Tetrapods
Teleosts
Euteleostomi
(bony vertebrates)
Chondrichtyhes
(cartilaginous fishes)