Mol Gen Genet (1995) 247:312–322 

ORIGINAL PAPER

Zsuzsanna Izsvák · Zoltán Ivics · Perry B. Hackett

# Characterization of a *Tc1*-like transposable element in zebrafish (*Danio rerio*)

Received: 7 September 1994 / Accepted: 13 December 1994

**Abstract** We have characterized *Tdr1*, a family of *Tc1*-like transposable elements found in the genome of zebrafish (*Danio rerio*). The copy number and distribution of the sequence in the zebrafish genome have been determined, and by these criteria *Tdr1* can be classified as a moderately repetitive, interspersed element. Examination of the sequences and structures of several copies of *Tdr1* revealed that a particular deletion derivative, 1250 bp long, of the transposon has been amplified to become the dominant form of *Tdr1*. The deletion in these elements encompasses sequences encoding the N-terminal portion of the putative *Tdr1* transposase. Sequences corresponding to the deleted region were also detected, and thus allowed prediction of the nucleotide sequence of a hypothetical full-length element. Well conserved segments of *Tc1*-like transposons were found in the flanking regions of known fish genes, suggesting that these elements have a long evolutionary history in piscine genomes. *Tdr1* elements have long, 208 bp inverted repeats, with a short DNA motif repeated four times at the termini of the inverted repeats. Although different from that of the prototype *C. elegans* transposon *Tc1*, this inverted repeat structure is shared by transposable elements from salmonid fish species and two *Drosophila* species. We propose that these transposons form a subgroup within the *Tc1*-like family. Comparison of *Tc1*-like transposons supports the hypothesis that the transposase genes and their flanking sequences have been shaped by independent evolutionary constraints. Although *Tc1*-like sequences are present in the genomes of several strains of zebrafish and in salmonid fishes, these sequences are not conserved in the genus *Danio*, thus raising the possibility that these elements can be exploited for gene tagging and genome mapping.

**Key words** Transposable element · Evolution
Zebrafish · Transposase · Inverted repeats

Communicated by J. A. Campos-Ortega

Z. Izsvák[1] · Z. Ivics · P. B. Hackett (✉)
Department of Genetics and Cell Biology,
University of Minnesota,
1445 Gortner Ave., St. Paul, MN 55108, USA

P. B. Hackett
Institute of Human Genetics, University of Minnesota,
St. Paul, MN 55108, USA

*Permanent address:*
[1] Institute of Biochemistry,
Biological Research Center of Hungarian Academy of Sciences,
Szeged, Hungary

## Introduction

The zebrafish (*Danio rerio*) is a superb model animal for studies of the molecular genetics of vertebrate development (Driever et al. 1994; Nüsslein-Volhard 1994; Rossant and Hopkins 1992). However, two methods of investigating the roles that genes play in development are lacking in zebrafish: a method of randomly tagging genes and of establishing transgenic fish lines that continue to express transgenic DNA over multiple generations. Hence, there is no method currently available for efficient integration of intact copies of foreign DNA into the zebrafish genome. Development of such a technology would not only be beneficial for the production of transgenic fish, but would also be useful for DNA-based insertional mutagenesis and gene tagging, which could be applied to generate and identify developmental mutants. A potential solution to the above problems would be to harness transposable elements as vectors for carrying specific DNA sequences into chromosomes for the purpose of mutagenesis, gene tagging and genetic engineering.

Since their discovery, repetitive elements with the characteristics of transposons have been found in the genomes of all organisms in which they were sought. Transposable elements are usually divided into two major groups according to their mechanism of transposition (Finnegan 1989). The first group is composed of DNA-based transposable elements that transpose via a DNA intermediate. The second group, the retroelements, move

by reverse transcription of an RNA intermediate. DNA transposons are common in bacteria, invertebrates and plants, whereas in vertebrates retrotransposons are more abundant (Henikoff 1992). In eukaryotes, DNA-based transposable elements typically range from 1 to 10 kb in length, have short inverted terminal repeats, and have sequences that may encode a transposase required for the transposition process. Elements of this sort include the *P* elements of *Drosophila melanogaster*, the *Ac* and *Spm* elements of maize and the *Tc1* transposon of *Caenorhabditis elegans*. Transposons exist in active and inactive forms (Hartl et al. 1992). Active transposable elements encode genes required in trans, as well as flanking sequences required in cis for mobility. Active elements are relatively rare, since they destabilize the genomes of their hosts by insertional mutagenesis. More common are inactive elements which often contain mutations in the sequence encoding the trans-acting transposase, thereby rendering these elements unable to transpose on their own. However, these inactive elements may still retain cis-active sequences needed for recognition by the transposase, and therefore can be mobilized by transposase expressed from an active element. Here, we use the term transposable element for sequences that have the hallmarks of either active or inactive transposons.

There is a growing number of repetitive elements known from fish species (Datta et al. 1988; Kido et al. 1991; Naruse et al. 1992; Winkfein et al. 1988; Wright 1989). In zebrafish, three types of repetitive elements have been reported: satellite-like sequences (Ekker et al. 1992); *Alu*-elements (He et al. 1992); and recently, a DNA transposon, *Tdr1* (Radice et al. 1994). *Tdr1* belongs to the family of *Tc1*-like transposable genetic elements, named after the prototype found in *C. elegans* (Emmons et al. 1983). Members of this transposon family share certain features including a length of about 1600 bp, conserved terminal nucleotides (TACAGT) of the inverted repeats and conserved sequence domains in the encoded transposase. *Tc1*-like sequences, as defined by the homology of their transposase genes, have been found in fungi (Daboussi et al. 1992), protozoa (Doak et al. 1994), and are widespread in animals including nematodes (Collins et al. 1989; Emmons et al. 1983; Prasad et al. 1991), insects (Brezinsky et al. 1990; Caizzi et al. 1993; Franz and Savakis 1991; Harris et al. 1988; Simmons and Merriman 1994), hagfish (Heierhorst et al. 1992), and teleost fishes such as zebrafish, channel catfish (*Ictalurus punctatus*; Wilson et al. 1990), rainbow trout (*Oncorhynchus mykiss*; Radice et al. 1994) and Atlantic salmon (*Salmo salar*; Goodier and Davidson 1994; Radice et al. 1994).

Transposable elements can be powerful tools for gene tagging and genetic mapping, especially if strains of the same organism contain small numbers of these sequences. Moreover, the transposase of active elements should serve as an excellent source for recombinase-mediated insertion of transgenic DNA (Ivics et al. 1993). Accordingly, we examined the *Tdr1*-like elements in various strains of zebrafish in order to find an active copy, and to determine whether a line of zebrafish could be found that had a small number of elements, and therefore would be suitable for transposon tagging and mutagenesis studies.

Here we present our characterization of several copies of *Tdr1*, the first DNA transposon characterized in zebrafish. All *Tdr1* elements we have analyzed exhibit variability in size and sequence, suggesting that *Tdr1* is not a recent invader of the zebrafish genome. The majority of *Tdr1* elements appear to be descendants of a single progenitor transposon with a deletion that removed the N-terminal region of the putative transposase. However, because sequences complementary to this missing portion can also be found in the genome, and a homologous element exists in salmon, the sequence of a hypothetical transposase gene can be predicted. A byproduct of this investigation is a preliminary phylogenetic analysis of sequences and structures of several *Tc1*-like transposons, which suggests that the *Tdr1* element may be a hybrid of two types of transposons.

## Materials and methods

### Source of fish DNAs

Wild-type zebrafish, "blue" danios, "giant" danios, "gold" danios and "leopard" danios were purchased in a local pet store. The C-32 homozygous diploid zebrafish line was obtained from the University of Oregon. Zebrafish from Singapore, Hong Kong and Indonesia were obtained from the Massachusetts General Hospital, Cardiovascular Research Center. The zebrafish pigmentation mutant is from our laboratory. Carp sperm was obtained from the Fish Research Station, Szarvas, Hungary. Chinook salmon DNA was prepared from the CHSE cell line, and northern pike DNA was extracted from the PG cell line.

### DNA preparation, blotting and hybridization

High molecular weight DNA was prepared from adult zebrafish, essentially as described (Ivics et al. 1993). Tissues were homogenized in 200 mM TRIS-HCl pH 8.0, 50 mM EDTA, 150 mM NaCl, 0.5% (w/v) SDS and 500 μg/ml Proteinase K overnight at 55°C. Samples were then extracted once with phenol-chloroform, once with chloroform and precipitated with 2 vol ethanol. DNA samples were prepared for dot-blotting by the addition of 1 vol 0.8 M NaOH, incubated at room temperature for 10 min, then 2 vol of ice-cold 2 M ammonium acetate was added, and the DNA samples were blotted to a Nytran (Schleicher & Schuell) nylon membrane with the help of a dot blot apparatus (BRL). Blots were then baked at 80°C for 2 h and irradiated with ultraviolet light for 30 s. Southern transfer was done according to standard procedures (Sambrook et al. 1989). An [$\alpha$-$^{32}$P]-labeled, approximately 700 bp internal fragment of *Tdr1* served as probe in DNA hybridizations (Fig. 5A). This fragment was subcloned from the *Tdr1* transposon isolated by Radice et al. (1994). High stringency hybridizations were done according to standard methods (Sambrook et al. 1989) in 50% (v/v) formamide, while low stringency hybridizations were in 30% (v/v) formamide. Stringent washings of the filters were done in 0.1×SSC, 0.1% (w/v) SDS at 65°C, whereas low-stringency washings were in 0.5×SSC, 0.1% (w/v) SDS at 42°C for 7 h. Autoradiograms were scanned with an AMBIS densitometry system.

PCR analysis

PCR was performed by a Perkin-Elmer Thermo-Cycler with a 6 min initial denaturation step at 94°C, then 30 cycles of 1 min at 94°C, 1 min at 50°C and 1 min at 72°C. To each reaction, 2 U Vent polymerase (New England Biolabs), approximately 500 ng zebrafish genomic DNA and primers to a final concentration of 1 μM were added. The sequences of oligonucleotide primers were as follows: Primer *a*: 5′-TACAGTTGAAGTCGGAAGTTTACAT-AC-3′; primer *b*: 5′-GCTTCCTTGTGTGACAACATGGG-3′; primer *c*: 5′-CTAGACAGTCAAAAGTTTACATA-3′; primer *d*: 5′-CCTTGCTGAGCGGCCTTTCAGGTTATGTCG-3′.

Sequence analysis

Double-strand sequencing of *Tdr1* clones was done with a Sequenase Version 2.0 kit (USB). DNA sequence analysis was carried out with the GCG sequence analysis package (University of Wisconsin Genetics Computer Group). Nucleotide sequences have been deposited in Genbank. Accession numbers of individual clones are given in Fig. 2A. Alignment of sequences was generated with the PILEUP program of GCG. Adjustments to the alignment were made by LINEUP. Computer searches were performed with the TFASTA program of GCG using the following amino acid query sequences: VFQQDNDPKHTS and LEWPSQSPDLN-PIENLW.

## Results

The *Tdr1* element discovered by Radice et al. (1994) was unusually short for a transposon of its type, only 1206 bp long, with a deletion of sequences corresponding to about 120 codons in the putative transposase coding region. Because the approximately 1600 bp *Tss1* elements they found in salmon were homologous to *Tdr1*, a consensus set of primers for PCR could be constructed that permitted an analysis of both the size distribution and number of *Tdr1* elements in zebrafish.

### The majority of *Tdr1* elements have a conserved deletion

We first determined the size distribution of *Tdr1* elements by PCR amplification from zebrafish genomic DNA (Fig. 1). Figure 1A shows a diagram of a *Tdr1* element and the PCR primers used for amplification. As shown in lane 1 of Fig. 1B, a dominant band of about 1250 bp could be detected using primer *a*, which is complementary to the terminal inverted repeats. Full-length *Tc1*-like transposons are about 1600 bp long, thus the elements detected in this PCR assay were about 350 bp shorter than expected, and of the same size as the original element found by Radice et al. (1994). In order to map the position of the missing sequence(s) in these amplified copies, two internal primers (Fig. 1A, primers *b* and *c*) were used to amplify the presumptive transposase coding region. A dominant band of approximately 700 bp was obtained with some diffuse sequences of higher and lower molecular weight (lane 2 in Fig. 1B). Since the putative transposase of salmonid *Tc1*-like transposons comprises 340 amino acids, encoded by a 1020 bp sequence (Radice et al. 1994), the size of the band in Fig. 1B, lane 2 suggests that there is a deletion, corresponding to about 120 codons, within the presumptive coding region of *Tdr1*. The size of the single *Tdr1* element reported by Radice et al. (1994) also had approximately 350 bp deleted from the presumptive transposase open reading frame. Thus, although there might be other forms of *Tc1*-like sequences in the genome, the use of two independent primer pairs in PCR amplifications in-

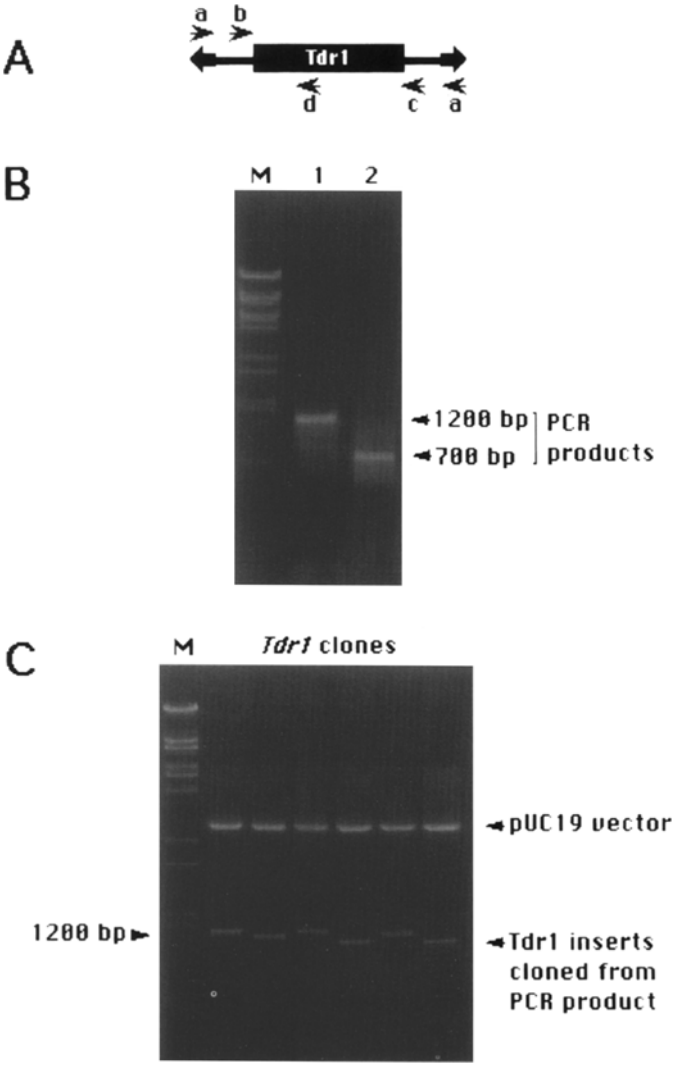



**Fig. 1A–C** PCR amplification and molecular cloning of genomic *Tdr1* sequences from zebrafish. **A** Schematic representation of a zebrafish *Tc1*-like element, with locations and orientations of oligonucleotide primers (*a*, *b*, *c* and *d*) used for the amplification of *Tdr1* sequences. The *filled box* represents the conceptual transposase coding region, while *arrows* at both ends of the transposon indicate terminal inverted repeats. **B** Ethidium-bromide-stained agarose gel showing *Tdr1* sequences amplified by PCR. Lanes: M, Lambda DNA cleaved with *Bst*EII as a size marker; 1, PCR products obtained with primer *a*, complementary to consensus terminal sequences of fish *Tc1*-like elements; 2, PCR products obtained with primers *b* and *c*, complementary to sequences flanking the putative transposase coding region. **C** Ethidium bromide-stained agarose gel showing individual *Tdr1* inserts released from plasmids by restriction digestion

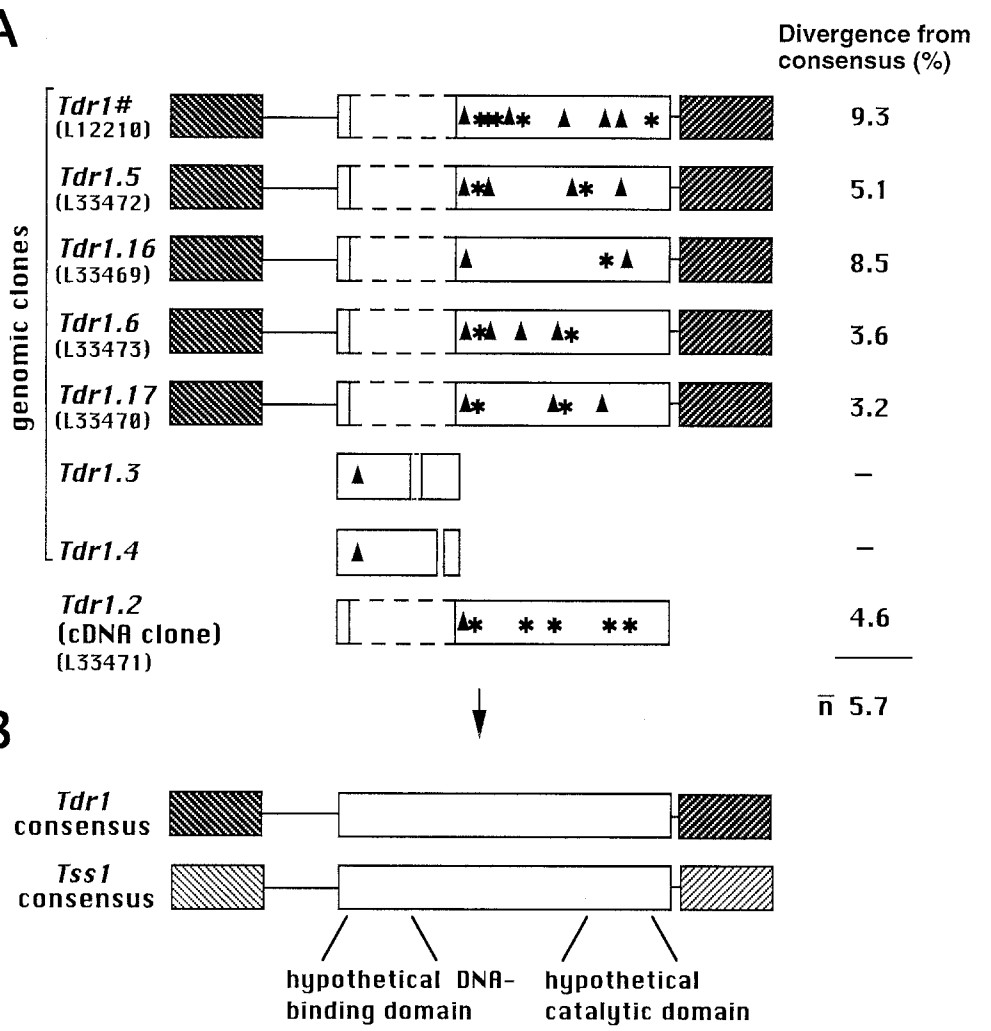



**Fig. 2A,B** Schematic representation of the structures of *Tdr1* sequences amplified by PCR. *Shaded boxes* at the ends of the transposons represent terminal inverted repeats, while open boxes indicate coding sequences. **A** Characteristics of *Tdr1* sequences amplified from genomic DNA or a cDNA library. Four genomic elements, two internal genomic *Tdr1* fragments and a cDNA clone were sequenced. *Filled triangles* mark frameshift mutations, and *asterisks* indicate stop codons in the transposase open reading frame. Genbank accession numbers are given below the clone names; L12210 is the *Tdr1* element described by Radice et al. (1994). **B** Schematic comparison of a consensus sequence of zebrafish *Tdr1* to a consensus *Tss1* transposon from Atlantic salmon. The eight *Tdr1* sequences shown in **A** were used to construct a putative full-length *Tdr1* sequence, which is given in Fig. 3. Relative positions of the putative DNA-binding and catalytic domains of the encoded transposases are indicated

dicates that the 1250 bp PCR products represent the majority of *Tdr1* elements.

The dominant, 1250 bp PCR products were isolated and cloned into a pUC19 vector. When individual *Tdr1* inserts were released from the plasmids, a variability in size of approximately 5% was observed (Fig. 1C).

As an alternative approach, we examined whether *Tdr1* transposase transcripts are present in early zebrafish embryos. In this experiment, an aliquot of a 0–24 h post-fertilization cDNA library was subjected to PCR amplification using primers flanking the putative open reading frame of the transposase polypeptide (primers *b* and *c* in Fig. 1A). An abundant, intensively hybridizing 700 bp PCR product, similar to that obtained from genomic amplifications, was isolated and cloned (data not shown). Production of a 700 nucleotide mRNA is consistent with transcription of a deleted *Tdr1* element from either its own promoter, or heterologous promoters of transcriptional units that contain a *Tdr1* sequence.

## Sequence of a putative full-length *Tdr1* transposon can be predicted

Having failed to find a *Tdr1* element of the length expected for an element with a complete transposase gene, we closely examined the sequences of several of the clones we obtained. The sequences of four genomic *Tdr1* clones (L33469-70 and L33472-73), a cDNA clone (L33471), and the original isolate of Radice et al. (1994) were used to construct a consensus sequence. Figure 2 schematically summarizes the data. On average, the sequenced elements are about 94% identical to the 1245 bp consensus sequence. All of the *Tdr1* elements analyzed contained stop codons, frameshift mutations and a deletion of approximately 350 bp, corresponding to codons encoding amino acids 5–121 of the transposase gene product of the *Tc1* family (Fig. 2A). Based on mapping of DNA-binding domains to the N-terminal region of the *C. elegans Tc1* transposase (Vos et al. 1993), and the conservation of transposase domains within the *Tc1*-like transposon family (Doak et al. 1994), the 350 bp deletion should remove transposase sequences that would probably be responsible for DNA-binding (Fig. 2B). The breakpoints of this deletion are conserved, suggesting that the *Tdr1*

**Fig. 3** Consensus DNA sequence of a putative, full-length *Tdr1* element. The terminal inverted repeats are displayed against a *black background*, and a 12 bp directly repeated motif is displayed in *italics*, on a *white background*. The first nucleotide of the hypothetical transposase translational start codon is marked by a *dot*, and the first nucleotide of the putative translational stop codon is marked by an *asterisk*. The putative poly(A) signal is *underlined*. Nucleotide positions are given in the left margin

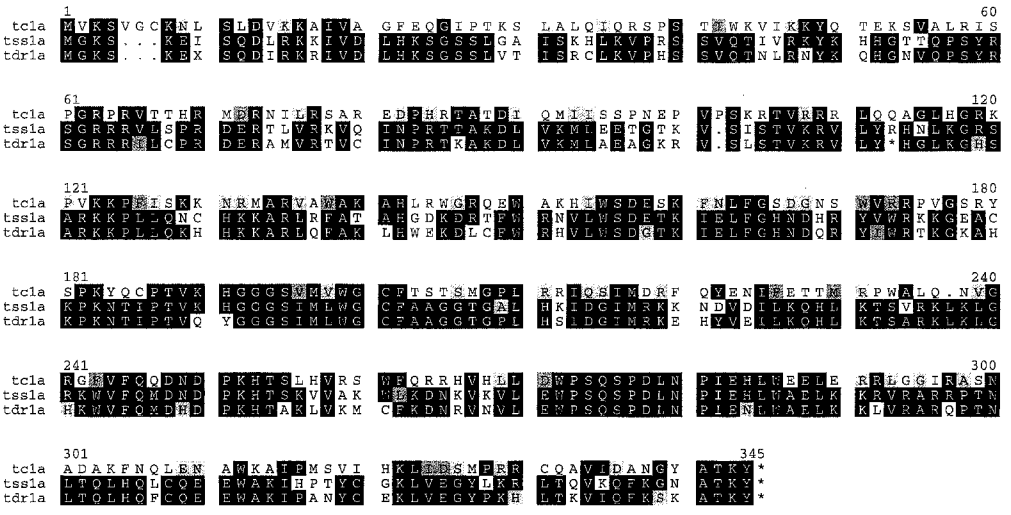
copies examined by us are all descendants of a single deletion derivative.

Genomic DNA was scanned for the putative sequences that are absent in the 1250 elements. Since sequences of full-length salmonid *Tc1*-like transposons were available (Radice et al. 1994), PCR primers were designed to amplify the first 350 bp of the transposase coding region, which included sequences within the conserved deletion (primers *b* and *d* in Fig. 1A). PCR products were sequenced (clones *Tdr1.3* and *Tdr1.4* in Fig. 2A) and found to represent sequences homologous to the corresponding region of salmonid *Tc1*-like elements. When these sequences were incorporated into the 1245 bp consensus, a putative full-length *Tdr1* element could be constructed (Figs. 2B and 3), which is approximately 72% identical at the nucleotide level, and 77% identical in its amino acid sequence to the salmon consensus sequence (Fig. 4).

## Determination of the number of *Tdr1* copies in the zebrafish genome

In order to determine whether transposon-tagging using *Tdr1* is feasible in zebrafish, the copy number of *Tdr1* elements in the zebrafish genome was determined by blotting serial dilutions of genomic DNA from a single wild-type zebrafish, and a 700 bp internal fragment of *Tdr1* as a standard (Fig. 5). The same *Tdr1* fragment was used as a probe (Fig. 5A). The radioactivity in each spot on the filter was quantified, and the linear range of data was used for copy number determination. Intensities of hybridization signals from zebrafish genomic DNA were about 2400 times less, per nanogram than those from the standard DNA (Fig. 5B). Assuming that the size of the

**Fig. 4** Amino acid alignment of the putative transposases encoded by *Tc1*, *Tss1* and *Tdr1* transposons. Shown are the transposase amino acid sequences encoded by *C. elegans Tc1* (Tc1A) (Vos et al. 1993), Atlantic salmon *Tss1* (Tss1A) (Radice et al. 1994) and zebrafish *Tdr1* (Tdr1A). The consensus amino acid sequence of a putative, full-length Tdr1A transposase was derived by conceptual translation of the consensus nucleotide sequence shown in Fig. 3. Conserved amino acids were highlighted by the PRETTYBOX program of GCG, and are shown as *white on a black background*. *Shaded boxes* indicate conservative amino acid substitutions. *Asterisks* indicate stop codons, and X in the Tdr1A sequence marks a conserved frameshift



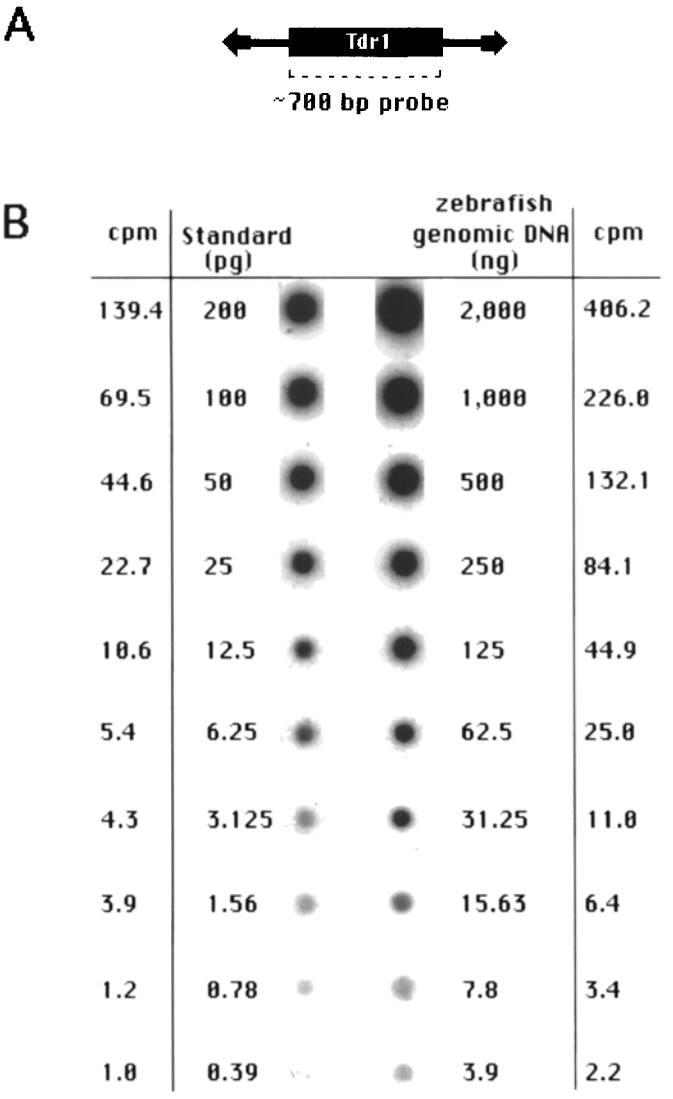



**Fig. 5A,B** Determination of the number of *Tdr1* elements in the zebrafish genome. **A** Schematic representation of a zebrafish *Tc1*-like element. The *filled box* represents the presumptive transposase coding region, while *arrows* at both ends of the transposon indicate terminal inverted repeats. The fragment used as standard for blotting and probe for hybridization is indicated. **B** Serial dilutions of genomic DNA from a wild-type zebrafish and the internal fragment of *Tdr1*, shown in **A** and used as a standard, were blotted and hybridized to the radioactively labeled *Tdr1*-specific probe. The filter was scanned and the radioactivity in each sample was quantified

zebrafish haploid genome is $1.6\times10^9$ bp (Driever et al. 1994), we estimate that the 700 bp internal portion of *Tdr1* is represented about 1000 times per haploid genome. This copy number is supported by our finding that 1.4% of the plaques in a zebrafish genomic library with an average insert size of 14 kb hybridized with the *Tdr1* probe. This result is consistent with data reported by Radice et al. (1994). Assuming that the majority of *Tdr1* elements are about 1250 bp in length, 1000 copies of the sequence would account for approximately 0.07% of the zebrafish genome.

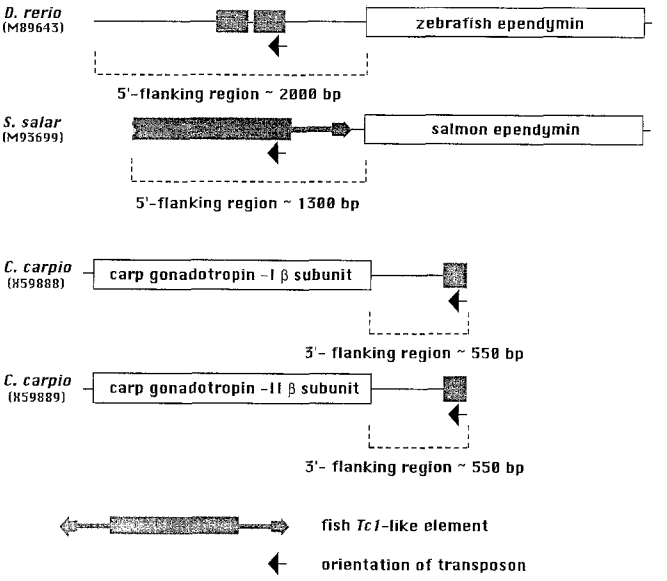



**Fig. 6** *Tc1*-like sequences are associated with fish genes. A schematic representation of the conservation, positions and orientations of *Tc1*-like sequences relative to the structural genes is shown. *Filled boxes* indicate sequences of *Tc1*-like origin, the *open boxes* represent coding regions of the salmon and zebrafish ependymin genes and the carp gonadotropin loci. Transcription of genes is in all cases from left to right. Genbank accession numbers are given below the species names

We employed pulsed-field gel electrophoresis (PFGE) followed by Southern hybridization to determine the genomic distribution of *Tdr1* in genomic DNA samples from zebrafish cells from various cell lines. When the ethidium bromide-stained gel and the autoradiogram were compared, the diffuse patterns of the fragments identified by staining and hybridization were similar (data not shown), suggesting that *Tdr1* elements are interspersed throughout the zebrafish genome.

## *Tc1*-like sequences are associated with fish genes

The above data suggest that *Tdr1* is present in high copy number in the zebrafish genome. If it is to be useful as a tagging agent, it must be capable of being inserted in the neighborhood of active genes, as are other transposons. To determine if this is the case with *Tdr1*, we conducted computer searches for *Tc1*-like sequences in the available databases. Besides the obvious homologies with various members of the *Tc1*-like transposon family, domains of *Tc1*-like transposases were detected in the 5′-flanking region of the zebrafish ependymin gene (Fig. 6). The homology extends only to limited stretches of amino acid sequences, notably the D(35)E motif believed to form the catalytic domain of a superfamily of transposase proteins (Doak et al. 1994). Beyond this motif, sequences were highly divergent. A well-conserved fragment of a *Tss1* element was also detected in the 5′-flanking region of the ependymin gene of the Atlantic salmon (Fig. 6), suggesting that this insertion occurred

**Table 1** Comparison of *Tdr1* to other *Tc1*-like elements. Comparison of the encoded transposases and the inverted repeats of *Tc1*-like transposons from fish, nematodes and insects. Amino acid sequence identity was determined by the alignment of the polypeptide products encoded by the transposons. References: *Tdr1* (Radice et al. 1994, and this paper); *Tss1* (Radice et al. 1994); *Tom1* (Radice et al. 1994); *IpTc1* (Wilson et al. 1990); *Tes1* (Heierhorst et al. 1992); *Tc1* (Vos et al. 1993); *Tcb1* and *Tcb2* (Prasad et al. 1991); *Tc3* (Collins et al. 1989); *S* element (Simmons and Merriman 1994); *Minos* (Franz et al. 1994); *HB1* (Harris et al. 1988); *Uhu* (Brezinsky et al. 1990); *Bari1* (Caizzi et al. 1993)

| | Transposon | Species | Amino acid sequence identity to *Tdr1* (%) | Length of IRS | Presence of IR-DR structure |
|---|---|---|---|---|---|
| Fish | *Tdr1* | *D. rerio* | 100 | 208 | Yes |
| | *Tss1* | *S. salar* | 77 | 226 | Yes |
| | *Tom1* | *O. mykiss* | 69 | 221 | Yes |
| | *IpTc1* | *I. punctatus* | 43 | 85 | No |
| Hagfish | *Tes1* | *E. stouti* | 34 | 66 | No |
| Nematodes | *Tc1* | *C. elegans* | 34 | 54 | No |
| | *Tcb1* | *C. briggsae* | 34 | 80 | No |
| | *Tcb2* | *C. briggsae* | 34 | 80 | No |
| | *Tc3* | *C. elegans* | 28 | 465 | No |
| Insects | *S* | *D. melanogaster* | 32 | 234 | Yes |
| | *Minos* | *D. hydei* | 32 | 255 | Yes |
| | *HB1* | *D. melanogaster* | 32 | 27 | No |
| | *Uhu* | *D. heteroneura* | 29 | 46 | No |
| | *Bari1* | *D. melanogaster* | 29 | 27 | No |

**Table 2** Sequence conservation of direct repeats located at the ends of the long inverted repeats

| Transposon | | Length of DRs |
|---|---|---|
| *Tdr1* | C<br>AAGTTTACATAC | 12 bp |
| *Tss1* | .AAGTTTACATACAC. | 16 bp |
| *Tom1* | AAGTTTACATACAC | 14 bp |
| *S* | .....GTTTACACA<br>C | 21 bp |

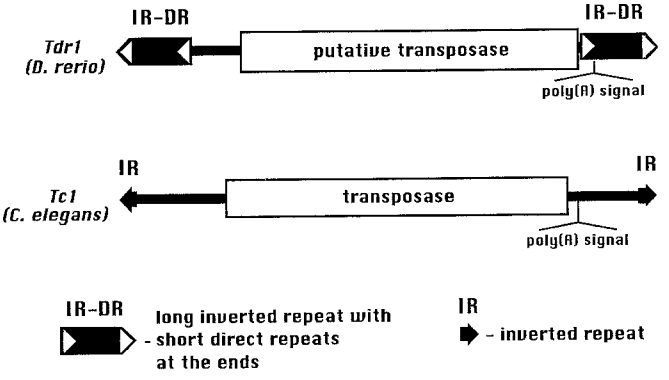



**Fig. 7** Structural comparison of zebrafish *Tdr1* and *C. elegans Tc1* transposons. The main structural features of *Tdr1* and *Tc1* are depicted to illustrate the differences in location of the open reading frames for the transposase with respect to the terminal inverted repeats, the length of the inverted repeats, and the presence or absence of short direct repeats within the inverted repeats (IR-DR). *Arrows* at the ends of the elements represent orientations of repeated sequences, and the *open boxes* indicate the transposase coding regions

in a common ancestor of these fish species. We also discovered a conserved domain of *Tc1*-like transposases in the 3′-flanking regions of two carp gonadotropin β-subunit genes (Fig. 6), indicating that a *Tc1*-like sequence was present in the gonadotropin locus before tetraploidization of the carp genome, and amplification of this gene, about 60 million years ago (Ferris and Whitt 1977).

## Structural comparison of *Tdr1* to other members of the *Tc1*-like transposon family

Although the consensus full-length *Tdr1* DNA sequence (Fig. 3) did not yield a single, uninterrupted open reading frame, the amino acid sequence of the putative *Tdr1* transposase can be predicted (Fig. 4). Comparison of this conceptual *Tdr1* transposase to other members of the *Tc1*-like family reveals that it shares the highest degree of amino acid identity with elements found in fish, although *Tdr1* is far more similar to the salmonid transposons than to *IpTc1* from catfish (Table 1). Moreover, Table 1 indicates that the presumptive fish transposases are about equally distant from those of insects and nematodes. The conserved protein domains characteristic of the *Tc1*-like family (Doak et al. 1994) can be detected in the hypothetical transposase encoded by *Tdr1*.

The consensus *Tdr1* nucleotide sequence (Fig. 3) contains imperfect (77% match) terminal inverted repeats that are 206 and 208 bp in length. Within the long inverted repeats a short 12 bp motif was identified. This motif is repeated four times: once at the beginning and once at the end of each of the long inverted repeats (Fig. 7). This arrangement of short, directly repeated sequences within long inverted repeats (IR-DR) is not typical of *Tc1*-like transposable elements. However, as shown in Table 1, the same structure was detected in *Tss1* of Atlantic salmon, in *Tom1* of rainbow trout, in *Minos*, a transposable element found in *Drosophila hydei*, and in the *S* element of *Drosophila melanogaster* (Simmons and Merriman 1994). Apparently, it is not the primary DNA sequence per se that has been conserved in these transposons, but the structure of the inverted repeats. However, the sequences of the direct repeats in the salmonid elements, *Tdr1* and the *S* element are related (Table 2). These results suggest that while the putative transposase gene products of the *Tc1*-type and the *Tdr1*-type elements are apparently related at the amino acid sequence level, sequences flanking the coding regions are different in base composition, sequence, and structure (Fig. 7).

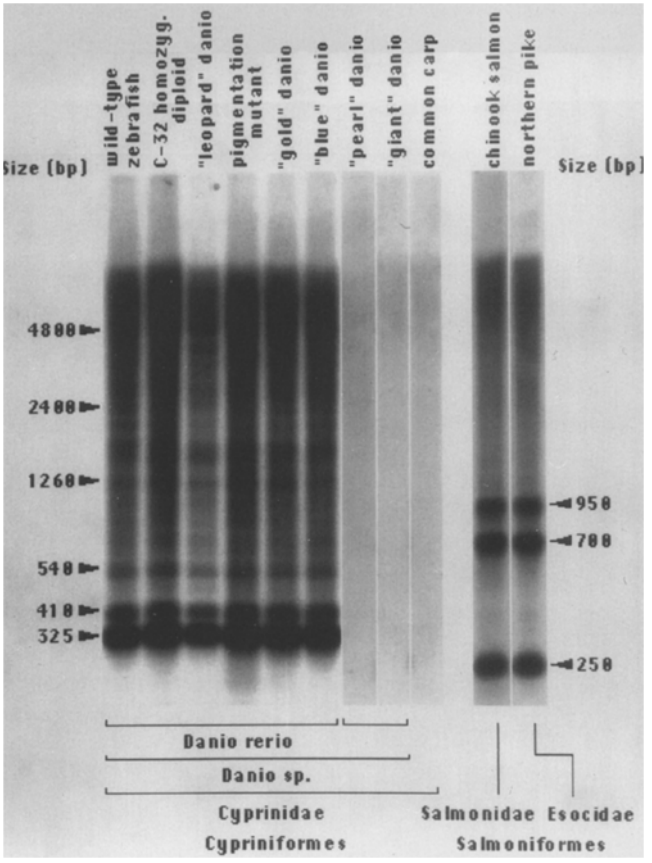



**Fig. 8** Distribution of *Tc1*-like sequences in the genus *Danio*. Equal amounts of genomic DNA samples from adult fish from various zebrafish strains, members of the genus *Danio*, common carp, chinook salmon (*Oncorhynchus tschawitscha*) and northern pike (*Esox lucius*) were digested with restriction endonuclease *Hin*dIII, electrophoresed, Southern-blotted, and hybridized to the radioactively labeled probe shown in Fig. 5A. Common names of the species are given on the top of each lane, while their phylogenetic classification is indicated at the bottom. Fragment sizes are shown in the margins

### Distribution of *Tc1*-like sequences in the genus *Danio*

As noted earlier, if *Tdr1* is to be an effective agent for insertional mutagenesis and gene tagging, the element should not be present in the genome in a high copy number. Because *Tdr1* is present in high copy number in wild-type *Danio rerio*, various zebrafish strains and other related fish species were examined for the prevalence of *Tc1*-like elements. Genomic DNA samples were isolated, digested with *Hin*dIII, Southern blotted, and hybridized to the *Tdr1* probe used earlier (Fig. 5A). When samples from various *Danio rerio* strains were compared, the patterns of hybridization showed an overall similarity (Fig. 8), suggesting widespread occurrence of *Tdr1* throughout these genomes. DNA samples of zebrafish from Hong Kong, Indonesia and Singapore were also analyzed. Although enhanced resolution of DNA fragments generated by *Hin*dIII exposed minor RFLPs between individuals (data not shown), the patterns suggest that the numbers and locations of most of the *Tdr1* sequences in these zebrafish strains are highly conserved. A few, intensively hybridizing bands were evident on the autoradiogram, suggesting either a tandemly repeated genomic arrangement of some *Tdr1* sequences, or "punchout" fragments released by *Hin*dIII from a subclass of *Tdr1* elements. A DNA fragment representing the strongest, 325 bp band was cloned, sequenced and found to be an internal *Hin*dIII fragment of *Tdr1*.

When samples of different *Danio* species were analyzed, the *Tdr1* probe did not appreciably hybridize to DNA prepared from pearl danio (*Danio albolineatus*) and giant danio (*Danio aequipinnatus*) (Fig. 8), species considered to be the closest relatives of zebrafish (Meyer et al. 1993). We observed no significant hybridization to DNA from carp (Fig. 8), which, like zebrafish, also belongs to the family *Cyprinidae*. Hybridization with a zebrafish probe for the eIF4E gene (S. Fahrenkrug, unpublished) showed that comparable amounts of nondegraded DNA were present in each lane (data not shown). When hybridization with the *Tdr1* probe was repeated at low stringency, there was evidence of some binding to giant danio DNA, and even less to pearl danio or carp DNAs (data not shown), suggesting that there may be related repetitive elements in these species. In contrast, intensive hybridization was obtained with DNA prepared from chinook salmon (*Oncorhynchus tschawitscha*) and northern pike (*Esox lucius*) (Fig. 8), phylogenetically distant species from the order *Salmoniformes*. Based on the hybridization signal intensities, the degree of sequence identity between the putative coding region of *Tdr1* and related elements from chinook salmon and northern pike is considerably greater than the similarity between *Tc1*-like sequences of zebrafish and those of pearl danio, giant danio and carp. As illustrated in Fig. 9, our results indicate that the distribution and conservation of *Tc1*-like sequences is not congruent with the established hierarchy of these fish species (Bernardi et al. 1993; Meyer et al. 1993).

## Discussion

The family of apparently defective *Tdr1* transposable elements found in zebrafish, is represented by about 1000 copies, accounting for about 0.07% of the haploid genome; this is within the range of from 0.02 to 0.5% found for members of the *Tc1* family in other organisms. Most *Tdr1* elements seem to be interspersed throughout the genome of zebrafish, and are about 1250 bp in length, with a deletion of sequences that encode the N-terminal portion of the putative transposase. In addition to this large deletion, all genomic copies and cDNA products we analyzed contained frameshift mutations and stop codons in the putative transposase coding region. The 1250 bp deletion derivatives do not seem to have been produced by independent deletion events, because sequences at the deletion breakpoint appear to be conserved. Instead, our data support the interpretation that a deleted form of *Tdr1* was mobilized, amplified and in-

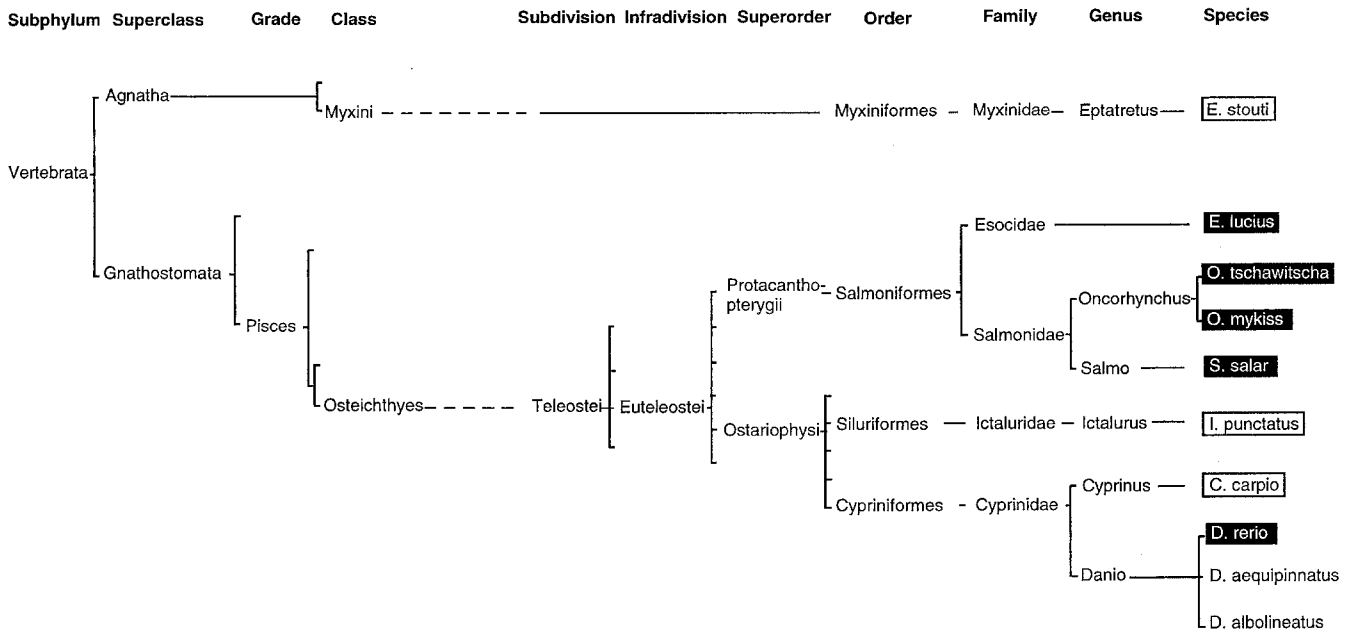



**Fig. 9** Phylogenetic distribution of *Tc1*-like sequences in fish. An abbreviated phylogenetic tree of fishes is shown to illustrate the relationships of species in which *Tc1*-like sequences have been identified. Only those taxonomical categories are shown that are relevant to the species examined in this study. *Black background* indicates species in which *Tc1*-like sequences have been detected with sequence similarity to *Tdr1* of greater than 65%, or which yield strong signals when hybridized to *Tdr1* under stringent conditions. *White boxes* mark those species in which *Tc1*-like sequences have been detected with sequence similarity to *Tdr1* of less than 65%. In the unmarked species, *Tc1*-like sequences cannot be unequivocally detected by Southern hybridization. References: *Eptatretus stouti* (Heierhorst et al. 1992); *E. lucius* (this paper); *Oncorhynchus tschawitscha* (this paper); *Oncorhynchus mykiss* (Radice et al. 1994); *Salmo salar* (Radice et al. 1994); *Ictalurus punctatus* (Wilson et al. 1990); *D. rerio* (Radice et al. 1994, and this paper); *D. aequipinnatus* (this paper); *D. albolineatus* (this paper); *Cyprinus carpio* (this paper)

serted into many loci in the zebrafish genome. For unknown reasons, this derivative of *Tdr1* has a selective advantage in transposition over other copies. Similar length homogeneities have been found for the maize transposons *Ds1* and *Tourist* (Bureau and Wessler 1992; MacRae and Clegg 1992), suggesting that size may influence the mobility of these elements. The dominance of a mutated form of a transposable element raises a question concerning its potential function. Truncated versions of transposases might have regulatory functions (Handler et al. 1993; Rio 1991). What role in the regulation of transposition this dominant form of *Tdr1* plays, if any, is unknown.

Besides the predominant 1250 bp form, sequences corresponding to the deleted region can be detected in the zebrafish genome. These sequences appear to be fragments of putative full-length elements that are (or were) capable of producing active transposase. However, no spontaneous mutation caused by a *Tdr1* insertion has been reported; therefore, the presence of active *Tdr1* elements in the genome of zebrafish still remains to be demonstrated experimentally. Although a mutant allele of the zebrafish *ntl* gene appears to be associated with an insertional element (Schulte-Merker et al. 1994), the inserted sequence and *Tdr1* do not share any apparent sequence similarity.

One of the goals of our work was to find a *Tdr1* element that encoded an active transposase. Despite intensive searches using a variety of PCR primers, we were unable to find a complete transposase coding region, either because none exists or because of the extremely high background of defective 1250 bp elements. However, based on the several sequences we have obtained (Fig. 2), reconstruction of an active *Tdr1* transposon capable of producing transposase should be feasible. Such an element would have the potential to be developed into a gene-tagging agent or a marker for genetic mapping in fish. Work is currently underway to engineer an active transposase protein. The engineered element can be used in species related to *Danio rerio*, e.g. pearl danio or giant danio, or, alternatively, an additional marker sequence can be added for the purposes of transposon tagging and recoverable insertional mutagenesis in zebrafish.

In the course of this work, several observations regarding the phylogeny of transposable elements were made. Comparison of *Tdr1* to other *Tc1*-like transposons allows us to draw three major conclusions about the evolution of these sequences. Firstly, *Tdr1* elements have long inverted terminal repeats and a short, directly repeated sequence within the inverted repeats. We propose that by the criterion of the structure of inverted repeats, the elements isolated from salmonid fishes, *Minos* from *Drosophila hydei*, the *S* element from *Drosophila melanogaster*, and *Tdr1* form a subgroup within *Tc1*-like transposable elements (IR-DR subgroup). The conservation of a particular inverted repeat structure suggests that it plays a role in the transposition process, and that mem-

bers of the IR-DR subgroup may transpose by similar mechanisms. Subterminal repeats of the maize *Ac* and *Spm* elements have been shown to serve as binding sites for transposase (Gierl et al. 1988; Kunze and Starlinger 1989), however, such a function remains to be tested for *Tdr1*. Secondly, the phylogeny of the hypothetical protein products encoded by *Tc1*-like transposons is not consistent with the phylogeny of the flanking regions of the transposase genes. Based on structural conservation of the inverted repeats in the IR-DR subgroup, we expected that the putative transposases encoded by these elements would also be similar. To our surprise, we found that the insect and the nematode proteins are about equally distant from *Tdr1*. This finding supports the hypothesis that the transposase coding regions and their flanking sequences may be shaped by independent evolutionary forces. Thirdly, phylogeny of these transposable elements does not follow the established phylogeny of the species that carry them. For example, the considerably higher similarity of *Tdr1* to *Tss1* of salmon than to *IpTc1* of catfish, which is phylogenetically closer to zebrafish, cannot be justified by the current classification of teleost fishes. Moreover, *Tc1*-like sequences of zebrafish seem to be more divergent from those of the pearl danio and giant danio – very close relatives of zebrafish – than those of phylogenetically distant salmonid fishes. A similar phenomenon has been reported for the *mariner* transposon in insects; distribution of this element is discontinuous in the *melanogaster* species subgroup of *Drosophila*, but it is present in the genus *Zaprionus* (Maruyama and Hartl 1991).

**Acknowledgements** We are grateful to Dr. Scott W. Emmons for sharing his results with us and thank him for helpful conversations and plasmids containing fish *Tc1*-like elements. We thank Dr. Michael Simmons for discussions and reading the manuscript. The zebrafish cDNA library was a kind gift from Drs. Jay White and Martin Petkovich, and the zebrafish genomic library was obtained from Dr. Walter Sauerbier. The Singapore, Hong Kong and Indonesia strains of zebrafish were from Dr. Didier Stainier. This work was supported by USDA grant No. 92-37205-7842 and NIH grant No. RR06625.

## References

Bernardi G, D'Onofrio G, Caccio S, Bernardi G (1993) Molecular phylogeny of bony fishes, based on the amino acid sequence of the growth hormone. J Mol Evol 37:644–649

Brezinsky L, Wang GV, Humphreys T, Hunt J (1990) The transposable element *Uhu* from Hawaiian *Drosophila* – member of the widely dispersed class of *Tc1*-like elements. Nucleic Acids Res 18:2053–2059

Bureau TE, Wessler SR (1992) *Tourist*: a large family of small inverted repeat elements frequently associated with maize genes. Plant Cell 4:1283–1294

Caizzi R, Caggese C, Pimpinelli S (1993) *Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics 133:335–345

Collins J, Forbes E, Anderson P (1989) The *Tc3* family of transposable genetic elements in *Caenorhabditis elegans*. Genetics 121:47–55

Daboussi MJ, Langin T, Brygoo Y (1992) *Fot1*, a new family of fungal transposable elements. Mol Gen Genet 232:12–16

Datta U, Dutta P, Mandal RK (1988) Cloning and characterization of a highly repetitive fish nucleotide sequence. Gene 62:331–336

Doak TG, Doerder FP, Jahn CL, Herrick G (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common "D35E" motif. Proc Natl Acad Sci USA 91:942–946

Driever W, Stemple D, Schier A, Solnica-Krezel L (1994) Zebrafish: genetic tools for studying vertebrate development. Trends Genet 10:152–159

Ekker M, Fritz A, Westerfield M (1992) Identification of two families of satellite-like repetitive DNA sequences from the zebrafish (*Brachydanio rerio*). Genomics 13:1169–1173

Emmons SW, Yesner L, Ruan KS, Katzenberg D (1983) Evidence for a transposon in *Caenorhabditis elegans*. Cell 32:55–65

Ferris SD, Whitt GS (1977) The evolution of duplicate gene expression in the carp (*Cyprinus carpio*). Experientia 33:1299–1301

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. Trends Genet 5:103–107

Franz G, Savakis C (1991) *Minos*, a new transposable element from *Drosophila hydei*, is a member of the *Tc1*-like family of transposons. Nucleic Acids Res 19:6646

Franz G, Loukeris TG, Dialektaki G, Thompson CRL, Savakis C (1994) Mobile *Minos* elements from *Drosophila hydei* encode a two-exon transposase with similarity to the paired DNA-binding domain. Proc Natl Acad Sci USA 91:4746–4750

Gierl SR, Lutticke S, Saedler H (1988) TnpA product encoded by the transposable element *En-1* of *Zea mays* is a binding protein. EMBO J 7:4045–4053

Goodier JL, Davidson WS (1994) *Tc1* transposon-like sequences are widely distributed in salmonids. J Mol Biol 241:26–34

Handler AM, Gomez SP, O'Brochta DA (1993) Negative regulation of *P* element excision by the somatic product and terminal sequences of *P* in *Drosophila melanogaster*. Mol Gen Genet 237:145–151

Harris LJ, Baillie DL, Rose AM (1988) Sequence identity between an inverted repeat family of transposable elements in *Drosophila* and *Caenorhabditis*. Nucleic Acids Res 16:5991–5997

Hartl DL, Lozovskaya ER, Lawrence JG (1992) Nonautonomous transposable elements in prokaryotes and eukaryotes. Genetica 86:47–53

He L, Zhu Z, Faras AJ, Guise KS, Hackett PB, Kapuscinski AR (1992) Characterization of *Alu*I repeats of zebrafish (*Brachydanio rerio*). Mol Marine Biol Biotech 1:125–135

Heierhorst J, Lederis K, Richter D (1992) Presence of a member of the *Tc1*-like transposon family from nematodes and *Drosophila* within the vasotocin gene of a primitive vertebrate, the Pacific hagfish *Eptatretus stouti*. Proc Natl Acad Sci USA 89:6798–6802

Henikoff S (1992) Detection of *Caenorhabditis* transposon homologs in diverse organisms. New Biol 4:382–388

Ivics Z, Izsvák Zs, Hackett PB (1993) Enhanced incorporation of transgenic DNA into zebrafish chromosomes by a retroviral integration protein. Mol Marine Biol Biotech 2:162–173

Kido Y, Aono M, Yamaki T, Matsumoto K, Murata S, Saneyoshi M, Okada N (1991) Shaping and reshaping of salmonid genomes by amplification of tRNA-derived retroposons during evolution. Proc Natl Acad Sci USA 88:2326–2330

Kunze R, Starlinger P (1989) The putative transposase of transposable element *Ac* from *Zea mays* L. interacts with subterminal sequences of *Ac*. EMBO J 8:3177–3185

MacRae AF, Clegg MT (1992) Evolution of *Ac* and *Ds1* elements in selected grasses (*Poaceae*). Genetica 86:55–66

Maruyama K, Hartl DL (1991) Evidence for interspecific transfer of the transposable element *mariner* between *Drosophila* and *Zaprionus*. J Mol Evol 33:514–524

Meyer A, Biermann CH, Orti G (1993) The phylogenetic position of the zebrafish (*Danio rerio*), a model system in developmental biology: An invitation to the comparative method. Proc R Soc Lond 252:231–236

Naruse K, Mitani H, Shima A (1992) A 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

Nüsslein-Volhard C (1994) Of flies and fishes. Science 266: 572–574

Prasad SS, Harris LJ, Baillie DL, Rose AM (1991) Evolutionarily conserved regions in *Caenorhabditis* transposable elements deduced by sequence comparison. Genome 34: 6–12

Radice AD, Bugaj B, Fitch DHA, Emmons SW (1994) Widespread occurrence of the *Tc1* transposon family: *Tc1*-like transposons from teleost fish. Mol Gen Genet 244: 606–612

Rio DC (1991) Regulation of *Drosophila P* element transposition. Trends Genet 7: 282-287

Rossant J, Hopkins N (1992) Of fin and fur: mutational analysis of vertebrate embryonic development. Genes Dev 6: 1–13

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Simmons M, Merriman P (1994) The *S* element: a *Drosophila* member of the *Tc1-mariner* superfamily of transposable elements. Abstracts of the 35th Annual Drosophila Res Conference, Chicago, Illinois, p 221

Schulte-Merker S, van Eeden FJM, Halpern ME, Kimmel CB, Nüsslein-Volhard C (1994) *no tail* (*ntl*) is the zebrafish homologue of the mouse *T* (*Brachyury*) gene. Development 120: 1009–1015

Vos JC, van Luenen HG, Plasterk RH (1993) Characterization of the *Caenorhabditis elegans Tc1* transposase in vivo and in vitro. Genes Dev 7: 1244–1253

Wilson MR, Marcuz A, van Ginkel F, Miller NW, Clem LW, Middleton D, Warr GW (1990) The immunoglobulin M heavy chain constant region gene of the channel catfish, *Ictalurus punctatus*: an unusual mRNA splice pattern produces the membrane form of the molecule. Nucleic Acids Res 18: 5227–5233

Winkfein RJ, Moir RD, Krawetz DA, Blanco J, States C, Dixon GH (1988) A new family of repetitive, retroposon-like sequences in the genome of the rainbow trout. Eur J Biochem 176: 255–264

Wright JM (1989) Nucleotide sequence, genomic organization and evolution of a major repetitive DNA family in tilapia (*Oreochromis mossambicus/hornorum*). Nucleic Acids Res 17: 5071–5079