The Tol2 transposable element of the medaka fish: an active DNA-based element naturally occurring in a vertebrate genome

Akihiko Koga and Hiroshi Hori*
Division of Biological Sciences, Graduate School of Science, Nagoya University, Nagoya 464-8602, Japan

Several DNA-based transposable elements are known to be present in vertebrate genomes, but few of them have been demonstrated to be active. The Tol2 element of the medaka fish is one such element and, therefore, is potentially useful for developing a gene tagging system and other molecular biological tools applicable to vertebrates. Towards this goal, analyses of the element at the molecular, cellular and population levels are in progress. Results so far obtained are described here.

Transposable elements as genetic tools. Transposable elements serve, in various organisms, as powerful tools for genetic analyses. The simplest form of application is as transformation vectors using an element carrying a gene or other DNA fragments (cf. Rubin and Spradling 1982). Inducing insertion mutations, followed by phenotype screening, is called gene tagging and allows quick cloning of causative genes (cf. Bingham et al. 1982, Searles et al. 1982). Similar usages are in promoter trapping and enhancer trapping, with an element that carries a marker gene lacking a promoter or containing a weak promoter (cf. O’Kane and Gehring 1987). These are applications common to both of the two major classes of transposable elements: RNA-mediated elements and DNA-based elements (class I and class II elements, respectively, of Finnegan 1992). A usage unique to the latter is to induce further changes such as reversion mutations due to precise excision or a range of mutations by imprecise excision (Salz et al. 1987). Another advantage of DNA-based elements is that transposition can be started, if the necessary strains have been prepared, merely by crossing the strains (cf. Cooley et al. 1988).

With advances in technology and the accumulation of information, a new method for detecting transposable elements was devised based on similarity of nucleotide sequences. Elements having sequence similarities with the Drosophila mariner element and the nematode Tc1 element were then found in a wide range of organisms and called the mariner/Tc1 family or the Tc1/mariner family (Plasterk 1996, Hartl et al. 1997). In addition to the wide distribution, it is a feature of this family that the copy number in the genome is high, which is significant for finding elements by methods that rely on sequence similarity.

Discovery of transposable elements. New transposable elements are most frequently encountered as mutations occurring in genes of host organisms. There are many examples of mutant genes for which structural analysis resulted in identification of extra DNA fragments. Such insertion appears more likely to happen in cases of high mutation rates. The Ac element of maize and the P element of Drosophila were first hypothesized to be present as genetic factors that cause high rates of mutant phenotypes of kernel color (McClintock 1948) and of fertility (Kidwell et al. 1979), respectively. P was actually identified and cloned as an insertion sequence in the Drosophila white locus (Rubin et al. 1982) and Ac in the maize Waxy locus (Fedorch et al. 1983).

Search for transposable elements in vertebrates. With advances in technology and the accumulation of information, a new method for detecting transposable elements was devised based on similarity of nucleotide sequences. Elements having sequence similarities with the Drosophila mariner element and the nematode Tc1 element were then found in a wide range of organisms and called the mariner/Tc1 family or the Tc1/mariner family (Plasterk 1996, Hartl et al. 1997). In addition to the wide distribution, it is a feature of this family that the copy number in the genome is high, which is significant for finding elements by methods that rely on sequence similarity.

The initial attempt to search a database for a nucleotide sequence having similarity with mariner/Tc1 family elements provided the first example, from the catfish (Heni-
The method was subsequently expanded to include PCR (polymerase chain reaction) and many more elements were discovered (Table 1). However, those found in the early stage were all inactive elements, not carrying a complete internal gene.

**Active elements found in medaka fish.** The medaka fish *Oryzias latipes* is a freshwater teleost species native to Asia, including China, Japan and Korea. This fish serves as a major material for research in genetics and many body color mutants have been identified. Among mutants stocked at Nagoya University, three albino strains having different alleles at the \(i^i\) (color interferer) locus were established by H. Tomita (Yamamoto 1969, Tomita 1975). Tyrosinase is the key enzyme in melanin biosynthesis, and many cases of albinism are known to be caused by its defects (Shibahara et al. 1990, Yokoyama et al. 1990). We first cloned the wild-type tyrosinase gene of the medaka fish (Inagaki et al. 1994) and then examined the structures of the genes of the albino mutant strains (Fig. 1). The \(i^1\) allele, which causes a complete albino phenotype, was shown to carry an extra 1.9-kb DNA fragment in the coding region (Koga et al. 1995). The element has structural characteristics of a DNA-based transposable element, and we named it Tol1 (Koga et al. 1995). The third allele, \(i^5\), causing a weak albino phenotype, was shown to have a Tol2 copy at a position different from that of the \(i^4\) allele (unpublished).

Of the two elements, Tol2 was demonstrated to be active soon after its discovery. We conducted PCR with primers encompassing the Tol2 insertion point. As expected, the length of the major PCR product was 5.2 kb, that is, the length between the primers (0.5 kb) plus the length of the Tol2 copy (4.7 kb). In addition to this major product, a 0.5 kb fragment was observed. Cloning and sequencing analyses of this short fragment revealed footprints of excision of Tol2. This was the first report of an active DNA-based element in a vertebrate. The method applied was what has long been the major approach for many model organisms, that is, analysis of the structure of mutant genes. Mutations, especially those that have newly arisen, are certainly a pointer of active elements. Tomita's lifetime, patient work on collecting spontaneous mutations was the basis for our detection of active elements.

**Structure of Tol2.** Tol2 carries imperfect terminal inverted repeats of 17 bp and 19 bp, and three subterminal repeats of about 30 bp located proximal to the right terminal inverted repeat. It also contains internal inverted repeats of about 300 bp, which is unique among DNA-based elements so far identified (Fig. 2).

### Table 1. Transposable elements discovered of demonstrated to be active in vertebrates

<table>
<thead>
<tr>
<th>Host</th>
<th>Element</th>
<th>Evidence for activity</th>
<th>Isolation of autonomous copy</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>frog</td>
<td>1723</td>
<td>No</td>
<td>No</td>
<td>Kay and Dawid (1983)</td>
</tr>
<tr>
<td>frog</td>
<td>Vi</td>
<td>No</td>
<td>No</td>
<td>Schubiger et al. (1985)</td>
</tr>
<tr>
<td>frog</td>
<td>Tx1</td>
<td>No</td>
<td>No</td>
<td>Garrett and Carroll (1986)</td>
</tr>
<tr>
<td>catfish</td>
<td>IpTc1</td>
<td>No</td>
<td>No</td>
<td>Henikoff (1992)</td>
</tr>
<tr>
<td>hagfish</td>
<td>Tes1</td>
<td>No</td>
<td>No</td>
<td>Heierhorst et al. (1992)</td>
</tr>
<tr>
<td>salmon</td>
<td>SALT1</td>
<td>No</td>
<td>No</td>
<td>Goodier and Davidson (1994)</td>
</tr>
<tr>
<td>zebrafish</td>
<td>Tbr1</td>
<td>No</td>
<td>No</td>
<td>Radice et al. (1994)</td>
</tr>
<tr>
<td>salmon</td>
<td>Tss1</td>
<td>No</td>
<td>No</td>
<td>Radice et al. (1994)</td>
</tr>
<tr>
<td>human</td>
<td>trcmMLE</td>
<td>No</td>
<td>No</td>
<td>Morgan (1995)</td>
</tr>
<tr>
<td>zebrafish</td>
<td>Tdr1</td>
<td>No</td>
<td>No</td>
<td>Izsák et al. (1995)</td>
</tr>
<tr>
<td>medaka fish</td>
<td>Tol1</td>
<td>No(^a)</td>
<td>No</td>
<td>Koga et al. (1995)</td>
</tr>
<tr>
<td>human</td>
<td>humar1</td>
<td>No</td>
<td>No</td>
<td>Oosumi et al. (1995)</td>
</tr>
<tr>
<td>zebrafish</td>
<td>Tdr2</td>
<td>No</td>
<td>No</td>
<td>Ivics et al. (1996)</td>
</tr>
<tr>
<td>salmon</td>
<td>Tss2</td>
<td>No</td>
<td>No</td>
<td>Ivics et al. (1996)</td>
</tr>
<tr>
<td>human</td>
<td>Tigger1</td>
<td>No</td>
<td>No</td>
<td>Smit and Riggs (1996)</td>
</tr>
<tr>
<td>human</td>
<td>MITE</td>
<td>No</td>
<td>No</td>
<td>Reiter et al. (1996)</td>
</tr>
<tr>
<td>frog</td>
<td>TXr</td>
<td>No</td>
<td>No</td>
<td>Lam et al. (1996b)</td>
</tr>
<tr>
<td>medaka fish</td>
<td>Tol2</td>
<td>Yes</td>
<td>Yes</td>
<td>Koga et al. (1996)</td>
</tr>
<tr>
<td>zebrafish</td>
<td>Tzf</td>
<td>Yes</td>
<td>No</td>
<td>Lam et al. (1996a)</td>
</tr>
<tr>
<td>salmon</td>
<td>Sleeping Beauty</td>
<td>Yes</td>
<td>Yes(^b)</td>
<td>Ivics et al. (1997)</td>
</tr>
<tr>
<td>nematode</td>
<td>Tc3</td>
<td>Yes</td>
<td>Yes</td>
<td>Raz et al. (1997)</td>
</tr>
<tr>
<td>Drosophila</td>
<td>mariner</td>
<td>Yes</td>
<td>Yes</td>
<td>Fadool et al. (1998)</td>
</tr>
<tr>
<td>zebrafish</td>
<td>Tdr2</td>
<td>No</td>
<td>No</td>
<td>Gottgens et al. (1999)</td>
</tr>
</tbody>
</table>

\(^a\) Transposition activity has now been demonstrated (unpublished).  
\(^b\) An autonomous copy was synthesized.
Fig. 1. Structures of medaka fish tyrosinase genes. The tyrosinase gene consists of five exons (Inagaki et al. 1998) that produce a protein of 540 amino acids (Inagaki et al. 1994), as shown in the wild-type gene structure. The black arrowheads indicate the locations of the first methionine codon and the stop codon of the wild-type gene. The \( i^1 \) albino strain exhibits a complete albino phenotype and its tyrosinase gene contains a \( Tol1 \) element in the first exon. The \( i^4 \) albino strain shows a quasi-albino phenotype and its tyrosinase gene carries a \( Tol2 \) element in the fifth exon. The open triangles indicate the locations of stop codons on the assumption that transcription continues into the transposable elements and translation proceeds along the transcripts.

The entire element is 4.7 kb in length and contains a gene, composed of four exons, for its transposase. The amino acid sequence of the gene has similarities, in some blocks, to the sequences of elements of the hAT family (Calvi et al. 1991, Atkinson et al. 1993), which includes hobo of Drosophila (McGinnis et al. 1983), Ac of maize (Fordoff et al. 1983) and Tam3 of snapdragon (Sommer et al. 1985). The amino acid blocks are located in three segments previously reported to be conserved among the hAT family elements (Feldmar and Kunze 1991).

**Transposase gene.** Four open reading frames were detected in the \( Tol2 \) copy first found in the tyrosinase gene, designated as \( Tol2-tyr \). These were expected to constitute a gene for a transposase of \( Tol2 \), as is the case for Ac and hobo, and this proved true. We isolated cDNAs for mRNAs that have sequence similarity with the open reading frames from medaka fish cells, and sequencing analysis revealed that the four open reading frames roughly correspond to exons, with two major forms of \( Tol2 \) mRNAs present in medaka fish cells (Koga et al. 1999). One form comprises all four exons and encodes a 685-amino acid polypeptide, assuming that translation is initiated at the first encountered AUG codon. The other mRNA is composed of exons 2 to 4 and includes a frame for 453 amino acids. We synthesized the longer mRNA by in vitro transcription and injected it into fertilized eggs together with a donor plasmid and a target plasmid, resulting in transposition from the donor to the target plasmid (Fig. 3). We also conducted a control experiment in which the two plasmids were injected without the mRNA and, in this case, transposition was not observed. The materials we used were \( O. latipes \) (medaka fish) and \( O. melastigma \) (Indian medaka fish). The latter species does not contain \( Tol2 \) in its genome. Therefore, our results clearly demonstrated that the gene carried by \( Tol2 \) is for a transposase and the transposase has entire activity for cut-and-paste transposition (Koga and Hori 2000).
Fig. 2. Structure of the Tol2 element. Tol2 is 4,682 bp in length and its total sequence is available in DDBJ/EMBL/GenBank under accession number D84375. The black arrows and arrowheads indicate repeated sequences. The white arrows show exons for the transposase gene. Components are not drawn to scale. The internal inverted repeats constitute another, short transposable element called Angel, which is present in $10^3$ to $10^4$ copies in the genome of zebrafish (Izsvák et al. 1999) but not in the medaka fish genome except in Tol2 (Koga et al. 2000).

Transposition activity was also shown in zebrafish for both excision (Kawakami and Shima 1999) and insertion (Kawakami et al. 2000).

**Target site duplication.** The method described above for detecting transposition from donor to target plasmids allowed us to collect many independent insertion products. We compared the sequences of target site duplications among 20 examples and determined the length of the duplication to be 8 bp in all cases, with no apparent tendency to favor specific nucleotides. We also examined the flanking sequences of 20 Tol2 copies present in the genome, and obtained the same results (unpublished results).

**Copy number and variation.** Genomic Southern blot analysis of ten species in the genus Oryzias demonstrated that Tol2 is present in two, O. latipes (medaka fish) and O. curvidus (Hainan medaka fish, inhabiting Southern China and Vietnam), with copy numbers, estimated by counting hybridization bands, of about 20 and about 30, respectively. Linkage analysis of the copies showed them to be dispersed throughout the genome (Koga et al. 2000).

It is generally observed that DNA-based elements comprise both “full-length” copies and defective, shorter copies. The latter are mostly the products of internal deletion. However, this does not apply to Tol2. We examined the variation in the structures of more than 200 copies of Tol2, and showed that most, and possibly all, of the copies are 4.7 kb in length. There appeared to be no or very limited internal deletions, as shown by restriction mapping with six restriction sites. An unexpectedly small amount of variation was observed even at the nucleotide sequence level; that is, no sequence variation was evident among five randomly chosen copies. We inferred from these results that all or most of the Tol2 copies, including Tol2-tyr, are autonomous (Koga and Hori 1999).

**Transposition in the germline.** Tol2 is not as active as Ac of maize or P of Drosophila, at least under usual laboratory conditions, because genomic Southern blots of fish of two consecutive generations have not shown any band unique to the second generation (Koga and Hori 1999, Koga et al. 2000). However, transposition activity in the germline is supported by the results of Southern blotting of random samples collected from natural populations. This analysis revealed that band patterns differ from one fish to another (Koga and Hori 1999). This result indicates that the “allele frequency” at the presence/absence “locus” of Tol2 is extremely low for all the Tol2 copies and that these copies are continuously moving, or have recently moved in an explosive fashion, in the germline.

Transposition in the germline from an injected Tol2 clone has already been observed. We injected fertilized...
eggs of *O. melastigma*, in which Tol2 is absent, with a plasmid clone of Tol2 and the transposase mRNA. The eggs were raised to adults and two of the 13 fish examined were shown to be positive after tail biopsy and PCR for Tol2. One of the two fish was crossed to another and five of their 46 offspring were shown to contain Tol2 in their genomes. Cloning and sequencing of Tol2 and its flanking chromosomal regions revealed a precise Tol2 insertion carried an 8-bp target site duplication (unpublished results). It is thus evident that transposition of Tol2 occurred in the germline of the first generation. Similar results were obtained for the zebrafish (Kawakami et al. 2000) that also lacks Tol2 in its genome, except for the internal inverted repeat regions.

**Development of genetic tools.** Tol2 is one of the few DNA-based elements so far demonstrated to be active in vertebrates. We have therefore concentrated attention on this element with a view to development of genetic tools.

Research for the same purpose is being conducted by several groups with different elements, and the most successful element in this respect appears to be Sleeping Beauty (Ivics et al. 1997, Izsvák et al. 2000). Autonomous copies of mariner/Tc1 elements have yet to be found in vertebrates, but these researchers have made an autonomous copy by artificially recombining nonautonomous copies from eight fish species. This element is active not only in fish but also in various other vertebrate genomes including the human genome. Other promising elements are the exogenous elements Tc3 and mariner from the nematode and Drosophila, respectively, for which mobility in zebrafish has already been demonstrated (Raz et al. 1997, Fadool et al. 1998).

The elements described above are all members of the *mariner/Tc1* family. Tol2, in contrast, belongs to the hAT family, elements of which are also widely distributed among organisms (cf. Koga et al. 1999). hAT family ele-
ments have features different from those of the mariner/Tc1 family. One difference is the element size, mariner/Tc1 elements being mostly 1 to 2 kb and autonomous hAT elements typically 3 to 6 kb. The larger size of the latter might be an advantage for carrying large DNA fragments. Another advantage might be expected concerning the distance between the "cut" and the "paste" sites. Studies of this distance with the maize Ac element showed that it transposes preferentially to nearby regions (Machida et al. 1997). If this is also the case for Tol2, it may be useful for local targeting, creating chromosomal deletions and inversions, and generating large numbers of mutations within a single gene.

The Tol1 element of the medaka fish also has potential regarding development of molecular biological techniques because its transposition activity has recently been shown. Spontaneous reversion mutation of the tyrosinase gene was observed in the i^1 albino strain, and PCR and sequencing analyses revealed excision footprints of the element (unpublished results). In addition, the particular Tol1 insertion that generated the i^1 allele has been shown to be a recent event (Koga and Hori 1997). It is not known which other element Tol1 might be related to, because an open reading frame has not been found in any of the Tol1 copies we have so far examined. The only inference we can make at present is that Tol1 is not a mariner/Tc1 element because its target site duplication is not the "TA" dinucleotide but rather 8-bp heterogeneous nucleotides. Efforts are now being made to identify an autonomous Tol1 copy.

**Evolution of Tol2.** Apart from the molecular mechanisms of transposition and the application to biotechnology, Tol2 has an interesting feature with regard to its evolution. Horizontal transfer is thought to be a significant factor for survival of transposable elements (Kidwell 1992, Lohe et al. 1995). Evidence of its presence is typically provided by observation of inconsistency between the phylogeny of the element and that of its host species. However, among many examples, only a few allow clear conclusions because in many cases explanations other than horizontal transfer cannot be excluded (cf. Capy et al. 1994). Tol2 was found to be highly homogeneous at the nucleotide sequence level among several samples from collection sites representative of the entire distribution area of O. latipes. High sequence homogeneity was shown even between the two Tol2-carrying species, O. latipes and O. curvinautus. Comparing this low sequence variation in Tol2 with that of a nuclear gene of the host species, we proposed horizontal transfer of Tol2 from one of the two species to the other or into both of them from an unknown common source (Koga et al. 2000). If our inference is true, the mechanism by which the element crossed the species barrier is an interesting theme from the viewpoint of genome evolution, because horizontal transfer of transposable elements raises the possibility of horizontal transfer of host factors carried by transposable elements.

The authors owe much to Professor Hideo Tomita of Nagoya University, who passed away in 1998, for his support of our work on medaka fish transposable elements, including Tol2. The staff of the World Medaka Aquarium of Nagoya City Hitachiyama Zoological Garden provided us with samples of various species of fish and this was of great assistance in conducting evolutionary studies.

**REFERENCES**


Capy, P., Anxolabéhère, D., and Langin, T. (1994) The strange Tol2 copies we have so far examined. The only inference we can make at present is that Tol2 is not a mariner/Tc1 element because its target site duplication is not the "TA" dinucleotide but rather 8-bp heterogeneous nucleotides. Efforts are now being made to identify an autonomous Tol2 copy.

**Evolution of Tol2.** Apart from the molecular mechanisms of transposition and the application to biotechnology, Tol2 has an interesting feature with regard to its evolution. Horizontal transfer is thought to be a significant factor for survival of transposable elements (Kidwell 1992, Lohe et al. 1995). Evidence of its presence is typically provided by observation of inconsistency between the phylogeny of the element and that of its host species. However, among many examples, only a few allow clear conclusions because in many cases explanations other than horizontal transfer cannot be excluded (cf. Capy et al. 1994). Tol2 was found to be highly homogeneous at the nucleotide sequence level among several samples from collection sites representative of the entire distribution area of O. latipes. High sequence homogeneity was shown even between the two Tol2-carrying species, O. latipes and O. curvinautus. Comparing this low sequence variation in Tol2 with that of a nuclear gene of the host species, we proposed horizontal transfer of Tol2 from one of the two species to the other or into both of them from an unknown common source (Koga et al. 2000). If our inference is true, the mechanism by which the element crossed the species barrier is an interesting theme from the viewpoint of genome evolution, because horizontal transfer of


Genetics 62, 797–809.