Occurrence of a short variant of the Tol2 transposable element in natural populations of the medaka fish

AKIHIKO KOGA1*, SHIN SASAKI2, KIYOSHI NARUSE3, ATSUKO SHIMADA4 AND MITSURU SAKAIZUMI5

1 Primate Research Institute, Kyoto University, Inuyama City 464-8506, Japan
2 Graduate School of Frontier Sciences, University of Tokyo, Kashiwa 277-0882, Japan
3 National Institute for Basic Biology, Okazaki 444-8585, Japan
4 Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan
5 Graduate School of Science and Technology, Niigata University, Niigata 950-2181, Japan

(Received 6 July 2010 and in revised form 16 September 2010; first published online 7 December 2010)

Summary

Tol2 is a member of the hAT (hobo/Activator/Tam3) transposable element family, residing as 10–30 copies per diploid genome in the medaka fish. We previously reported that this element is highly homogeneous in structure at both the restriction map level and the nucleotide sequence level. It was, however, possible that there is variation of such a low frequency as not to have been detected in our previous surveys, in which samples from 12 geographical locations were used. In the present study, we first conducted searches of genome sequence databases of medaka, and found a 119-bp-long internal deletion. We then conducted a survey of samples from 58 locations for this deletion by performing PCR preceded by restriction enzyme digestion to increase the sensitivity to this deletion. We found that copies suffering this deletion have spread, or have been generated by multiple origins, in the northern-to-central part of mainland Japan. Thus, although the high homogeneity in structure is a distinct feature of Tol2, variation does exist at low frequencies in natural populations of medaka. The current status of Tol2 is expected to provide information with which results of future surveys can be compared for clarification of determinants of population dynamics of this DNA-based element.

1. Introduction

Transposable elements are dispersed repetitive sequences that are, or were at some time in the past, capable of moving from one chromosomal location to another. DNA-based elements comprise one major class of transposable elements. Elements of this class are transposed directly from DNA to DNA in a ‘cut and paste’ manner, while retrotransposable elements move via RNA intermediates by ‘copy and paste’ mechanisms. The transposition reaction of a DNA-based element is catalysed by an enzyme called transposase, and the gene encoding the enzyme is usually located inside the element itself. In the ‘cut’ process of the transposition reaction, although the transposase enzyme excises the entire element from the chromosome, it recognizes only the terminal regions of the element. Because of this property, a copy of the element in which the transposase gene is defective, or even missing, can be transposed as long as the enzyme is provided by a complete, transposase-gene-carrying copy coexisting in the same host cell. Copies of complete and defective forms are thus called autonomous and non-autonomous copies, respectively.

The hAT family is one of the major families of DNA-based elements, and was named after the hobo, Activator and Tam3 elements of Drosophila, maize and snapdragon, respectively (Calvi et al., 1991; Atkinson et al., 1993). Elements of this family have been found in a wide range of organisms, including animals, plants and fungi (Kempken & Windhofer, 2001; Rubin et al., 2001). In elements of this family, deletion of internal regions is the most common cause...
of transition from an autonomous copy to a non-autonomous copy (Fedoroff et al., 1983; Streck et al., 1986; Warren et al., 1994), and the underlying mechanism of such deletions is considered to be premature interruption of gap repair after excision of the element (Rubin & Levy, 1997). With one exception, all elements of the hAT family known to date comprise both autonomous and non-autonomous copies, or only non-autonomous copies, the latter case constituting the great majority. The exception is the Tol2 element of medaka Oryzias latipes, in which no copy suffering a deletion was found among more than 200 copies carried by medaka originating from a local population (Koga & Hori, 1999), or among more than 200 copies in medaka collected at another 12 geographical locations (Koga et al., 2000). We proposed, as the reason for this exceptional situation regarding an hAT family element, that Tol2 is a recent invader of the medaka genome. The almost complete lack of nucleotide sequence variation among Tol2 copies supports this view (Koga et al., 2000).

The scale of our previous surveys appears to have been sufficient for proposing the above hypothesis that Tol2 is a young element in its host species. It was, however, possible that structural variation is present at such a low frequency that it could not be detected in our previous surveys. If a low-frequency variant were found and had features that turned out to lead to future proliferation in natural populations, the current status of the variant would serve as useful information for comparison with the results of future studies. One big difference in the research infrastructure between the time of our previous surveys and now is the availability of genome sequence databases of medaka. In the present study, we first examined the databases for structural variation of Tol2, and obtained information suggesting the presence of a copy suffering an internal deletion. After confirming that the deletion exists in the fish genome and is not an artefact that occurred at the stage of library preparation for the databases, we examined by PCR the distribution of this deletion in natural medaka populations. We surveyed samples from 58 locations (about five times the number of locations sampled in our previous surveys), and found that eight of these samples carried at least one copy of the deletion variant. Possible causes of the occurrence of this internal deletion in limited, but wide, areas are discussed.

2. Materials and methods

(i) Database search

Searches for variation in the nucleotide sequence of the Tol2 element were conducted using the BLAST program of Ensembl (http://www.ensembl.org) and the medaka genome browser of the National Institute of Genetics (http://dolphin.lab.nig.ac.jp/medaka/). The query sequence for initial searches was the entire sequence of the Tol2 element (EMBL D84375).

(ii) Fish samples

We used fish samples from 58 mass-mating stocks originally collected at various geographical locations (Fig. 1). We first conducted surveys using the 12 samples that had been materials for our previous surveys (Koga et al., 2000), and the 32 samples that had been used in another study to examine the distribution of the Tol1 element (Koga et al., 2009). Having found the short version of Tol2 in some samples originating from local populations in northern Japan, we added six samples collected in or around these regions, and eight more samples from locations in areas where the sampling density had previously been relatively low. Two commonly used laboratory strains, HNI and AA2, were also included in our analysis.

Medaka shows geographical variation. According to data on isozyme frequencies (Sakaizumi, 1986) and nucleotide sequences (Takehana et al., 2005), there are four geographical populations: Northern Japan, Southern Japan, East Korea, and China and West Korea. This population structure is also illustrated in Fig. 1.

Philippine medaka Oryzias luzonensis is a species closely related to Oryzias latipes (Naruse 1996; Takehana et al., 2005). This species does not harbour the Tol2 element in its genome (Koga et al., 2000). We used genomic DNA of this species in a test for the detection ability of our PCR analysis.

(iii) Preparation of genomic DNA

We extracted genomic DNA from whole bodies of adult fish by the method described in Koga et al. (1995). The DNA concentration in the solvent (10 mm Tris (pH 8.0)) was adjusted to 20 ng/μl. Restriction enzyme digestion was conducted by treating 1 μg of genomic DNA with 20 units of enzyme. After 1 h of incubation for digestion, DNA was precipitated with ethanol and dissolved in 50 μl of the solvent, 20 ng/μl being the expected concentration on the assumption of no loss during treatments.

(iv) PCR analysis

We used ExTaq polymerase (Takara Bio Inc., Otsu, Japan) and reagents supplied with the enzyme for our PCR analysis. The reaction mixture was 25 μl in total volume, containing 0.4 μM of each primer, 2.0 mm MgCl2, 1× buffer, 0.1 μl of the polymerase, 100 ng (5 μl) of genomic DNA and extra DNA if necessary. The primers used were as follows: 1L (nt 1757–1784 of D84375), 1R (nt 2315–2288), 2L (nt 1–28) and 2R...
The PCR conditions were, unless otherwise noted, as follows: (120 s at 94°C), 32 cycles of (10 s at 98°C, 15 s at 64°C, 15 s at 72°C) and (30 s at 72°C). After the completion of PCR, 5 μl of the reaction solution was electrophoresed on a 90-mm-long agarose gel.

3. Results

(i) Identification and confirmation of an internally deleted Tol2 copy

It is a general feature of genome sequence databases that repetitive sequences, especially those of 1 kb or more in length, tend to be underrepresented. This is because assembly of shotgun sequence reads for contigs is harder with repetitive than with non-repetitive sequences. In fact, a BLAST search against the sequence files of the medaka Hd-rR strain, using the entire sequence of the Tol2 element (4682 bp) as a query, detected no contigs including the entire sequence, and there were only two hits showing alignment of more than 100 bp (data not shown) despite the presence of 13 ‘full-length’ Tol2 copies per haploid genome in this strain (Koga et al., 2006). For this reason, we could not expect clear evidence for the existence, if present, of structural variation of Tol2. We then analysed individual shotgun reads for differences from the Tol2 sequence. This analysis suggested that a Tol2 copy lacking a 119-nt block (nt 1793–1911) is present in the genome of the HNI strain. The ‘full-length’ Tol2 element carries a pair of internal inverted repeats (‘internal’ is prefixed here to distinguish from the terminal inverted repeats; designated as IIRs) of 302 bp (nt 1434–1735) and 303 bp (nt 1786–2088) and the 119-nt block is included in the right repeat unit (Figs 2 and 3). The two units are separated by 50 bp (nt 1736–1785), possibly forming a hairpin structure (Izsvak et al., 1999).

Because inverted repeat structures are known to be fragile during bacterial artificial chromosome (BAC) amplification in host bacteria (Doherty et al., 1993), there was a possibility that the deletion was an artefact that had been generated during the process of the genome library preparation. To determine if this had
happened, we performed PCR to amplify the region of the right repeat unit from genomic DNA, using primers 1L and 1R. The expected length of a PCR product was 532 bp if it had originated from the ‘full-length’ Tol2 copy, and 413 bp if from a Tol2 copy that had suffered the 119-bp deletion. PCR amplification from genomic DNA of the HNI strain resulted in two bands with mobilities corresponding to these sizes (Fig. 4, lane 2). We cloned and sequenced the product fragments (0.53 kb and 0.41 kb), and thereby confirmed that the deletion was exactly at the 119-nt block. Thus, the deletion was not an artefact that occurred during library preparation, but a structural change that occurred in one or more Tol2 copies present in the HNI strain. We designate this 119-bp-long deletion as D1, and a Tol2-copy suffering this deletion as Tol2-D1. When we need to specify a ‘full-length’ Tol2 copy, we use the designation Tol2-F.

(ii) Preparation of virtual medaka genome containing a single Tol2-D1 copy

Tol2 constitutes multicopy sequences dispersed in the medaka genome. We wanted to attempt to detect Tol2-D1 copies by PCR, but we thought that Tol2-F copies coexisting in the genomic DNA would act as competitors for primers. An ideal survey of fish samples for the presence/absence of Tol2-D1 would be one having the ability to detect a single Tol2-D1 copy coexisting with many Tol2-F copies. As material for a test of the detection ability, we prepared a virtual...
medaka genome containing a single Tol2-D1 copy and 39 Tol2-F copies.

The reason for setting 40 as the total number of Tol2 copies was as follows. We had already examined the copy numbers per diploid genome by genomic Southern blotting (Koga & Hori, 1999; Koga et al., 2000). We analysed a total of 24 fish samples, obtaining an average copy number of 17, the highest copy number 30 and the lowest copy number 10. Thus, the great majority of medaka samples were considered to contain 40 or fewer Tol2 copies.

We first amplified the entire Tol2 region from genomic DNA of an HNI fish by PCR using primers 2L and 2R, which represented the terminal regions of Tol2. In this PCR, we set the time period of the extension step at 6 min. Two bands 4.7 and 4.6 kb in length were observed, as expected, on an electrophoresis gel (data not shown). We cloned these bands into a plasmid (pT7Blue-2; 3.1 kb in length) by the TA-cloning method.

The haploid genome size of medaka has been estimated to be 0.68–0.85 Gb (Tanaka, 1995). Here we use the upper limit of the estimate, 0.85 Gb. If we assume 39 copies of Tol2-F (4.7 kb) in a diploid genome, the Tol2 regions correspond to a fraction of 1.1 × 10^{-4} [(39 × 4.7 kb)/(2 × 0.85 Gb)] of the genome. Similarly, one copy of Tol2-D1 (4.6 kb) corresponds to a fraction of 2.7 × 10^{-6} [(4.6 kb)/(2 × 0.85 Gb)]. We thus added, to 10 µg of genomic DNA of O. luzonensis, 1.8 ng [(1.1 × 10^{-4}) × (7.8/4.7) × 10 µg] of DNA of the clone of Tol2-F, and 45 pg [(2.7 × 10^{-6}) × (7.7/4.6) × 10 µg] of DNA of the clone of Tol2-D1. We designate the DNA sample prepared in this way ‘1/40’.

(iii) Test for detection ability

We conducted PCR to amplify the region between primers 1L and 1R from DNAs of the HNI strain, the AA2 strain, and the ‘1/40’ sample (Fig. 4). Before PCR amplification, genomic DNA was treated in three ways: no restriction enzyme digestion, digestion with PvuII and digestion with TaqI.

PvuII and TaqI have several recognition sites in the Tol2 sequence (Fig. 2). In the region of D1, there is one site for TaqI. It can thus be expected that digestion of DNA with TaqI would prevent amplification of fragments from Tol2-F copies and lead to a higher efficiency of amplification from Tol2-D1 copies. PvuII does not have a recognition site in the region between 1L and 1R.

PCR amplification from undigested DNA of the HNI fish (Fig. 4, lane 2) yielded two bands (0.53 and 0.41 kb), indicating that the HNI fish contains both types of Tol2 copies. Only the 0.53-kb band was observed in the lane for the 1/40 sample. It was thus likely that the PCR was not sufficiently powerful to detect a single Tol2-D1 copy coexisting with 39 Tol2-F copies. The AA2 fish also exhibited only the 0.53-kb band, but this does not necessarily mean that all copies in this fish were Tol2-F copies.

PCR using PvuII-digested DNA (lanes 5–7) resulted in band patterns with no detectable difference from those of undigested DNA (lanes 2–4). We can infer from this result that restriction enzyme digestion and subsequent manipulations (such as ethanol precipitation that might cause loss of short restriction fragments) had no positive or negative effects on the detection ability of the PCR analysis.

Digestion with TaqI resulted in clear differences in band patterns. The HNI fish (lane 8) exhibited a 0.41-kb band and no 0.53-kb band, and the 0.41-kb band was broader than that in lane 2. The 1/40 sample (lane 10) exhibited a 0.41-kb band, in contrast to the absence of a band of this size in lane 4. Although this band was narrower than the corresponding band of the HNI fish (lane 8), it was distinguishable as a clear band. The AA2 fish exhibited a faint 0.53-kb band and no 0.41-kb band. A plausible explanation for this result is that all Tol2 copies in this fish are Tol2-F copies. The faint 0.53-kb band can be thought to be a product from a small amount of DNA fragments left uncut after the TaqI treatment, for reasons such as DNA methylation at the TaqI-recognition site or stochastic non-association of enzyme and substrate.

These results indicate that digestion of genomic DNA with TaqI has a positive effect on the ability to detect a D1-suffering copy in subsequent PCR analysis, and that our method is sufficiently powerful to detect a single Tol2-D1 copy coexisting with 39 Tol2-F copies.

(iv) Distribution of deletion

We examined 58 fish samples for the presence of Tol2-D1 in their genomes. Fig. 5a shows the results of PCR amplification in which genomic DNA was used without restriction enzyme digestion. The primary purpose of this analysis was to confirm that the quality of genomic DNA was high enough for PCR and that the PCR worked. All samples exhibited a 0.53-kb band, and a 0.41-kb band was observed as an additional band in five samples (5, 16, 19, 21 and 23). Fig. 5b shows the results of PCR amplification from TaqI-digested genomic DNA. The 0.41-kb band was observed in all five samples that exhibited this band in Fig. 5a, and in three additional samples (6, 10 and 14). Thus, these eight samples carry in their genomes at least 1 Tol2-D1 copy together with Tol2-F copies, and the other 50 samples can be thought to have only Tol2-F copies. Five of the eight samples exhibiting the 0.41-kb band originated from the Northern Japan population, and the other three samples from the Southern Japan population.
Sequencing analysis of PCR products

We cloned and sequenced the 0.41-kb bands from the eight samples, and confirmed that their deletion breakpoints were identical to that of the 0.41-kb band of the HNI strain (Fig. 2).

4. Discussion

(i) Origin of D1

In the present study, we first searched genome sequence databases of medaka for variation in the sequence of Tol2, and obtained information suggesting the presence of an internally deleted Tol2 copy. We confirmed, by cloning and sequencing, the occurrence of the deletion, which we named D1. We then established an effective method to detect D1, and surveyed fish samples originally collected at 58 locations. Eight of these samples were found to carry Tol2-D1 copies.

Medaka inhabits East Asia, including Japan, Korea and China. Although the fish samples we used in this study may not represent well the four geographical populations of medaka, we can see a clear feature of the distribution of D1: the occurrence of D1 is restricted to the Northern Japan population and the eastern part of the Southern Japan population. This is, however, not a small area, considering that medaka is a freshwater fish species, mountain ranges divide the two populations, and medaka eggs and embryos are not tolerant to drying. The maximum distance between collection sites in which D1 was found is 600 km (between 14 and 16), and this distance appears to be too far for fish to migrate in a short time. The age of the deletion is likely to be fairly old if we assume that the Tol2-D1 copies we found originated from a single mutational event. Even if this is true, however, the age of D1 is not likely to be as old as that of the species. This is because Tol2 is highly homogeneous at the nucleotide sequence level (Koga et al. 2000): the Tol2 sequences (4682 bp, including exons of a total of 2319 bp) were virtually identical between samples from the Northern Japan and Southern Japan populations, while 49 nucleotide substitutions have accumulated in a 2051-bp-long region for the tyrosinase gene (exons, 1045 bp; introns, 1006 bp).

An explanation alternative to single origin is that D1 was generated multiple times. However, the results we have obtained so far appear to support the
possibility of single origin. First, the breakpoint of D1 was identical among the eight locations. Second, D1 was not found in samples in western parts of the Southern Japan population (14 collection points) or Korea (13 collection points). If there is a nucleotide substitution specifically found in Tol2-D1 copies, the hypothesis of a single origin could be tested in detail. There is, however, little expectation of such a situation at present because Tol2 is highly homogeneous in nucleotide sequence (Koga & Hori, 1999; Koga et al., 2000).

(ii) Autonomy of Tol2-D1

The transposase gene of Tol2 contains four exons, and all three introns have nucleotide blocks that fit the consensus sequences of splicing donor sites (/GTRAGT; Alberts et al., 2007) and splicing acceptor sites (Y_A, NYAG; Alberts et al., 2007), just with one exceptional nucleotide in the third intron (Koga et al., 1999). D1 is located inside the first intron (Fig. 2). In the Tol2-D1 sequence, there are no nucleotide blocks that appear to fit the above consensus sequences at or around the breakpoint. For this reason, it is not likely that D1 affects the structures of mature mRNAs for the Tol2 transposase. It is, however, not known if D1 creates a novel enhancer or abolishes an existing enhancer. Gene expression experiments with cloned Tol2-D1 fragment will be required to determine whether Tol2-D1 is autonomous.

(iii) Possible mechanisms for spread of D1

On the assumption that Tol2-D1 copies had a single origin with respect to D1, we discuss below possible mechanisms for the spread of D1 in natural medaka populations.

It is a commonly accepted idea that natural selection acts against the transposition activity of transposable elements, and selection pressure is relatively weak against non-autonomous copies (Hartl et al., 1997). If Tol2-D1 is non-autonomous, weakened natural selection may be a factor contributing to the spread of D1, in addition to stochastic changes in copy number. An effect of D1 on the secondary structure of DNA might be another factor. Tol2-F has IIRs in the first intron of the transposase gene (Fig. 2). The repeats are considered to have originated from insertion of a pair of MITEs (miniature inverted-repeat transposable elements) called Angel, and energy calculations have suggested that they form a stable hairpin structure (Izsvak et al., 1999). In fact, one often encounters unusual events while handling clones of Tol2 in molecular biology experiments. For example, faint extra bands often appear in photos of gel electrophoresis, extension of complementary strands by DNA polymerase is often halted in the IIR region in reactions for DNA sequencing, and plasmid clones carrying IIRs are frequently rearranged unless one uses host bacterial strains that carry the uvrC and umuC mutations. If the IIR regions also form hairpin structures in host cells, they may give rise to a high rate of scission of chromosomal DNA at their positions, and a decrease in the efficiency of DNA duplication. D1 is located inside the right repeat unit of the IIRs. One conceivable factor that might have facilitated the spread of D1 is reduced instability of DNA sequences of this region due to a decreased size of the hairpin structure. The scale of possible hairpin structures differs between Tol2-F and Tol2-D1. The stem regions of Tol2-F1 are about 300 bp in length, with a nucleotide identity of 96.1%. The loop region is 50 bp in length. In the case of Tol2-D1, the stem regions are about 180 bp, the nucleotide identity is 95.6%, and the loop region is 177 bp. Thus, in Tol2-D1, the stem regions are shorter, and the hairpin structure is expected to be less rigid. These changes may lead to a decrease in the instability of chromosome DNA at the IIR regions of inhabitant Tol2 elements. Thus, with Tol2-D1, the survival rate of host cells per Tol2 copy may be higher, or the energy cost for chromosome duplication may be lower. To test this speculation, experiments examining DNA structures using Tol2 clones will be needed, and continuous future surveys of natural medaka populations for the distribution of D1 will also be required. The current status of D1 reported here is expected to provide information with which results of future surveys can be compared for determination of factors controlling population dynamics of this DNA-based element.

(iv) Possibility of participation of a vector

All the hypotheses we have discussed so far are based on the assumption that Tol2-D1 copies, whether they are of a single origin or multiple origins, have spread only by migration of fish and inheritance by fish. Now below we will discuss another scenario by challenging this assumption.

We now assume participation of a vector that carries the transposable element from fish to fish or from an external source to fish. We also assume that the vector is capable of moving over a long distance in a short time. Viruses and parasitic organisms are examples of such vectors. With these assumptions, we can explain all of the following results: medaka exhibits a high level of nucleotide sequence variation (Sakaizumi et al., 1986; Takehana et al., 2005; Kasahara et al., 2007), Tol2 is homogeneous in nucleotide sequence among different medaka populations and even among different species (Koga & Hori, 1999; Koga et al., 2000), and Tol2-D1 is found in different medaka populations (this study). An example
of a scenario along this line is that variation accumulated in the medaka genome while medaka inhabited East Asia for a long time, Tol2 was introduced recently into medaka through wide-range invasion by a Tol2-carrying vector, and the variant Tol2-D1 was already present at the time of the invasion.

Recent bioinformatic studies have revealed that horizontal transfer of DNA-based transposable elements has occurred frequently in vertebrates (Ray et al., 2008; Pace et al., 2008; Novick et al., 2010). For many of them, there is evidence for invasion of multiple species. The most striking example with respect to the width of the host range would be the Space Invaders element (Pace et al., 2008), which is found in the genomes of several mammals (such as bat and opossum), and in lizard and frog. These authors considered that DNA viruses, especially poxviruses, are good candidates for a vector. There are also several reports suggesting horizontal transfer of transposable elements mediated by DNA viruses (Friesen & Nissen, 1990; Jehle et al., 1998; Piskurek & Okada, 2007).

As for medaka and the Tol2 element, there is no information about an exogenous vector at present. Nevertheless, the scenario invoking such a vector is, in our opinion, worth considering because Tol2 is highly active in a wide range of organisms, once artificially introduced into genomes (Balciunas et al., 2006; Hamlet et al., 2006; Keng et al., 2009).

We are grateful to Drs Elizabeth Nakajima, Lira Yu, Hiroshi Hori and Shinichi Morishita for helpful discussions. The fish samples were obtained from the medaka division of the National BioResource Project of Japan. This work was supported by grant no. 19570003 from the MEXT of Japan to A.K., and by Global COE program A06 of Kyoto University.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2007). Molecular Biology of the Cell, 5th edn. London: Taylor and Francis Group.

Atkinson, P. W., Warren, W. D. & O’Brochta, D. A. (1993). The hobo transposable element of Drosophila can be cross-mobilized in houseflies and excises like the Ac element of maize. Proceedings of the National Academy of Sciences of the USA 90, 9693–9697.

Balciunas, D., Wangensteen, K. J., Wilber, A., Bell, J., Geurts, A., Sivasubbu, S., Wang, X., Hackett, P. B., Largespa, D. A., McIvor, R. S. & Ekker, S. C. (2006). Harnessing a high cargo-capacity transposon for genetic applications in vertebrates. PLoS Genetics 2, e169.

Calvi, B. R., Hong, T. J., Findleys, D. & Gelbert, W. M. (1991). Evidence for a common evolutionary origin of inverted repeat transposons in Drosophila and plants: hobo, Activator, and Tam3. Cell 66, 469–471.

Doherty, J. P., Lindeman, R., Trent, R. J., Graham, M. W. & Woodcock, D. M. (1993). Escherichia coli host strains SURF and SRB fail to preserve a palindrome cloned in lambda phage: improved alternate host strains. Gene 124, 29–35.

Fedoroff, N., Wessler, S. & Shure, M. (1983). Isolation of the transposable maize controlling elements Ac and Ds. Cell 35, 235–242.

Friesen, P. D. & Nissen, M. S. (1990). Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. Molecular and Cellular Biology 10, 3067–3077.

Hamlet, M. R., Yergeau, D. A., Kuliyev, E., Takeda, M., Taira, M., Kawakami, K. & Mead, P. E. (2006). Tol2 transposon-mediated transgenesis in Xenopus tropicalis. Genesis 44, 438–445.

Hartl, D. L., Lohe, A. R. & Lozovskaya, E. R. (1997). Modern thoughts on an ancient marine: function, evolution, regulation. Annual Review of Genetics 31, 337–358.

Izvsk, Z., Ivics, Z., Shimoda, N., Mohn, D., Okamoto, H. & Hackett, P. B. (1999). Short inverted-repeat transposable elements in teleost fish and implications for a mechanism of their amplification. Journal of Molecular Evolution 48, 13–21.

Jehle, J. A., Nickel, A., Vlak, J. M. & Backhaus, H. (1998). Horizontal escape of the novel Tcl-like lepidopteran transposon TcP3.2 into Cydia pomonella granulovirus. Journal of Molecular Evolution 46, 215–224.

Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., Shimizu, N., Hashimoto, Y., Yang, J., Lee, Y., Matsushima, K., Sugano, S., Sakaizumi, M., Narita, T., Ohishi, K., Haga, S., Ohta, F., Nomoto, H., Nogata, K., Morishita, T., Endo, T., Shin, I.-T., Takeda, H., Morishita, S. & Kohara, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. Nature 447, 714–719.

Kempfen, F. & Windhofer, F. (2001). The hAT family: a versatile transposon group common to plants, fungi, animals, and man. Chromosoma 110, 1–9.

Keng, V. W., Ryan, B. J., Wangensteen, K. J., Balciunas, D., Schmidt, C., Ekker, S. C. & Largespa, D. A. (2009). Efficient transposition of Tol2 in the mouse germline. Genetics 183, 1565–1573.

Koga, A. & Hori, H. (1999). Homogeneity in the structure of the medaka fish transposable element Tol2. Genetic Research Cambridge 73, 7–14.

Koga, A., Iida, A., Hori, H., Shimada, A. & Shima, A. (2006). Vertebrate DNA transposon as a natural mutator: the medaka fish Tol2 element contributes to genetic variation without recognizable traces. Molecular Biology and Evolution 23, 1414–1419.

Koga, A., Inagaki, H., Hesho, Y. & Hori, H. (1995). Insertion of a novel transposable element in the tyrosinase gene is responsible for an albino mutation in the medaka fish, Oryzias latipes. Molecular and General Genetics 249, 400–405.

Koga, A., Shimada, A., Shima, A., Sakaizumi, M., Tachida, H. & Hori, H. (2000). Evidence for recent invasion of the medaka fish genome by the Tol2 transposable element. Genetics 155, 273–281.

Koga, A., Suzuki, M., Maruyama, Y., Tsutsui, M. & Hori, H. (1999). Amino acid sequence of a putative transposase protein of the medaka fish transposable element Tol2 deduced from mRNA nucleotide sequences. FEBS Letters 461, 295–298.

Koga, A., Wakamatsu, Y., Sakaizumi, M., Hamaguchi, S. & Shimada, A. (2009). Distribution of complete and defective copies of the Tol2 transposable element in natural
populations of the medaka fish *Oryzias latipes*. *Genes and Genetic Systems* **84**, 345–352.

Naruse, K. (1996). Classification and phylogeny of fishes of the genus *Oryzias*. *Fish Biology Journal Medaka* **8**, 1–10.

Novick, P., Smith, J., Ray, D. & Boissinot, S. (2010). Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. *Gene* **449**, 85–94.

Pace, J. K. 2nd, Gilbert, C., Clark, M. S. & Feschotte, C. (2008). Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proceedings of the National Academy of Sciences of the USA* **105**, 17023–17028.

Piskurek, O. & Okada, N. (2007). Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. *Proceedings of the National Academy of Sciences of the USA* **104**, 12046–12051.

Ray, D. A., Feschotte, C., Pagan, H. J., Smith, J. D., Pritham, E. J., Arensburger, P., Atkinson, P. W. & Craig, N. L. (2008). Multiple waves of recent DNA transposon activity in the bat, *Myotis lucifugus*. *Genome Research* **18**, 717–728.

Rubin, E. & Levy, A. A. (1997). Abortive gap repair: underlying mechanism for *Ds* element formation. *Molecular and Cellular Biology* **17**, 6294–6302.

Rubin, E., Lithwick, G. & Levy, A. A. (2001). Structure and evolution of the *hAT* transposon superfamily. *Genetics* **158**, 949–957.

Sakaizumi, M. (1986). Genetic divergence in wild populations of Medaka, *Oryzias latipes* (Pisces: Oryziatidae) from Japan and China. *Genetica* **69**, 119–125.

Streck, R. D., MacGaffey, J. E. & Beckendorf, S. K. (1986). The structure of *hobo* transposable elements and their insertion sites. *EMBO Journal* **5**, 3615–3623.

Takehana, Y., Naruse, K. & Sakaizumi, M. (2005). Molecular phylogeny of the medaka fishes genus *Oryzias* (Beloniformes; Adrianichthyidae) based on nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution* **36**, 417–428.

Tanaka, M. (1995). Characterization of medaka genes and their promoter regions. *Fish Biology Journal Medaka* **7**, 11–14.

Warren, W. D., Atkinson, P. W. & O’Brochta, D. A. (1994). The *Hermes* transposable element from the house fly, *Musca domestica*, is a short inverted repeat-type element of the *hobo*, *Ac*, and *Tam3* (*hAT*) element family. *Genetical Research Cambridge* **64**, 87–97.