Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome

Corinne Dasilva*, Hajer Hadji*, Catherine Ozouf-Costaz†, Sophie Nicaud*, Olivier Jaillon*, Jean Weissenbach*, and Hugues Roest Crollius**

*Genoscope and Centre National de la Recherche Scientifique, Unité Mixte de Recherche 8030, 2 Rue Gaston Crémieux, 91057 Evry Cedex, France; and †Muséum National d’Histoire Naturelle, Centre National de la Recherche Scientifique, FR 1541, 43 Rue Cuvier, 75231 Paris Cedex 05, France

Edited by Margaret G. Kidwell, University of Arizona, Tucson, AZ, and approved August 9, 2002 (received for review May 13, 2002)

*Tetraodon nigroviridis* is among the smallest known vertebrate genomes and as such represents an interesting model for studying genome architecture and evolution. Previous studies have shown that *Tetraodon* contains several types of tandem and dispersed repeats, but that their overall contribution is >10% of the genome. Using genomic library hybridization, fluorescent in situ hybridization, and whole genome shotgun and directed sequencing, we have investigated the global and local organization of repeat sequences in *Tetraodon*. We show that both tandem and dispersed repeat elements are compartmentalized in specific regions that correspond to the short arms of small subtelocentric chromosomes. The concentration of repeats in these heterochromatic regions is in sharp contrast to their paucity in euchromatin. In addition, we have identified a number of pseudogenes that have arisen through either duplication of genes or the retro-transcription of mRNAs. These pseudogenes are amplified to high numbers, some with more than 200 copies, and remain almost exclusively located in the same heterochromatic regions as transposable elements. The sequencing of one such heterochromatic region reveals a complex pattern of duplications and inversions, reminiscent of active and frequent rearrangements that can result in the truncation and hence inactivation of transposable elements. This tight compartmentalization of repeats and pseudogenes is absent in large vertebrate genomes such as mammals and is reminiscent of genomes that remain compact during evolution such as *Drosophila* and *Arabidopsis*.

The reasons for which some eukaryote genomes accommodate large quantities of apparently unnecessary DNA, whereas others seem to either engage in a more efficient “pruning” process or are equipped with better protective mechanisms against parasitic elements remains a mystery (1). It is, however, largely accepted that transposable elements and pseudogenes are two categories of such sequences that generate large amounts of “ballast” DNA and thus contribute to increasing genome size. The genomes of *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Arabidopsis thaliana* comprise between 100 and 125 Mb of euchromatin, and this fraction has now been nearly fully sequenced (2–4). In the euchromatin of these small genomes, only between 3% and 10.5% of nucleotides are contributed by transposable elements (5). Only 10 pseudogenes have been documented in Flybase (6), 432 in Wormbase (7), and 750 in the *A. thaliana* genome (4). The heterochromatic fraction (centromeres, short chromosome arms or dispersed heterochromatin) varies between approximately 10 and 60 Mb for *A. thaliana* and *D. melanogaster*, respectively, and concentrates a large proportion of the transposable elements of these genomes (β heterochromatin) while containing few active genes (4, 8, 9), which are often interspersed with alternating blocks of tandem repeats (α heterochromatin). The majority of this sequence has not yet been sequenced and/or assembled, mostly because the repetitive nature of these regions is likely to make any shotgun sequence assembly extremely difficult.

The only repeat-rich genome that has been extensively sequenced and analyzed is the human genome (5). Here the density and distribution of transposable elements and pseudogenes in the euchromatin contrasts severely with the situation in sequenced small genomes. At least 45% of human euchromatin is contributed by transposable elements, and the annotation of chromosomes 20, 21, and 22 (10–12) shows that ~20% of annotated features are pseudogenes, which provides an estimate of 6,000 pseudogenes in the entire genome. Repeats and pseudogenes are scattered in the euchromatin as well as in the heterochromatin, with variations in local densities but with a relative overall uniformity. Is this situation seen in human general for vertebrates or is it typical of a genome that is much less efficient at the pruning of transposable elements and pseudogenes than a compact genome?

To investigate this issue, we have studied the distribution of transposable elements and pseudogenes in a compact vertebrate genome. Among vertebrates, the pufferfish *Tetraodon nigroviridis* possesses the smallest known genome (13), and we have previously characterized the sequence and genomic localization of several tandem repeat sequences (including the centromeric and subtelo-centric satellite sequences) and established a preliminary catalog of transposable elements (14). This initial analysis based on 50 Mb of random shotgun reads indicated that the genome contains very few repeats (~7%, of which 1% is contributed by transposable elements) compared with other known vertebrates such as mammals, but that most families of mobile elements, including LTR retrotransposons, non-long interspersed elements (LINEs), and DNA transposons are represented. *Tetraodon* is 18–30 million years distant from *Takifugu rubripes*, a marine pufferfish for which a whole genome shotgun (WGS) assembly is now available (15).

Here, we show that the distribution of transposable elements and pseudogenes in *Tetraodon* is highly compartmentalized between heterochromatin and euchromatin. The genomic landscape is reminiscent of that seen in compact genomes across other phyla and contrasts with the larger mammalian genomes. Based on these results we discuss the implications that compact eukaryotic genomes may deal with transposable elements using common mechanisms.

Materials and Methods

Colony Filter Hybridizations. The construction of two *Tetraodon* bacterial artificial chromosome (BAC) libraries has been described (14). An equal number of clones from library A (pBACe.3.6, *EcoRI*) and library B (pBeloBAC11, *HindIII*) were spotted on nylon membranes (Appligene, Strasbourg, France) by using a robotic arrayer (QBOT, Genetix, New Milton, U.K.). Membranes

Abbreviations: WGS, whole genome shotgun; BAC, bacterial artificial chromosome; FISH, fluorescent in situ hybridization; iSET, isolated SET; TRAP, translocase-associated protein; E2H2, enhancer of Zeste homolog 2; LINE, long interspersed element.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AJ457054, AJ313481, AJ313482, and AJ313483).

*To whom correspondence should be addressed. E-mail: hrc@geneoscope.cns.fr.*
were placed on 2YT agar overnight at 37°C, and colonies were then lysed and DNA was fixed as described (16). Each membrane contains 27,648 clones in duplicate (55,296 colonies) or 10 equivalents of the *Tetraodon* genome. Transposable element hybridization probes were PCR products amplified with incorporation of digoxigenin-labeled nucleotides (Roche Diagnostics) from the *Tetraodon* BAC clone C0AA029L14 (GenBank accession no. AJ457054; see below) and correspond to the following positions on this sequence: Dm-Line, 5' end of putative coding region between base pairs 62656 and 62902; TC1-like, nearly the entire coding sequence between base pairs 82039 and 83147; Dr-Line, 5' end of putative coding region between base pairs 108586 and 109120; and Copia-like, central region of the putative coding region between base pairs 66851 and 67038. The 10-bp satellite probe was prepared as described (14). After denaturation, all probes from transposable elements were hybridized without competition by using a nonspecific protocol as described (17).

**BAC Sequencing and Assembly.** The insert of BAC clone C0AA029L14 was sequenced by using standard protocols for shotgun sequencing of BAC inserts (GenBank accession no. AJ457054). Because of the particular complexity of this sequence, each of the 1,299 reads was manually edited and assembled using the Staden package (18), taking into account the pairing and estimated distances between end-clones. The assembly was verified and confirmed by restriction digest and electrophoresis of the BAC DNA with six enzymes expected to cut frequently (and confirmed by restriction digest and electrophoresis of the BAC clone C0AA029L14), exons of *SET* (H9251), translocom-associated protein (H9262), enhancer of Zeste homolog 2 (H1032), and 3.3% of the clones arrayed on the membranes (Fig. 1 A). The four transposable elements include Dr-Line, a LINE element similar to a *Dario rerio* (zebrafish) LINE; Dm-Line, which is a different LINE element homologous to *Drosophila* factor I; a TC1-like, and a copia-like sequences. Together the five probes (10-mer satellite, Dr-Line, Dm-Line, TC1-like, and Copia-like) identify 1,743 clones, of which 1,562 are identified with at least one transposable element and 76 clones hybridize with all four elements. The Dm-Line (B) and the TC1-like (C) probes were cohybridized by FISH to the same *Tetraodon* metaphase chromosome preparations and show distinct signals in the same pairs of small subtelocentric chromosomes. (Magnification: ×750.)

**Fluorescent in Situ Hybridization (FISH).** Biotin- and digoxigenin-labeled Dm-Line and TC1-like probes were mixed in Quantum-Appligene high stringency Hybrisol VI at 10 ng/μl each. Probes were hybridized and detected on *Tetraodon* freshly thawed chromosome preparations without any pretreatment, according to the Quantum-Appligene protocol for repetitive probes (19). Preparations were mounted in 1.2 ng/μl 4',6-diamidino-2-phenylindole in Antifade (Vector Laboratories) and analyzed by using Genus FISH-imaging equipment and software for animal chromosomes from Applied Imaging (Santa Clara, CA).

**Sequence Comparisons.** The sequence of BAC clone C0AA029L14 was annotated by comparison to nucleotide and protein databases using BLAST (20). WGS sequencing of the *Tetraodon* genome at Genoscope and the Whitehead Institute for Genome Research, (Cambridge, MA) will be described in full elsewhere. To estimate the copy numbers of enhancer of Zeste homolog 2 (EZH2), translocasm-associated protein α (TRAPα), Trazepe, and isolated SET (iSET), exons of ~100 bp were selected and compared by using BLASTN to 2.6 and 2.9 genome equivalent of shotgun reads from Genoscope and the Whitehead Institute, respectively. Alignments that were retained contained at least 20 consecutive identical bases (W = 20) and covered the entire length of the exons, and consecutive mismatches were not allowed (X = 5). The number of alignments was then adjusted to one genome equivalent for comparison purposes. Substitutions and insertion/deletion analysis of the *Trazepe* sequence was performed by first identifying all shotgun sequences that contained exons 5–8 and exons 9–12. Each set of sequences was aligned by usingPILEUP in GCG version 10–2 (21) and manually edited. Substitutions, deletions, and insertions were then automatically identified by comparison to the original EZH2 and TRAPα sequences with purpose-built scripts.

**Results**

**Tetraodon Transposable Elements Are Highly Compartmentalized in Heterochromatin.** To investigate the distribution of transposable elements in the *Tetraodon* genome, we hybridized four probes to a 10-genome equivalent *Tetraodon* BAC library. Each probe was specific for a different element, which belongs to the two main classes of transposable elements: class I retrotransposons with and without LTR (probe copia-like, Dm-Line, and Dr-Line), and DNA transposons (probe TC1-like). Each probe identifies between 1.6% and 3.3% of the clones arrayed on the membranes (Fig. 1 A), in agreement with previous results that showed that the *Tetraodon* genome contains few transposable elements compared with other vertebrate genomes (14). Strikingly, the four probes hybridize to an unexpectedly high fraction of clones in common. For instance, of 922 clones that hybridize to the Dm-Line probe (a retrotransposon), 423 clones (46%) also hybridize to the TC1-like probe (a DNA transposon). This finding shows that the four transposable elements are clustered in close proximity to each other in regions that cover ~5% of the genome (1,562 clones of 27,648).

The BAC library was then hybridized with a 10-bp tandem repeat that we previously localized specifically in short heterochromatic arms of approximately 10 pairs of subtelocentric chromosomes (19). Of the 259 positive clones, 30% are shared with the set of clones identified by the Dr-Line element, instead of the 0.015% expected if the two sequences were independently distributed in the genome. This finding clearly suggests that the satellite and the LINE element are located in the same regions of the genome. To confirm this result, we cohybridized the Dm-Line and the TC1-like probes to metaphase chromosomes by double-color FISH (Fig. 1 B and C). The two probes specifically hybridize to the same short heterochro-
matic arms of six pairs of subtelocentric chromosomes, where the 10-mer satellite resides.

**Frequent Rearrangements in Heterochromatin.** We further investigated the local organization of *Tetraodon* heterochromatin by sequencing one of the 76 BAC clones that was positive with all four probes tested on the library (BAC clone C0AA029L14, Fig. 1A). A comparison of the 115-kb insert sequence against itself shows that 34% of the sequence is duplicated at least once either in tandem or inverted orientation, in multiple fragments ranging from a few hundred base pairs to 15 kb (Fig. 2). This organization suggests that frequent rearrangements take place between such sequences in this region. The insert contains 18 transposable elements (Table 1) of which only two are complete. These are duplicate copies of a TC1-like transposon with the same in-frame stop codon within the gene, indicating that both are nonfunctional and that one originates from a local duplication. The abundance of transposable elements in this BAC clone is in sharp contrast with their average frequency in the human genome. Indeed, because there is no evidence of more than one original complete copy of the amplification

Table 1. Transposable elements, genes, and pseudogenes identified in BAC C0AA029L14

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>Copies</th>
<th>Total</th>
<th>% of BAC DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transposable</td>
<td>Copia-Ten</td>
<td>3</td>
<td>18</td>
<td>16.6</td>
</tr>
<tr>
<td>elements</td>
<td>Rex1-Ten</td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>TC1-Ten</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Factori-Ten</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maui-Ten</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BUSTER2 like</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>iSET domains</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Protein kinases</td>
<td>5</td>
<td>13</td>
<td>8.2</td>
</tr>
<tr>
<td></td>
<td>Trapeze</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unknown ORF</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Fig. 2.** Structure and annotation of BAC clone C0AA029L14 positive with all four transposable elements tested on the BAC library. Analysis of duplicated sequences larger than 1,000 bases with Miropeats (45) shows that >34% of the sequence is not unique. Arched lines join the beginning and end of duplicated regions that are represented as thick horizontal lines. The Trapeze pseudogene (filled box) and the six iSET pseudogenes (open boxes) are represented on either side of the scale bar (forward and reverse strand).
of sequences derived from EZH2 and TRAPα must be subsequent to the fusion event.

**iSET**, a Retrotransposed Pseudogene, Is Highly Amplified in the Genome. The C-terminal region of human EZH2, mouse EZH1, and fly E(z) proteins contain a highly conserved SET domain (23). This domain spans the last five exons of the predicted Tetraodon EZH2 protein. Because Trapeze is similar only to the first eight exons of EZH2, it does not contain the SET sequence. However, the sequence of the BAC clone C0AA029L14 contains six dispersed copies of an iSET domain. The six 125-aa sequences are nearly identical to each other but only 57% similar and 39% identical to the Tetraodon EZH2 SET domain (see Fig. 5, which is published as supporting information on the PNAS web site, www.pnas.org). The iSET sequence is intronless and contains a stop codon in position 77, implying that it is not part of a functional protein. A comparison of iSET to TREMBL (24) shows that the best alignment is obtained with the SET domain of the Tetraodon EZH2 domain. Taken together, these observations suggest that iSET is a pseudogene that originates from the reverse transcription of a truncated 3′ EZH2 mRNA. The sequences flanking the six copies of iSET in the insert of the BAC clone are always identical up to at least a few hundred base pairs upstream and downstream, indicating that iSET does not multiply as an isolated sequence but rather as part of larger segmental duplications. The minimal region of sequence identity between the six loci containing iSET is 2.3 kb. Interestingly the sequences flanking iSET sequences always include the same 3′ segment of a LINE element 900 bp upstream, pointing toward a possible mechanism for the appearance of this pseudogene (see Discussion). No poly(A) tail in the immediate downstream region of iSET could be identified. The high level of divergence between iSET and the EZH2 SET domain indicates a very ancient retrotranscription event whereas the high similarity between six copies that are presumably under no selection indicate a very recent serial duplication event. When searched against the two available Tetraodon WGS databases, iSET identifies an average of 240 sequences per genome, and thus is ∼5 times more abundant than Trapeze (Table 2).

![Fig. 3. Exon-intron structure of the Trapeze pseudogene (GenBank accession no. AJ313483; Middle) and the native Tetraodon EZH2 gene (GenBank accession no. AJ313481, Top) and TRAPα genes (GenBank accession no. AJ313482, Bottom). Filled exons indicate EZH2 exons both in the native gene and Trapeze, and open exons indicate TRAPα exons, both in Trapeze and the native gene. Exon 9 of Trapeze is the result of a fusion between part of exon 9 of EZH2 and part of exon 5 of TRAPα. The Tetraodon EZH2 gene spans <4 kb, a 10-fold reduction in size compared with the human EZH2 gene (46). Underlined exons were used to evaluate the number of copies of each gene in the Tetraodon genome by sequence comparison. The region in Trapeze underlined with a thick dashed line was used as probe in FISH and BAC library hybridization, and the region underlined with a thin dashed line was used for deletion rate measurements. The dashed box spans the exons that code for the SET domain of EZH2.](image)

**Table 2. Copies of EZH2, TRAPα, Trapeze, and iSET exons identified in Tetraodon shotgun data**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Number of copies in 1× Genosope shotgun</th>
<th>Number of copies in 1× WICGR shotgun</th>
</tr>
</thead>
<tbody>
<tr>
<td>TRAP_exon2</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>EZH2_exon12</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Trapeze_exon3</td>
<td>100</td>
<td>31</td>
</tr>
<tr>
<td>Trapeze_exon12</td>
<td>75</td>
<td>28</td>
</tr>
<tr>
<td>iSET</td>
<td>192</td>
<td>318</td>
</tr>
</tbody>
</table>

WICGR, Whitehead Institute for Genome Research.

![Fig. 4. Hybridization results from the iSET domain and the Trapeze pseudogene, in comparison with those of transposable elements. Of the 1,136 BAC clones identified by iSET, 1,028 were already identified by one of the four transposable elements previously hybridized to the library. Similarly, of the 396 clones positive with Trapeze, 368 are also positive with iSET, indicating that both sequences are always physically associated in the genome.](image)
Table 3. Substitutions, insertions, and deletions in exons and introns of Trapeze pseudogenes

<table>
<thead>
<tr>
<th></th>
<th>Substitutions</th>
<th>Insertions</th>
<th>Deletions</th>
<th>Average insertion sizes, bp</th>
<th>Average deletion sizes, bp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exons</td>
<td>803</td>
<td>44</td>
<td>82</td>
<td>1.3</td>
<td>8.3</td>
</tr>
<tr>
<td>Introns</td>
<td>481</td>
<td>21</td>
<td>9</td>
<td>1.8</td>
<td>2.7</td>
</tr>
<tr>
<td>Total</td>
<td>1,282</td>
<td>65</td>
<td>91</td>
<td>1.5</td>
<td>7.7</td>
</tr>
</tbody>
</table>

To enable comparisons between exons and introns, values for introns are adjusted to correct for the fact that the total size of introns is smaller than the total size of exons. Values for exons are absolute values. The value of DNA loss rate (D; see text) was, however, calculated across all exons and introns by using uncorrected values.

Discussion

The distribution of transposable elements and pseudogenes in the compact pufferfish T. nigroviridis genome is in sharp contrast to the situation seen in the larger genome of other vertebrates such as human or mouse. In Tetraodon, transposable elements are tightly clustered in a specific compartment where abundant satellite sequences are found and they constitute a minimal fraction of the genome (<10%). These regions are the short heterochromatic arms of approximately 10 pairs of small subtelocentric chromosomes. In addition, although pseudogenes were previously thought to be an extremely rare feature in pufferfishes with a single documented case in Takifugu rubripes (30), we show that the Tetraodon genome contains several hundred copies of a few highly amplified pseudogenes, found in close association with transposable elements.

A number of studies have already shown that transposable elements are preferentially clustered in or around centric and telomeric heterochromatin in other sequenced genomes such as that of Drosophila and Arabidopsis (4, 8, 31–34). Although some plant and insect genomes may reach enormous sizes in some species, that of Arabidopsis and Drosophila each represent a very compact version of the genome in their respective phylum, as does the genome of Caenorhabditis elegans among nematodes. It is striking that the genome of Tetraodon, the smallest known vertebrate genome, shares a number of features with these compact genomes, such as a preferential clustering of transposable elements in heterochromatin and the paucity of euchromatic pseudogenes. This feature is not shared with other vertebrate genomes such as mouse or human, where transposable elements and pseudogenes are abundant and well dispersed in euchromatin (5). Thus, we can delineate across several phyla a common trend in four small genomes: they are not immune to repeat invasions but the amplification and dispersion of parasitic sequences are somehow contained and show similar distribution profiles.

In Tetraodon, we show that some pseudogenes are highly amplified but remain exclusively in heterochromatin. The sequencing of a BAC clone from such a region shows that the multiple copies amplify by duplications of large segments as opposed to being restricted to the pseudogene sequences only. In all copies found, the iSET pseudogene, which corresponds to the last four exons of the EZH2 gene, is systematically downstream of and on the same segment as the 3′ end of a LINE element highly similar to the Drosophila factor I element (35) (Dm-Line in the hybridization experiments). It is thus possible that this LINE element was once inserted in intron 14 (Fig. 3) of the EZH2 gene and generated the iSET pseudogene via a transduction or “read through” mechanism where the LINE RNA polymerase skipped its own 3′ end to instead use the poly(A) signal of the EZH2 gene (36–38). Because the current Tetraodon EZH2 gene does not show any trace of this LINE element, this event from the intron of the EZH2 gene. We believe that the emergence of the first copy of the iSET pseudogene and the amplification of the different copies seen in the genome today are two independent events separated by a long evolutionary distance, judging by the high sequence similarity between the different copies (98% amino acid similarity), but the high divergence between these sequences and the original EZH2 exons (57% similarity).

Trapeze, the second highly prevalent pseudogene in T. nigroviridis heterochromatin, is not retroprocessed but is a chimeric structure between part of the EZH2 gene and part of the TRAPc gene. The much higher sequence similarity between the homologous parts of Trapeze and EZH2 compared with iSET and EZH2 suggests that the emergence of Trapeze occurred later than that of iSET.

The analysis of the spectrum of substitutions, insertions, and deletions in Trapeze is in agreement with previous studies showing that the number of deletions is higher than insertions in neutrally evolving DNA (27, 39). However, although most studies have focused on defunct transposable elements or retroprocessed pseu-
heterochromatin, or (the genome, (seem to reside almost exclusively in heterochromatin, one question the size of a genome (1, 40). Because Trapeze and ISET sequences seem to reside almost exclusively in heterochromatin, one question is whether their absence from euchromatin is the result of (i) a more intense deletion rate of parasitic sequences in this compartment of the genome, (ii) a transfer of such sequences from euchromatin to heterochromatin, or (iii) an amplification that is itself confined to heterochromatin. Any of these hypotheses would require a mechanism that is not yet known to occur in eukaryotic cells.

It is intriguing that the EZH2 gene is at the origin of both Trapeze and ISET pseudogenes, albeit via different routes. Strong experimental evidence links EZH2 to a conserved mechanism of eukaryotic gene silencing (41), and this protein has been shown to form a complex with the embryonic ectoderm development protein (42) and histone deacetylases that mediate the repression of gene transcription (43). More generally, SET domain proteins are seen as multifunctional chromatin regulators with roles that include telomeric and centromeric gene silencing (23) and possibly the determination of chromosome architecture. This observation raises the question of a potential link between the presence of high copy number EZH2-derived pseudogenes in transcriptionally repressed heterochromatin and the known function of EZH2 in eukaryotic cells.

In the human (5), Drosophila (44), and Takifugu (15) genome assembly projects, heterochromatin has been left out from the initial assembly for technical reasons, although many clones and shotgun sequences are available. This compartment is, however, likely to be a dynamic and vital part of the nucleus, and essential for understanding many aspects of genome evolution and nuclear organization. We would like to draw attention to the vast amount of shotgun sequences that remain to be studied in human, Drosophila, Takifugu, and probably also mouse after the initial assembly, as a way to access and study the heterochromatic compartment in different eukaryotic genomes as was done here for Tetraodon. In particular, comparisons with Takifugu heterochromatic sequences should reveal whether the same clustering of transposable elements occurs and whether Trapeze and ISET pseudogenes also exist and are similarly amplified.

Our results for the Tetraodon genome contrast with well-known mammalian genomes and underline the fact that organisms from different phyla (plants, invertebrates, and vertebrates) with small genomes appear to share common features in terms of transposable element and pseudogene content and distribution. This finding may suggest common mechanisms in dealing with such sequences and may guide further studies on the molecular evolution of pseudogenes and transposable elements for better understanding of how the smallest known vertebrate genome achieves this extraordinary efficiency in packaging genetic information.

We thank Fiona Francis for critical reading of the manuscript, Frederic Brunet, Pierre Capy, and an anonymous reviewer for insightful comments on an earlier version of the text, and the Whitehead Institute Center for Genome Research for early access to part of the Tetraodon shotgun sequences used in this study.