



An Active Non-LTR Retrotransposon With Tandem Structure in the Compact Genome of the Pufferfish *Tetraodon nigroviridis*

Laurence Bouneau, Cécile Fischer, Catherine Ozouf-Costaz, et al.

Genome Res. 2003 13: 1686-1695

Access the most recent version at doi:[10.1101/gr.726003](https://doi.org/10.1101/gr.726003)

References

This article cites 46 articles, 9 of which can be accessed free at:
<http://genome.cshlp.org/content/13/7/1686.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero cape and mask, and the Cellecta logo, which consists of a cluster of green dots.

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

An Active Non-LTR Retrotransposon With Tandem Structure in the Compact Genome of the Pufferfish *Tetraodon nigroviridis*

Laurence Bouneau,^{1,4} Cécile Fischer,^{1,4} Catherine Ozouf-Costaz,² Alexander Froschauer,³ Olivier Jaillon,¹ Jean-Pierre Coutanceau,² Cornelia Körting,³ Jean Weissenbach,¹ Alain Bernot,¹ and Jean-Nicolas Volff^{3,5}

¹Genoscope/Centre National de Séquençage and CNRS-UMR 8030, F-91057 Evry Cedex 06, France; ²Laboratoire d'Ichtyologie and Service de Systématique Moléculaire, CNRS IFR 1541, Muséum National d'Histoire Naturelle, F-75231 Paris Cedex 05, France; ³BioFuture Research Group "Evolutionary Fish Genomics", Physiologische Chemie I, Biozentrum, University of Würzburg, D-97074 Würzburg, Germany

The fish retrotransposable element *Zebulon* encodes a reverse transcriptase and a carboxy-terminal restriction enzyme-like endonuclease, and is related phylogenetically to site-specific non-LTR retrotransposons from nematodes. *Zebulon* was detected in the pufferfishes *Tetraodon nigroviridis* and *Takifugu rubripes*, as well as in the zebrafish *Danio rerio*. Structural analysis suggested that *Zebulon*, in contrast to most non-LTR retrotransposons, might be able to retrotranspose as a partial tandem array. *Zebulon* was active relatively recently in the compact genome of *T. nigroviridis*, in which it contributed to the extension of intergenic and intronic sequences, and possibly to the formation of genomic rearrangements. Accumulation of *Zebulon* together with other retrotransposons was observed in some heterochromatic chromosomal regions of the genome of *T. nigroviridis* that might serve as reservoirs for active elements. Hence, pufferfish compact genomes are not evolutionarily inert and contain active retrotransposons, suggesting the presence of mechanisms allowing accumulation of retrotransposable elements in heterochromatin, but minimizing their impact on euchromatic regions. Homologous recombination between partial tandem sequences eliminating active copies of *Zebulon* and reducing the size of insertions in intronic and intragenic regions might represent such a mechanism.

[The sequence data from this study have been submitted to GenBank/EMBL under accession nos. AL808032, AY135221, AJ496734, AJ496221, AJ496222, AJ496223, AJ496224, and AJ496225.]

The different classes of autonomous retrotransposable elements with flanking long-terminal repeats (LTRs) are related at both the structural and phylogenetic levels (Xiong and Eickbush 1990). LTRs are of primordial importance for retrotransposition and are involved in transcription initiation and termination, in synthesis of double-stranded DNA from the RNA intermediate, and are bound by the integrase (Boeke and Chapman 1991). Vertebrate retroviruses as well as *Ty1/Copia*, *Ty3/Gypsy*, and *BEL* retrotransposons have LTRs in direct orientation, but inverted repeats and split direct repeats have been observed in the *Difs1* class of retrotransposons (Goodwin and Poulter 2001).

In contrast, the absence of long flanking sequences is characteristic of non-LTR retrotransposons (also called LINES or autonomous retrotransposons). These elements are frequently truncated at their 5' end by incomplete reverse transcription of their mRNA. Several non-LTR retrotransposons, frequently telomeric, are arranged in a head-to-tail fashion, in which neighboring copies are separated by either poly(A) stretches

or target repeat oligomers (Danilevskaya et al. 1997; Takahashi et al. 1997; Arkhipova and Morrison 2001). Interestingly, the promoter of the *Drosophila* telomeric retrotransposon *HeT-A* is not located at the 5' end, but at the 3' end of the element, and drives the transcription of the downstream *HeT-A* copy in tandem arrays (Danilevskaya et al. 1997). Hence, the 3' region of *HeT-A* retrotransposons presents some functional analogy with LTRs (Pardue and Debaryshe 2000).

Both LTR and non-LTR retrotransposons encode a reverse transcriptase. In contrast, the enzymes required for the cleavage and integration at new genomic target sites can be very different. Whereas LTR retrotransposons encode either an integrase, related to the transposase of DNA transposons, or a λ recombinase-like protein, non-LTR retrotransposons can encode an apurinic/apyrimidinic endonuclease related to cellular DNA repair enzymes, or a restriction enzyme-like (REL) endonuclease (Feng et al. 1996; Malik and Eickbush 1999; Yang et al. 1999; Goodwin and Poulter 2001). Elements not clearly related to either LTR or non-LTR retrotransposons encode an Uri endonuclease that is also found in group I introns (Lyozin et al. 2001; Volff et al. 2001a).

The release of transposable element copy number constraints appears to be a major characteristic of large genomes (Kidwell 2002). Non-LTR retrotransposons, short interspersed nuclear elements (SINEs), and retrovirus-like sequences to-

⁴These authors contributed equally to this work.

⁵Corresponding author.

E-MAIL volff@biozentrum.uni-wuerzburg.de; FAX (0) 931-888-4150.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.726003>. Article published online before print in June 2003.

gether make up more than 40% of the first draft of the human genome (International Human Genome Sequencing Consortium 2001). Certain other vertebrate species have much more compact genomes characterized by small intronic and intergenic sequences and a low percentage of repetitive sequences. With 380–400 Mb, the genomes of the marine Japanese pufferfish *Takifugu rubripes* (Fugu) and the freshwater green-spotted pufferfish *Tetraodon nigroviridis* are about eight times smaller than the human genome. For this reason, both are objects of (almost) completed genome-sequencing projects (Brenner et al. 1993; Crollius et al. 2000; Fischer et al. 2000; Roest Crollius et al. 2000; Aparicio et al. 2002). Although both

pufferfish species are separated by ~20–30 million years of evolution, the compaction of their genome has been conserved. To understand this phenomenon, it is of primordial importance to characterize the diversity and activity of retrotransposable elements in pufferfish genomes. Interestingly, and maybe surprisingly, numerous families of retrotransposons have been identified to date in the genome of *T. rubripes* (Aparicio et al. 2002) and/or *T. nigroviridis*, including LTR elements from the *Ty3/Gypsy* (Poulter and Butler 1998; Volff et al. 2001b; Goodwin and Poulter 2002), *Ty1/Copia* (Crollius et al. 2000), *BEL* (Frame et al. 2001), and *Dirs1* classes (Goodwin and Poulter 2001). Non-LTR retrotransposons encoding apurinic-apyrimidinic (Duvrenell and Turner 1998; Poulter et al. 1999; Volff et al. 1999, 2000, 2001d) and restriction enzyme-like endonucleases (Volff et al. 2001c), as well as *Uri* retrotransposons (Volff et al. 2001a) have been identified in pufferfish compact genomes too. We report here the presence of a novel REL endonuclease-encoding non-LTR retrotransposon, called *Zebulon*, recently active in the genome of *T. nigroviridis*. Interestingly, this non-LTR element frequently displays a tandem structure.

RESULTS

Zebulon Is a Novel Vertebrate REL Retrotransposon

In the course of a survey of the retrotransposable element content in the compact genome of the pufferfish *T. nigroviridis*, we identified reverse-transcriptase-encoding sequences with no obvious close relationship to any known vertebrate retroelement. A 3711-bp consensus sequence that we interpreted as the complete version of a novel fish retrotransposon called *Zebulon* was reconstructed (AY135221; Figs. 1–3). This element was not identified in the recent analysis of the genome of the Japanese pufferfish *T. rubripes* (Aparicio et al. 2002). Five *T. nigroviridis* genomic library plasmids containing *Zebulon* were sequenced completely (pG1167O21, pG16H23, pG556A21, pG976B21, and pG109H19; Fig. 1). *Zebulon* was also detected in genomic inserts from different bacterial artificial chromosome (BAC) genomic clones from *T. nigroviridis* (AC113583, AJ496734, AC117942, and AL808032; Fig. 1). Analysis of the 3' extremity of different *Zebulon* insertions suggested that this element ended by a short (0–21) poly(A)

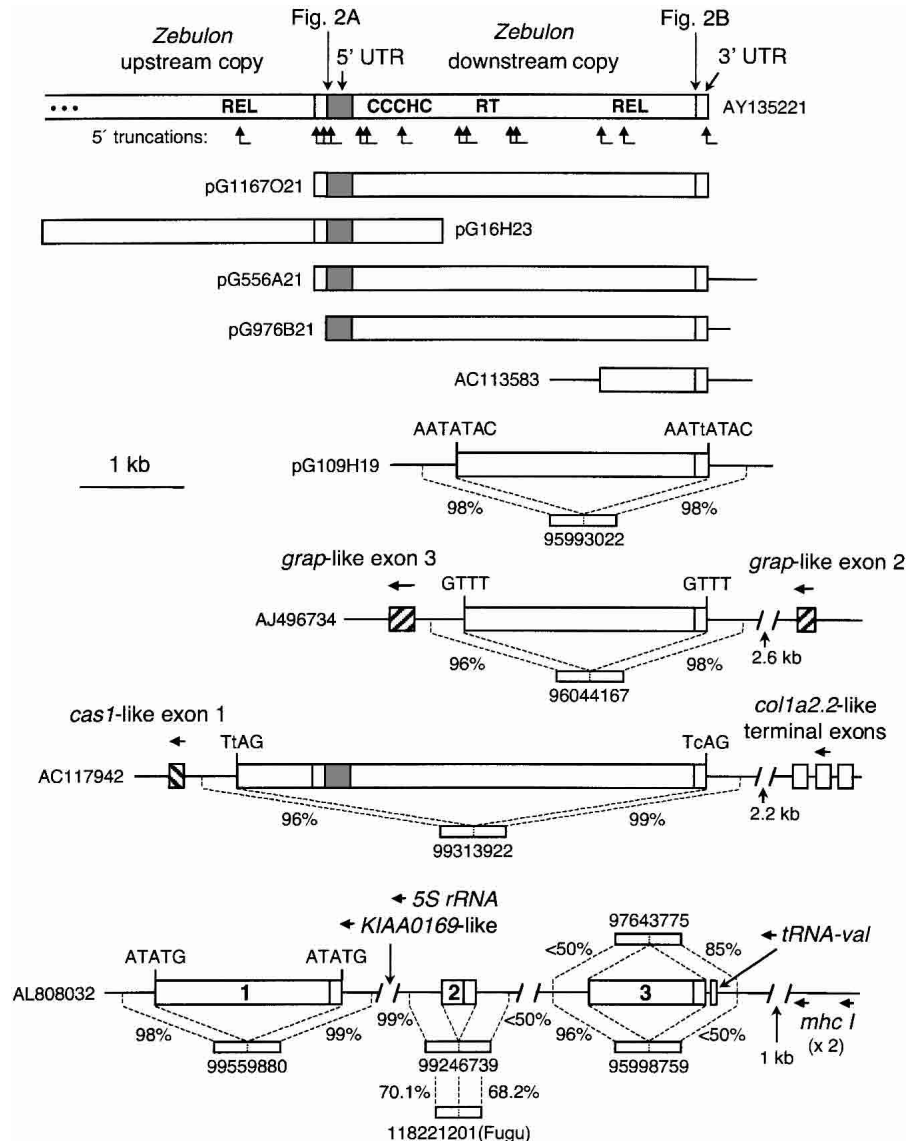


Figure 1 Genomic structure of *Zebulon* elements in *T. nigroviridis*. The position of 5' truncations is shown by broken arrows. Most of them are from single sequences and could therefore not be assigned either to the upstream or to the downstream copy; they were arbitrarily positioned in the downstream copy. Trace sequences (<http://www.ncbi.nlm.nih.gov/blast/tracemb.html>) reflecting the genomic site before insertion are shown, and their percentage of nucleotide identity with the site sequences directly flanking *Zebulon* insertions is given. Horizontal arrows show the direction of gene transcription. Putative flanking target site duplications are shown. (RT) Reverse transcriptase; (CCCHC) putative CXC₃CX₈HX₅C zinc finger-like domain.

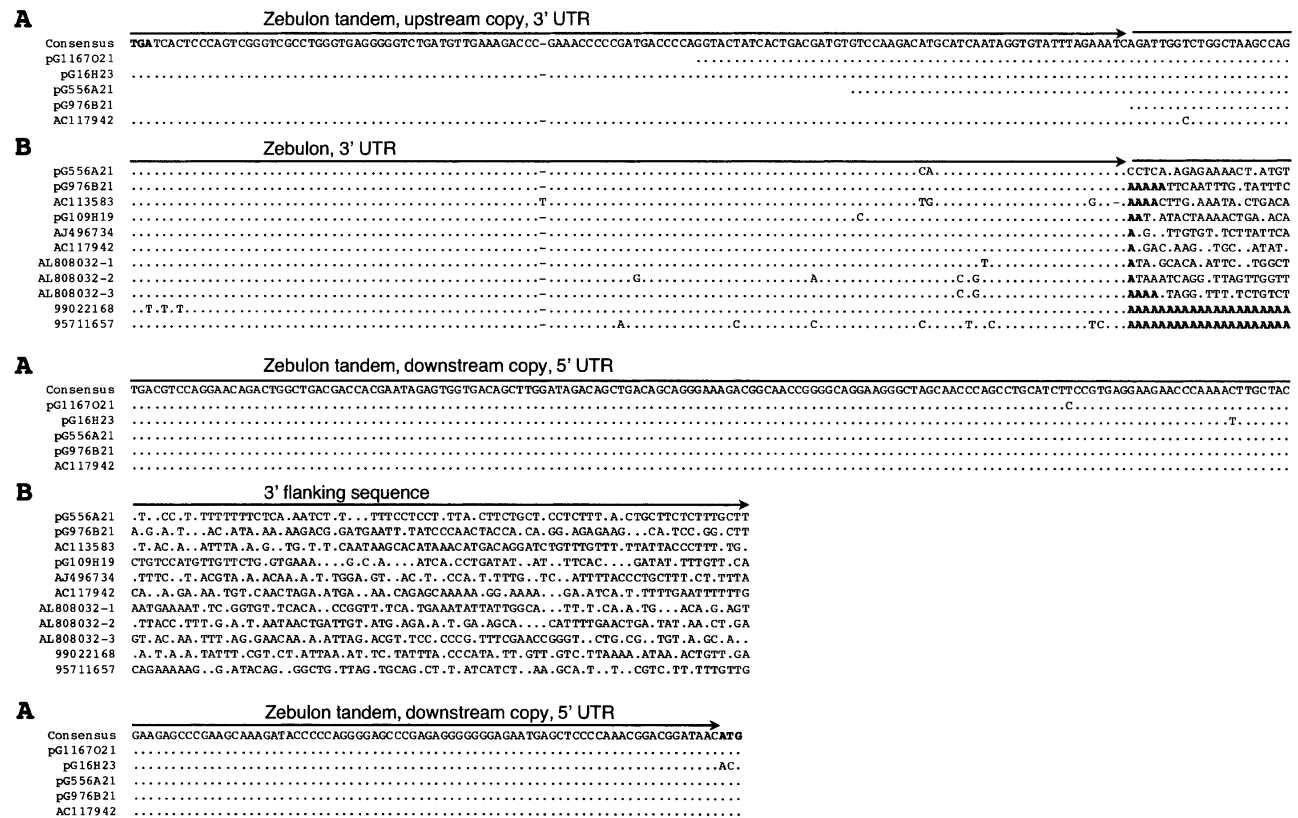


Figure 2 Sequence comparison between upstream–downstream junctions in tandem arrays (A) and between 3' ends (B) of *Zebulon* elements shown in Fig. 1. The part of the tandem array sequences shown extend from the stop codon of the upstream copy to the putative start codon of the downstream copy.

stretch (Fig. 2). We could identify in the *T. nigroviridis* NCBI trace database, sequences with >95% nucleotide identity to both 5' and 3' site sequences flanking some *Zebulon* insertions (pG109H19, AJ496734, AC117942, and the first copy of AL808032; Fig. 1), but lacking the intervening *Zebulon* element. These sequences, which are likely to reflect the genomic site before integration, allowed the identification of short target-site duplications flanking *Zebulon* insertions, AAT(t)ATAC for pG109H19, GTTT for AJ496734, TYAG for AC117942, and ATATG for the first copy of AL808032 (Fig. 1).

Using the reconstructed nucleotide consensus sequence as a query, *Zebulon* was detected in ~0.2% of whole-genome shotgun (WGS) sequences from *T. nigroviridis* (4969 of 2,049,513 sequences with Expect value $E < 10^{-10}$; 4549 sequences with $E < 10^{-20}$; Altschul et al. 1990). A precise copy number could not be estimated from genomic data because of the high variability in copy size, but Southern blot analysis was compatible with the presence of multiple copies of *Zebulon* in the genome of *T. nigroviridis* (Fig. 4).

Zebulon Is Related to Nematode Site-Specific Non-LTR Retrotransposons

The unique ORF of *Zebulon* encodes a putative 1112 amino-acid protein containing a reverse transcriptase domain and a carboxy-terminal restriction enzyme-like endonuclease (REL; Yang et al. 1999) (AY135221; Fig. 3). All amino-acid residues

characteristic of REL endonucleases were detected, including a CCHC zinc finger-like domain probably involved in nucleic acid binding (Yang et al. 1999). A second putative (C)CCHC domain is present in the amino-terminal part of the protein upstream from the reverse transcriptase domain, which might correspond to the putative Gag-like domain found in numerous other non-LTR retrotransposons (Fig. 3). Although CCHH zinc finger-like domains are more frequently present at similar positions in REL retrotransposons (see Malik and Eickbush 2000), a CCHC domain was also reported in several elements from trypanosomes (for review, see Gabriel et al. 1990).

Phylogenetic analysis using the reverse transcriptase domain (Malik et al. 1999) supported a relationship between *Zebulon* and the non-LTR retrotransposon *NeSL-1* from the nematode *Caenorhabditis elegans* (Malik and Eickbush 2000) (data not shown). This was confirmed by using together both the reverse transcriptase and the REL domains as done by Burke et al. (2002) (Fig. 5). The phylogeny obtained confirmed the relationship between *NeSL-1* and the *R4* clade of REL retrotransposons proposed by Burke et al. (2002). Nevertheless, as the invertebrate elements *R4* and *Dong* are clearly more closely related to *Rex6*, another fish element, than to *Zebulon*, *R4* and *NeSL-1* are likely to correspond to two distinct (sister) clades.

Despite their phylogenetic relationship, *Zebulon* and *NeSL-1* present several essential differences. The cysteine protease domain identified in *NeSL-1* (Malik and Eickbush 2000) is apparently absent from *Zebulon*. In addition, *NeSL-1* specifi-

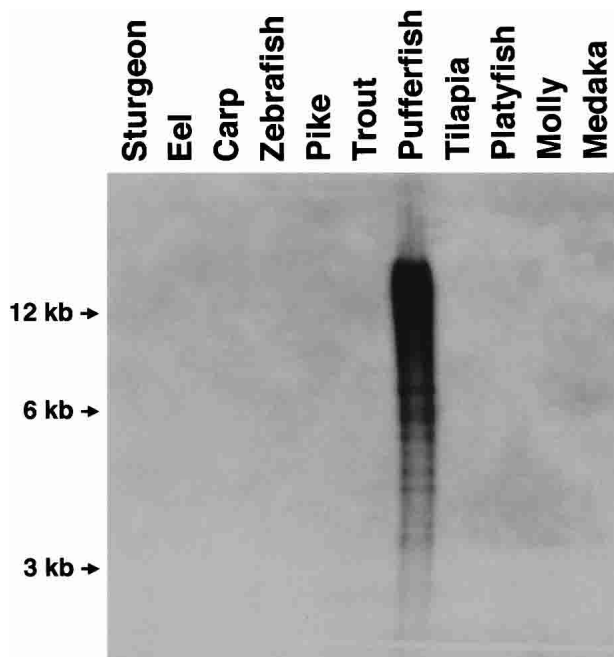


Figure 4 Southern blot analysis of the distribution of *Zebulon* in fish. Genomic DNA was cut with *Hind*III (does not cut in *Zebulon*). The probe used is pG16H23. Origin of fishes is given in Crollius et al. (2000) and Volff et al. (2000).

and the first copy of AL808032 (AATtATAC, GTTT, TYAG, and ATATG, respectively; Fig. 1).

Zebulon Copies With a Tandem Structure

During the reconstruction of a *T. nigroviridis* consensus element by assembling shotgun genomic sequences, it became evident that the 3711-bp *Zebulon* unit was flanked frequently on its 5' side by a sequence corresponding to the 3' end of the element. This tandem structure was, for example, observed in genomic inserts of BAC clone AC117942 and plasmids pG1167021, pG16H23, and pG556A21 (Fig. 1). Strikingly, the junction between the 3' end of the upstream copy and the 5' end of the downstream copy was exactly at the same position in all database tandem sequences analyzed (Fig. 2; data not shown). This was probably not resulting from overrepresentation of a particular tandem element in databases, as identical junctions were found in *Zebulon* tandem elements corresponding to different insertions (for example, BAC clone AC117942, plasmids pG556A21, and pG16H23, Figs. 1–2; shotgun sequences AL251620 and AL191204 with strongly 5' truncated upstream copy; data not shown). After analysis of the 5' extremity of copies without 5' duplicated region, no evidence for an alternative structure at the 5' end of *Zebulon* could be found; all of these copies corresponded to truncated versions of the 3711-bp unit, the position of the truncation varying between different copies (Fig. 1). The upstream copy was also 5' truncated in tandem arrays, generating a sort of LTR-like structure. The position of the 5' truncation in the upstream copy was different in different tandem arrays of *Zebulon* (e.g., pG16H23, AC117942, Fig. 1; AL251620 and AL191204; data not shown). No 5' truncation in the downstream copy was detected when an upstream copy was present.

Zebulon Extends Intronic and Intergenic Regions in the Compact Genome of *T. nigroviridis*

Zebulon can integrate into intronic and intergenic sequences in the genome of *T. nigroviridis*. One copy of *Zebulon* (AJ496734, Fig. 1) is integrated only ~400 nucleotides away from the third exon of a gene encoding a protein homologous to Grap, an adaptor protein coupling tyrosine kinases to the Ras pathways in human (Q13588). In AC117942, *Zebulon* is present between two genes, ~400 nucleotides upstream from the first exon of a gene encoding a product homologous to the human Cas1p O-acetyltransferase (AAL33538) and ~2.2 kb away from the terminal exon of a gene related to the type I collagen α 2 chain gene *col1a2.2* from chum salmon (BAB79230) (Fig. 1). In AL808032, copy 3 of *Zebulon* is integrated only 54 bp from a putative tRNA-Val gene, and at the proximity of MHC class I gene duplicates. All of these copies were integrated in an opposite transcriptional orientation compared with the neighboring exons. These observations indicate that *Zebulon* is a retrotransposon occasionally contributing to the extension of intronic and intergenic regions in the compact pufferfish genome.

Preferential Localization of *Zebulon* in Some Heterochromatic Regions

According to Fischer et al (2000), DAPI brightly stains, after denaturation treatment, heterochromatic regions in the chromosomes of *T. nigroviridis*, that is, short arms of subtelocentric chromosomes and pericentromeric regions. These regions correspond mostly to satellite repeats and other kinds of repetitive sequences (Crollius et al. 2000; Dasilva et al. 2002). *Zebulon* hybridizes mainly in those areas, showing major regions of accumulation in at least five chromosome pairs (Fig. 6a1,a2). Other weaker signals were usually detected at the end of the arms of subtelocentric chromosomes and in pericentromeric regions. Moreover, when the pG16H23 probe was cohybridized on *T. nigroviridis* chromosomes with a plasmid containing the non-LTR retrotransposon *Rex3* (Volff et al. 1999; C. Fischer, L. Bouneau, and C. Ozouf-Costaz, unpubl.), signals were, in most cases, overlapping, particularly for the major signals of *Zebulon* (Fig. 6b1–b4).

Recent Activity of *Zebulon* in *T. nigroviridis*

Even if some more divergent copies with only 80% nucleotide identity are present in the *T. nigroviridis* trace database, comparison of the 11 different *Zebulon* elements shown in Figure 1 revealed a general high level of nucleotide identity (between 94.1% and 99.7%, average 97.6%). Particularly, the degree of nucleotide identity between the clearly different insertions in genomic sequences AJ496734 and AC117942 was as high as 99.7%. The complete ORF of the downstream copy of AC117942 was intact and its putative translation product displayed only two conservative differences over 1112 aminoacids compared with *Zebulon* consensus protein sequence. *Zebulon* in genomic sequence AJ496734 was truncated at its 5' end (Fig. 1). Nevertheless, the remaining part of the ORF was still intact and its conceptual product showed only one conservative and one nonconservative replacement over 738 aminoacids. Hence, the very high degree of sequence identity between different *Zebulon* insertions added to the presence of noncorrupted, possibly functional copies, indicate that this element retrotransposed relatively recently and might be still active in the compact genome of the pufferfish *T. nigroviridis*.

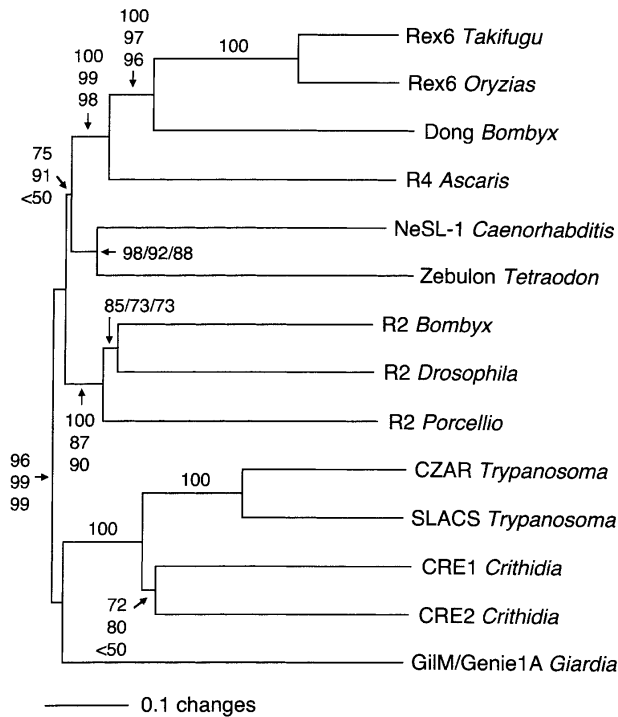


Figure 5 Phylogenetic relationship between *Zebulon* and the nematode site-specific retrotransposon *NeSL-1*. Phylogeny was performed using together the reverse transcriptase and REL domains (Burke et al. 2002). The tree (neighbor-joining) is unrooted. Branches with <50% support have been collapsed. Bootstrap values using neighbor-joining (1000 replicates, first values) and maximum parsimony analyses (100 replicates, third values), as well as reliability values for maximum likelihood analysis (quartet puzzling, 10,000 puzzling steps, second values) are given. Accession numbers are the same as in Malik et al. (1999) and Burke et al. (2002).

Involvement of *Zebulon* in Genomic Rearrangements?

Using the genomic sequences flanking copies 2 and 3 from BAC clone AL808032 as queries against *T. nigroviridis* sequence databases, corresponding sequences without *Zebulon* insertion were identified (e.g., NCBI trace sequences 99246739 and 95998759 for copies 2 and 3, respectively; Fig. 1). Strikingly, these unoccupied sites showed >95% nucleotide identity to the 5' sequence flanking directly the insertions (identity ending exactly at the position of the *Zebulon* insertion), but no significant identity to the 3' sequence flanking the insertion, or to any other sequence present in AL808032. A sequence corresponding to *T. nigroviridis* sequence 99246739 and also presenting significant nucleotide identity only to the 5' sequence flanking insertion 2 in AL808032 was identified in *T. rubripes* (trace sequence 118221201; Fig. 1). In addition, *T. nigroviridis* sequence 97643775 presented 85% nucleotide identity to the 3' sequence directly flanking copy 3 in AL808032, but showed no significant identity to the 5' flanking sequence (Fig. 1). To exclude that the structure observed for copies 2 and 3 in AL808032 was the result of cloning or assembling artifacts having eliminated the intervening sequence between two nonallelic copies of *Zebulon*, both copies 2 and 3 were amplified by PCR from *T. nigroviridis* genomic DNA using primers matching their 5' and 3' flanking sequences. The size of the obtained PCR fragments and their sequence confirmed that

the structure of copies 2 and 3 in AL808032 is also found in *T. nigroviridis* genome (data not shown). The structure observed for copy 2 and 3 might have been created by ectopic homologous recombination between two nonallelic copies of *Zebulon*. For example, recombination between two copies, one inserted in a 95998759-like site and one integrated in a 97643775-like site (Fig. 1), may have generated a hybrid element with flanking sequences originating each from different genomic sites. Hence, *Zebulon* might be involved in the formation of rearrangements in the genome of *T. nigroviridis*. Alternatively, the structure observed in copy 2 and 3 might be the result of deletions having affected the 5' genomic sequence flanking *Zebulon* insertions. Such deletions are associated with the retrotransposition of *LI* in transformed human cells (Gilbert et al. 2002; Symer et al. 2002), but retrotransposition-independent deletions having included both the 5' part of the element and its 5' flanking sequence might generate the same type of structure.

Is *Zebulon* Fish Specific?

The distribution of *Zebulon* was studied by Southern blot analysis and homology searching of sequence databases. Using pG16H23 from *T. nigroviridis* as a probe in Southern blot hybridization, no significant signal could be detected even under low-stringency conditions in 10 other fish species (Fig. 4; the Japanese pufferfish *T. rubripes* was not included).

In contrast, *Zebulon* sequences presenting an average 67.5% nucleotide identity (from 63.1% to 74.5%) to *T. nigroviridis* elements were identified in *T. rubripes* by database analysis. A short *Zebulon* element is located ~850 bp upstream of the second exon of a gene encoding the rho-type GTPase-activating protein rhoGAPX-1 (AF012274). *Zebulon* elements are also present in at least 10 of 12,403 WGS scaffolds from the genome draft of *T. rubripes* (<http://fugu.hgmp.mrc.ac.uk/>; Aparicio et al. 2002), all of them with ORFs corrupted by 5' truncations, frameshifts, and/or stop codons. *Zebulon* copies identified in *T. rubripes* presented to each other a level of nucleotide identity ranging from 87.9% to 96.5% (94.1% on average; more divergent sequences with only ~75.0% identity are also present in the trace database). An almost complete 3.6-kb *Zebulon* sequence could be reconstructed from different genomic scaffolds and was used as a query against the *T. rubripes* WGS trace database (1,877,457 sequences with an average size of 920 nucleotides). A total of 246 sequences (0.013% vs. 0.24% for *T. nigroviridis*) showed significant nt identity to *Zebulon* ($E < 10^{-10}$, a threshold allowing the detection of copies with <80% nucleotide identity). This suggested that the *T. nigroviridis* genome contains 15–20 times more copies of *Zebulon* than the genome of *T. rubripes*. This conclusion was not modified by choosing a more stringent threshold ($E < 10^{-20}$; 0.0093% for *T. rubripes* vs. 0.22% for *T. nigroviridis*).

Zebulon was also detected in the genome of the zebrafish *Danio rerio*, which diverged from pufferfishes ~150 million years ago. Elements truncated at their 5' end and presenting various other kinds of corrupting mutations were identified in at least 10 different database genomic sequences (e.g., AL627164, AL929152, and AL928790). These copies shared from 61.1% to 67.6% nucleotide identity (average 63.9%) with *Zebulon* elements from *T. nigroviridis*, probably explaining the absence of signal in Southern blot hybridization (Fig. 4). *Zebulon* copies of zebrafish showed an average 82.2% nucleotide identity to each other (from 72.2% to 89.8%). *Ze-*

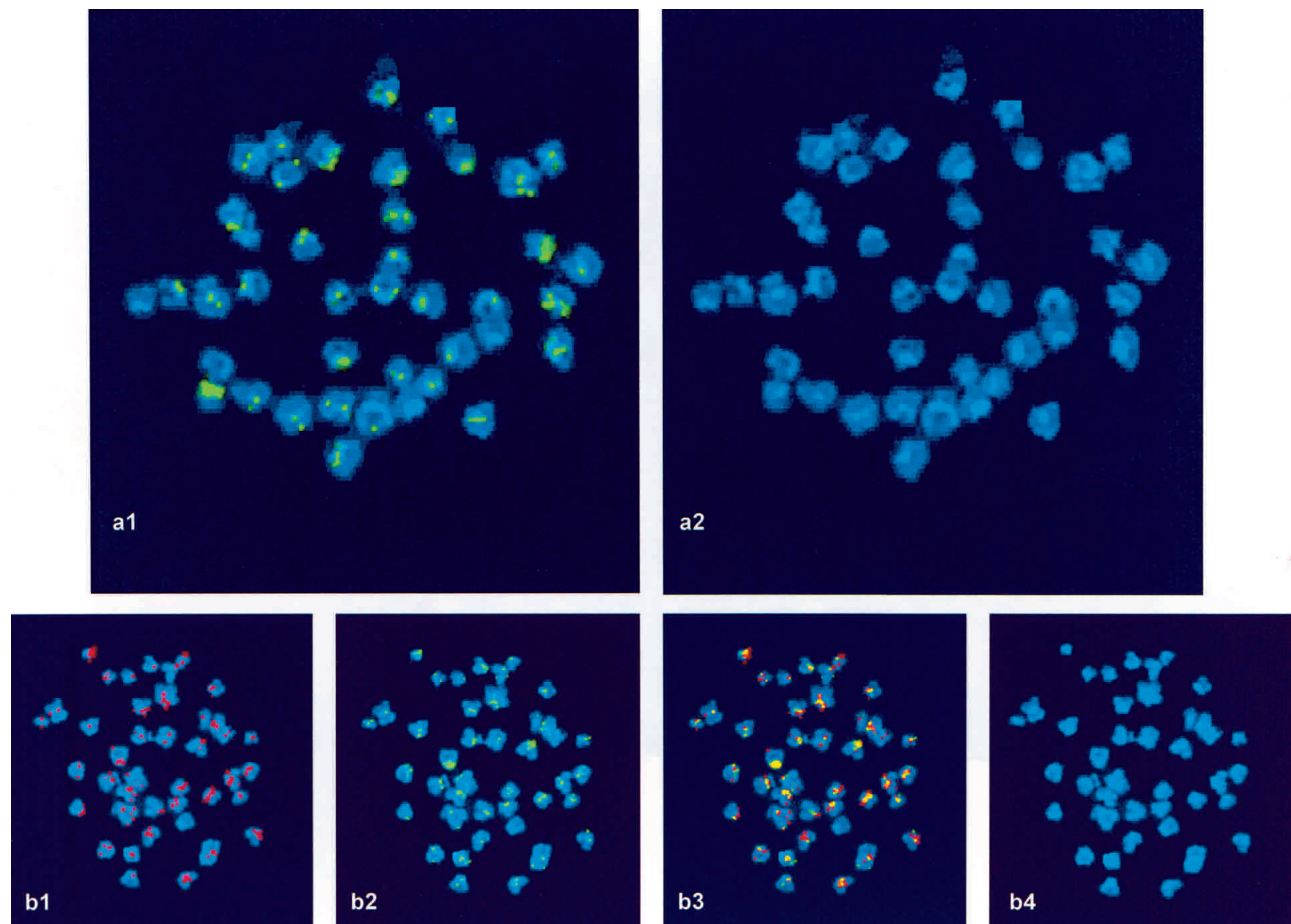


Figure 6 Chromosomal localization of *Zebulon* in the genome of *T. nigroviridis* by FISH. Weak, scattered spots have been removed by electronic thresholding in order to retain only major regions of accumulation. The genomic areas in which *Zebulon* preferentially localizes (*a1*) mostly correspond to heterochromatic, DAPI-positive regions as shown in this over-denaturated metaphase (*a2*). Double FISH between *Zebulon* (DIG-labeled pG16H23, *b1*) and *Rex3* (biotin-labeled, *b2*), a non-LTR retrotransposon abundant in the genome of *T. nigroviridis* (C. Fischer, L. Bouneau, and C. Ozouf-Costaz, unpubl.) shows superimposed signals (*b3*) corresponding to common regions of accumulation in DAPI-positive regions (*b4*).

bulon was also detected in about 90 of 158,689 contigs from the zebrafish WGS assembly 06 (http://www.ensembl.org/Danio_reio/blastview). An almost complete 3.5-kb copy of *Zebulon* was identified in WGS contig z06s014441 and used as a query against the NCBI zebrafish WGS trace database (11,453,550 sequences with an average length of 700 nucleotides). The results indicated that 0.005%–0.007% of the zebrafish WGS sequences contained *Zebulon* (781 sequences with $E < 10^{-10}$; 587 sequences with $E < 10^{-20}$), a value much lower than that obtained for *T. nigroviridis*. We could not establish without ambiguity whether the tandem structure observed in *T. nigroviridis* was also present in *T. rubripes* and *D. rerio*.

Zebulon was not detected outside of the fish lineage in the huge amount of sequences present in databases. Particularly, *Zebulon* was not present within the public draft of the human genome. Hence, if we assume a mode of vertical transmission, *Zebulon* might have been lost from some vertebrate lineages.

DISCUSSION

Retrotransposons encoding a restriction enzyme-like endonuclease have been identified originally in insects and other

invertebrates (Yang et al. 1999). After the *Rex6* element from fish (Volf et al. 2001c), *Zebulon* is the second retrotransposon of this type to be identified in vertebrates. *Zebulon* was not detected in the human genome by sequence database analysis. Other instances of retrotransposable elements active in fish, but apparently either absent from mammals or present as inactive molecular fossils, have been reported already (Volf et al. 2001e). These observations suggest a greater diversity of active retrotransposable elements in fish compared with human and probably other mammals. Thinking in terms of competition, the extinction of some families of retrotransposons in the mammalian lineage might have allowed, or alternatively might have been caused by the formidable expansion of both *L1* non-LTR retrotransposons and vertebrate endogenous retroviruses (International Human Genome Sequencing Consortium 2001).

Most families of retrotransposable elements described in teleost fish are present in the genome of the pufferfishes *T. nigroviridis* and *T. rubripes*. Despite the presence of multiple families of retroelements, a strong compaction of the genome (eight times smaller than the human genome) has been maintained for unknown reasons in both pufferfishes since their

divergence 20–30 millions years ago. Even if exceptional genes exist (Aparicio et al. 2002), small intergenic and intronic regions and a low percentage of repetitive sequences are characteristic of both pufferfish compact genomes. Using *Zebulon* as an example, we could show that pufferfish genomes contain retrotransposons having been very recently (and probably still) active. *Zebulon* was apparently more successful in the freshwater pufferfish *T. nigroviridis* than in the Japanese pufferfish *T. rubripes* or even than in the zebrafish *D. rerio* having an approximately three times larger genome. *Zebulon* is a factor contributing to the extension of intergenic and intronic sequences. If there is a selection maintaining genome compaction in pufferfishes, it should act strongly against *Zebulon* retrotransposition in gene-rich regions.

If we assume a vertical modus of inheritance, putative mechanisms might explain the maintenance of *Zebulon* activity in the compact genome of *T. nigroviridis*. As revealed by FISH experiments, *Zebulon* preferentially concentrates within some heterochromatic regions, generally within short chromosome arms or pericentromeric regions. Very recently, this phenomenon has been also reported for other tandem and dispersed repeat elements in the same fish species (Dasilva et al. 2002). Preferential localization of retrotransposable elements in heterochromatin has been reported frequently in other genomes (Dimitri and Junakovic 1999; Bartolomé et al. 2002), but its significance remains controversial. Generally, heterochromatic retrotransposable elements are defective (for example, see Vaury et al. 1989). On the other hand, such gene-poor heterochromatic regions might serve as reservoirs that are tolerated by the genome, and can maintain active copies of *Zebulon* (an advantageous role of retrotransposons in heterochromatin has even been proposed, see Dimitri and Junakovic 1999). Interestingly, *Zebulon* colocalized very frequently in FISH experiments with *Rex3*, another abundant non-LTR retrotransposon, suggesting the presence of general heterochromatic reservoirs for retrotransposable elements. Nevertheless, the reservoir theory implies that these retrotransposons can use promoters that are active in the generally gene-silencing heterochromatin, as reported for the *HeT-A* retrotransposon in *Drosophila* (Danilevskaya et al. 1997; Pardue and Debaryshe 2000).

Which mechanisms might be responsible for the uneven distribution of *Zebulon* in heterochromatic and euchromatic regions? *Zebulon* might possess some kind of (non-strict) specialization for heterochromatic regions, as observed for telomeric retrotransposons in some organisms (Danilevskaya et al. 1997; Takahashi et al. 1997; Arkhipova and Morrison 2001). Nevertheless, we could not observe any target sequence specificity for *Zebulon*, indicating that if a preference is present, it is probably not driven by the primary sequence of the target site. On the other hand, drastic preferential elimination of retrotransposons in euchromatin might occur, maintaining the compaction of gene-rich regions in the pufferfish. This might be particularly achieved by natural selection against individual insertions, against genomic rearrangements mediated by ectopic homologous recombination between non-allelic copies, and/or against retrotransposition itself if it occurs at the cost of the host (Bartolomé et al. 2002; Eickbush and Furano 2002, and references therein).

A possible advantage of *Zebulon* in the compact genome of the pufferfish is suggested by the observation that this non-LTR retrotransposon frequently displays a partial tandem structure with variable 5' truncations of the upstream copy. Homologous recombination between the 3' ends of both up-

stream and downstream copies might lead to the elimination of active elements and reduce the size of *Zebulon* insertions in a mechanism reminiscent of that generating solo LTRs from LTR retrotransposons and retroviruses. This might minimize the effect of *Zebulon* on the extension of intergenic and intronic sequences and maintain the number of active copies to a number tolerable by pufferfish euchromatin.

The mechanism of formation of *Zebulon* tandem arrays remains unknown. Particularly, we do not know whether they are the result of successive events of retrotransposition, or whether the tandem array itself can be retrotransposed. In non-LTR retrotransposons arranged in head-to-tail arrays, the tandem structure is generated by successive events of retrotransposition, and the different units are either separated by poly(A) stretches of different lengths, or by a variable number of copies of the repeated sequence serving as targets (e.g., telomeric repeats). Alternatively, some retrotransposons can create tandem arrays by jumping into themselves, generally at different positions inside of the target element (Higashiyama et al. 1997). In contrast, the identity of the junctions between upstream and downstream copies in different elements suggests that *Zebulon* tandem arrays might function as a retrotransposition unit. The structure of the tandem array in sequence AC117942 with short flanking sequence duplications is compatible with a single integration event.

The promoter(s) driving the transcription of *Zebulon* remains to be identified. A promoter located within the upstream copy might be able to promote the transcription of partial tandem arrays, in a manner reminiscent of that reported for the telomeric retrotransposon *HeT-A* in *Drosophila* (Danilevskaya et al. 1997). Because of the almost impossibility of performing functional analysis in *T. nigroviridis* due to the absence of laboratory strains, cell lines, and transgenesis technology, we are not able at the moment to provide any information about the promoter region(s) driving the transcription of *Zebulon*.

The use of a 3' promoter, coupled to variable degrees of 5' truncation by incomplete reverse transcription, might generate the tandem structures of variable lengths observed in some copies of *Zebulon*. Alternatively, nonreproducible truncations of the upstream copy in tandem arrays might be due to the use of alternative transcription starts from a same promoter, or to the presence of different promoters in the upstream copy. Finally, tandem arrays might have been generated by the massive transcription of a single tandem fortuitously integrated at the neighborhood of an exogenous strong promoter. Because functional analyses are almost impossible in pufferfish, functional *Zebulon* elements have now to be identified and characterized in alternative fish model systems to elucidate the mechanism of retrotransposition and the genomic impact of this interesting retroelement.

METHODS

Plasmids and DNA Manipulation

T. nigroviridis genomic libraries and sequencing procedures have been described elsewhere (Crollius et al. 2000; Fischer et al. 2002). *Zebulon*-containing plasmids pG109H19, pG1167O21, pG16H23, pG556A21, and pG976B21 (Fig. 1; AJ496221–AJ496225) were identified by end-sequencing in a plasmid genomic library with average insert size of 4 kb, and sequenced subsequently to completion. Genomic DNA isolation and Southern blot analysis were performed according to standard protocols (Volf et al. 1999, and references therein).

Southern blot hybridization was performed in 35% formamide at 42°C, the filter was washed with 2× SSC/1% SDS at 50°C.

FISH Analysis

Zebulon-containing plasmids were digoxigenin (DIG) or biotin labeled for FISH analysis by nick translation (Roche). Labeled probes were purified using the Qiaquick PCR purification kit (QIAGEN), ethanol precipitated, and mixed up again in QBIO-gene high-stringency Hybrisol VI at 20 ng/μL each. Probes were hybridized and detected on *T. nigroviridis* freshly thawed chromosome preparations without any pretreatment, according to the protocol of QBIO-gene for repetitive probes (Crollius et al. 2000). Preparations were counterstained simultaneously, mounted with 1.2 ng/μL DAPI in Antifade (Vector Laboratories), and analyzed using Genus FISH-imaging equipment and software for animal chromosomes (Applied Imaging). For unequivocal chromosome localization of *Zebulon*, double FISH was performed with two different plasmids (biotin-labeled pG109H19 and DIG-labeled pG16H23), and their correct overlapping was checked. Only results obtained with pG16H23 are shown.

Sequence Analysis

Multiple sequence alignments were generated using PileUp of the GCG Wisconsin package (Version 10.0, Genetics Computer Group) and ClustalX (Thompson et al. 1997). Phylogenies were determined with PAUP* (D.L. Swofford, Smithsonian Institution) by bootstrap analysis using maximum parsimony (100 replicates) and neighbor-joining (1000 replicates; Saitou and Nei 1987). Maximum likelihood analysis was performed by quartet puzzling using TREE-PUZZLE 5.0 (Schmidt et al. 2002). Gene structure was analyzed using programs available at the NIX server (<http://menu.hgmp.mrc.ac.uk/menu-bin/Nix>). Pufferfish and zebrafish genome survey and trace sequences were obtained using the NCBI BLAST server (<http://www.ncbi.nlm.nih.gov/BLAST>). *Zebulon* nondegenerated consensus sequence (AY135221) was reconstructed by assembling trace sequences showing overlaps with >95% nucleotide identity.

ACKNOWLEDGMENTS

We thank the Genoscope production teams (cloning, sequencing, and finishing), Corinne Cruaud (Genoscope) for sequencing assistance and Muriel Ronsin (Genoscope) for technical work. This work was supported by the French Museum National d'Histoire Naturelle, the Centre National de la Recherche Scientifique (CNRS) and the Ministère de la Recherche et de la Technologie (to MNHN and Genoscope), and by the BioFuture program of the German Ministry for Research and Education (BMBF) (to J.N.V.).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.

Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A.F., et al. 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* **297**: 1301–1310.

Arhipova, I.R. and Morrison, H.G. 2001. Three retrotransposon families in the genome of *Giardia lamblia*: Two telomeric, one dead. *Proc. Natl. Acad. Sci.* **98**: 14497–14502.

Bartolomé, C., Maside, X., and Charlesworth, B. 2002. On the abundance and distribution of transposable elements in the genome of *Drosophila melanogaster*. *Mol. Biol. Evol.* **19**: 926–937.

Boeke, J.D. and Chapman, K.B. 1991. Retrotransposition mechanisms. *Curr. Opin. Cell Biol.* **3**: 502–507.

Brenner, S., Elgar, G., Sandford, R., Macrae, A., Venkatesh, B., and Aparicio, S. 1993. Characterization of the pufferfish (*Fugu*) genome as a compact model vertebrate genome. *Nature* **366**: 265–268.

Burke, W.D., Malik, H.S., Rich, S.M., and Eickbush, T.H. 2002. Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, *Giardia lamblia*. *Mol. Biol. Evol.* **19**: 619–630.

Crollius, H.R., Jaillon, O., Dasilva, C., Ozouf-Costaz, C., Fizames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., et al. 2000. Characterization and repeat analysis of the compact genome of the freshwater pufferfish *Tetraodon nigroviridis*. *Genome Res.* **10**: 939–949.

Danilevskaia, O.N., Arkhipova, I.R., Traverse, K.L., and Pardue, M.L. 1997. Promoting in tandem: The promoter for telomere transposon *HeT-A* and implications for the evolution of retroviral LTRs. *Cell* **88**: 647–655.

Dasilva, C., Hadji, H., Ozouf-Costaz, C., Nicaud, S., Jaillon, O., Weissenbach, J., and Crollius, H.R. 2002. Remarkable compartmentalization of transposable elements and pseudogenes in the heterochromatin of the *Tetraodon nigroviridis* genome. *Proc. Natl. Acad. Sci.* **99**: 13636–13641.

Dimitri, P. and Junakovic, N. 1999. Revising the selfish DNA hypothesis. New evidence on accumulation of transposable elements in heterochromatin. *Trends Genet.* **15**: 123–124.

Duvernell, D.D. and Turner, B.J. 1998. *Swimmer 1*, a new low-copy-number LINE family in teleost genomes with sequence similarity to mammalian *L1*. *Mol. Biol. Evol.* **15**: 1791–1793.

Eickbush, T.H. and Furano, A.V. 2002. Fruit flies and humans respond differently to retrotransposons. *Curr. Opin. Genet. Dev.* **12**: 669–674.

Feng, Q., Moran, J.V., Kazazian Jr., H.H., and Boeke, J.D. 1996. Human *L1* retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**: 905–916.

Fischer, C., Ozouf-Costaz, C., Roest Crollius, H., Dasilva, C., Jaillon, O., Bouneau, L., Bonillo, C., Weissenbach, J., and Bernot, A. 2000. Karyotype and chromosome location of characteristic tandem repeats in the pufferfish *Tetraodon nigroviridis*. *Cytogenet. Cell Genet.* **88**: 50–55.

Fischer, C., Bouneau, L., Ozouf-Costaz, C., Crnogorac-Jurcevic, T., Weissenbach, J., and Bernot, A. 2002. Conservation of the T-cell receptor α/δ linkage in the teleost fish *Tetraodon nigroviridis*. *Genomics* **79**: 241–248.

Frame, I.G., Cutfield, J.F., and Poulter, R.T. 2001. New *BEL*-like LTR-retrotransposons in *Fugu rubripes*, *Caenorhabditis elegans*, and *Drosophila melanogaster*. *Gene* **263**: 219–230.

Gabriel, A., Yen, T.J., Schwartz, D.C., Smith, C.L., Boeke, J.D., Sollner-Webb, B., and Cleveland, D.W. 1990. A rapidly rearranging retrotransposon within the minixion gene locus of *Crithidia fasciculata*. *Mol. Cell. Biol.* **10**: 615–624.

Gilbert, N., Lutz-Prigge, S., and Moran, J.V. 2002. Genomic deletions created upon *LINE-1* retrotransposition. *Cell* **110**: 315–325.

Goodwin, T.J. and Poulter, R.T. 2001. The *DIRS1* group of retrotransposons. *Mol. Biol. Evol.* **18**: 2067–2082.

———. 2002. A group of deuterostome *Ty3/gypsy*-like retrotransposons with *Ty1/copia*-like pol-domain orders. *Mol. Genet. Genomics* **267**: 481–491.

Higashiyama, T., Noutoshi, Y., Fujie, M., and Yamada, T. 1997. *Zepp*, a LINE-like retrotransposon accumulated in the Chlorella telomeric region. *EMBO J.* **16**: 3715–3723.

International Human Genome Sequencing Consortium. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.

Kidwell, M.G. 2002. Transposable elements and the evolution of genome size in eukaryotes. *Genetica* **115**: 49–63.

Lyozin, G.T., Makarova, K.S., Velikodvorskaja, W., Zelentsova, H.S., Khechumian, R.R., Kidwell, M.G., Koonin, E.V., and Evgen'ev, M.B. 2001. The structure and evolution of *Penelope* in the virilis species group of *Drosophila*: An ancient lineage of retroelements. *J. Mol. Evol.* **52**: 445–456.

Malik, H.S. and Eickbush, T.H. 1999. Modular evolution of the integrase domain in the *Ty3/Gypsy* class of LTR retrotransposons. *J. Virol.* **73**: 5186–5190.

———. 2000. *NeSL-1*, an ancient lineage of site-specific non-LTR retrotransposons from *Caenorhabditis elegans*. *Genetics* **154**: 193–203.

Malik, H.S., Burke, W.D., and Eickbush, T.H. 1999. The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16**: 793–805.

- Pardue, M.L. and Debaryshe, P.G. 2000. *Drosophila* telomere transposons: Genetically active elements in heterochromatin. *Genetica* **109**: 45–52.
- Poulter, R. and Butler, M. 1998. A retrotransposon family from the pufferfish (fugu) *Fugu rubripes*. *Gene* **215**: 241–249.
- Poulter, R., Butler, M., and Ormandy, J. 1999. A LINE element from the pufferfish (fugu) *Fugu rubripes* which shows similarity to the *CR1* family of non-LTR retrotransposons. *Gene* **227**: 169–179.
- Roest Crolius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fizames, C., Wincker, P., Brottier, P., Quetier, F., et al. 2000. Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nat. Genet.* **25**: 235–238.
- Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- Schmidt, H.A., Strimmer, K., Vingron, M., and von Haeseler, A. 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**: 502–504.
- Symer, D.E., Connelly, C., Szak, S.T., Caputo, E.M., Cost, G.J., Parmigiani, G., and Boeke, J.D. 2002. Human *L1* retrotransposition is associated with genetic instability *in vivo*. *Cell* **110**: 327–338.
- Takahashi, H., Okazaki, S., and Fujiwara, H. 1997. A new family of site-specific retrotransposons, *SART1*, is inserted into telomeric repeats of the silkworm, *Bombyx mori*. *Nucleic Acids Res.* **25**: 1578–1584.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. 1997. The ClustalX windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**: 4876–4882.
- Vaury, C., Bucheton, A., and Pelisson, A. 1989. The β heterochromatic sequences flanking the *I* elements are themselves defective transposable elements. *Chromosoma* **98**: 215–224.
- Volff, J.-N., Körting, C., Sweeney, K., and Scharl, M. 1999. The non-LTR retrotransposon *Rex3* from the fish *Xiphophorus* is widespread among teleosts. *Mol. Biol. Evol.* **16**: 1427–1438.
- Volff, J.-N., Körting, C., and Scharl, M. 2000. Multiple lineages of the non-LTR retrotransposon *Rex1* with varying success in invading fish genomes. *Mol. Biol. Evol.* **17**: 1673–1684.
- Volff, J.-N., Hornung, U., and Scharl, M. 2001a. Fish retrotransposons related to the *Penelope* element of *Drosophila virilis* define a new group of retrotransposable elements. *Mol. Genet. Genomics* **265**: 711–720.
- Volff, J.-N., Körting, C., Altschmied, J., Duschl, J., Sweeney, K., Wichert, K., Froschauer, A., and Scharl, M. 2001b. *Jule* from the fish *Xiphophorus* is the first complete vertebrate *Ty3/Gypsy* retrotransposon from the *Mag* family. *Mol. Biol. Evol.* **18**: 101–111.
- Volff, J.-N., Körting, C., Froschauer, A., Sweeney, K., and Scharl, M. 2001c. Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J. Mol. Evol.* **52**: 351–360.
- Volff, J.-N., Körting, C., Meyer, A., and Scharl, M. 2001d. Evolution and discontinuous distribution of *Rex3* retrotransposons in fish. *Mol. Biol. Evol.* **18**: 427–431.
- Volff, J.-N., Körting, C., and Scharl, M. 2001e. *Ty3/Gypsy* retrotransposon fossils in mammalian genomes: Did they evolve into new cellular functions? *Mol. Biol. Evol.* **18**: 266–270.
- Xiong, Y. and Eickbush, T.H. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9**: 3353–3362.
- Yang, J., Malik, H.S., and Eickbush, T.H. 1999. Identification of the endonuclease domain encoded by *R2* and other site-specific, non-long terminal repeat retrotransposable elements. *Proc. Natl. Acad. Sci.* **96**: 7847–7852.

WEB SITE REFERENCES

- <http://fugu.hgmp.mrc.ac.uk/>; The Fugu Genomics site at the UK HGMP Resource Centre.
- <http://menu.hgmp.mrc.ac.uk/menu-bin/Nix>; The Bio-informatics Application Server at the UK HGMP Resource Centre.
- http://www.ensembl.org/Danio_rerio/blastview; The Zebrafish BLAST server at the Wellcome Trust Sanger Institute.
- <http://www.ncbi.nlm.nih.gov/BLAST>; The BLAST server at the National Center for Biotechnology Information.
- <http://www.ncbi.nlm.nih.gov/blast/tracemb.html>; The Trace server at the National Center for Biotechnology information.

Received August 21, 2002; accepted in revised form April 18, 2003.