J.-N. Volff · U. Hornung · M. Schartl

Fish retroposons related to the Penelope element of Drosophila virilis define a new group of retrotransposable elements

Received: 13 November 2000 / Accepted: 2 February 2001 / Published online: 7 April 2001 © Springer-Verlag 2001

Abstract Poseidon and Neptune are two ancient lineages of retroposons related to the Penelope element from Drosophila virilis. They have been identified in various teleost fish species, including the medaka fish (Oryzias latipes), and the pufferfishes Fugu rubripes and Tetraodon nigroviridis, whose genomes are currently being sequenced. Some of these elements are highly reiterated in fish genomes. Penelope-related elements were also identified in blood fluke, shrimp, sea urchin, cichlid fish and frog, showing that they are widespread in animals. Penelope-related retroposons were not detected among sequences from the Drosophila melanogaster and human genome projects, suggesting that they have been lost from certain animal lineages. A sequence encoding a putative Uri (also called GIY-YIG) endonuclease domain was detected downstream from the gene for reverse transcriptase. To the best of our knowledge, this type of endonuclease sequence has previously been identified in group I introns and in genes for prokaryotic excinucleases but not in retrotransposable elements. Penelope-related elements are frequently truncated at their 5’ ends and can also be flanked by long terminal repeat-like structures. Phylogenetic analysis of the reverse transcriptase domain failed to assign Penelope-related retroposons to one of the major groups of retroelements. Overall, therefore, the evidence strongly suggests that these sequences represent a new group of retrotransposable elements.

Keywords Poseidon · Penelope · Teleosts · Retroposon · Uri endonuclease

Introduction

Reverse transcriptase-encoding retrotransposable elements, which use an RNA transcript as a retrotransposition intermediate, are important constituents of eukaryotic genomes, where they are generally considered as being selfish and parasitic. Because of their repetitive nature and their ability to move into new genomic locations, to mobilize adjacent sequences and non-autonomous retroposons, to generate processed pseudogenes and to influence the expression of flanking coding sequences, they are certainly important motors in gene and genome evolution (Britten 1997; Brosius 1999; Moran et al. 1999; Esnault et al. 2000; Kidwell and Lisch 2000; Pickeral et al. 2000). Some retrotransposable elements even have a beneficial effect on their hosts. Two retroposons constitute the telomeres of Drosophila chromosomes (Pardue and DeBaryshe 1999), and useful cellular functions might have evolved from autonomous retroelements (Eickbush 1997; Mi et al. 2000; Volff et al. 2001a).

Retrotransposons (also called LTR retrotransposons) are structurally, mechanistically and phylogenetically related to vertebrate retroviruses (Xiong and Eickbush 1990; Boeke and Chapman 1991). They bear flanking long terminal repeats (LTRs) involved in controlling transcription, in synthesis of the double-stranded DNA from the RNA intermediate and in binding of the integrase. Retrotransposon integrases that have the sequence signature D-D$_{35}$E are related to transposases from DNA transposons and are necessary for integration of the double-stranded DNA into the target site.

Autonomous retroposons (also called LINES or non-LTR retrotransposons) lack LTRs and are frequently truncated at their 5’ ends due to incomplete reverse transcription of their RNA template. Coding sequences for two kinds of endonucleases have been identified to date in retroposons: apurinic-apyrimidinic (AP) endonucleases, first described in the L1 human retroposon (Feng et al. 1996), and REL (restriction enzyme-like) endonucleases, related to type IIS restriction enzymes and found in
several site-specific retrotransposons (Yang et al. 1999). Retrotransposon endonucleases nick one strand of the target sequence. The 3’ hydroxyl group exposed by this nick is used to prime the reverse transcription of the RNA intermediate (Luan et al. 1993). Some bacterial, mitochondrial and chloroplast reverse transcriptase-encoding group II introns related to retrotransposons encode an endonuclease containing a so-called H-N-H motif. Such endonucleases are also found in group I introns, a group of mobile DNA elements (Gorbalenya 1994; Shub et al. 1994).

Fishes, which make up more than half of the 48,000 species of living vertebrates, represent an attractive model for the study of the evolution of retroelements in vertebrate genomes. Several reverse transcriptase-encoding retrotransposable elements have already been identified in fish genomes. AP endonuclease-encoding retrotransposons have been found in the medaka fish (Oryzias latipes), the Japanese pufferfish (Fugu rubripes), the platyfish (Xiphophorus maculatus) and other teleost species (Duvernell and Turner 1998; Poulter et al. 1999; Volff et al. 1999, 2000, 2001b). The first vertebrate endonuclease found to encode a restriction enzyme-like endonuclease has recently been isolated from the medaka fish and from related teleost species (Volff et al. 2001c). The only two complete retrotransposons described to date in vertebrates have been reported from two teleost species (Poulter and Butler 1998; Volff et al. 2001d). We report here on the presence in teleost genomes of the first vertebrate retrotransposable elements related to the Penelope element of Drosophila viridis.

Materials and methods

Fish stocks

The following fish species were from stocks maintained at the University of Würzburg: Xiphophorus maculatus (platyfish, Rio Jamapa), Heterandria bilamellata (Tierra Blanca), Poecilia formosa (Amazon molly, Tampico), Poeciliopsis gracilis (porthole livebearer, Rio Jamapa), Girardinus falcatus (aquarium stock), Gambusia affinis (Western mosquitofish, Pena Blanca), Fundulus heteroclitus (killifish, Laguna de Labradorres), Oryzias latipes (medaka fish, strain Carbio), Oreochromis niloticus (Nile Tilapia, Rio Purificación) and Danio rerio (zebrafish, strain m14). Rainbow trout (Oncorhynchus mykiss), pike (Esox lucius), common carp (Cyprinus carpio), European eel (Anguilla anguilla) and sturgeon (Acipenser sturio) were obtained from a local fish farm near Würzburg.

DNA manipulations

The 3.3-kb EcoRI fragment containing the partial Poseidon element from the medaka fish Oryzias latipes (strain Carbio) was isolated from a partial library of male genomic DNA cloned into pBlue-script SK+. This fragment is located on the Y chromosome of the medaka fish (Schartl and Hornung, unpublished). The DNA probe used in Southern hybridization experiments is a 647-bp XhoI-XhoI subfragment of the original EcoRI fragment containing only the partial Poseidon element. Genomic DNA was isolated as described (Schartl et al. 1996). For Southern analysis, genomic DNA was blotted, after EcoRI digestion and electrophoretic fractionation, onto positively charged nylon membranes and hybridized with the probe at 42°C in 35% formamide, 0.1% sodium pyrophosphate, 50 mM TRIS-HCl (pH 7.5), 5×SSC, 1% SDS, 5×Denhardt’s solution, and 100 μg/ml calf thymus DNA. Filters were washed with 2×SSC/1% SDS at 50°C (low stringency).

DNA and protein sequence analysis

Sequences were analyzed using the GCG Wisconsin package (Version 10.0, Genetics Computer Group, Madison). Multiple sequence alignments used for phylogenetic analysis were generated using the PileUp subroutine of GCG. Phylogenetic analysis was done with PAUP* (Swofford 1989), part of the GCG package, using neighbour-joining bootstrap analysis (1000 replicates; Saitou and Nei 1987). Protein domains were identified on the BLAST server using the Conserved Domain Database (http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi; Autschal et al. 1997).

Reconstruction of pufferfish Penelope-related elements

Partial retrotransposons related to Penelope, which were obtained from pufferfish, were extended and reconstructed using sequences generated by the Japanese pufferfish (F. rubripes; Elgar et al. 1999; http://fugu.hgmp.mrc.ac.uk) and the freshwater pufferfish (T. nigroviridis; Crollius et al. 2000; http://www.genoscope.cns.fr) genome projects: Neptune-F. rubripes (Nep-F), cosmids 041K15 and 045F03; Neptune-T. nigroviridis B (Nep-Tb), sequences AL284190 and AL326206; Neptune-T. nigroviridis A (Nep-Ta), sequences AL224686, AL306644, AL248379, AL285735, AL265878 and AL298980; Poseidon-F. rubripes (Pos-Fr), cosmids 008520, 092003, 008F14, 1778B0, 084A06 and 123102; Poseidon-T. nigroviridis (Pos-Tn), sequences AL218342, AL227991, AL249772, AL192326, AL225633 and AL273431. These sequences are genome survey sequences (random “single-pass read” and cosmid/BAC end sequences (http://www.ncbi.nlm.nih.gov/dbGGs/index.html). They might therefore correspond to different genomic copies of the same retrotransposable element and contain sequencing errors. Sequences used in the same reconstruction displayed at least 90% identity at the nucleotide level. Consensus sequences were obtained using the GelMerge fragment assembly program in the GCG Wisconsin package. Manual editing was performed when necessary. The reconstructed sequences are available on request.

Data deposition

The nucleotide sequence of the partial medaka fish retrotransposon Poseidon has been deposited in the EMBL nucleotide database under Accession No. AJ293655.

Results

Identification of Penelope-related elements in fish genomes

Sequencing of a 3.3-kb EcoRI fragment from the Y chromosome of the medaka fish O. latipes (Schartl and

---

Fig. 1 Sequence alignment of the conceptual translation products of Penelope-related elements. Identical residues are shown in black, conservative substitutions in gray (drawn using MacBoxshade). The asterisks mark the aspartate and glutamate residues previously proposed to form the catalytic triad of the putative D-D15-E inteigrase of Penelope from D. viridis (Evgen’ev et al. 1997). Abbreviations: Dy, Drosophila viridis; EN, putative Uri-like endonuclease; Fr, Fugu rubripes; Nep, Neptune; Ol, Oryzias latipes; Pen, Penelope; Pos, Poseidon; RT, reverse transcriptase; Tn, Tetraodon nigroviridis. The Accession No. for Penelope is AAA92124. The origins of pufferfish sequences are given in Materials and methods.