First publ. in: Molecular Biology and Evolution 18 (2001), pp. 427-431

Evolution and Discontinuous Distribution of *Rex3* Retrotransposons in Fish

Jean-Nicolas Volff, Cornelia Körting,* Axel Meyer,† and Manfred Schartl**

*Physiological Chemistry I, Biocenter, University of Würzburg, Würzburg, Germany; and †Department of Biology, University of Konstanz, Konstanz, Germany

The fish non–long-terminal-repeat (non-LTR) retrotransposon *Rex3* has recently been isolated from the platyfish *Xiphophorus maculatus* (Volff et al. 1999). Complete versions of *Rex3* encode a reverse transcriptase (RT) and an apurinic/apyrimidinic endonuclease (fig. 1). *Rex3* belongs to the RTE family of non-LTR retrotransposons (Malik and Eickbush 1998; Volff et al. 1999). From all autonomous fish retrotransposons reported to date, *Rex3* has the widest distribution observed in teleosts and is present in fish species having diverged 150–200 MYA. We report here a large PCR- and Southern blot–based survey of *Rex3* evolution including 21 representative teleost species (fig. 1) and 115 *Rex3* partial reverse transcriptase sequences (fig. 2). The species chosen include a panel of economically important fishes (salmon *Salmo salar,* trout *Oncorhynchus mykiss,* carp *Cyprinus carpio,* sturgeon *Acipenser sturio,* mandarin fish *Siniperca chuatsi*) and several small aquarium teleosts used as models for developmental biology, cancer research and evolutionary studies (zebrafish *Danio rerio,* medakafish *Oryzias latipes,* platyfish *Xiphophorus maculatus,* and other Poeciliidae), as well as the genome project fish, the Japanese pufferfish *Fugu rubripes.*

Rex3 partial RT-encoding sequences could be amplified by PCR from the majority of fish species with at least one of the nine primer combinations tested, but not from the more divergent nonteleost *Acipenser sturio* (sturgeon) and not from both salmonid species tested (*O. mykiss* and *S. salar*) (fig. 1). Using cloned *Rex3* probes from *X. maculatus* and from *Anguilla anguilla* (European eel) in Southern blot experiments, no specific signal could be detected in *A. sturio, O. mykiss,* and *S. salar* even under low-stringency conditions (not shown), confirming the PCR analysis results. According to classical morphological and molecular fish phylogenies (fig. 1; Nelson 1994; Forey et al. 1996; Orti and Meyer 1996), the distribution of *Rex3* in teleosts is discontinuous. Hence, *Rex3* was lost (or, alternatively, diverged extremely rapidly) in the Oncorhynchus/Salmo lineage after its divergence from the Esox lineage.

There are about 1,000 *Rex3* copies in the haploid genome of Xiphophorus species (Volff et al. 1999). All other Poeciliidae species included in this study and the related Fundulus displayed a high level of *Rex3* reiteration as well (data not shown). *Rex3* is present in high copy numbers in the genomes of *O. latipes, Oreochromis niloticus* (Nile tilapia), *Batrachocottus baikalensis, A. anguilla* (Volff et al. 1999), *Esox lucius* (pike), *Cichlasoma labridens,* and *Siniperca chuatsi* (not shown). The *Rex3* copy numbers per haploid genome were estimated by quantitative slot blot as described (Volff et al. 1999) and found to be approximately 50 for the carp *C. carpio* and 500 for the related zebrafish *D. rerio* using carp- and zebrafish-specific probes, respectively (data not shown).

A total of 115 unique sequences covering a common 420-nt part of the RT-encoding domain were obtained for phylogenetic analysis (fig. 2), including three Xiphophorus sequences isolated from a genomic cosmid library (Volff et al. 1999; AF125981–AF125983), one database sequence from an intron of the membrane guanylyl cyclase gene of the medakafish *O. latipes* (AB021490), and one database sequence from the immunoglobulin heavy-chain gene cluster of the pufferfish *F. rubripes* (AF108422), as well as several sequences from the *F. rubripes* genome project (Elgar et al. 1996; http://fugu.hgmp.mrc.ac.uk). According to morphological and molecular fish phylogenies (Nelson 1994; Forey et al. 1996; Orti and Meyer 1996), *Rex3 A. anguilla* sequences were chosen as the outgroup in phylogenetic analysis because they were the most divergent elements.

In most cases, one sequence was more related to the other sequences from the same fish species than to sequences from other species. This showed the occurrence of numerous independent bursts of retrotransposition from distinct master copies during teleost genome evolution. Different waves of retrotransposition were detected even between different members of the same family, as observed, for example, between *O. niloticus* (an Old world cichlid) and *C. labridens* (a New world cichlid). In comparisons between phylogenetic groups of sequences having different last common ancestors, the rate of substitutions between *Rex3* RT genes was clearly higher at synonymous sites (K_s) than at nonsynonymous sites (K_a) (table 1). This indicated purifying selection maintaining *Rex3* RT activity and again suggested frequent retrotransposition of *Rex3* during fish evolution. The K_s/K_a ratio was closer to unity in comparisons between closely related sequences. This indicated that *Rex3* elements were first influenced by pseudogene-like evolution after retrotransposition, as observed for other retroelements (McAllister and Werren 1997).

Using *Rex3* sequences from 10 different cosmids sequenced by the pufferfish genome project (Elgar et al. 1996; http://fugu.hgmp.mrc.ac.uk), two different *Rex3* populations were detected in *F. rubripes* (fig. 2). Eight of these Fugu elements and another database sequence (Fugu immunoglobulin heavy chain gene cluster, AF108422) were found to be more related to Fundulus/Poeciliidae elements (sequences C, only two genome

Key words: non-LTR retrotransposon, reverse transcriptase, phylogeny, Xiphophorus, teleost evolution.

Address for correspondence and reprints: Jean-Nicolas Volff, Physiological Chemistry I, Biocenter, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany. E-mail: volff@biozentrum.uni-wuerzburg.de.

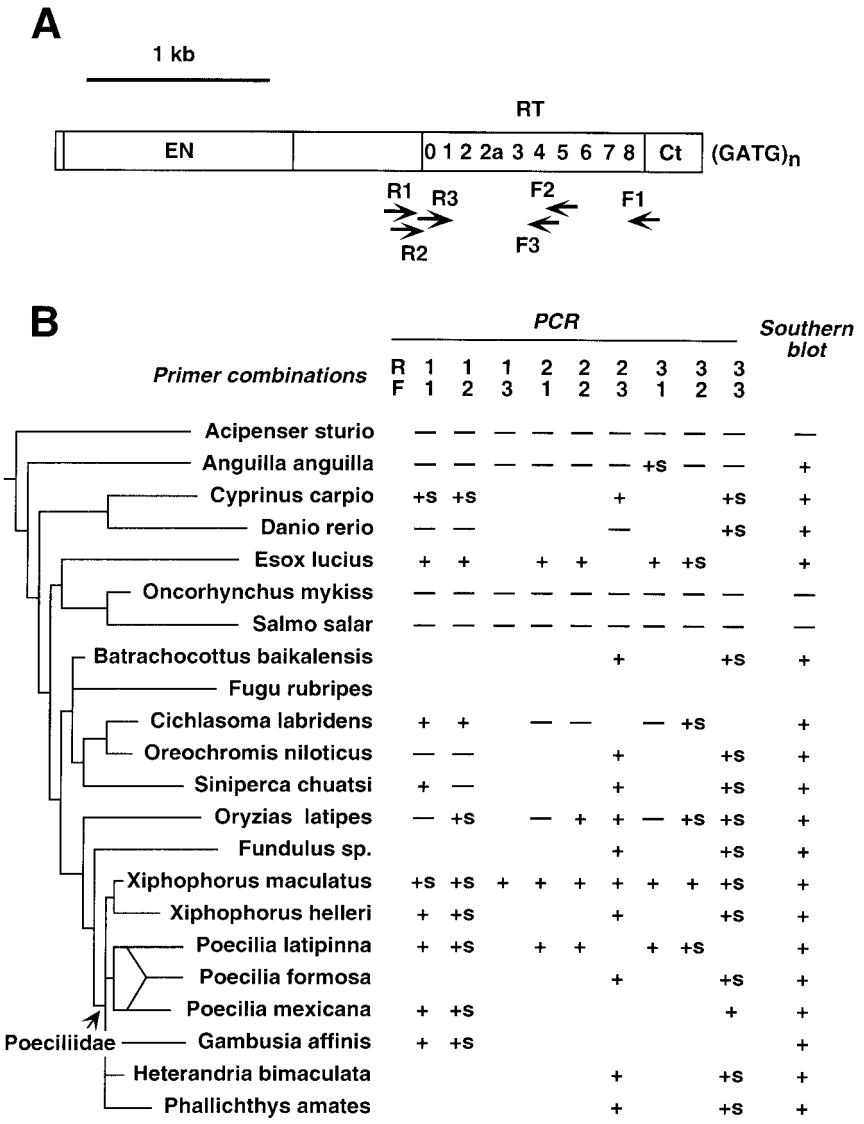

FIG. 1.—PCR- and Southern blot–based detection of *Rex3* elements in different teleost species. *A*, Structure of *Rex3* with positions of the PCR primers (arrows). EN = endonuclease-encoding region; RT = reverse transcriptase–encoding region; Ct = C-terminal domain–encoding region; 0. . . 8 = sequences encoding the conserved RT domains (Malik, Burke, and Eickbush 1999; Volff et al. 1999). All PCR reactions were done in 5% DMSO with an annealing temperature of 55°C with 200 ng genomic DNA using primers RTX3-R1 (5′-aaagttcctcggtggcaagg-3′), RTX3-R2 (5′-ccrggggtggatgarrtccgccc-3′), and RTX3-R3 (5′-tggcagachggggtggtggt-3′) in combination with primers RTX3-F1 (5′-tacggagaaaacccatttcg-3′), RTX3-F2 (5′-aacaccttggctgcgcctag-3′), and RTX3-F3 (5′-cggtgayaaagggcacccctg-3′). *B*, Success (+) or failure (−) in detecting *Rex3* by PCR and Southern blot hybridization. ''s'' indicates that these PCR products have been cloned, sequenced, and included in the phylogenetic analysis. A simplified phylogeny including the 21 teleost species and one nonteleost (*Acipenser sturio*) is given according to Nelson (1994). The origin of the fish material can be found elsewhere (Volff et al. 1999; Volff, Körting, and Schartl 2000). DNA manipulations were performed according to Volff et al. (1999).

project elements from cosmids 153A18 and 130C01 and sequence AF108422 are shown) than to a group of two other Fugu sequences (sequences D from cosmids 007P07 and 053O19) (fig. 2). Within Poeciliidae, no interspecific phylogenetic difference was found in most cases. Nevertheless, elements called ''B'' (fig. 2) were detected exclusively in the genus Poecilia (*P. mexicana*, *P. latipinna*, and *P. formosa*). As these elements clearly form a distinct group, they have probably been generated through retrotransposition of a new master sequence after the divergence of the sister genera Poecilia and Xiphophorus but before the separation of *P. mexicana* and *P. latipinna* (*P. formosa* is a hybrid between these two species).

The phylogeny of *Rex3* (fig. 2) diverges significantly from the classical fish phylogeny ($P < 0.0001$, Kishino-Hasegawa test; Kishino and Hasegawa 1989). In particular, *D. rerio*, *C. carpio*, and *E. lucius* elements should have diverged before those of *O. niloticus*, *B. baikalensis*, and *C. labridens*, and the Poeciliidae/Fundulus sequences should be more related to the *O. latipes* elements than to the *F. rubripes* C sequences. Taking different segments of the 420-nt sequence for phylogenetic analysis (the first, last, and middle 200 nt) did not

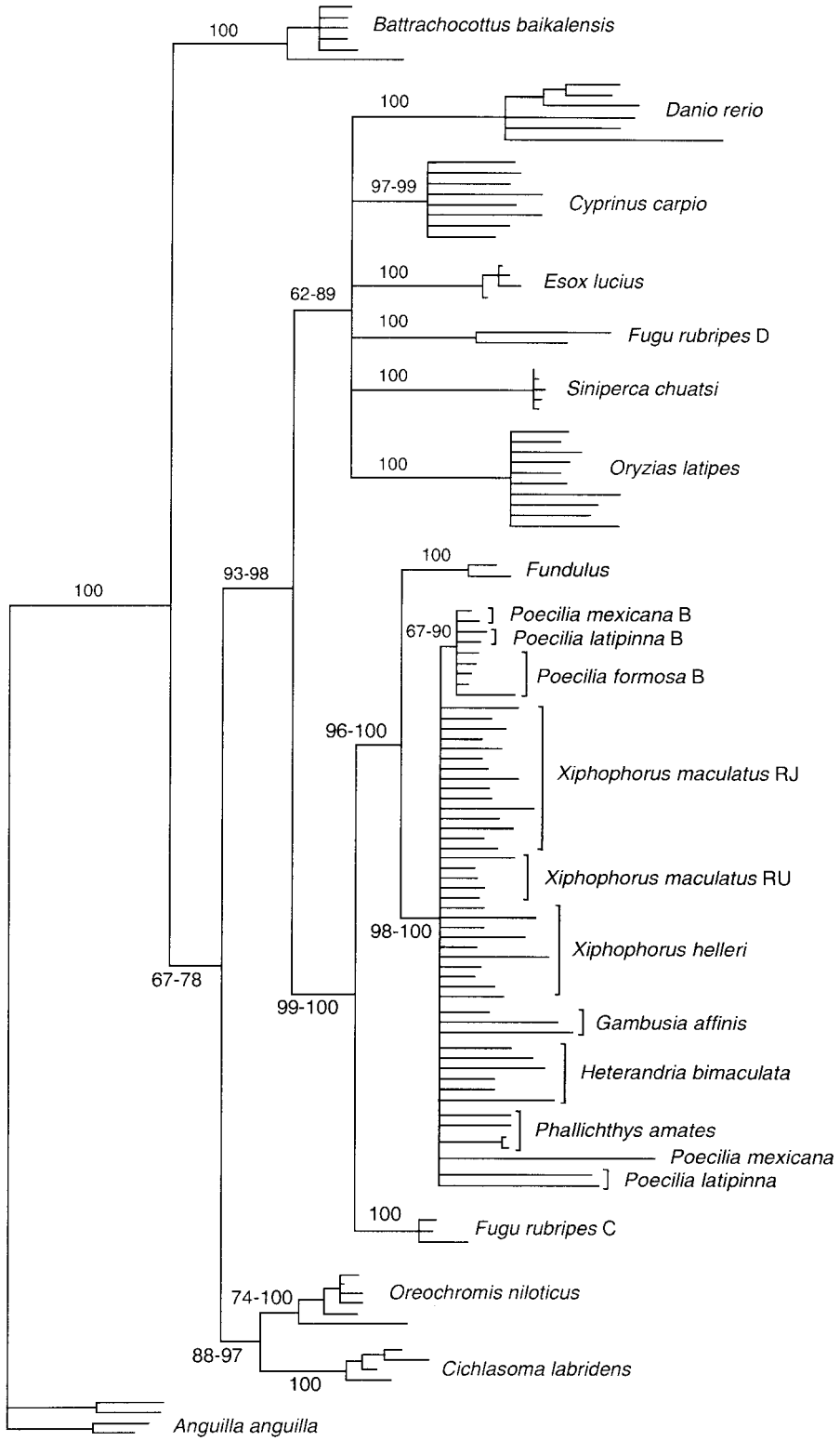

FIG. 2.—Phylogenetic analysis of teleost *Rex3* sequences. The 50% majority-rule consensus tree shown here has been rooted on the *Anguilla anguilla* sequences. Nucleotide sequences were analyzed using programs of the GCG Wisconsin package, version 10.0 (Genetics Computer Group, Madison, Wis.), as described Volff et al. (1999) and Volff, Körting, and Schartl (2000). Phylogenetic analyses were done with PAUP* (Swofford 1989) from the GCG package. Maximum parsimony and distance (minimum evolution) were used as optimality criteria. Using these methods, bootstrap analyses (100 replicates) were performed with only 46 representative *Rex3* sequences because of computer limitations. Neighbor-joining bootstrap analysis (1,000 replicates) was done with all *Rex3* sequences. Minimal and maximal bootstrap values obtained using the different methods are given. Partial *Rex3* sequences have been deposited in the EMBL database under accession numbers A400357–AJ400462. The alignment of *Rex3* partial sequences used to establish *Rex3* phylogeny has been deposited in the EMBL nucleotide sequence database (accession number ds42169). RJ = Rio Jamapa; RU = Rio Usumacinta (two different geographical populations of *Xiphophorus maculatus*).

Table 1
Average Ratios of Synonymous Versus Nonsynonymous Substitution (K_s/K_a) for *Rex3* Reverse Transcriptase Partial Nucleotide Sequences

	Ang	Cyp	Dan	Eso	Bat	FugC	FugD	Cic	Ore	Sin	Ory	Fun	XimJ	Pol	PolB
Ang	0.9	**2.2**	**2.2**	**2.9**	**3.7**	**3.2**	**2.9**	**3.5**	**2.9**	**2.7**	**3.1**	**3.4**	**3.4**	**3.3**	**3.5**
Cyp		1.4	**2.6**	**3.9**	**4.3**	**3.6**	**2.7**	**3.9**	**2.5**	**3.5**	**4.1**	**3.8**	**3.7**	**3.2**	**3.8**
Dan			1.6	**3.1**	**3.8**	**2.7**	**2.6**	**3.1**	**2.8**	**4.6**	**4.5**	**3.4**	**3.1**	**2.8**	**3.3**
Eso				ND	**5.2**	**4.6**	**3.7**	**4.3**	**3.7**	**5.5**	**4.4**	**4.7**	**4.7**	**3.7**	**4.3**
Bat.					1.5	**5.4**	**4.5**	**3.4**	**4.2**	**4.7**	**4.3**	**5.1**	**4.8**	**4.2**	**4.8**
FugC . . .						0.9	**3.1**	**3.7**	**3.8**	**5.6**	**5.5**	**4.9**	**4.8**	**3.4**	**4.4**
FugD . . .							1.6	**3.3**	**2.4**	**3.5**	**4.3**	**3.1**	**3.4**	**2.7**	**3.8**
Cic.								2.0	**5.4**	**4.1**	**4.6**	**4.3**	**4.0**	**3.1**	**3.9**
Ore									1.7	**3.3**	**4.4**	**4.7**	**4.6**	**3.1**	**4.3**
Sin.										ND	**6.2**	**3.6**	**4.1**	**3.1**	**4.3**
Ory											1.5	**4.1**	**4.1**	**3.4**	**4.2**
Fun												1.8	**4.5**	**2.8**	**4.2**
XimJ . . .													1.4	1.1	**3.0**
Pol.														0.8	1.3
PolB . . .															ND

NOTE.—Values greater than 2 are shown in bold. ND = not determinable because of a high standard deviation, indicating that the value was not significant. Ang = *Anguilla anguilla*; Bat = *Battrachocottus baikalensis*; Cic = *Cichlasoma labridens*; Cyp = *Cyprinus carpio*; Dan = *Danio rerio*; Eso = *Esox lucius*; FugC/D = *Fugu rubripes* C/D sequences (fig. 2); Fun = *Fundulus* sp.; Ore = *Oreochromis niloticus*; Ory = *Oryzias latipes*; Pol = *Poecilia latipinna* (B, B sequences; fig. 2); Sin = *Siniperca chuatsi*; XimJ = *Xiphophorus maculatus* Rio Jamapa.

introduce any significant change compared with the phylogeny of the whole sequence and did not result in a phylogeny more compatible with classical phylogenies.

Such phylogenetic discrepancies can have different causes (Capy, Anxolabéhère, and Langin 1994; Cummings 1994). The topology of *Rex3* phylogeny might be explained by the presence of several different ancient *Rex3* lineages that diverged before their actual host genomes did. Loss or nondetection by PCR of certain *Rex3* lineages could lead to comparison between paralogous sequences and introduce major differences between host and transposon phylogenies. Accordingly, two lineages were detected in the PCR-independent Fugu sequences (fig. 2). Nevertheless, we observed that the rates of synonymous substitutions between *Rex3* elements from different species were generally not higher than those for other nuclear genes presenting similar levels of codon bias, as observed for the Drosophila non-LTR retrotransposon *R1* (Lathe et al. 1995), and were frequently even lower (data not shown). This suggests that if several *Rex3* lineages are present in teleost genomes, they are probably not very ancient. Alternatively, horizontal transfer of *Rex3* retrotransposons might be compatible with the generally low rates of synonymous substitutions found for *Rex3*. Nevertheless, although horizontal transfer of a non-LTR retrotransposon has recently been suggested in teleosts (Volff, Körting, and Schartl 2000), the large number of interspecific transmissions necessary to explain the *Rex3* phylogeny appears inconsistent with the presumed extreme rarity of such events (Malik, Burke, and Eickbush 1999). Apparent anomalies observed in transposon phylogenies could also result from differences in evolutionary rates. Sequences with higher rates of evolution can typically be ''pulled down'' to the root in a tree and will give the impression that they diverged from a lineage earlier than they evolutionarily did. Finally, it also appears possible that several mechanisms have acted together during the evolution of *Rex3* sequences in fish genomes.

Acknowledgments

We are grateful to G. Schneider, H. Schwind, and P. Weber for fish maintenance, to S. Chen (University of Würzburg, Germany, and Huanghai Fisheries Research Institute, Qingdao, China), Y. Hong, and C. Winkler (University of Würzburg) and S. Kirilchik and M. Grachev (Institute of Limnology, Irkutsk, Russia) for fish organs and DNA, to K. T. Cuong and A. K. Meyer for DNA sequencing, to Y. Hong for helping to prepare fish for DNA isolation, and to the Fugu genome project for making genome survey sequences public. This work was supported by the Deutsche Forschungsgemeinschaft (Sonderforschungsbereich 465 ''Entwicklung und Manipulation pluripotenter Zellen''), the European Commission (FAIR Project PL 97-3796, ''Basis of sex determination and gonadal sex differentiation for sex control in aquaculture''), and the Fonds der Chemischen Industrie.

LITERATURE CITED

CAPY, P., D. ANXOLABÉHÈRE, and T. LANGIN. 1994. The strange phylogenies of transposable elements: are horizontal transfers the only explanation? Trends Genet. **10**:7–12.

CUMMINGS, M. P. 1994. Transmission patterns of eukaryotic transposable elements: arguments for and against horizontal transfer. Trends Ecol. Evol. **9**:141–145.

ELGAR, G., R. SANDFORD, S. APARICIO, A. MACRAE, B. VENKATESH, and B. BRENNER. 1996. Small is beautiful: comparative genomics with the pufferfish (*Fugu rubripes*). Trends Genet. **12**:145–150.

FOREY, P., D. T. J. LITTLEWOOD, P. RITCHIE, and A. MEYER. 1996. Interrelationships of elopomorph fishes. Pp. 175–191 *in* M. L. J. STIASSNY, L. R. PARENTI, and G. D. JOHNSON, eds. The interrelationships of fishes. Academic Press, San Diego.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J. Mol. Evol. **29**:170–179.

LATHE, W. C., W. D. BURKE, D. G. EICKBUSH, and T. H. EICKBUSH. 1995. Evolutionary stability of the *R1* retrotransposable element in the genus *Drosophila*. Mol. Biol. Evol. **12**: 1094–1105.

MCALLISTER, B. F., and J. H. WERREN. 1997. Phylogenetic analysis of a retrotransposon with implication for strong evolutionary constraints on reverse transcriptase. Mol. Biol. Evol. **14**:69–80.

MALIK, H. S., W. D. BURKE, and T. H. EICKBUSH. 1999. The age and evolution of non-LTR-retrotransposable elements. Mol. Biol. Evol. **16**:93–805.

MALIK, H. S., and T. H. EICKBUSH. 1998. The RTE class of non-LTR retrotransposons is widely distributed in animals and is the origin of many SINEs. Mol. Biol. Evol. **15**:1123–1134.

NELSON, J. S. 1994. Fishes of the world. 3rd edition. Wiley, New York.

ORTI, G., and A. MEYER. 1996. Molecular evolution of ependymin and the phylogenetic resolution of early divergences among euteleostean fishes. Mol. Biol. Evol. **13**:556–573.

SWOFFORD, D. L. 1989. PAUP: phylogenetic analysis using parsimony. Portable version 4.0.0d55 for Unix. Smithsonian Institution, Washington, D.C.

VOLFF, J.-N., K. KÖRTING, K. SWEENEY, and M. SCHARTL. 1999. The non-LTR retrotransposon *Rex3* from the fish *Xiphophorus* is widespread among teleosts. Mol. Biol. Evol. **16**:1427–1438.

VOLFF, J.-N., K. KÖRTING, and M. SCHARTL. 2000. Multiple lineages of the non-LTR retrotransposon *Rex1* with varying success in invading fish genomes. Mol. Biol. Evol. **17**: 1673–1684.

THOMAS EICKBUSH, reviewing editor

Accepted October 31, 2000