Nucleic Acids Research, 1996, Vol. 24, No. 16 **3167–3172**

A functional role for some *Fugu* introns larger than the typical short ones: the example of the gene coding for ribosomal protein S7 and snoRNA U17

Francesco Cecconi, Claudia Crosio, Paolo Mariottini[1], Gianni Cesareni, Marcello Giorgi, Sydney Brenner[2] and Francesco Amaldi*

Dipartimento di Biologia, Università di Roma 'Tor Vergata', I-00133 Rome, Italy, [1]Dipartimento di Biologia, Terza Università di Roma, I-00154 Rome, Italy and [2]King's College, Cambridge CB2 IST, UK

Received May 9, 1996; Revised and Accepted June 27, 1996

EMBL accession no. X94942

ABSTRACT

The compact genome of *Fugu rubripes*, with its very small introns, appears to be particularly suitable to study intron-encoded functions. We have analyzed the *Fugu* gene for ribosomal protein S7 (formerly S8, see Note), whose *Xenopus* homolog contains in its introns the coding sequences for the small nucleolar RNA U17. Except for intron length, the organization of the *Fugu S7* gene is very similar to that of the *Xenopus* counterpart. The total length of the *Fugu S7* gene is 3930 bp, compared with 12691 bp for *Xenopus*. This length difference is uniquely due to smaller introns. Although short, the six introns are longer than the ~100 bp size of most *Fugu* introns, as they host U17 RNA coding sequences. While four of the six U17 sequences are 'canonical', the remaining two represent diverged U17 pseudocopies. In fact, microinjection in *Xenopus* oocytes of *in vitro* synthesized *Fugu* transcripts containing the 'canonical' U17f sequence results in efficient production of mature U17 RNA, while injection of a transcript containing the U17ψb sequence does not.

INTRODUCTION

The coding sequences for most small nucleolar RNAs (snoRNAs) are localized in the introns coding for ribosomal proteins (r-proteins) or for other housekeeping proteins involved in the production and function of the ribosome (for reviews see 1–3). These intron-encoded snoRNAs are not produced by independent transcription but by processing of the host gene pre-mRNA (for references see 4–9). One of these snoRNAs, U17 RNA, has been found to be encoded in two introns of the human gene for RCC1 (7) and in each of the six introns of the *Xenopus* gene for r-protein S7 (formerly S8, see Note; 8). It seemed of interest to study the organization of the U17 RNA coding sequence in the genome of the fish *Fugu rubripes*, a powerful model system for studying intron-encoded functions. In fact, it has a particularly small genome, ~400 Mb in length, about 7.5 and 10 times smaller than the human and *Xenopus* genomes respectively (10). This genome compactness reflects short gene distances and, more relevant for our purposes, short gene size due to particularly small introns (10,11). This being the rule, some notable exceptions have been found of long introns comparable with human ones (12–14). It has been suggested that these exceptions are due to specific structural or functional properties of some intron sequences.

We present here an analysis of the *S7/U17* gene organization and show that the compact *Fugu* genome with its short introns can be particularly useful for identifying and studying intron-nested snoRNA coding sequences.

MATERIALS AND METHODS

DNA and RNA analysis

Unless stated otherwise, all techniques used for preparation, analysis and manipulation of DNA and RNA were performed according to standard laboratory manuals (15). Sequencing was performed by the dideoxy chain termination method (16) on both strands of overlapping fragments.

DNA cloning

A cosmid library (11) has been probed with *Xenopus S7* cDNA (17) and snoRNA U17 (8) fragments. The selected cosmid 85F12 was digested by *Eco*RI to generate a 3526 bp fragment containing the central portion of the gene and by *Dra*I to generate an overlapping ~2500 bp fragment containing the 5′-region of the gene. These fragments have been cloned in the pEMBL18 and Bluescript KS(+) vectors respectively (clones pF-S7.1 and pF-S7.2). The 3′-region of the gene has been obtained by inverse PCR using two oligonucleotide primers corresponding to nt 3850–3868 and 3882–3899 (Fig. 1); as template we used cosmid 85F12 DNA, previously digested with *Bam*HI and circularized. The ~450 bp fragment obtained was cloned in the PCRII vector (Invitrogene) to generate the pF-S7.3 clone. Plasmid pF-S7.11, containing U17c sequence, has been previously described (18). Plasmid pFS7.10, containing the U17ψb sequence, was obtained by cloning a DNA fragment encompassing the second intron, including the U17ψb sequence and the two flanking exons of the *Fugu S7* gene, into the PCRII vector. This fragment was generated by PCR amplification of the region between nt 392 and

* To whom correspondence should be addressed

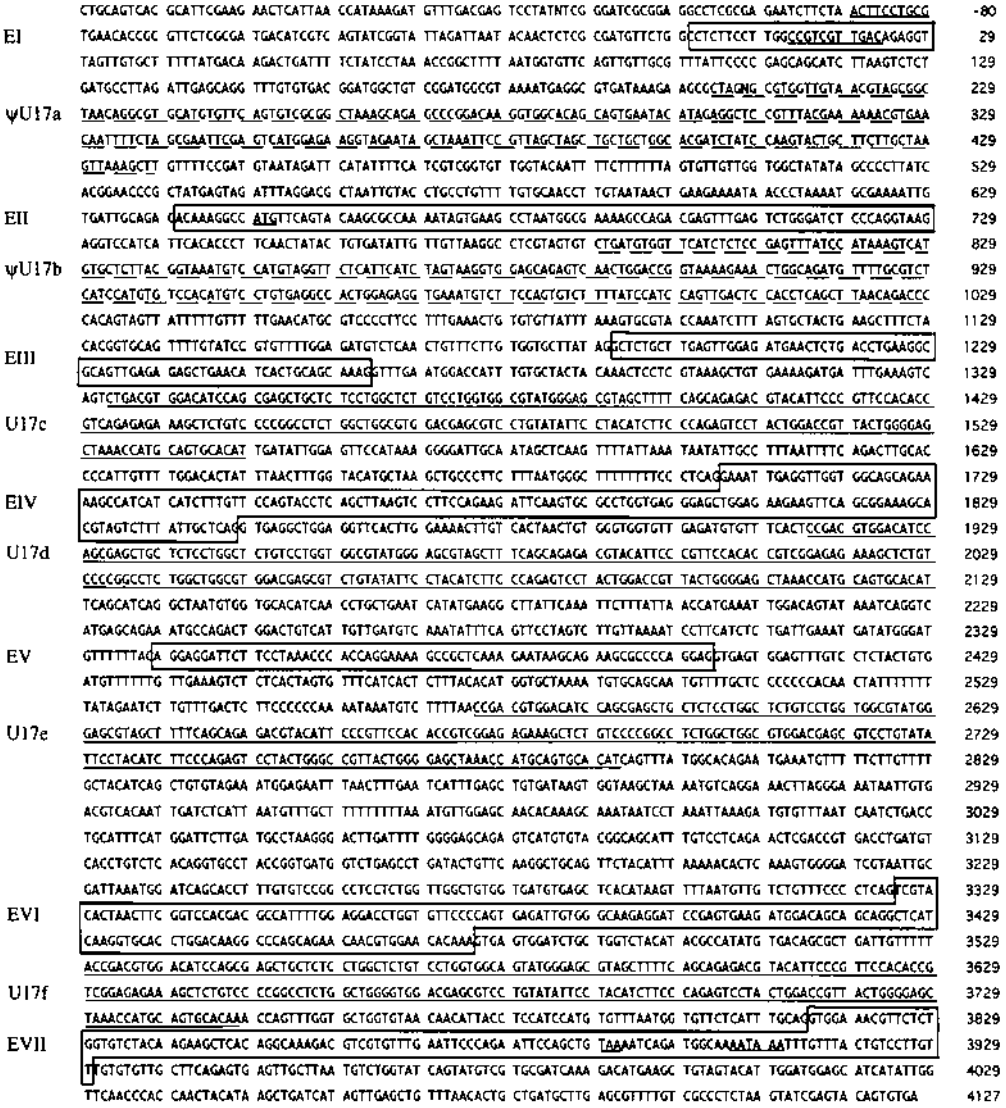

Figure 1. Nucleotide sequence of the *F.rubripes S7/U17* gene and its flanking regions. The sequence is numbered relative to the transcription start point. The seven exons are boxed and marked EI–EVII. Canonical and pseudo U17 sequence copies are underlined (continuous and dashed respectively) within the introns and marked U17c–f and ψU17a and b respectively. Promoter sequence motifs are underlined upstream and within the first exon. The start codon is underlined within the second exon, the stop codon and the polyadenylation signal are underlined within the last exon. The EMBL Data Library accession no. of this sequence is X94942.

1313 of the *S7* gene (see Fig. 1). For sequencing, plasmids pF-S7.1 and pF-S7.2 were digested with various restriction enzymes to generate overlapping fragments and subcloned in Bluescript KS(+). A fragment of the rRNA gene encompassing the 18S rRNA region complementary to U17 RNA was PCR amplified using two primers corresponding to regions well conserved among vertebrates, nt 568–589 and 1092–1112 according to the *Xenopus* 18S rRNA gene sequence (19).

In vitro synthesis of radioactive transcripts

To obtain the transcripts to be used as processing substrates in microinjected oocytes, 1 µg pF-S7.11 DNA was digested with *Eco*RI and transcribed with T7 RNA polymerase, while 1 µg pF-S7.10 was digested with *Xba*I and transcribed with SP6 RNA polymerase, in the presence of 50 µCi [α-^{32}P]UTP as described (20). Transcripts of 1048 nt (IIi) and 592 nt (IIIi) were obtained respectively from plasmid pF-S7.10 and pF-S7.11. After transcription and DNase digestion, the RNAs were purified by phenol/chloroform/isoamyl alcohol (50:50:1) extraction and ethanol precipitation and resuspended in H_2O for microinjection.

RNA microinjection into *Xenopus* oocytes

Isolation of *Xenopus* stage V–VI oocytes and microinjection of RNA into the germinal vesicle were carried out essentially as previously described (21). Oocytes were injected with 40 nl H_2O containing ~80 000 c.p.m. (corresponding to 10–40 ng) of the *in vitro* transcribed RNAs and incubated for increasing time intervals at 22°C. After incubation, nuclei from pools of 10 oocytes were manually prepared (21) and then lysed in 300 µl 100 mM Tris, pH 7.5, 300 mM NaCl, 10 mM EDTA, 2% SDS, containing

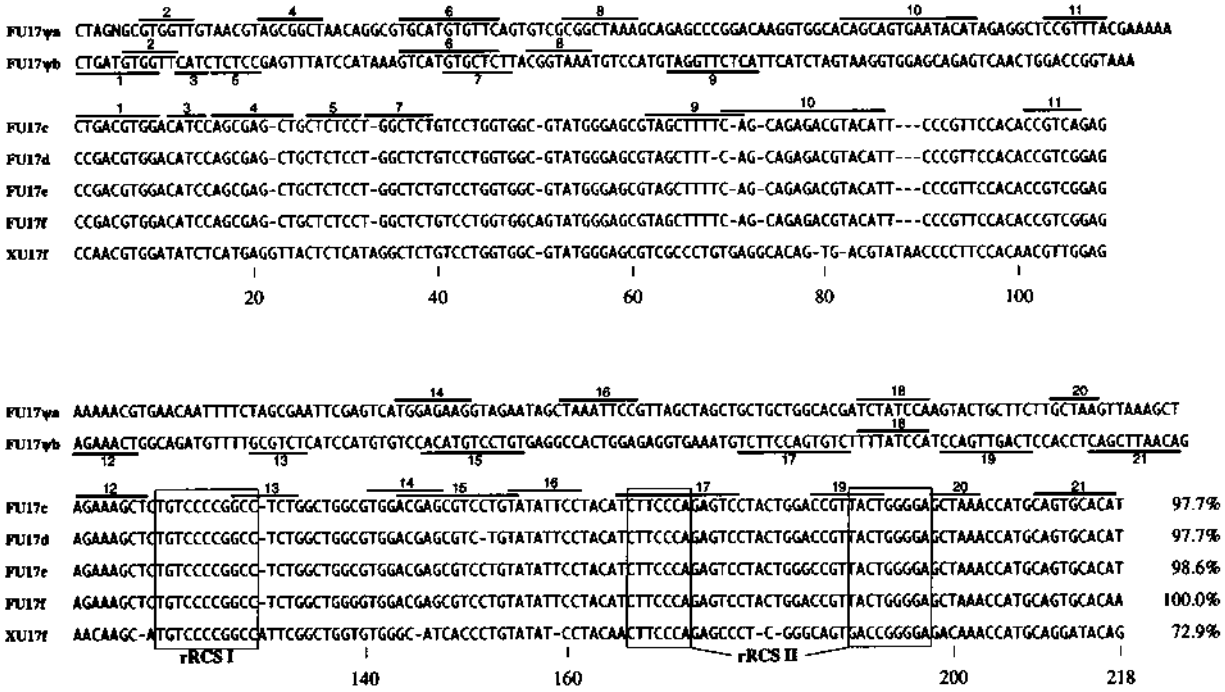

Figure 2. Comparison of the nucleotide sequences of the *Fugu* and *Xenopus* U17 snoRNA gene copies. The rRNA complementary sequences (rRCSI and II) are boxed; nucleotide sequence homologies with respect to the *Fugu* f copy (FU17f) are shown at right. The two pseudocopies U17ψa and U17ψb are not aligned with the canonical copies, due to the many insertions/deletions; to help comparison some homologous blocks are marked and numbered along the sequence.

1 mg/ml proteinase K (22). RNA was extracted and analyzed by 6% polyacrylamide–8 M urea gel electrophoresis according to standard procedures.

RESULTS

Isolation and sequencing of *S7/U17* genomic clones

A *F.rubripes* cosmid library (11) was probed at low stringency with a *Xenopus* cDNA specific for r-protein S7 (17) and a *Xenopus* genomic fragment containing a U17 snoRNA gene copy (8). The finding of clones hybridizing to both probes suggested that the U17 RNA coding sequence is hosted in the *S7* gene of *Fugu*, as occurs in *Xenopus*. The selected cosmid 85F12 was digested with *Eco*RI and analyzed by Southern blot with the same probes as above. The positive *Eco*RI 3.5 kb fragment, corresponding to most of the gene, was subcloned and sequenced. Completion of the 5′- and 3′-regions was obtained as described in Materials and Methods. The sequence of the entire *Fugu rp-S7* gene with its flanking sequences is shown in Figure 1.

The r-protein *S7* gene

Sequence comparison with the *Xenopus* r-protein *S7* cDNA (17) and gene (23) allowed precise determination of the exon–intron boundaries in the *Fugu* gene. The 5′-end of the first exon (transcription start site) and the 3′-end of the last exon have not been experimentally determined, however, the very good sequence conservation in the regions surrounding the two sites allows their tentative identification in *Fugu* by comparison with *Xenopus*. The *Fugu S7* gene is made up of seven exons and six introns, as is its *Xenopus* homolog, and the positions of the six introns are also perfectly conserved between the two species. The overall size of the *Fugu S7/U17* gene is 3930 bp, compared with the 12691 bp *Xenopus* counterpart. Exon size ranges from 27 to 151 bp, as in *Xenopus*, while introns range from 339 to 920 bp, shorter than the 1057–4645 bp observed in *Xenopus*. Thus, the more compact *Fugu* gene organization is uniquely due to the smaller size of the introns, while the coding regions for the r-protein S7 are identical in the two species. Comparison of the coding sequences (not shown) indicates a very high homology between *Xenopus* and *Fugu* at the protein level (95.9%) and a somewhat lower homology at the nucleotide level (82.4%), due to several silent nucleotide substitutions, mainly in the third position of codons.

All vertebrate r-protein genes analyzed up to now have their transcription start site located within a 12–25 nt pyrimidine sequence, so as to transcribe mRNAs always starting at their 5′-end with a 6–12 pyrimidine sequence, which has been implicated in the translational regulation of this class of mRNAs (for references see 24,25). The *Fugu S7* gene also follows this rule: the presence of two G residues ‘contaminating’ the pyrimidine sequence is not unusual, as a similar situation has also been described for other r-protein genes.

Comparison of the 5′-region of the *Fugu S7* gene with the promoter regions of other vertebrate r-protein genes has revealed the presence of at least two relevant sequences, indicated in Figure 1. At position –72 the sequence 5′-ACTTCCTGCG is present, also found in the promoters of other vertebrate r-protein genes and shown to be responsible for binding of a transcription factor, called β in the mouse (26,27) or XrpFI in *Xenopus* (28). Moreover, the sequence 5′-GGCCGTCGTT at +11 shows high homology to an element located at the same position in the *Xenopus S7* gene and described in the promoters of some mouse

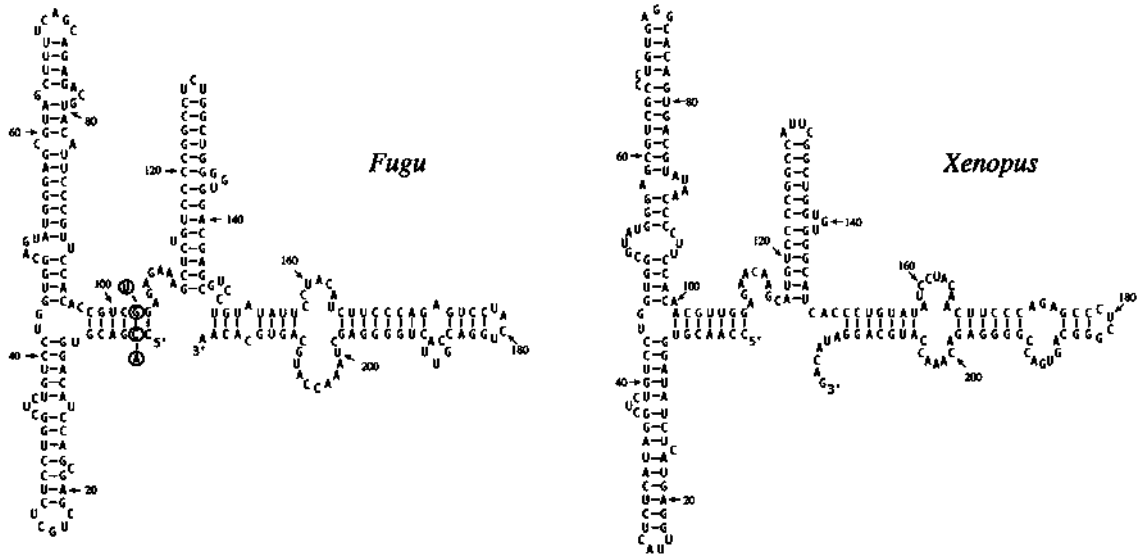

Figure 3. Secondary structure of *Fugu* and *Xenopus* U17 snoRNA f copies. RNA structures were generated with the MFOLD and SQUIGGLES programs of GCG (31). The two compensatory substitutions present in the *Fugu* c copy, with respect to the *Fugu* f copy, are circled.

r-proteins genes, where it has been shown to be responsible for binding of the δ transcription factor (26,27).

The U17 RNA coding sequences

As already suggested by the hybridization data (see above), the introns of the *S7* gene host the coding sequences for U17 RNA. Figure 1 shows, underlined, four easily identifiable U17 sequence copies, one in each of the last four introns (copies c–f, named according to their intron localization). The four sequences are compared in Figure 2. Their homology is very high (~98% using copy f as reference) and somewhat lower (~73%) in comparison with the f copy of *Xenopus* U17 sequence.

The computer-derived secondary structure of *Fugu* U17 RNA, shown in Figure 3, is in excellent agreement with that proposed for *Xenopus* U17 RNA (8). Two of the very few nucleotide differences among the four *Fugu* sequences represent compensatory substitutions that leave the secondary structure unaltered. In fact, as indicated in Figure 3, in U17c nucleotide 2 is a U, matching with an A at position 101, while in the other three U17 copies there is a C base paired with a G. The other few substitutions occurred in unpaired regions. The comparison between the *Fugu* and *Xenopus* sequences shows many more differences; in this case also they either involve nucleotides not implicated in base pairing or are compensatory nucleotide changes.

Several snoRNAs contain regions of complementarity to rRNA, probably implicated in interactions at these sites. In particular, it was proposed that *Xenopus* U17 snoRNA has two boxes complementary to regions of the 18S rRNA and of the ETS respectively (8). These sequences are conserved in the four copies of *Fugu* U17 sequence and are boxed in Figure 2 (rRCSI and rRCSII, rRNA complementary sequences I and II). We have also cloned and sequenced a fragment of the *Fugu* 18S rRNA gene encompassing the rRCSI complementary sequence (not shown) that is also conserved between *Xenopus* and *Fugu*. Thus the complementarity between U17 RNA and 18S rRNA is maintained.

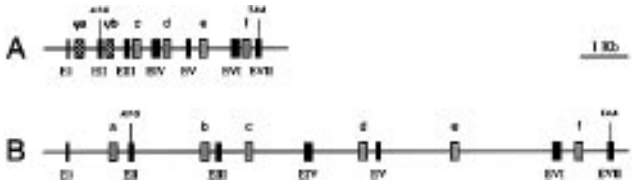

Figure 4. A comparison of *Fugu* (**A**) and *Xenopus* (**B**) *S7*/*U17* genes. Exons are represented by filled boxes and named according to Figure 1. Gray and striped boxes represent respectively canonical and pseudo copies of U17 sequences, named as in Figure 1. Start and stop codons are indicated in the second and last exons.

Careful inspection of the first two introns revealed the presence of two degenerate U17 sequences: copy ψa in the first intron and ψb in the second. The sequences are dashed underlined in Figure 1 and compared with the canonical U17 sequences in Figure 2. The relation to the U17 sequence is indicated by the presence of some conserved sequence blocks: copy ψb presents somewhat better matches than copy ψa. Notice that some blocks are more conserved between the two pseudocopies than between these and the other copies. Attempts to generate computer-derived secondary structures for these two U17 pseudocopies comparable with the canonical one have failed. The absence or poor conservation of the rRCS elements also strongly suggests that these two sequences represent pseudogene copies.

A comparison of the organization of the *Fugu S7*/*U17* gene with that of the corresponding *Xenopus* gene is shown in Figure 4.

U17 RNA processing from *Fugu* transcripts injected in *Xenopus* oocytes

The correct and efficient processing of various intron-encoded snoRNAs by heterologous systems indicates that the processing mechanism is evolutionarily very well conserved among all vertebrates (2). In particular, we have shown that U17 RNA

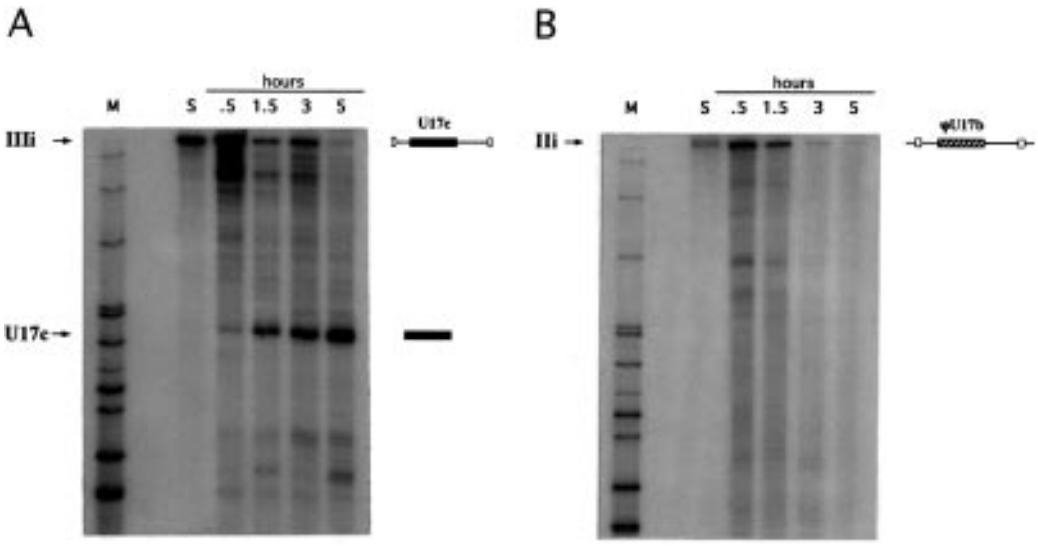

Figure 5. Analysis of *Fugu* U17 snoRNA production by RNA microinjection into *Xenopus* oocyte nuclei. (**A**) An *in vitro* synthesized 592 nt radioactive RNA (IIIi), corresponding to part of the third exon, the entire third intron bearing the U17c copy and part of the fourth exon of the *Fugu S7* gene, was injected in *Xenopus* oocytes. After incubation for increasing time intervals (0.5–5 h), total RNA was extracted and analyzed by gel electrophoresis and autoradiography. (**B**) An *in vitro* synthesized 1048 nt long radioactive RNA (IIi), corresponding to part of the second exon, the entire second intron bearing the ψU17b copy and part of the third exon of the *Fugu S7* gene, was utilized and analyzed as in (A). M, RNA size marker; S, transcript substrate. Arrows point to the intact injected transcripts and to the mature product. Schematic representations of the RNA molecules are shown on the right.

production by processing of precursor transcripts, mainly due to 5′ and 3′ exonucleolytic activities, is conserved between fish and amphibians (18). In fact, a radioactive RNA precursor containing a *Fugu* U17 sequence is correctly and efficiently processed to yield mature U17 RNA when microinjected into the germinal vesicle of *Xenopus* oocytes. Now, to verify that the two degenerate U17ψa and ψb sequences, present in the first two introns, are indeed pseudogenes, we microinjected the following radioactive RNAs: (i) a 1048 bp transcript (IIi) containing the entire intron 2, including the less degenerate U17ψb copy sequence and part of the flanking exons; (ii) a 592 bp transcript (IIIi) containing the entire intron 3, including the canonical U17c copy sequence and part of the flanking exons. Figure 5A shows that the injected transcript IIIi is correctly processed to produce mature U17 RNA; in contrast, transcript IIi does not produce any stable RNA and is completely degraded.

Expression of the *Fugu* U17 snoRNA gene in different tissues

A Northern blot analysis has been carried out on total RNA isolated from different *Fugu* tissues (not shown). The results obtained prove that U17 RNA is expressed in the adult fish and provide an estimate of its relative abundance that seems to be more or less the same in all the somatic and germinal tissues examined. As for most other snoRNAs, the presence of U17 RNA in normal tissues and its conservation in evolution are the only indications that it has a functional role, the specific function remaining to be determined.

DISCUSSION

It has been shown that the fish *F.rubripes* has, among vertebrates, a particularly compact genome, approximately eight times smaller than that of mammals (10,11). One of the reasons for this compactness is the small size of most introns, which have a modal length of <100 nt. It seemed to us that this situation is particularly suitable for the study of the structural organization of those genes which contain, nested in their introns, the coding sequences for snoRNAs. Since the intron coding arrangement was revealed for mouse U14 RNA, the number of intron-encoded snoRNA genes in various vertebrates has grown to >15 during the last 3 years and is still growing fast (for references see 1–3). All these snoRNA sequences are hosted in the introns of genes coding for r-proteins or for other proteins involved in the production and function of the translation apparatus. In some cases specific snoRNAs are encoded in different host genes in different organisms. In particular, U17 RNA coding sequences have been found in all six introns of the gene for r-protein S7 in *Xenopus* (8) and in the first two introns of the human *RCC1* gene (7). We have shown here that in *Fugu* the same snoRNA is encoded, as in *Xenopus*, in the introns of the r-protein *S7* gene. However, only four copies (c–f, in introns 3–6) appear to be canonical on the basis of their conserved sequence and computer-derived secondary structure. For the c copy, we have shown that the corresponding precursor transcript, when microinjected into *Xenopus* oocytes, is correctly processed to produce mature U17 RNA (18). On the other hand, two other copies, ψa and ψb, located in the first two introns of the host *S7* gene, appear to be degenerate copies of the U17 sequence, ψb being somewhat better conserved than ψa. This structural divergence is accompanied by a loss of the ability to be correctly processed; in fact, microinjection into *Xenopus* oocytes of a transcript corresponding to the second intron and containing the U17ψb sequence did not result in the production of any stable RNA, but in complete degradation of the injected precursor.

The presence of these U17 coding sequences seems to be the reason for the relatively large size of these six introns, ranging between 339 and 920 bp; although much smaller than in *Xenopus*,

this size is larger than that reported for the majority of *Fugu* introns (10). Thus analysis of the gene structure in this species could facilitate the study of intron-encoded functions, in part because of the reduced cloning and sequencing workload, but mainly because the intron length itself, if substantially exceeding 100 nt, could be indicative of the presence of an intron-specific function.

The results presented here also bear on the problem of evolution of genome organization. At present it is difficult to decide if the short intron-containing genome of *Fugu* represents an ancestral situation from which longer intron-containing genomes of other vertebrates originated or it is the result of progressive reduction of an ancient large intron-containing genome. The first view is consistent with the finding that since the time the two U17ψa and ψb sequences started to degenerate their two host introns have maintained a length similar to that of the four introns containing conserved U17 RNA sequences and several times larger than the average *Fugu* intron. However, one might object that not enough time has yet elapsed for an appreciable reduction in intron size to have occurred.

NOTE

In previous papers we have used the ribosomal protein numbering system introduced in our first study of *Xenopus* r-proteins (29). The large amount of sequencing data now accumulated allows us to adopt, as a unified nomenclature, the rat system (30). Thus the r-protein that we previously designated S8 is now identified as S7 for both *Xenopus* and *Fugu*.

ACKNOWLEDGEMENTS

This work was supported by grants from Progetto Finalizzato Ingegneria Genetica, CNR and from Ministero Università e Ricerca Scientifica e Tecnologica. Part of this work has been carried out in the Department of Medicine of the University of Cambridge Clinical School, UK, by G.Cesareni, with the support of a CNR fellowship.

REFERENCES

1 Sollner-Webb,B. (1993) *Cell*, **75**, 403–405.
2 Maxwell,E.S. and Fournier,M.J. (1995) *Annu. Rev. Biochem.*, **35**, 897–934.
3 Bachellerie,J.-P., Michot,M., Nicoloso,A., Balakin,J., Ni,J. and Fornier,J.M. (1995) *Trends Biochem. Sci.*, **20**, 261–264.
4 Leverette,R.D., Andrews,M.T. and Maxwell,E.S. (1992) *Cell*, **71**, 1215–1221.
5 Fragapane,P., Prislei,S., Michienzi,A., Caffarelli,E. and Bozzoni,I. (1993) *EMBO J.*, **12**, 2921–2928.
6 Tycowsky,K.T., Shu,M.-D. and Steitz,J.A. (1993) *Genes Dev.*, **78**, 1176–1190.
7 Kiss,T. and Filipowicz, W. (1993) *EMBO J.*, **12**, 2913–2920.
8 Cecconi,F., Mariottini,P., Loreni,F., Pierandrei-Amaldi,P., Campioni,N. and Amaldi,F. (1994) *Nucleic Acids Res.*, **22**, 732–741.
9 Pellizzoni,L., Crosio,C., Campioni,N., Loreni,F. and Pierandrei-Amaldi,P. (1994) *Nucleic Acids Res.*, **22**, 4607–4613.
10 Brenner,S., Elgar,G., Sandford,R., Macrae,A., Venkatesh,B. and Aparicio,S. (1993) *Nature*, **366**, 265–268.
11 Baxendale,S., Abdulla,S., Elgar,G., Buck,D., Bercks,M., Micklem,G., Durbin,R., Bates,G., Brenner,S., Beck,S. and Lehrach,H. (1995) *Nature Genet.*, **10**, 67–75.
12 Mason,P.J., Stevens,D.J., Luzzatto,L., Brenner,S. and Aparicio,S. (1995) *Genomics*, **26**, 587–591.
13 Macrae,A. D. and Brenner,S. (1995) *Genomics*, **25**, 436–446.
14 Aparicio,S., Morrison,A., Gould,A., Gilthorpe,J., Chaudhuri,C., Rigby,P., Krumlauf,R. and Brenner,S. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 1684–1688.
15 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
16 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
17 Mariottini,P., Bagni,C., Annesi,F. and Amaldi,F. (1988) *Gene*, **67**, 69–74.
18 Cecconi,F., Mariottini,P. and Amaldi,F. (1995) *Nucleic Acids Res.*, **23**, 4670–4676.
19 Ajuh,P.M., Heeney,P.A. and Maden,B.E.H. (1991) *Proc. R. Soc. Lond. B*, **245**, 65–71.
20 Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
21 Bozzoni,I., Fragapane,P., Annesi,F., Pierandrei-Amaldi,P., Amaldi,F. and Beccari,E. (1984) *J. Mol. Biol.*, **180**, 987–1005.
22 Pierandrei-Amaldi,P., Beccari,E., Bozzoni,I. and Amaldi,F. (1985) *Cell*, **42**, 317–323.
23 Mariottini,P., Bagni,C., Francesconi,A., Cecconi,F., Serra,M. J., Chen,Q.-M., Loreni,F. and Amaldi,F. (1993) *Gene*, **132**, 255–260.
24 Pierandrei-Amaldi,P. and Amaldi,F. (1994) *Genetica*, **94**, 181–193.
25 Meyuhas,O., Avni,D. and Shama,S. (1996) In Hershey,J.W.B. *et al.* (eds), *Translational Control of Ribosomal Protein mRNAs in Eukaryotes*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 363–388.
26 Hariharan,N., Kelley,D.E. and Perry,R.P. (1989) *Genes Dev.*, **3**, 1789–1800.
27 Sáfrány,G. and Perry,R.P. (1995) *Eur. J. Biochem.*, **230**, 1066–1072.
28 Lagna,G., Loreni,F., Beccari,E. and Carnevali, F. (1990) *Nucleic Acids Res.*, **18**, 5811–5816.
29 Pierandrei-Amaldi,P. and Beccari, E. (1981) *Eur. J. Biochem.*, **106**, 603–611.
30 Wool,I.G., Endo,Y., Chan,Y. L. and Glück,A. (1990) In Hill,W.E. *et al.* (eds), *The Ribosome: Structure, Function and Evolution*. American Society for Microbiology, Washington, DC, pp. 203–214.
31 Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.