

Structure and Organization of the Engraulidae Family U2 snRNA: An Evolutionary Model Gene?

Hicham Chairi^{1,2} · Laureana Rebordinos Gonzalez¹

Received: 20 November 2014 / Accepted: 27 March 2015 / Published online: 3 April 2015
© Springer Science+Business Media New York 2015

Abstract The U2 snRNA multigene family has been analyzed in four species of the Engraulidae family—*Engraulis encrasicolus*, *Engraulis mordax*, *Engraulis ringens*, and *Engraulis japonicus*—with the object of understanding more about the structure of this multigene family in these pelagic species and studying their phylogenetic relationships. The results showed that the cluster of this gene family in the *Engraulis* genus is formed by the U2–U5 snRNA with highly conserved sequences of mini- and micro-satellites, such as (CTGT)_n, embedded downstream of the transcription unit; findings indicate that this gene family evolved following the concerted model. The phylogenetic analysis of the non-transcribed spacer of cluster U2–U5 snDNA in the 4 species showed that the sequences of the species *E. encrasicolus* and *E. japonicus* are closely related; these two are genetically close to *E. mordax* and slightly more distant from *E. ringens*. The data obtained by molecular analysis of U2–U5 snDNA and their secondary structure, with the presence of the micro-satellite (CTGT)_n and mini-satellites, show clearly that the species *E. encrasicolus* and *E. japonicus* are closely related and would be older than *E. mordax* and *E. ringens*.

Keywords Multigene family · U2 snRNA · U5 snRNA · Concerted evolution · Engraulidae

Introduction

The multigene snRNA family plays a key role in the splicing of the vast majority of pre-mRNA by removing introns and giving rise to a mature mRNA, through a catalytic active complex called the spliceosome. The U1, U2, U4, U5, and U6 snRNA, together with a group of seven proteins known as Sm-proteins, form a complex called “small nuclear ribonucleic proteins” (snRNP); this complex plays an important role in the pairing of bases with the branching sequences of pre-mRNA.

The coding sequence of U2 snRNA has been highly conserved in all eukaryotes. However, the number and organization of U2 genes differ considerably among diverse species, reflecting a wide variation in the genes that encode the U2 snRNA (Hernandez 2001; Yu et al. 2001). For example, in hemiascomycete yeasts, the size and organization of the gene diverge widely, taking a large size (1177 bp) in *Saccharomyces cerevisiae*, and slightly smaller (1100 bp) in *S. servazzii*, but a much smaller size (195 bp) in *Yarrowia lipolytica*, a less related yeast species; this latter is similar to the size of the human U2 snRNA-188 bp (Bon et al. 2003).

With respect to transcription, this family of genes is very interesting because all of its members have very similar promoters, although some of them are transcribed by polymerase II. Genes encoding for snRNA have served as a model system for comparing transcriptional mechanisms of RNA polymerase II and III (Hernandez 2001). Studies of pre-mRNA have also provided abundant evidence of the important function of alternative or

✉ Laureana Rebordinos Gonzalez
laureana.rebordinos@uca.es

¹ Laboratorio de Genética, Facultad de Ciencias del Mar y Ambientales, CACYTMAR, Universidad de Cádiz, Polígono Río San Pedro, s/n, 11510, Puerto Real, Cádiz, Spain

² Present Address: Laboratoire Pluridisciplinaire, Faculté Polydisciplinaire, Université Abdelmalek Essaadi, B.P. 745, Poste Principale, 92004 Larache, Morocco

differential splicing in the development and differentiation of tissues (Arystarkhova et al. 2002).

Most tandemly repeated multigene families evolve according to a concerted model, which means that all copies of the gene family are homogenized (Liao 2000; Nei and Rooney 2005; Drouin and Tsang 2012). Unequal crossover and gene conversion have been proposed as mechanisms that might explain the sequence homogenization during concerted evolution (Pavelitz et al. 1995; Liao et al. 1997; Rooney 2004). Although many studies have demonstrated that the majority of multigene families evolve in a strictly concerted model, the null hypothesis of evolution through a homogenization process has been rejected in H1, H3, and H4 histone gene families (Piontkivska et al. 2002; Rooney et al. 2002; Eirin-Lopez et al. 2004). It has been proposed that the long-term evolution of these families follows the birth-and-death model (Rooney 2004; Rooney and Ward 2005). Merlo et al. (2012) have described a possible transition from a birth-and-death to a concerted evolution model in *Halobatrachus didactylus*. Under this process, a multigene family both expands because of gene duplication and contracts due to the loss of genes. Thus, copies of different genes accumulate differences, leading some of them to degenerate into pseudogenes, and some others are completely eliminated (Vierna et al. 2009; Eirín-López et al. 2012; Rebordinos et al. 2013).

The Engraulidae or anchovy family is composed of 17 genera and 147 species. The *Engraulis* genus comprises nine species, five of them, *Engraulis ringens*, *Engraulis japonicus*, *Engraulis encrasicolus*, *Engraulis capensis*, and *Engraulis mordax*, are caught in large volumes and are thus of major commercial importance (FAO 2012; Eschmeyer 2014). Fish of these species are small, of size between 10 and 20 cm. Generally, they are coastal fish and their catches represent at least 15 % of the 90 million tons of fish caught annually in the world (FAO 2012).

From the anatomical perspective, eight species of *Engraulis* were divided into two groups by Whitehead et al. (1988) in their famous Engraulidae catalog. One group, known as the “Old World” species, contains *E. encrasicolus*, *E. eurystole*, *E. capensis*, *E. japonicus*, and *E. australis*; a second group, the “New World” species, is subdivided into subgroup 2a containing *E. ringens* and *E. anchoita*, and subgroup 2b containing *E. mordax*. The most significant morphological differences between these two main groups correspond to the size from the tip of the maxilla to the pre-operculum, and the pseudobranch size compared to the eye.

The aims of the present study are the analysis of structure and organization of the U2 snDNA multigene family in four species of *Engraulis*, *E. encrasicolus*, *E. mordax*,

E. ringens, and *E. japonicus*, and study of the possible evolution of this gene in the four species.

Materials and Methods

Animal Sampling and DNA Isolation

For the analysis of the U2 snDNA multigene family in the four species of the Engraulidae family, the following were used: fresh samples of *E. encrasicolus* from the Gulf of Cadiz; samples of *E. mordax* from Baja California in Mexico; *E. ringens* from Valparaiso in Chile; and *E. japonicus* from the Sea of Japan. All samples were collected and preserved in absolute ethanol. Total genomic DNA was extracted from 150 mg of muscle sections with the FastDNA kit for 40 s, at a speed of 5 m/s, in a FP120 Fastprep instrument (Bio101, Inc., Vista, CA).

PCR Amplification, Cloning, and Sequencing

The U2 snDNA gene was amplified with the two universal primers U2F: CAAAGTG TAGTATCTGTTCTTATCAGC and U2R: CTTAGCCAAAAGGCCGAGA designed by (Cross and Rebordinos 2005). PCR was performed in a final volume of 50 μ L containing approximately 150 ng of genomic DNA, 3 mM MgCl₂, 200 μ M dNTP (Fermentas), 0.2 μ M of each primer, 2 units of *Taq* polymerase (Euro-Clone, Italy), and the appropriate buffer for the enzyme. The reaction was carried out using an initial cycle of 5 min at 94 °C and 35 cycles of 45 s at 94 °C, 45 s at 59 °C, and 1 min at 72 °C and a final extension of 10 min at 72 °C. Products were visualized by agarose gel 2 % with ethidium bromide (0.5 μ g/mL) after electrophoresis. The PCR products obtained were purified using the NucleoSpin[®] Extract II kit (Macherey–Nagel, Inc., Bethlehem, PA) and cloned using the pGEM cloning kit from T Easy Vector Systems (Promega, Madison, WI). The sequencing was carried out using a sequencing kit (BigDye Terminator v 3.1 Cycle Sequencing Kit, Applied Biosystems).

Data Analysis

Sequencing peaks were analyzed with Chromas 2.0, and the sequences obtained were aligned using the ClustalX (Thompson et al. 1997) and BioEdit (Hall 1999) programs. Evolutionary divergence between sequences is estimated by the number of base differences per site from comparison of the sequences. A phylogenetic analysis was conducted in MEGA6 (Tamura et al. 2013), using the Neighbor-joining method, and the reliability of the resulting topologies was tested using the bootstrap method (Felsenstein 1995). The

evolutionary distances were computed using the p-distance method (Nei and Kumar 2000).

Results

The gene encoding the U2 snRNA was amplified using universal primers in three individuals of each of the four species of *Engraulis*. Cloning was carried out using the pGEM kit, and 12 clones were sequenced for each species. The sequences obtained were submitted to BLASTN (Altschul et al. 1990) (NCBI) to determine the similarities of the sequences obtained to other sequences from the GenBank database.

The alignment of the sequences of the clones of each species using the ClustalX software revealed that, for *E. encrasicolus*, their sizes vary between 724 and 788 bp, and, for *E. ringens*, clones range in size between 819 and 859 bp. For *E. mordax* and *E. japonicus*, clones of size between 704 and 760 bp, and between 728 and 781 bp, were obtained, respectively (Table 1). This difference in length is due mainly to the presence of the micro-satellite (CTGT)_n with a variable number of repetitions, the presence of poly A sequences of different sizes, and to the presence of a mini-satellite sequence of 23 bp that also varies in the number of repetitions.

Figure 1 shows the alignment of the coding region of the U2 snRNA of the 4 species. These sequences are highly conserved within all clones. The region shows a high degree of conservation, especially in the first 80 bp and in the protein binding sites or RNA–protein and protein–protein interaction, such as *Sm*, *C-hnRNP* and *U2A'–U2B'* and *Branch point* intron in the pre-mRNA. The *E. mordax* sequence has one more base in position 111; it also has an identical termination to *E. ringens*, and both are different from *E. encrasicolus* and *E. japonicus*. These last two species have the same coding region of the U2 snRNA. The nucleotide substitution pattern between *E. ringens* sequences and the other three species is almost 50 % transition and 50 % transversion, while nucleotide substitution between *E. encrasicolus*/*E. japonicus* and *E. mordax* is 75 % transition and 25 % transversion.

Table 1 Sizes of clones and their GenBank accession number, for the 4 species studied

Species	Sizes of clones (pb)	GenBank accession number
<i>E. encrasicolus</i>	724–788	KP067917–KP067925
<i>E. mordax</i>	704–760	KP067926–KP067934
<i>E. ringens</i>	819–859	KP067935–KP067943
<i>E. japonicus</i>	728–781	KP067944–KP067951

Most variation in U2 snRNA secondary structure occurs between the IIB and IV helical regions (Fig. 2): at the portion of stem 86–88, the medial portion of helix III 117/141, and the medial portion of helix IV 156–157/175–176. For the paired sequences of *E. encrasicolus* and *E. japonicus*, nucleotide U-117 was paired with nucleotide A-141; whereas in *E. ringens* and *E. mordax*, C-117 was paired with G-141. Furthermore, the nucleotide C-157 for *E. encrasicolus*, *E. japonicus*, and *E. mordax* was paired with nucleotide G-175, while the result for *E. ringens* sequences is A-157/U-175. Covariation of these nucleotides was apparent within the four species' sequences, with double compensatory mutation. Where the base pairs U-156/G-176 appear for *E. encrasicolus* and *E. japonicus*, the base pairs result for *E. ringens* and *E. mordax* is U-156/A-176, with evidence of a single compensatory mutation.

The hairpin IV corresponds to the protein–protein interaction site of *U2A'–U2B'* which has greater polymorphism compared to other protein binding sites. Comparative analysis of this hairpin IV between the four species of this work, considered together, and other species has revealed that practically all have the loop in common (Fig. 3). In the first five base pairs of the stem, the hairpin IV of the species of this study has the nucleotide U-2 paired with A/G-22, which corresponds to a particular pairing for these species, because this position is occupied by the pairing C-G for *D. sargus*, *D. melanogaster*, and *H. sapiens*. The same occurs for the nucleotide A-17, which is only present for *E. ringens* in comparison with the above taxa.

The base pair equivalent to U-2/A-22 in *E. mordax* and *E. ringens* is a U-G base pair in *E. encrasicolus* and *E. japonicus*, and C-G in several other taxa such as *X. laevis*, *M. musculus*, *S. senegalensis*, *G. gallus*, *C. angulata*, *D. sargus*, *H. sapiens*, and *D. melanogaster*. This suggests that the U-G base pair present in *E. encrasicolus* and *E. japonicus* is ancestral within the four species of *Engraulis*, and is retained in *E. mordax* and *E. ringens* by G to A transition.

In respect of the U5 snRNA secondary structure, *E. encrasicolus* and *E. japonicus* have a nucleotide A-4, while *E. mordax* and *E. ringens* have G-4. In the complementary strand, the four species have U-79. The result is an A-U pair in *E. encrasicolus* and *E. japonicus*, and a G-U pair in *E. mordax* and *E. ringens*.

The secondary structure of the U2 snRNA and U5snRNA genes is confirmed as a functional structure because most of the pairs have been subjected to canonical covariation to maintain it.

Using the BioEdit program, a consensus was generated of the snDNA sequences of the four species. The alignment of these sequences revealed that the sequences of *E. encrasicolus* and *E. japonicus* are very similar. With the

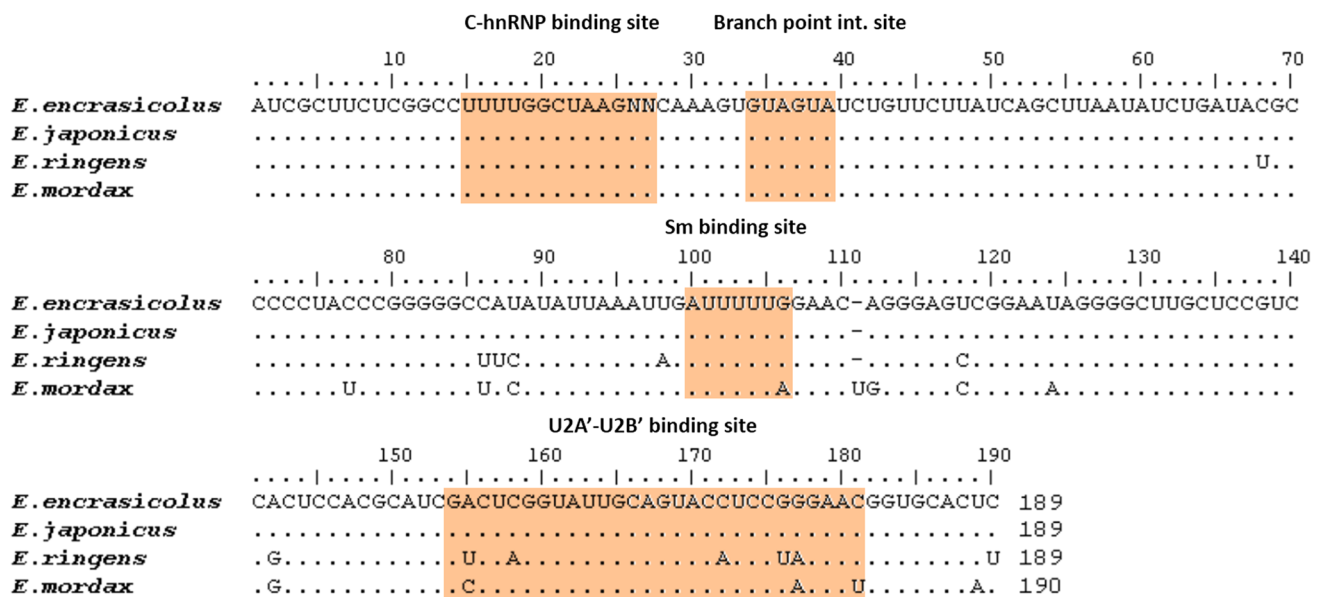


Fig. 1 Alignment of the coding region of the U2 snRNA of the four species studied. Colored boxes indicate protein binding sites, RNA–protein and protein–protein interaction Sm, C-hnRNP, U2A'-U2B', and Branch point introns (Color figure online)

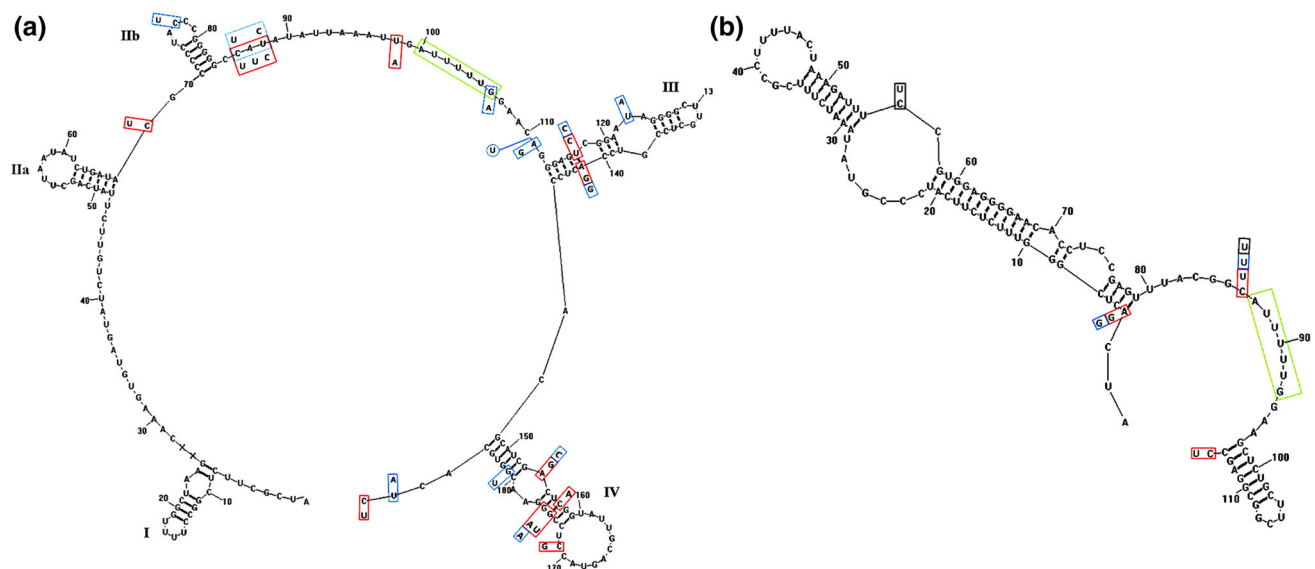


Fig. 2 Conserved snRNA structures and covariation evidence within the four species of *Engraulis*: **a** *Engraulis encrasicolus* and *Engraulis japonicus* secondary structure of U2 snRNA and **b** *Engraulis encrasicolus* secondary structure of U5 snRNA. Red, blue, and black boxes denote nucleotide substitutions in the sequences of *Engraulis*

ringens, *Engraulis mordax*, and *Engraulis japonicus*, respectively. Green boxes highlight the Sm protein binding site. The numbers I, IIa, IIb, III, and IV represent the different hairpin loops. The nucleotide surrounded (U) is present only in *E. mordax* sequences (Color figure online)

same program, the percentage of similarity calculated between the sequences is observed to be 96 % between *E. encrasicolus* and *E. japonicus*, followed by 81 % between *E. encrasicolus* and *E. mordax*, and 71 % between *E. encrasicolus* and *E. ringens*. Analysis of non-transcribed spacers (NTSs) using MEGA 6 revealed a high degree of

homology between *E. encrasicolus* and *E. japonicus*. However, greater evolutionary divergence is observed between these two species, on the one hand, and *E. mordax* and *E. ringens* on the other (Table 2).

The micro-satellite (CTGT)_n is intact in the sequences of the species *E. encrasicolus*, *E. japonicus*, and *E. mordax*

Fig. 3 Secondary structures of the hairpin IV from the four species of this study and other organisms (as indicated). Colored boxes denote phylogenetic comparison of nucleotides comprising the hairpin IV (Color figure online)

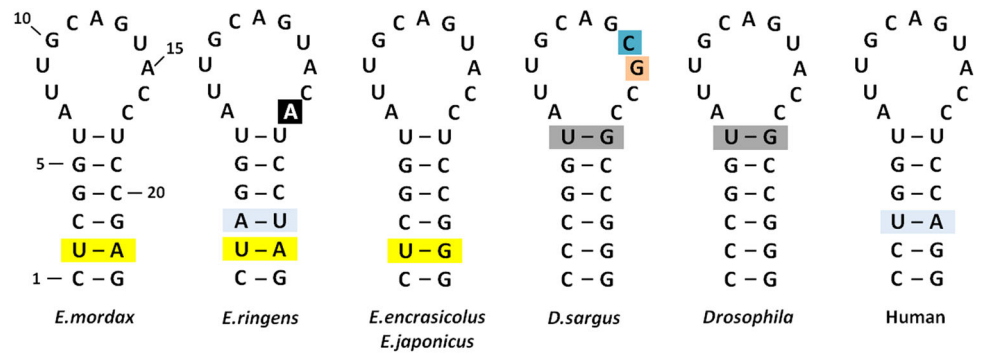


Table 2 Average evolutionary divergence of NTSs between species

	<i>E. encrasicolus</i>	<i>E. japonicus</i>	<i>E. mordax</i>	<i>E. ringens</i>
<i>E. encrasicolus</i>		0.005	0.011	0.014
<i>E. japonicus</i>	0.019		0.011	0.014
<i>E. mordax</i>	0.112	0.113		0.015
<i>E. ringens</i>	0.169	0.167	0.185	

Values, in bold, in lower diagonal, and standard deviation in upper diagonal

but interrupted in the sequences of *E. ringens*. The presence is also revealed of a mini-satellite of 23 bp, which differs in size and number of repetitions between the species, and presents polymorphism in the unit of repetition, like the sequences that precede the units of repetition indicated in Fig. 4.

The mini-satellite occurs in all four species immediately preceding the micro-satellite (CTGT)_n, around bases 227 and 231 (Fig. 5), which suggests that this location of the mini-satellites may have been conserved.

In the NTS of the U2 of all four species, a highly conserved sequence of 115 bp was found by submitting it to the BLASTN (NCBI); this finding was obtained with the sequence of a gene that encodes for the U5 snRNA. The alignment of these sequences shows differences in the termination of the gene for *E. ringens* (Fig. 6). A phylogenetic analysis was also conducted in MEGA 6, using the neighbor-joining method; the reliability of the resulting topologies was then tested using the bootstrap method (Felsenstein 1995). A phylogenetic tree was generated using the sequences of all clones of the U2–U5 snDNA of the four species. The clones of U2–U5 snRNA of *E. encrasicolus* and *E. japonicus* are

```

E. encrasicolus  AATGTTTGCTAAACTGTGATGGG
E. japonicus   .....C.....T....
E. mordax     .....C.....T....
E. ringens    ---.....GA.....-----
    
```

Fig. 4 Consensus sequence of mini-satellites found in the NTSs of the U2 snDNA of *E. encrasicolus*, *E. japonicus*, *E. mordax* and *E. ringens*, and phylogenetic comparison

mixed, confirming the high degree of similarity found between the sequences of U2–U5 snDNA of these 2 species; they also share the same branch with clones of *E. mordax*. The clones of *E. ringens* are, however, well separated from each other (Fig. 7).

Using the Multiple Alignment Construction and Analysis Workbench (MACAW) program (Schuler et al. 1991), many blocks were found to be shared between the sequences of the U2–U5 NTSs of the four species, but only those that are statistically significant have been reported here. A total of four blocks are shown in Fig. 8. The NTSs in the U2–U5 of *E. encrasicolus* and *E. japonicus* share all the blocks found, and in the same order and in the same position; this once again confirms their close genetic proximity. *E. mordax* shares three of these blocks; and *E. ringens* shares three blocks with *E. encrasicolus* and *E. japonicus*, and two blocks with *E. mordax*, in the same position.

Discussion

Study of the U2 snDNA has served two purposes: first, this family of genes has been analyzed and characterized for the first time in a species of pelagic fish, specifically in the important Engraulidae family; and second, a useful tool is provided for studying the relationships between closely related species, strengthening the hypothesis of evolution of the four species studied, given that there have been very few studies of this multigene family in any species of fish. The results obtained, mainly in respect of *E. encrasicolus* and *E. japonicus*, are very similar to those obtained for the

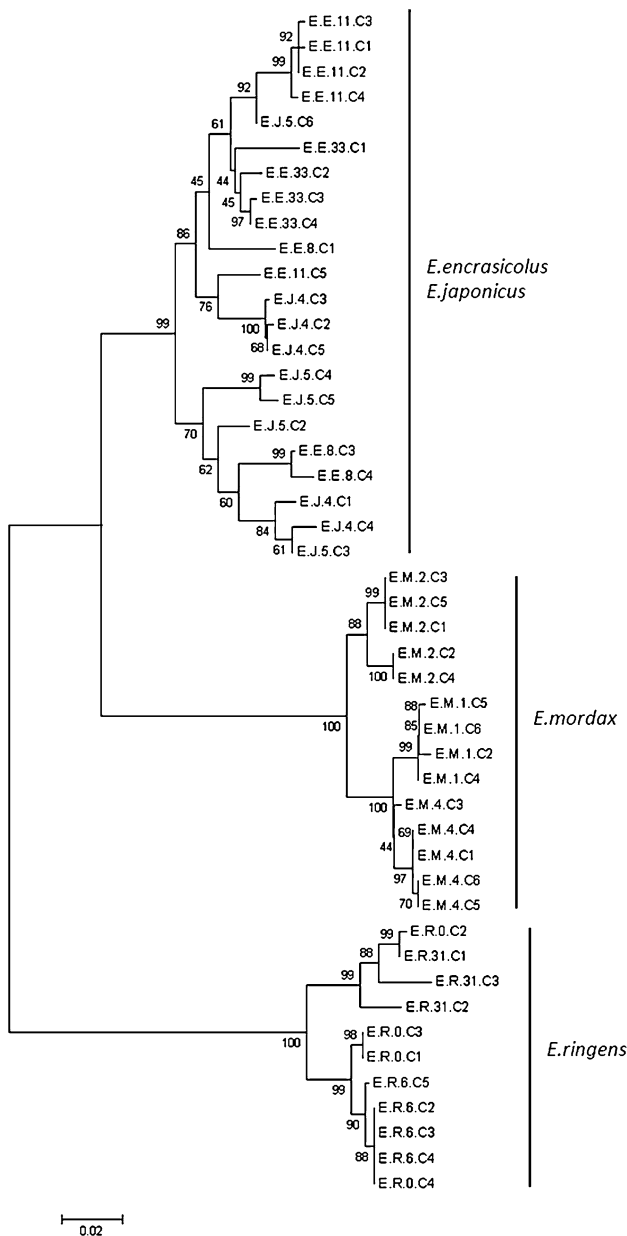


Fig. 7 Phylogenetic tree of all the NTSs of the U2–U5 snDNA of the four species studied. *E.E.* *E. encrasicolus*, *E.J.* *E. japonicus*, *E.M.* *E. mordax*, and *E.R.* *E. ringens*

the repetition unit (Liao and Weiner 1995). Unequal crossing-over and gene conversion have been proposed as mechanisms that could explain the sequence homogenization during concerted evolution of tandemly repeated multigene families (Pavelitz et al. 1995, 1999; Liao et al. 1997; Nei and Ronney 2005).

Several tandemly repeated multigene families contain micro-satellite sequences downstream of the transcribed region, like micro-satellite (CT)_n in the U2 snRNA (Liao and Weiner 1995; Pavelitz et al. 1995; Cross and Rebordinos 2005). A similar micro-satellite has also been found

in the human U1 snRNA (Lund and Dahlberg 1984); the (GT)_n micro-satellite has been found in the 5S rDNA (Little and Bratten 1989); and the same micro-satellite is downstream of the transcription region for 45S rRNA (Dickson et al. 1989).

A study in humans has demonstrated that the micro-satellite (CT)_n stabilizes a long series of tandem integration of the transcription of the U2 snRNA genes (Bailey et al. 1998). Thus, it is suggested that the function of these repeating units in multigene families is to conserve the tandem array (Liao and Weiner 1995). Cross and Rebordinos (2005) have proposed that the micro-satellite (CT)_n may be considered a spacer between the repetition units of the U2 snRNA, rather than a micro-satellite inserted within a repetition unit.

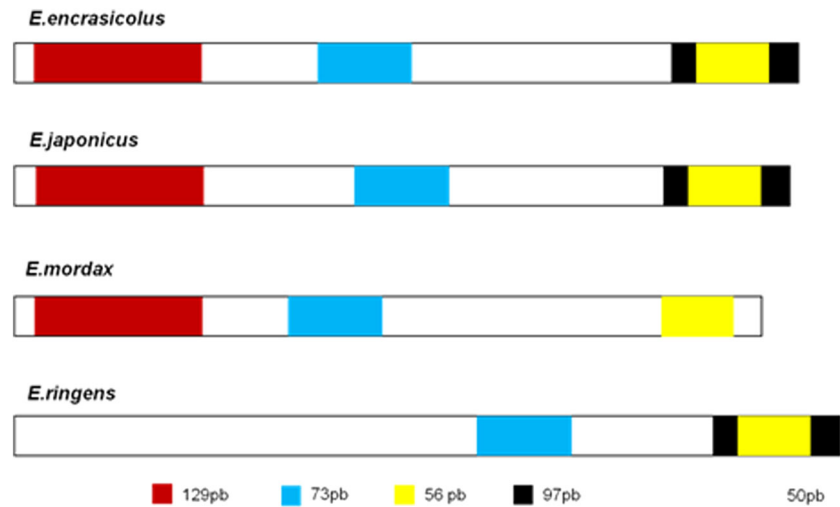
The presence of the mini-satellite sequences downstream of the transcribed region and immediately following the micro-satellite (CTGT)_n in all four of the *Engraulis* species studied may not be random: one function of these units may be to maintain the sequences repeated in tandem. In addition, the repeated sequences of this mini-satellite of 23 bp, which is conserved in *E. encrasicolus*, *E. japonicus*, and *E. mordax*, but which presents the loss of almost half of its sequence in *E. ringens*, demonstrate the evolution of the mini-satellite, since it is known that the mechanisms of concerted evolution produce homogenization and fixation of the new organization, but its loss in other related species (Drouin and Moniz de Sa 1995). These arguments again reinforce the evolutionary relationship of these species.

Correct secondary structure prediction can be used to discriminate between functional genes and pseudogenes, which are often found in eukaryotic genomes (Barciszewska et al. 2001). Phylogenetic analysis, based on the secondary structure, by examining the homologous RNA, confirms regions which maintain equivalent helical pairs despite variations in their primary sequence. Our data suggested that the substitutions occurred among paired bases, a result of selection for compensatory mutations that maintain the secondary structure stable.

The hairpin IV corresponds to a protein–protein interaction site of U2A′–U2B′ which has greater polymorphism compared to other protein binding sites. In several studies with the U2A′ binding site and mutant U2B′, it has been described that at least enough sequence remains so that there is interaction with proteins (Scherly et al. 1990; Boelens et al. 1991).

Although the need for compensatory mutations is simple to understand in terms of maintaining the secondary structure of RNA, the accumulation of pairs of compensatory mutations is more difficult to model in terms of evolution (Hancock et al. 1988). This is because a multigene family like the snRNA contains a variable number of tandem repeat copies that undergo gains and losses of

Fig. 8 Common blocks found in the NTSs region of the U2–U5 snDNA of *E. encrasicolus*, *E. japonicus*, *E. mordax*, and *E. ringens*. Colored blocks indicate their sizes (Color figure online)



sequences stochastically between different species as the result of a concerted evolution.

The evolution of U2 snRNA (the RNU2 locus) has been widely studied in humans and primates (Pavelitz et al. 1995, 1999; Liao et al. 1997, 1998), and the results indicate that this gene evolved according to the concerted model without changes for more than 35 million years (Matera et al. 1990; Pavelitz et al. 1999). In addition to the presence of the (CT)_n micro-satellite sequences in locus RNU2, another repeat unit known as a LTR (long terminal repeat) has been found. It has been suggested that *cis*-acting elements within the U2 snRNA repeat unit may play an important role in the development of the matrix arranged in tandem; and one such possible *cis* element is the micro-satellite (CT)_n embedded within each U2 snRNA repeat unit (Liao and Weiner 1995). It has also been proposed that the main driving force of concerted evolution of tandem U2 genes is intra-chromosomal homogenization (Liao et al. 1997).

Together, these data suggest that the micro-satellite (CTGT)_n, together with the mini-satellite sequences present in the gene coding for the U2–U5 snRNA of the *Engraulis* species, may be *cis*-acting elements involved in the formation and/or maintenance of tandem arrays. The U2 snDNA gene has been found to be linked to the U5 snDNA gene; this seems to be the common situation among the fishes so far studied with this gene family.

Using only certain molecular markers to establish the hypothesis of a relationship between species is considered problematic because the resulting trees can represent the evolutionary history of the genes rather than that of the species. A tree of the gene can differ from a tree of the species due to the duplication of genes and the extinction, hybridization or the horizontal transfer of genes (Maddison 1997). However, several authors have argued that the

analysis of multiple genes in combination will result in a stronger support for the estimated phylogeny of species (Colgan et al. 1998; Manchado et al. 2006b; Merlo et al. 2013).

For example, study of the linkage between U2 and U1 and their NTSs was shown to be useful for differentiating between 6 Soleidae species (Manchado et al. 2006a); and U2 snRNA was used as a marker to study the interrelationships of the major groups of arthropods (Colgan et al. 1998); these examples suggest that the snRNA class can be very useful for phylogenetic purposes.

Our results obtained by the molecular analysis of U2–U5 snDNA with the presence of the micro-satellite (CTGT)_n and the mini-satellites, and analysis of the secondary structure of both these genes, together with our work on the 5S rDNA in the same species of *Engraulis* (Chairi and Rebordinos 2014), confirm the results obtained by Whitehead et al. (1988). These findings make clear that the species *E. encrasicolus* and *E. japonicus* are closely related, and would be older than *E. mordax* and *E. ringens*.

Acknowledgments This study has been supported by grants from the Junta de Andalucía (Spain) to the PAI BIO-219 group.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Arystarkhova E, Donnet C, Asinovski NK, Sweadner KJ (2002) Differential regulation of renal Na, K-ATPase by splice variants of the subunit. *J Biol Chem* 277(12):10162–10172
- Bailey AD, Pavelitz T, Weiner AM (1998) The micro-satellite sequence (CT)_n z (GA)_n promotes stable chromosomal integration of large tandem arrays. *Mol Cell Biol* 18(4):2262–2271
- Barciszewska MZ, Szymański M, Erdmann VA, Barciszewski J (2001) Structure and functions of 5S rRNA. *Acta Bioch Pol* 48:191–198

- Boelens W, Scherly D, Beijer RP, Jansen EJ, Dathan NA, Mattaj IW, Van Venrooij WJ (1991) A weak interaction between the U2A' protein and U2 snRNA helps to stabilize their complex with the U2B'' protein. *Nucleic Acids Res* 19:445–460
- Bon E, Casaregola S, Blandin G, Llorente B, Neuveglise C, Munsterkotter M, Guldener U, Mewes HW, Van Helden J, Dujon B, Gaillardin C (2003) Molecular evolution of eukaryotic genomes: hemiascomycetous yeast spliceosomal introns. *Nucleic Acids Res* 31:1121–1135
- Chairi H, Rebordinos L (2014) A rapid method for differentiating four species of the *Engraulidae* (Anchovy) family. *J Agric Food Chem* 62:2803–2808
- Colgan DJ, McLauchlan A, Wilson GDF, Livingston SP, Edgecombe GD, Macaranas J, Cassis G, Gray MR (1998) Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Austr J Zool* 46(5):419–437
- Cross I, Rebordinos L (2005) 5S rDNA and U2 snRNA are linked in the genome of *Crassostrea angulata* and *Crassostrea gigas* oysters: does the (CT)_n(GA)_n micro-satellite stabilize this novel linkage of large tandem arrays? *Genome* 48:1116–1119
- Dickson KR, Braaten DC, Schlessinger D (1989) Human ribosomal DNA: conserved sequence elements in a 4.3-kb region downstream from the transcription unit. *Gene* 84:197–200
- Drouin G, Sá MMD (1995) The concerted evolution of 5 s ribosomal genes linked to the repeat units of other multigene families. *Mol Biol Evol* 12(3):481–493
- Drouin G, Tsang C (2012) 5S rRNA gene arrangements in protists: a case of non-adaptive evolution. *J Mol Evol* 74:342–351
- Eirín- López JM, Rebordinos L, Rooney AP, Rozas J (2012) The birth- and death evolution of multigene families revisited. *Genome Dyn* 7:170–196
- Eirín-López JM, Ruiz MF, González-Tizón AM, Martínez A, Sánchez L, Méndez J (2004) Molecular evolutionary characterization of the mussel *Mytilus* histone multigene family: first record of a tandemly repeated unit of five histone genes containing an H1 subtype with “Orphon” features. *J Mol Evol* 58:131–144
- Eschmeyer WN (2014) Catalog of fishes: genera, species, references. <http://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp>. Electronic version accessed 12 may 2014
- FAO, The State of World Fisheries and Aquaculture; Fisheries and Aquaculture Department: Rome, 2012
- Felsenstein J (1995) PHYLIP (Phylogeny inference package), version 3.57c. Department of Genetics, SK-50, University of Washington, Seattle, WA
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41:95–98
- Hancock JM, Tautz D, Dover GA (1988) Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Mol Biol Evol* 5:393–414
- Hernandez N (2001) Small nuclear RNA genes: a model system to study fundamental mechanisms of transcription. *J Biol Chem* 276:26733–26736
- Liao D (2000) Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA Genes in bacteria and archaea. *J Mol Evol* 51:305–317
- Liao D, Weiner AM (1995) Concerted evolution of the tandemly repeated genes encoding primate U2 small nuclear RNA (the RNU2 locus) does not prevent rapid diversification of the (CT)_n(GA)_n micro-satellite embedded within the U2 repeat unit. *Genomics* 30:583–593
- Liao D, Pavelitz T, Kidd JR, Kidd KK, Weiner AM (1997) Concerted evolution of the tandemly repeated genes encoding human U2 snRNA (the RNU2 locus) involves rapid intrachromosomal homogenization and rare interchromosomal gene conversion. *EMBO J* 16(3):588–598
- Liao D, Pavelitz T, Weiner AM (1998) Characterization of a novel class of interspersed LTR elements in primate genomes: structure, genomic distribution, and evolution. *J Mol Evol* 46:649–660
- Little RD, Braaten DC (1989) Genomic organization of human 5 S rDNA and sequence of one tandem repeat. *Genomics* 4:376–383
- Lund E, Dahlberg JE (1984) True genes for human U1 small nuclear RNA. *J Bio Chem* 259(3):2013–2021
- Maddison WP (1997) Gene trees in species trees. *Syst Biol* 46(3):523–536
- Manchado M, Rebordinos L, Infante C (2006a) U1 and U2 small nuclear RNA genetic linkage: a novel molecular tool for identification of six sole species (*Soleidae*, Pleuronectiformes). *J Agric Food Chem* 54:3765–3767
- Manchado M, Zuasti E, Cross I, Merlo A, Infante C, Rebordinos L (2006b) Molecular characterization and chromosomal mapping of the 5S rRNA gene in *Solea senegalensis*: a new linkage to the U1, U2, and U5 small nuclear RNA genes. *Genome* 49:79–86
- Matera AG, Weiner AM, Schmid CW (1990) Structure and evolution of the U2 small nuclear RNA multigene family in primates: gene amplification under natural selection? *Mol Cell Biol* 10(11):5876–5882
- Merlo MA, Cross I, Palazón JL, Úbeda-Manzanaro M, Sarasquete C, Rebordinos L (2012) Evidence for 5S rDNA horizontal transfer in the toadfish *Halobatrachus didactylus* (Schneider, 1801) based on the analysis of three multigene families. *BMC Evol Biol* 12:201
- Merlo MA, Cross I, Manchado M, Cárdenas S, Rebordinos L (2013) The 5S rDNA high dynamism in *Diplodus sargus* is a transposon-mediated mechanism. comparison with other multigene families and sparidae species. *J Mol Evol* 76:83–97
- Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York
- Nei M, Rooney AP (2005) Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet* 39:121–152
- Pavelitz T, Rusche L, Matera AG, Scharf JM, Weiner AM (1995) Concerted evolution of the tandem array encoding primate U2 snRNA occurs in situ, without changing the cytological context of the RNU2 locus. *EMBO J* 14(1):169–177
- Pavelitz T, Liao D, Weiner AM (1999) Concerted evolution of the tandem array encoding primate U2 snRNA (the RNU2 locus) is accompanied by dramatic remodeling of the junctions with flanking chromosomal sequences. *EMBO J* 18(13):3783–3792
- Piontkivska H, Rooney AP, Nei M (2002) Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol* 19(5):689–697
- Rebordinos L, Cross I, Merlo A (2013) High evolutionary dynamism in 5S rDNA of fish: state of the art. *Cytogen Genome Res* 141:103–113
- Rooney AP (2004) Mechanisms underlying the evolution and maintenance of functionally heterogeneous 18S rRNA genes in apicomplexans. *Mol Biol Evol* 21(9):1704–1711
- Rooney AP, Ward TJ (2005) Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. *PNAS* 102(14):5084–5089
- Rooney AH, Piontkivska H, Nei M (2002) Molecular evolution of the nontandemly repeated genes of the histone 3 multigene family. *Mol Biol Evol* 19:68–75
- Santaclara FJ, Cabado AG, Vieites JM (2006) Development of a method for genetic identification of four species of anchovies: *e. encrasicolus*, *E. anchoita*, *E. ringens* and *E. japonicus*. *Eur Food Res Technol* 223:609–614
- Scherly D, Boelens W, Dathan NA, van Venrooij WJ, Mattaj IW (1990) Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B and their cognate RNAs. *Nature* 345:502–506

- Schuler GD, Altschul SF, Lipman DJ (1991) A workbench for multiple alignment construction and analysis. *Proteins Struct Funct Genet* 9:180–190
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Vierna J, González-Tizón AM, Marínez-Lage A (2009) Long-term evolution of 5S ribosomal DNA seems to be driven by birth-and-death and selection in *Ensis* razor shells (Mollusca: bivalvia). *Biochem Genet* 47:635–644
- Whitehead PJP, Nelson GJ, Wongratana T (1988) FAO species catalogue. Vol. 7. Clupeoid fishes of the world (Suborder Clupeoidei). An annotated and illustrated catalogue of the herrings, sardines, pilchards, sprats, shads, anchovies and wolf-herrings. Part 2 - Engraulidae. FAO Fish. Synop. 7(125)Pt. 2:579 p
- Yu Y-T, Shu M-D, Narayanan A, Terns RM, Terns MP, Steitz JA (2001) Internal modification of U2 small nuclear (sn)RNA occurs in nucleoli of *Xenopus* oocytes. *J Cell Biol* 152(6):1279–1288