



Trends in evolution of 5S rRNA of deuterostomes: bases and homogeneous clusters

Sandra Maria Rodrigues Subacius*¹ and Romeu Cardoso Guimarães²

¹*Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, Brazil.*

²*Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil.*

Abstract

Evolution of metazoan 5S rRNA sequences was analyzed through base composition and types, location and frequency of clustered bases. Characters from sequences of protostomes did not show regular trends as compared with paleontology dating or organism complexity. Trends of increasing G and C, stronger in G clusters, and decreasing A and U, were detected in deuterostomes, in parallel with evolution of complexity. The multifunctional domain 71-104 was highlighted among conserved stretches. Clusters of C were typical of helices. Those of G were longer, extending from helices into loops or related to bulges, which is suggestive of functional significance. Deuterostomian trends were installed early in the lineage and reached full development in aquatic organisms, not increasing further after reptiles. It can be suggested that ribosomal RNA structures participated in deuterostomian high regulatory complexity, either specifically or as part of the widespread processes of chromosomal regionalization.

Key words: ribosomal RNA, evolution, repetitive sites, deuterostomes, protostomes.

Received: March 5, 2002; accepted: June 17, 2002.

Introduction

Early molecular evolution studies utilized 5S rRNA sequences intensively due to their small sizes and conserved secondary structure, scarcity of modified bases and the availability of a large data bank (Fox *et al.*, 1987). Phylogenies based on 5S rRNA produced inconsistencies (Steele *et al.*, 1991; Halanych, 1991) attributed to conservatism of the molecule and close proximity of interdependent functional domains, which lowered the number of nearly neutral sites. Otherwise, 5S rRNA remained informative for wide range evolutionary studies (Gomes *et al.*, 1985; Guimarães and Erdmann, 1992) and for understanding its structure and functions (Digweed *et al.*, 1986; Guo-Rong *et al.*, 1988; Subacius, 1994; Guimarães *et al.*, 1997; Rezek-Ferreira, 1997).

Homogeneous clusters of nucleotides are interesting hot-spots for recombinatory or replicative slippage and as sites with peculiar interactions: redundancy lowers infor-

mational content of sequences but, on the other hand, would guarantee availability for binding of a base which may be slightly displaced, that is, flexibility with positional ambiguity. The frequency of adenine clusters was shown to increase in organelles and mycoplasmas, decreasing in lineages from simple to complex organisms (Guimarães and Erdmann, 1992).

A survey of all types of clusters in 5S rRNA of metazoans is reported here. Clear trends originating early in the deuterostomes lineage were depicted.

Material and Methods

The Berlin RNA Databank as of August 1987 (BRDb, containing 513 molecules of 5S rRNA; see Guimarães and Erdmann, 1992) was used, with some additions of sequences from metazoans (Rice *et al.*, 1993; see also Szymanski *et al.*, 1997). Sequences from deuterostomes totaled 23 in the 1987 database (18 aquatic and 5 terrestrial organisms, the latter from reptiles to mammals) and 34 in the 1993 database (samples from aquatic organisms increased to 29, being 25 from somatic tissues [echinoderms to mammals] and nine from oocytes [cyclostomes to amphibians]). There were 49 sequences from protostomes (21 arthropods), plus 7 from cnidarians. Four pseudogene

*Present address: Faculdade de Ciências Biológicas, Exatas e Experimentais, Universidade Presbiteriana Mackenzie, São Paulo, SP. E-mail: ear_@uol.com.br.

Send correspondence to R.C. Guimarães. Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil. E-mail: romeucg@icb.ufmg.br.

sequences from amphibians were included only for mutation analyses.

Frequencies of nucleotides (N) and homodinucleotides (DiN) were calculated relative to the total number of nucleotides of a molecule. When the repetition of a base was >2 (TriN or longer), the number of DiN in the cluster was total size minus 1. This way of counting DiN in homogeneous clusters considers a unit to be equivalent to the region between two identical bases, with partial overlap of contiguous DiN.

For graphical comparisons, averages of each taxon were normalized into standard deviation units (SDU) through the equation (Z-score): $Z = (X - X_g) / SD_g$, where X = average of a taxon, X_g = average of the whole sample, SD_g = SD of the whole sample. For any variable, X_g became the zero baseline and taxon averages were plotted in SDU.

Regression analyses were carried out with non-transformed variables and included all gene sequences. Paleontology dating corresponded to the most remote fossil evidence of organisms phenotypically similar to their modern analogues (Subacius, 1998).

Positions refer to individual bases or the first base of a DiN, according to the eukaryote model of Szymanski *et al.* (1997). Location of clusters considered the three types of secondary structure; borders indicated extension of clusters from helices into loops. Frequency of DiN in a position, among molecules in a group, was a measure of the variability of the character. Frequency and types of base changes were computed per molecule of deuterostomes *versus* the molecule of mammals; no size alteration occurred. Bases are presented in the hydrophathy order (Guimarães, 1998).

Results

Protostomes and deuterostomes

Among all molecules in the databank, G and C were the most abundant bases, U and A less frequent. Most of the occurrences of U and A were isolated while about half of the content of G and C occurred in homogeneous clusters. The correlation indicated above was not maintained in the G and C couple, the former being more abundant along the whole molecule, but C clustered more frequently than G (Table I). Among metazoans, deuterostomes showed highest levels of G and C clusters (Figures 1 and 2). A temporal trend in base and cluster composition was detected in the deuterostomes lineage (Figure 1). Protostomes presented an earlier paleontology record and contained the better sampled arthropod lineage, but depicted no trend, showing variations around the baseline (Figure 2). Few unbalanced molecules were noted in platyhelminths and tentaculates, with high U and low DiC. In deuterostomes, A and U decreased, and G and C increased with paleontology dating and organism complexity, but changes were most intense in

Table I - Isolated and clustered bases in 5S rRNA.

	Total	Isolated (%)	Dinucleotides
A	27.13 ± 3.39	16.32 ± 2.59 (60.15)	6.05 ± 2.13
G	34.67 ± 3.58	17.47 ± 2.99 (50.39)	10.10 ± 3.06
C	32.00 ± 3.96	13.55 ± 2.52 (42.34)	10.48 ± 2.76
U	25.14 ± 4.21	16.97 ± 3.07 (67.50)	4.32 ± 2.48
Total	118.94	64.31 (54.07)	30.95

Dinucleotides are total per molecule, counted as in Material and Methods. Average ± standard deviation in the whole sample of 513 molecules.

G than in other bases. DiG contents increased up to reptiles, thereafter becoming constant. DiC increased up to cyclostomes, somewhat further into amphibians but decreased slightly afterwards. Decreases in DiU were more regular than those in DiA. Continuous increase in isolated G was depicted by birds and mammals, after DiG had stabilized.

Transitions were indicated, from hemi- and urochordates to cyclostomes and fishes, and from these to the reptile emergence after amphibians (Figure 1). These and other species-specific variations, however, did not produce large deviations from the linear fit between purine base composition, throughout the entire range of data, or DiG, in the range up to reptiles, and paleontology dating. Increases in DiG were steeper than in G (Figure 3).

Location of G and C clusters in deuterostomes

Clusters of C were typical of helices but those of G usually extended into more complex arrangements (Figure 4, Table II). Each type of cluster occurred 11 times, but the former were detected only once at a border (helix α I-loop A) and once in loop C, this being only in aquatics, and variably. The latter were longer (3 TriC, 6 TriG) and involved with the borders of loops A and E (TriGs with invariant Gs), the border of loop D (TriG), and all bulges: bulge 49-50 was inserted in a TriG, and bulges 63 and 83 were contiguous to a TriG and a DiG, respectively. These sites were fixed either in all deuterostomes or in terrestrials; loop B was the only one generally free of G or C clusters. Contiguity of different types of clusters generated mixed G and C dinucleotides, which were more frequently CG (7 sites: 29, 30, 40, 46, 47, 69, 105) than GC (2 sites: 8, 66).

Molecules from terrestrials were nearly homogeneous in G and C cluster sites. Only one site was variable (DiC 2 at helix α I), while such variability was abundant in aquatics. Excess of G over C clusters in terrestrials was concentrated in helices α I, β II and γ V (Table II).

Base substitutions in deuterostomes

Comparing sequences of earlier deuterostomes with those of mammals (Table III), the least changes were shown by G sites, a characteristic also observed in

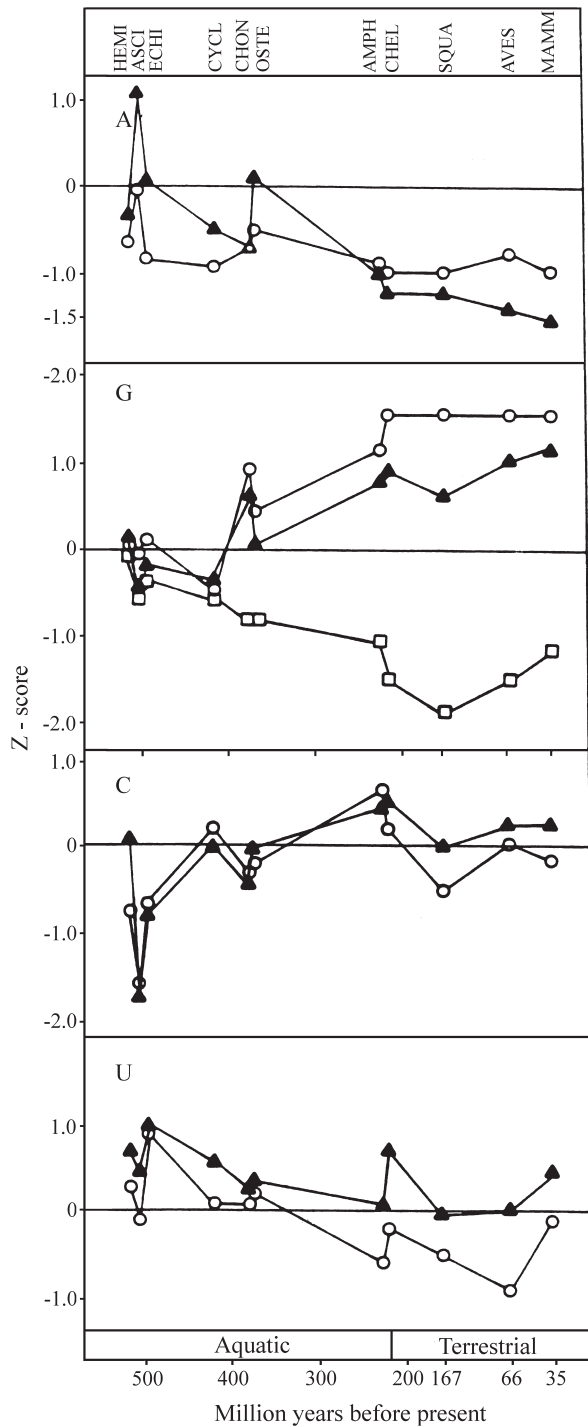


Figure 1 - Frequency of bases and dinucleotides in the 5S rRNA of deuterostomes vs. paleontology dating. Total of a base (Δ), isolated G (\square); dinucleotides (O) and z-scores calculated as in Material and Methods. HEMIchordates, ASCIdian, ECHI noderms, CYCLOstomes, CHONdri- chians, OSTEichthians, AMPHibians, CHEL onian, SQUA mate, MAM- Malians.

pseudogenes. Gene sequences showed a higher rate (10x) of A \rightarrow G transitions than pseudogenes (7 x) and reversed the trend of higher C \rightarrow U than U \rightarrow C transitions of pseudogenes. Addition of higher rates of C \rightarrow G and U \rightarrow G

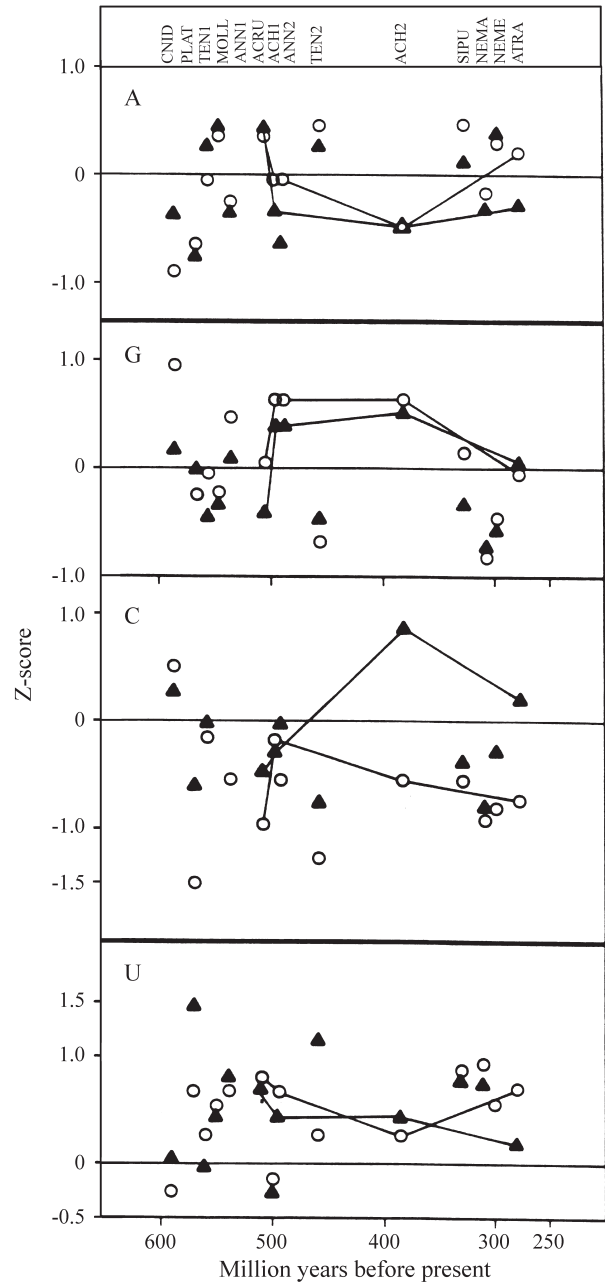


Figure 2 - Frequency of bases and dinucleotides in the 5S rRNA of protostomes vs. paleontology dating. Total of a base (Δ); dinucleotides (O) and z-scores calculated as in Material and Methods. The ARTHropod lineage (ACRUstaceans, ACHelicerates [ACH1 = Merostomates, ACH2 = Arachnids], ATRAcheids) is united by lines. CNIDarians, PLATyhel- minths, TENTaculates (TEN1 = Brachiopods, TEN2 = Bryozoans), MOL- Luses, ANNelids (ANN1 = Polychaetes, ANN2 = Oligochaetes), SIPUnculid, NEMAthelminths, NEMertine.

transversions in somatic molecules to their A \rightarrow G transitions made them enriched in G and C, while oocyte molecules were only enriched in C.

Conserved regions were concentrated in the core of helix α 1, parts of the complex helix β III, all loops but one side of B, all bulges but site 63, and the whole region 71-104 of domain γ (Figure 5).

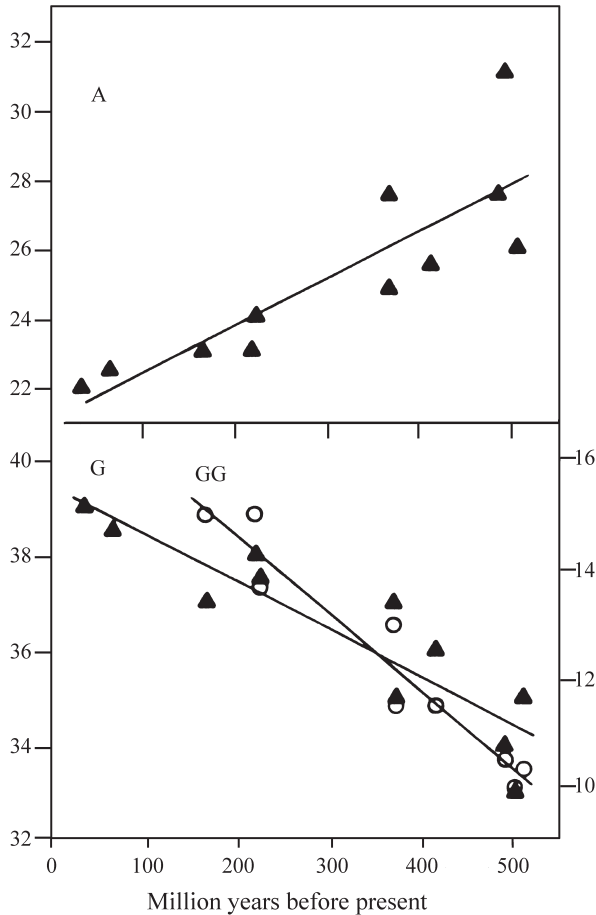


Figure 3 - Trends in purine base composition and in guanine dinucleotide frequency in the 5S rRNA of deuterostomes. Linear regression of group averages vs. paleontology dating. Dinucleotide data beyond squamates not utilized, due to having become constant (see Figure 1). $A = 21.070 + 0.013 x$, $r = 0.848$; $G (\blacktriangle, \text{left ordinate}) = 39.453 - 0.010 x$, $r = -0.904$; $GG (O, \text{right ordinate}) = 17.521 - 0.014 x$, $r = -0.966$.

Discussion

Study of the evolution of adenine clusters in 5S rRNA informed on their concentration in loops, abundance in simpler organisms, increase in simplified organelles and mycoplasmas, and decrease in all routes of increased organism complexity (Guimarães and Erdmann, 1992). The necessary complement of other base clusters, now presented in metazoans, highlighted the regular increase of G and C clusters, in contrast with those of A and U clusters.

The already low number of bases could be halved into simple indices such as the G + C contents, based on the long known parallelism between complementary bases, largely necessary for structuring helical segments, or into the purine / pyrimidine balance (Subacius and Bussab, 1998), possibly necessary also for maintaining adequate sequence complexity. Such indicators might be interesting for studies of heterogeneous sequences or very wide range evolution. Nonetheless, these were considered too rough for the pres-

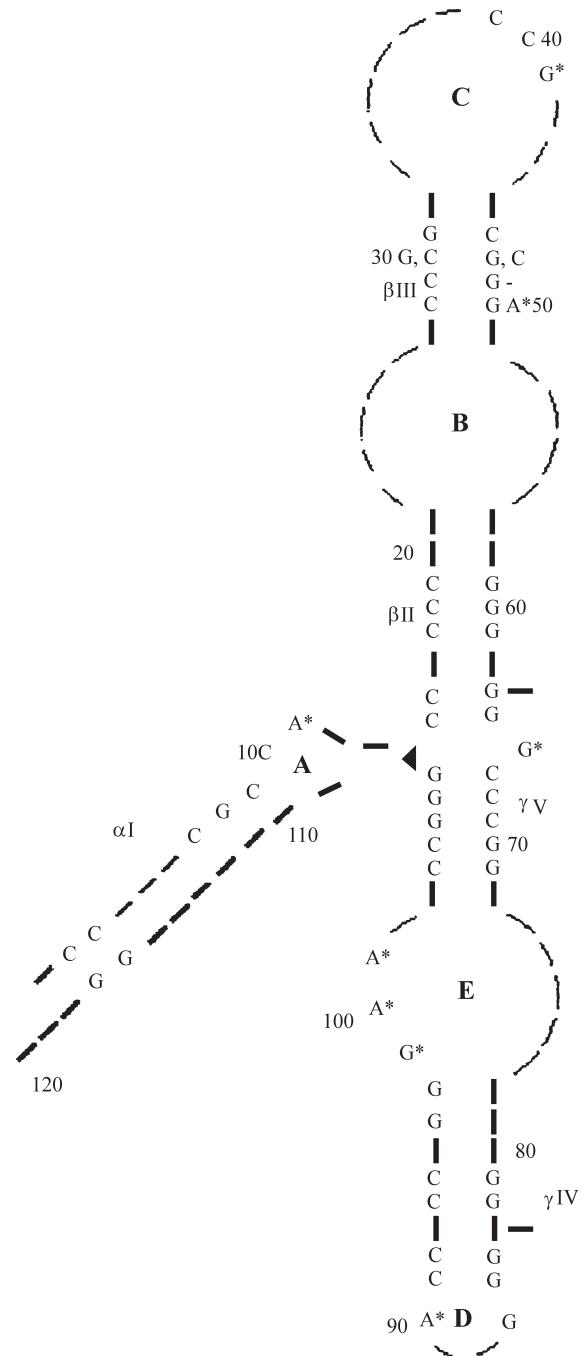


Figure 4 - Location of G and C clusters along the secondary structure of 5S rRNA in deuterostomes. *, invariant positions. Bases 49 and 50, 63, and 83 are bulged. The complex loop A is formed by bases 10-13, 66 and 109. Drawing according to the model of Szymanski *et al.* (1997).

ent purpose of describing details in a small RNA sequence, where single stranded stretches are abundant, and inside one division of metazoans. Accordingly, changes in G were singled out among all bases.

Variations among protostomes did not depict inter- or intragroup regularities, being distributed around the overall

Table II - Location and frequency of G and C clusters in aquatic and terrestrial deuterostomes.

Position	Aquatic	Terrestrial
Helix α I		
DiC 2:	53	37
DiG 116	65	100
DiG 7	65	100
DiC 9 (border loop A)	100	100
Helix β II		
DiC 14:	75	100
TriG 64 (border loop A, bulge 63)	78	100
65	87	100
TriC 17:	53	100
TriG 59	56	100
DiG 21 (border loop B, apomorphic in aves)		
Helix β III		
TriC 28	100	100
29:	63	100
TriG 47 (inserted bulge 49-50)	56	100
48	100	100
DiG 30:	34	0
DiC 46	35	0
DiC 39 (loop C)	53	0
Helix γ V		
TriC 67	79	87
68:	14	0
TriG 106	18	87
107	100	100
DiG 70:	87	100
DiC 104	87	87
Helix γ IV		
DiC 78 (apomorphic in amphibian oocytes)		
DiG 81: (border bulge 83)	91	100
DiC 94	91	100
TriG 85: (border loop D)	100	100
DiC 91	91	87
TriG 97 (border loop E)	100	100

Frequency (%) among molecules in the group; ±, paired clusters; see also Figure 4 for locations.

baseline. These can be described as non-biased and compensated as to base composition, and cluster sites were found distributed as apomorphisms (not shown). Deuterostomian complexification was paralleled by increases in G and C clusters. Both were typical of helices but G clusters were related to complex features of the molecule such as helix-loop borders and bulge sites. It was indicated that loops remained the most complex segments of the molecule, due to their scarcity in all types of clusters and wide conservatism.

Functional derivation from RNA structures, in the context of nucleoprotein complexes, should consider their interdependence with proteins. Data on deuterostomes

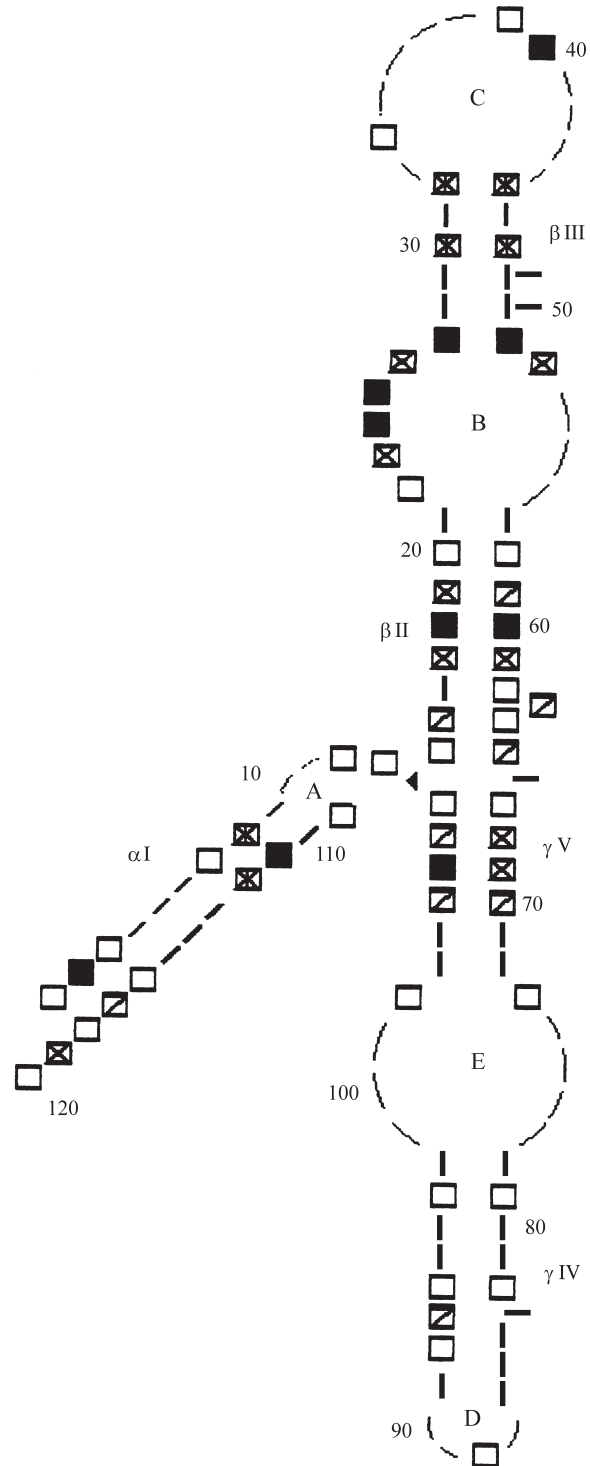


Figure 5 - Location and frequency of bases different in other deuterostomes from the 5S rRNA of mammals. Frequency (%) per molecule: -, zero; □, 0.04 - 0.08; ✓, 0.13 - 0.17; x, 0.21 - 0.29; ⊠, 0.33 - 0.38; ■, >0.42.

would indicate evolution towards stronger participation of RNA in helix stabilization and for protein recognition and binding, while those on protostomes indicated that endogenous RNA helix thermal stability was not so important for

Table III - Base substitutions in the 5S rRNA of deuterostomes.

Non-mammalian		Mammalian				Total
		A	G	C	U	
Somatic	A	-	1.27	0.46	0.24	1.97
	G	0.13	-	0.30	0.02	0.45
	C	0.02	0.26	-	0.67	0.95
	U	0.21	0.24	1.10	-	1.55
	Total	0.36	1.77	1.86	0.93	4.92
Oocytic	A	-	0.81	0.18	0.06	1.05
	G	0.06	-	0.06	0.06	0.18
	C	0.02	0.04	-	0.83	0.89
	U	0.06	0.02	0.94	-	1.02
	Total	0.14	0.87	1.18	0.95	3.14
Pseudogene	A	-	3.50	1.75	0	5.25
	G	0.50	-	0	0.75	1.25
	C	0.50	0.50	-	4.25	5.25
	U	1.00	0.25	3.00	-	4.25
	Total	2.0	4.25	4.75	5.00	16.00

Differences, per molecule, from non mammalian to the mammalian. Rows read bases lost from non-mammalians, columns bases gained in the molecule of mammals. Transitions bold.

guaranteeing functionality, which could be helped by proteins whenever necessary.

Early installation of the deuterostome trend of increased G and C cluster content, already fully developed in aquatics, should be related to an intrinsic property of the lineage, as opposed to environmental influences. Forces behind the process would point to their increasing loss of egg determinations and the corollary dependence on intercellular induction (Raff and Kaufman, 1983; Kauffman, 1993; Gerhart and Kirschner, 1997), which should be at the root and basis for further complexification. It is indicated that 5S rRNA interactions with proteins and the transcription and translation machinery reflected and participated fully in the organism trends.

Localization of 5S rRNA in early replicating chromosomal regions, together with the clustering of most housekeeping genes (Gilbert, 1986; Guinta *et al.*, 1986; Holmquist, 1987), would be a mechanism enabling its acquisition of G and C, which are supposedly not among the most abundant in nucleotide pools (see Guimarães and Erdmann, 1992).

If, as indicated elsewhere (Bernardi *et al.* 1985 ; Bernardi and Bernardi, 1991; and many others), endothermy participated in further enhancement of the trend, leading to the clear demarcation of chromosomal heavy isochores, it would have been a superimposed late process.

The transition for depicting chromosomal regionalization in endotherms would not be well represented in rRNA due to its early response to the basic complexification forces and, especially in the case of 5S rRNA, to its

precocious filling with G and C sites (isolated or clustered), as shown by the high and stabilized DiG content already in reptiles. The 5S rRNA might also be peculiar, as indicated by its enrichment in CG sites, as opposed to GC sites, which is contrary to the genome tendency, found in eukaryotes, of reducing CG doublets.

The long conserved stretches depicted by our mutation analyses might contribute to the mapping of deuterostomian multifunctional overlapping domains, since they corresponded approximately to some previous determinations (Braun *et al.* 1992; Lee *et al.*, 1995): loop B (53-56) and the invariant A of bulge 50, included in the TriG 47, matched the region of the internal promoter (50-83) and the TFIIIC binding site (box A, 50-61); region 71-104 matched precisely the Neurospora box C (73-103), which was found shorter in other eukaryote genes (yeast 81-94, human 80-97, Drosophila 'internal control region IV' 78-98) but longer in the RNA binding site for TFIIIA (53-105); the bacterial U89 of loop D was found to interact with C2475, at the peptidyl-transferase site of 23S rRNA (Dokudovskaya *et al.*, 1996).

Acknowledgments

CNPq and FAPEMIG support to RCG; SMRS is specially grateful to Dr. Gabriel Padilla (ICB-USP) and FAPESP for the post-doctoral fellowship.

References

Bernardi G and Bernardi G (1988) Compositional constraints and genome evolution. *J Mol Evol* 24:1-11.

- Bernardi G and Bernardi G (1991) Compositional properties of nuclear genes from cold-blooded vertebrates. *J Mol Evol* 33:57-67.
- Braun BR, Bartholomew B, Kassavetis GA and Geiduschek EP (1992) Topography of transcription factor complexes on the *S. cerevisiae* 5S RNA gene. *J Mol Biol* 228:1063-1077.
- Digweed M, Pieler T and Erdmann VA (1986) A comparative analysis of structural dynamics in 5S rRNA. In: Van Knippenberg PH and Hilbers CW (eds). *Structure and Dynamics of RNA*. Plenum Press, New York, pp 205-217.
- Doktydovskaya S, Dontsova O, Shpanchenko O, Bogdanov A and Brimacombe R (1996) Loop IV of 5S ribosomal RNA has contacts both to domain II and to domain V of the 23S rRNA. *RNA* 2:146-152.
- Fox GE, Stackebrandt RB, Hespel RB *et al.* (16 co-authors) (1987) The phylogeny of prokaryotes. *Science* 209:457-463.
- Gerhart J and Kirschner M (1997) *Cells, embryos and evolution*. Blackwell, Malden MA, USA.
- Gomes E, Guimarães RC, Padovani CA and Bloch DP (1985) The universal dinucleotide asymmetry rule does not hold for the 5S rRNA. *Arq Biol Tecnol* 28:3.
- Guimarães RC (1998) Genetic code: hydrophobic, dinucleotide type and aminoacyl-tRNA synthetase class organization. In: Chela-Flores, J and Raulin F (eds) *Exobiology - matter, energy and information in the origin and evolution of life in the universe* Kluwer, Dordrecht, pp 157-160.
- Guimarães RC and Erdmann VA (1992) Evolution of adenine clustering in 5S rRNA. *Endocyt. Cell Res* 9:13-45.
- Guimarães RC, Trifonov EN and Lagunez-Otero J (1997) Taxonomy of 5S ribosomal RNA by the linguistic technique: Probing with mitochondrial and mammalian sequences. *J Mol Evol* 45:271-277.
- Guinta DR, Tso JY, Narayanswami S, Hamkalo BA and Korn LJ (1986) Early replication and expression of oocyte-type 5S RNA genes in a *Xenopus* somatic cell lineage carrying a translocation. *Proc Nat Acad Sci USA* 83:5150-5154.
- Guo-Rong Q, Gong-Jie C, Peng J, Xiao-Li F and Xiang-Rong G (1988) Studies on the sites expressing evolutionary changes in the structure of eukaryotic 5S ribosomal RNA. *J Mol Evol* 27:336-340.
- Halanych KM (1991) 5S ribosomal RNA sequences inappropriate for phylogenetic reconstruction. *Mol Biol Evol* 8:249-253.
- Holmquist GP (1987) Role of replication time in the control of tissue-specific gene expression. *Am J Hum Gen* 40:151-173.
- Kauffman SA (1993) *The origins of order. Self-organization and selection in evolution*. Oxford University Press, Oxford, UK.
- Lee Y, Erkin AM, VanRyk DJ and Nazar RL (1995) *In vivo* analyses of the internal control region in the 5S rRNA genes from *S. cerevisiae*. *Nucl Acids Res* 23:634-640.
- Raff RA and Kaufman TC (1983) *Embryos, genes and evolution*. Macmillan, New York.
- Rezek-Ferreira RH (1997) *Análise linguística e taxonômica do RNAr 5S de fungos*. PhD Thesis, Universidade Federal de Minas Gerais, Belo Horizonte - MG, Brasil.
- Rice CM, Fuchs R, Higgins DG, Stoehr PJ and Cameron GN (1993) The EMBL data library. *Nucl Acids Res* 21:2967-2971.
- Steele KP, Holsinger KE, Jansen RK and Taylor DW (1991) Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. *Mol Biol Evol* 8:240-248.
- Subacius SMR (1994) *Evolução do RNAr 5S*. PhD Thesis, Universidade de São Paulo, São Paulo - SP, Brasil.
- Subacius SMR and Bussab WO (1998) Purine and pyrimidine composition in 5S rRNA and its mutational significance. *Genet Mol Biol* 21:255-258.
- Subacius SMR (1998) Dating the emergence of the deuterostomes during the Proterozoic Eon based on 5S rRNA. *An Acad Bras Ci* 70:491-500.
- Szymáński M, Barciszewska MZ, Barciszewska J, Specht T and Erdmann VA (1997) Compilation of ribosomal 5S ribonucleic acid nucleotide sequences: eukaryotic 5S rRNAs. *Biochim Biophys Acta* 1350:75-79.