J Mol Evol (1999) 49:798–805

JOURNAL OF MOLECULAR EVOLUTION

Analysis of the Primary Sequence and Secondary Structure of the Unusually Long SSU rRNA of the Soil Bug, *Armadillidium vulgare*

Chong Pyo Choe,[1] John M. Hancock,[2] Ui Wook Hwang,[1,*] Won Kim[1]

[1] Department of Molecular Biology, College of Natural Sciences, Seoul National University, Seoul 151-742, Korea
[2] MRC Clinical Sciences Centre, Imperial College School of Medicine, Hammersmith Hospital, London, UK

Received: 2 March 1999 / Accepted: 22 July 1999

Abstract. The complete nucleotide sequence of the SSU rRNA gene from the soil bug, *Armadillidium vulgare* (Crustacea, Isopoda), was determined. It is 3214 bp long, with a GC content of 56.3%. It is not only the longest SSU rRNA gene among Crustacea but also longer than any other SSU rRNA gene except that of the strepsipteran insect, *Xenos vesparum* (3316 bp). The unusually long sequence of this species is explained by the long sequences of variable regions V4 and V7, which make up more than half of the total length. RT-PCR analysis of these two regions showed that the long sequences also exist in the mature rRNA and sequence simplicity analysis revealed the presence of slippage motifs in these two regions. The putative secondary structure of the rRNA is typical for eukaryotes except for the length and shape variations of the V2, V4, V7, and V9 regions. Each of the V2, V4, and V7 regions was elongated, while the V9 region was shortened. In V2, two bulges, located between helix 8 and helix 9 and between helix 9 and helix 10, were elongated. In V4, stem E23-3 was dramatically expanded, with several small branched stems. In V7, stem 43 was branched and expanded. Comparisons with the unusually long SSU rRNAs of other organisms imply that the increase in total length of SSU rRNA is due mainly to expansion in the V4 and V7 regions.

Key words: Soil bug — *Armadillidium vulgare* — Crustacea — Small-subunit ribosomal RNA — Secondary structure — Variable regions — Increase in length — GC content — Slippage motif

Introduction

The small-subunit rRNA (SSU rRNA) gene consists of regions with varying degrees of sequence divergence, divided into the slow-evolving conserved regions and fast-evolving variable regions. Because of these features, SSU rRNA has been used extensively as a molecular chronometer to infer the evolutionary relationships among both closely and remotely related taxa. Though SSU rRNA genes of metazoan animals usually range from 1800 to 1900 bp, showing small size variations, unusually long SSU rRNA genes have been reported so far from some insect taxa such as *Aonidiella aurantii* (Campbell et al. 1994), *Acyrthosiphon pisum* (Kwon et al. 1991), *Pealius kelloggii* (Campbell et al. 1994), *Stylops melittae* (Chalwatzis et al. 1995), and *Xenos vesparum* (Chalwatzis et al. 1995) and a crustacean water flea, *Daphnia pulex* (Crease and Colbourne 1998). Among these, the longest SSU rRNA is found in the strepsipteran insect, *X. vesparum* (3316 bp). It is generally recognized that the elongation of rRNA genes is caused mainly by sequence insertions, which may take place within conserved or variable regions. However, sequence insertions in the conserved regions are restricted to introns (Bhattacharya et al. 1994; DePriest and Been 1992; De Wachter et al. 1992; Sogin and Edman 1989; Wilcox et

Present address: Department of Parasitology, Yonsei University College of Medicine, Seoul 120-752, Korea; e-mail:uwhwang@yumc.yonsei.ac.kr
Correspondence to: Dr. Won Kim; *e-mail:* wonkim@plaza.snu.ac.kr

al. 1992) and therefore elongations are not seen in the mature rRNA. In contrast, elongation in variable regions of rRNA genes occurs in the mature rRNA and causes the variable regions to be greatly expanded (Kwon et al. 1991; Campbell et al. 1994; Hinkel et al. 1994; Chalwatzis et al. 1995; Crease and Colbourne 1998).

While sequencing extensively the representative crustacean peracaridan taxa (such as amphipods and isopods) to study their evolutionary relationships, we found that the soil bug, *Armadillidium vulgare*, has an extremely long SSU rRNA gene. Here we present the complete nucleotide sequence of *A. vulgare* SSU rRNA and its putative secondary structure. We also analyze the primary sequence and secondary structure and discuss some features specific to *A. vulgare* compared with those of other eukaryotes as well as the crustacean water flea, *Daphnia pulex.*

Table 1. The primers used for sequencing the SSU rDNA

	Forward	
328	5′-TACCTGGTTGATCCTGCCAG-3′	1–20
B	5′-ACGGGTAACGGGGAATCA-3′	414–431
A	5′-AGGGCAAGTCTGGTGAAG-3′	612–630
V4B	5′-AAAGGGGGTCGTCGGGAGG-3′	792–810
V4D	5′-GCGTCGTCCCCTCTCTTCCTC-3′	1069–1089
G	5′-GAGGTGAAATTCTTGGAC-3′	1956–1973
13	5′-GGGTGGTGGTGCATGGCCGT-3′	2324–2343
16	5′-AACGAGGAATTCCCAGTA-3′	2978–2995
	Reverse	
B	5′-ATTCCCCGTTACCCGT-3′	414–429
A	5′-CTTGCACCAGACTTGCCCT-3′	612–630
V4F	5′-CGCGAATCGCTTCACCCGCG-3′	1335–1353
V4E	5′-GAGGGGGAAGAGGGAGAGGC-3′	1600–1610
V4C	5′-CTGTTCCATCATTCCATGCG-3′	1848–1867
G	5′-GTCCAAGAATTTCACCTC-3′	1956–1973
13	5′-ACGGCCATGCACCACCACCC-3′	2324–2343
D	5′-TCTAAGGGCATCACAGACCTG-3′	2832–2852
329	5′-TAATGATCCTTCCGCAGGTT-3′	3195–3214

Materials and Methods

DNA Isolation, PCR, and Sequencing

Genomic DNA was isolated from one frozen individual of *A. vulgare* according to Gustincich et al. (1991). The extracted genomic DNA was dialyzed in 1× TE to remove pigments that may prevent the subsequent PCR reaction. Concentration and quality of the isolated genomic DNA were analyzed on 1% agarose gels. The SSU rRNA coding regions were selectively amplified using PCR with two primers that recognize conserved sequences proximal to the 5′ and 3′ termini (Nelles et al. 1984) (1–19, 5′-CCTGGTTGATCCTGCCAG-3′, 1848–1868, 5′-TAATGATCCTTCCGCAGGTTA-3′: the numbers indicate the positions of human SSU rDNA; EMBL accession number M10098). Approximately 50 ng of genomic DNA as template, a 0.2 m*M* concentration of dNTPs, a 20 pM concentration of each primer, 2.5 m*M* $MgCl_2$, 1× reaction buffer (Promega Co.), and 2.5 U of Taq DNA polymerase (Promega Co.) were used in a 50-μl amplification reaction. Cycling was carried out in a Thermo Cycler (Perkin Elmer Co.): 2 min at 94°C for initial denaturation, then 10 cycles with 10 s at 94°C, 30 s at 50°C for annealing, and 4 min at 68°C for elongation, 20 cycles with 10 s at 94°C, 30 s at 50°C, 4 min at 68°C, 30 s at 4°C, and pause for 20 s. The amplified products were examined on a 1% agarose gel and then purified with a PCR product purification kit (QIAGEN Co). The purified PCR products were inserted into the T-easy vector system (Promega Co.), then the ligased plasmid was transformed into the *E. coli* strain DH5α. Double-stranded recombinant plasmid DNA was purified using a plasmid purification kit (QIAGEN Co.). Sequencing was performed with the Big Dye Terminator kit (Perkin–Elmer Co.) and analyzed on an Applied Biosystems 310 Automated Sequencer. The primers used to sequence the PCR products are listed in Table 1. Alignment was carried out using the CLUSTAL X program applying default settings (Thompson et al. 1997). And then it was adjusted with eyes considering the SSU rRNA secondary structure.

RNA Isolation and RT-PCR Analysis

Total cellular RNA was isolated from one frozen individual of *A. vulgare* using Trizol (GibcoBRL Co.). RT-PCR was performed with approximately 20 ng of total RNA as template using the Titan one-tube reverse transcriptase (RT)–PCR system (Boehringer Mannheim). Reactions were carried out in a total volume of 25 μl with a 0.2 m*M* concentration of each dNTP, a 20 pM concentration of each primer, 1× Titan reaction buffer, and 1 U of RT-PCR enzyme mix. To generate cDNA, the sample was incubated at 50°C for 30 min before amplification. The subsequent PCR reaction was performed under the following condition: 10 cycles with 30 s at 94°C, 30 s at 55°C, and 45 s at 68°C followed by 20 cycles with 30 s at 94°C, 30 s at 60°C, and 50 s at 68°C with an extension of 5 s/cycle. Primers to amplify the partial V4 regions (615–634, 5′-TTGGAGGGCAAGTCTGGTGC-3′; and 988–1005, 5′-ATAGCCACCGCCGACACC-3′) and the complete V7 regions (2337–2356, 5′-GGGTGGTGGTGCATGGCCGT-3′; and 2846–2866, 5′-TCTAAGGGCATCACAGACCTG-3′) were designed and these primers used to sequence the RT-PCR products. The V4 primers are located in unalignable hypervariable regions and the V7 primers in conserved regions. Prior to the synthesis of the first-strand cDNA, the total RNA was treated with RNase free DNase I (Promega Co.) to remove the possibility of DNA contamination. Then it was heated to remove the remaining enzyme activity. As a negative control, V4 and V7 regions were directly amplified from total RNA using the same conditions as were used for amplifying those regions from genomic DNA with Taq DNA polymerase.

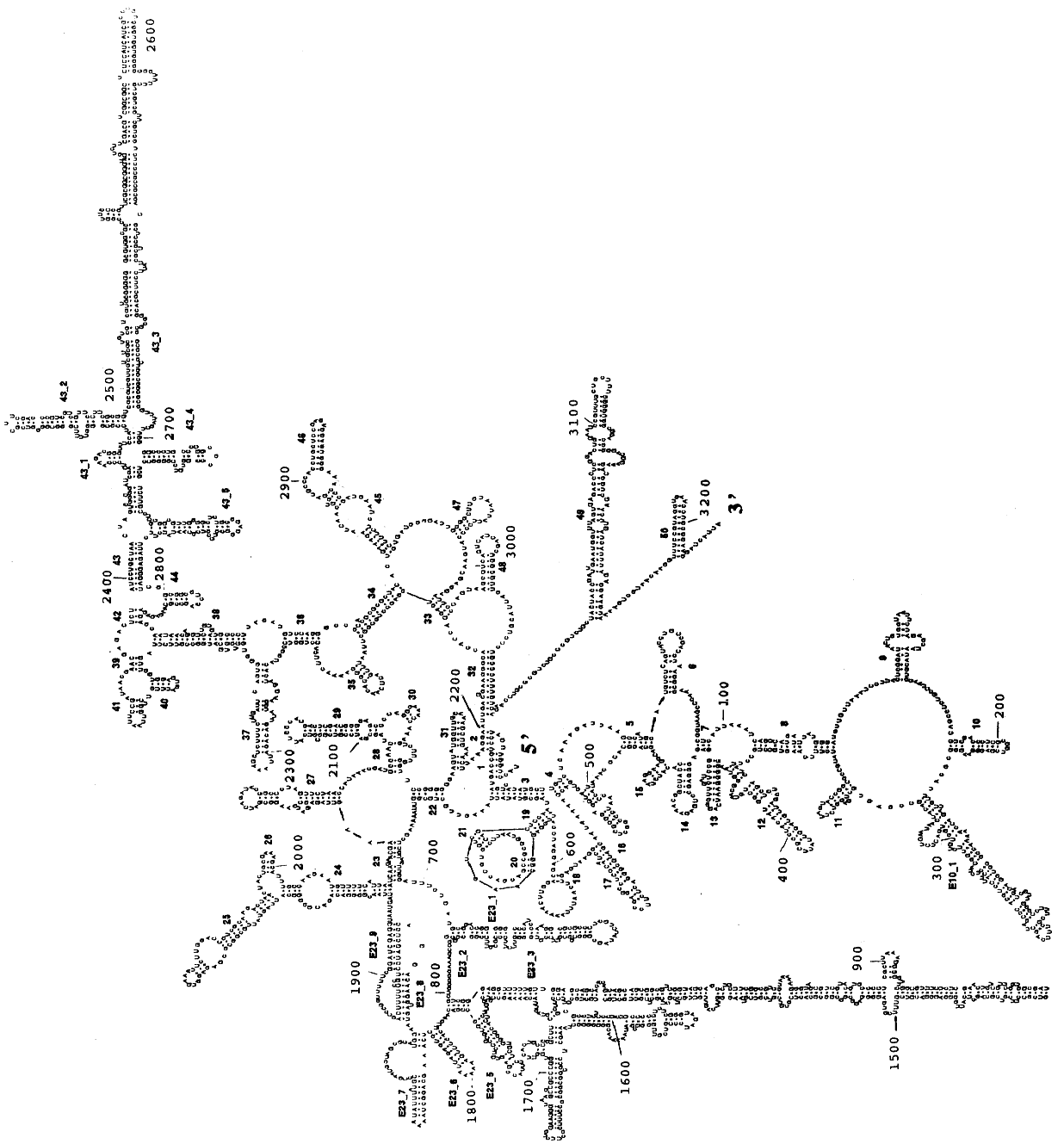
RNA Secondary Structure and Sequence Simplicity Analysis

The overall secondary structure of SSU rRNA of *A. vulgare* was constructed on the basis of the model of Van de Peer et al. (1998). In addition, partial secondary structures for the greatly expanded V4 and V7 regions were predicted using the MacDNASIS V 3.0 program (Hitachi Software Engineering Co., Japan). These two structures were then combined. The universal helices and the eukaryotic specific helices were numbered according to the model of Neefs et al. (1993). Sequence simplicity analysis of the complete SSU rRNA sequence was carried out using the SIMPLE34 program (Hancock and Armstrong 1994). Dot matrix analysis of the complete sequence was carried out using Dotty Plotter (Gilbert 1990) using a stringency of 19 matches in a 35 bp window as described previously (Hancock and Dover 1988).

Results

Primary Sequence and Secondary Structure

The primary sequence of *A. vulgare* SSU rRNA is 3214 bp long with a GC content of 56.3%. This is the longest


5'
3'
100
200
300
400
500
600
700
800
900
1500
1600
1700
1800
1900
2000
2100
2200
2300
2400
2500
2600
2700
2800
2900
3000
3100
3200
E23_1
E23_2
E23_3
E23_5
E23_6
E23_7
E23_8
E23_9
E10_1
43_1
43_2
43_3
43_4
43_5

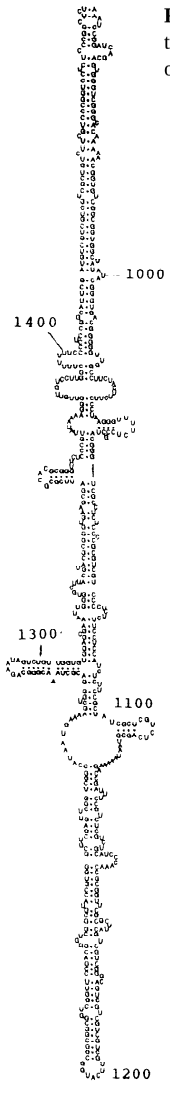


Fig. 1. A putative secondary structure of the small-subunit ribosomal RNA of *Armadillidium vulgare*. This is constructed based on the general eukaryote model of Van de Peer et al. (1998). Nucleotide positions were numbered every 100 nucleotides. The numbering of helices is also after that of Van de Peer et al. (1998).

SSU rRNA gene reported from metazoan animals except that of the strepsipteran insect, *X. vesparum*, which is 3316 bp long.

A putative secondary structure of *A. vulgare* SSU rRNA is shown in Fig. 1. All universal helices according to the general eukaryote model of Van de Peer et al. (1998) as well as ones specific to *A. vulgare* are predicted. Among the three greatly expanded variable regions (V2, V4, and V7), the V2 and V4 regions represent a typical structure for a eukaryote. However, the V7 region shows multibranched pattern similar to that of the strepsipteran insect, *X. vesparum* (Choe et al. 1999), instead of forming an expanded single helix.

RT-PCR Analysis of Expanded Variable Regions V4 and V7

RT-PCR analysis was performed for two expanded variable regions, the most expanded portion (390 bp) of V4 and the whole region of V7 (529 bp), to exclude the possibility that the extremely long sequences may be due to the presence of introns or pseudogenes. Products with lengths expected from the rDNA sequences were exactly obtained from RT-PCR analysis (Fig. 2). No products were observed on the agarose gel when a negative control experiment was carried out by using the total RNA without generation of the first-strand cDNA as a PCR template. Therefore the result indicates that the products obtained from RT-PCR amplification did not originate from contaminating genomic DNA. Furthermore, the nucleotide sequences of these PCR products were identical to the corresponding sequence of the SSU rRNA gene. Thus RT-PCR analysis directly demonstrated the presence of the expanded V4 and V7 regions in the mature SSU rRNA.

Sequence Simplicity Analysis

Sequences were analyzed for sequence repetition in two ways. Low-stringency dot-matrix analysis was used to identify regions showing internal sequence repetition and sequence coevolution (Hancock and Dover 1988). This analysis showed two regions of internal repetition, corresponding to variable regions V4 and V7 (data not shown). Simple sequence motifs within the sequence were characterized using the SIMPLE34 program (Hancock and Armstrong 1994), taking a threshold of 0.95 for the probability of high-scoring windows not having occurred by chance (corresponding to $p < 0.05$ of the score occurring by chance). This revealed two regions of high sequence simplicity within the sequence. The first of these spanned bases 1174–1231 and corresponded primarily to the tip of stem E23-3 within variable region V4. The second, and more prominent, repetitive region corresponded to two adjacent regions, bases 2433–2517 and 2595–2685. These regions encompassed most of stems 43-1 to 43-3 in variable region V7. Both regions contained similar repetitive motifs as defined by SIMPLE34 (Table 2).

Table 2. Frequency of occurrence of different repetitive motifs in *A. vulgare* SSU rRNA as measured by the program SIMPLE34

Motif[a]	V4	V7
GTC	5	10
TCG	2	7
CGT	1	5
GGC	2	2
CGG	—	1
CGC	—	1

[a] Significantly simple motif as defined by SIMPLE34 at a stringency of 0.95 ($p < 0.05$).

Discussion

Primary Sequence, GC Content, and Simplicity Analysis

Recently Crease and Colbourne (1998) reported that the unusually long SSU rRNA genes of arthropods are due to the expansion of the V2, V4, and V7 regions. Although these three regions occupy about 30% of the total length of the typical SSU rRNA of arthropods, they occupy more than 40% in the unusually long SSU rRNA genes (longer than 2000 bp), and nearly 60% in the extremely long SSU rRNA genes (longer than 3000 bp). In *A. vulgare* the V2, V4, and V7 regions also affect the total gene length. The V2, V4, and V7 regions are 264, 1219, and 400 bp long, respectively, and each of these three regions is longer than their average lengths in arthropods. The expansions of V4 and V7 are especially noticeable because the regions are longer than any others except for those of the strepsipteran insect, *X. vesparum.* These two regions occupy more than 50% of the total gene length and thus account for the elongation of the total SSU rRNA gene of *A. vulgare.* The expansion has also been reported of variable region D3 of the large subunit (LSU) rRNA gene in the same species (Nunn et al. 1996), suggesting coevolution of SSU and LSU rRNA genes. On the other hand, the length of the V9 region in *A. vulgare* is 92 bp, shorter than its average length in arthropods. It is reported that the V9 region does not generally contribute to the elongation of the total SSU rRNA gene and that its length is usually similar regardless of the different total lengths of SSU rRNA genes among major arthropod groups except for crustacea, in which V9 ranges from 92 to 150 bp (Crease and Colbourne 1998).

The average GC content of *A. vulgare* SSU rRNA is 56.3%, with 59.4% in V4 and 69.2% in V7. Figure 3 shows that the V4 and V7 regions have higher GC contents than any other regions. Therefore the high overall

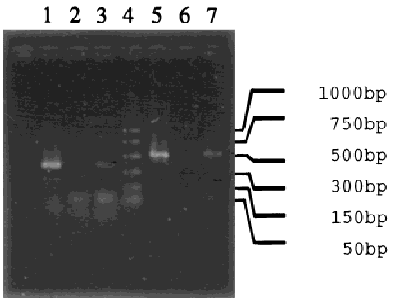


Fig. 2. Reverse transcriptase (RT)–PCR analysis of V4 and V7 regions in small-subunit ribosomal RNA of *Armadillidium vulgare*. Lanes 1–3 are results of amplifications of V4 from RNA via RT-PCR, RNA via RNA, and DNA. Lane 4 is the PCR marker (Promega Co.; Catalog No. G3161). Lanes 5–7 are results of amplifications of V7 from RNA via RT-PCR, RNA via RNA, and DNA.

GC content is explained by the high GC contents of the V4 and V7 regions. In general, the average GC contents of the total and V4 and V7 regions from SSU rRNA genes of the Crustacea and Insecta are approximately 50.8, 50.9, and 48.5 and 54.4, 47.8, and 44.5%, respectively. Thus, V4 and V7 regions are normally more AT rich than the rest of the SSU rRNA gene. In the unusually long SSU rRNA genes, the average GC content of the total gene is higher than that of other SSU rRNA genes, and those of V4 and V7 are higher than that of the total gene with the exception of the strepsipteran insect, which is AT rich (Table 3). It is noticeable that some different patterns exist in the increase in the overall GC content of the unusually long SSU rRNA genes between crustaceans and insects. In crustaceans, the GC content of V7 is higher than that of V4, thus, the V7 region contributes more than V4 to the increase in the overall GC content. On the other hand, in insects, the GC content of V4 is similar to or a little higher than that of V7 and these two regions contribute almost equally to the increase in the overall GC content (Table 3).

Hancock (1995) reported that repetitive sequences can be located in the variable regions of SSU rRNA (particularly in V4 and V7) and showed that simplicity of sequences increased as the bias in nucleotide composition increased. Among unusually long SSU rRNA genes, slippage motifs were found in the V4 and V7 regions of the aphid, *A. pisum* (Kwon et al. 1991). In addition, a 61-nt repeat instead of short slippage motifs was seen in V7 of the amoeba, *Phreatamoeba balamthi* (Hinkel et al. 1994), although unequal crossing-over would be expected to be a more important source of variation for motifs of that length. However, Crease and Colbourne (1998) did not notice any simple sequence motif occurring in V4 and V7 of *D. pulex* but noticed them in C7. They explained the absence of simple sequence motifs in V4 and V7 by noting that slippage could have contributed in the distant past but the repetitive sequence might have decayed by subsequent sequence divergence.

Both dot-matrix and SIMPLE34 analysis of the *A. vulgare* sequence revealed repetitive regions in the V4 and V7 regions but not elsewhere. This indicates a contribution from slippage to the evolution of these regions, particularly the motif CGT (Table 2). Both regions detected by SIMPLE34 corresponded to evolutionarily highly variable stem–loop regions in our secondary structure model (E23-3 and 43-1 to 43-3, respectively). Association of repetitive motifs with elements of secondary structure has given rise to the suggestion that a process of compensatory slippage takes place in such regions, whereby complementary slippage-derived motifs accumulate and result in expansion of particular structures (Hancock and Dover 1990). Compensatory slippage has been suggested to make a significant contribution to the evolution of stem E23-3 in tiger bettle (Cicindelidae) SSU rRNAs (Vogler et al. 1997; Hancock and Vogler 1998). In the case of *A. vulgare* the overwhelming majority of repetitive motifs were part of a circularly permutable group (GTC/TCG/CGT). This motif is partially self-complementary, as it contains the dinucleotide CG and could form stable structures, as the unpaired U's are small and can be accommodated relatively comfortably in RNA helices. It is therefore possible that stems E23-3 and 43-1 to 43-3 in *A. vulgare* V4 and V7 have evolved by compensatory slippage.

Examples mentioned above indicate a possible interrelationship among an increase in length, a high GC content, and high simplicity in V4 and V7 regions of unusually long SSU rRNA genes. That is, as GC-rich motifs slip repeatedly in variable regions, the length and GC content of these regions are also increased. A typical example is found in *A. vulgare*, where the length and the

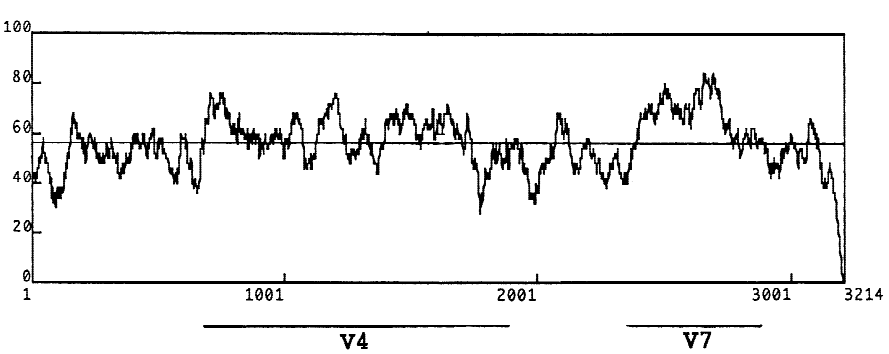


Fig. 3. Local GC ratios of the small subunit ribosomal RNA of *Armadillidium vulgare*. GC percentage was calculated per each 50 bases along the SSU rRNA of *A. vulgare*. The average GC content was represented by the central line. V4 and V7 regions are represented by the *horizontal lines* below the graph.

Table 3. Lengths and GC contents of the V4, V7, and total regions of unusually long SSU rRNA in Crustacea and Insecta

		V4		V7		Total	
Taxon	Order	Length	GC content	Length	GC content	Length	GC content
Crustacea							
Armadillidium vulgare	Isopoda	1219 nt	59% (725 nt)	400 nt	69% (277 nt)	3214 nt	56.3%
Daphnia pulex	Cladocera	393 nt	54% (213 nt)	303 nt	57% (174 nt)	2293 nt	53.7%
Insecta							
Mesoperlina percircai	Plecoptera	443 nt	56% (249 nt)	73 nt	48% (35 nt)	2081 nt	51.4%
Acyrthosiphon pisum	Hemiptera	546 nt	69% (377 nt)	327 nt	71% (233 nt)	2469 nt	59.4%

GC content are increased by GC-rich slippage motifs in V4 and in V7.

Secondary Structure

Because the expansion events are concentrated in three variable regions (V2, V4, and V7), the overall secondary structure of *A. vulgare* shows typical features of previously published eukaryote models. This conserved secondary structure implies that the unusually long SSU rRNA gene of *A. vulgare* is not pseudogene (Buckler et al. 1997). Insertions in V2 expand bulges located between stem 8 and stem 9 and between stem 9 and stem 10, but do not expand stems 9, 10, and E10-1. This differs from the case of *D. pulex*, in which insertions expand mainly stem 10 and E10-1, and from the case of *X. vesparum*, which forms a new helix, E10-2 (Crease and Colbourne 1998; Choe et al. 1999). There seems to exist no general pattern of insertion affecting the secondary structure in V2. In the V4 region, although stem E23-3 has four small branched stems, a single main stem is dramatically expanded by insertions. There exist two long bifurcated stems (E23-3 and E23-5) caused by stem E23-2 in *A. vulgare*. This bifurcating structure is also seen in *X. vesparum* and some tiger beetle species (Choe et al. 1999; Hancock and Vogler 1998). Crease and Colbourne (1998) reconstructed the same region of V4 in *D. pulex* as a unique structure having three helices. However, our reconstruction of this region in *D. pulex* reveals a secondary structure with the same pattern as that of *A. vulgare*. Figure 4 represents the reconstructed secondary structure of helices E23-2, E23-3, and E23-5 for *D. pulex*. Except for *X. vesparum*, which forms a long expanded helix between E23-8 and E23-9, insertions in V4 are therefore likely to occur mainly in helices E23-2 to E23-5. The V7 region of *A. vulgare* forms a unique structure unlike V2 and V4. Basically, helix 43 is not only expanded but also folded into four individual stems by insertions. It is quite different from the general tendency of forming one stable helix 43. However, this pattern is similar to that of *X. vesparum*, which consists of two expanded and branched helices made by long insertions (Choe et al. 1999). In *D. pulex*, however, only one branched helix exists in spite of the expansion of helix

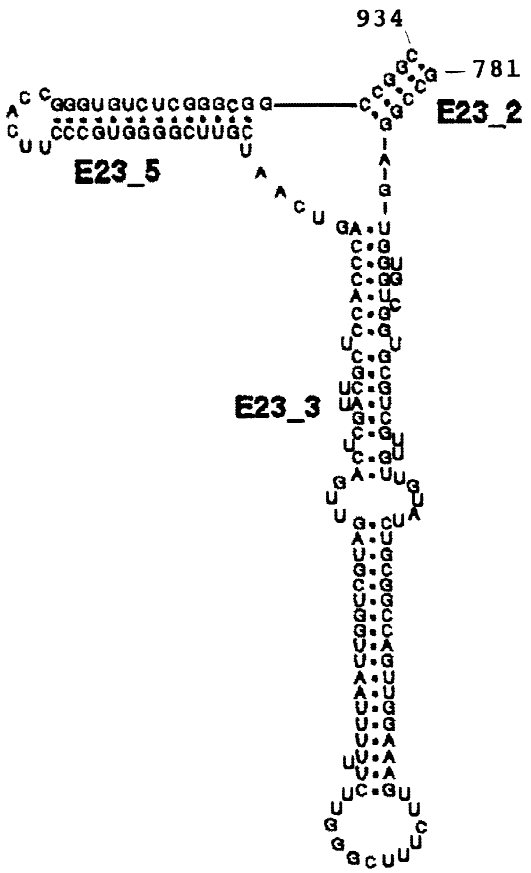


Fig. 4. Reconstructed secondary structure model of helices E23-2, E23-3, and E23-5 from the small subunit ribosomal RNA of *Daphnia pulex*. The free energy value is −75.5 kcal/mol.

43. Therefore further investigations are needed to see whether any general common features exist in the secondary structure of the V7 region and to obtain meaningful information from such characteristic structures.

The unusually long SSU rRNA gene of *A. vulgare* is the second report, in Crustacea, following that of *D. pulex*. Gerbi (1986) suggested that the lack of functions in these regions allows the expansion of V4 and V7 regions and this, in turn, is responsible mainly for the increase in the total length. On the other hand, Crease and Colbourne (1998) suggested that the coordinated increases in these two regions may be due to a functional relationship. According to our recent comparative analysis of the

secondary structures of these two hypervariable regions, V4 and V7, in insect 18S rRNAs, it can be deduced that their expansion events did not occur simultaneously but independently at different periods during the insect evolution (Hwang et al. in press). It may support that V4 and V7 regions are not functionally correlated. This analysis, however, was performed excluding unusually long SSU rRNA genes. Thus it is premature to infer any kind of relationship between expanded regions as a rule. The phylogenetic distribution of unusually long SSU rRNA genes is still unknown because unusually long SSU rRNA genes have been known from only a few taxa so far. Additional sequence data of many kinds of taxa with unusually long SSU rRNA genes are necessary to unravel the possible functional relationship between the expanded regions and to elucidate evolutionary mechanisms of such variable regions of SSU rRNA.

Acknowledgements. This work was in part supported by grants from KOSEF in 1995–1997 (95-0401-04-01-3), SRC (98-4-2), and Ministry of Education of Korea (Institute for Molecular Biology and Genetics) in 1996–1998.

References

Bhattacharya DB, Surek B, Rusing M, Damberger S, Melkonian M (1994) Group I introns are inherited through common ancestry in the nuclear-encoded rRNA of Zygnematales (Charophyceae). Proc Natl Acad Sci USA 91:9916–9920

Buckler ES 4th, Ippolito A, Holtsford TP (1997) The evolution of ribosomal DNA: Divergent paralogues and phylogenetic implications. Genetics 145:821–832

Campbell BC, Steffen-Campbell JD, Gill RJ (1994) Evolutionary origin of whiteflies (Hemiptera: Sternorrhyncha:Aleyrodidae) inferred from 18S rDNA sequences. Insect Mol Biol 3:175–194

Chalwatzis N, Baur A, Stetzer E, Kinzelbach R, Zimmermann FK (1995) Strongly expanded 18S rRNA genes correlated with a peculiar morphology in the insect order Strepsiptera. Zoology 98: 115–126

Choe CP, Hwang UW, Kim W (1999) Putative secondary structures of unusually long strepsipteran SSU rRNAs and its phylogenetic implications. Mol Cells 9:191–199

Crease TJ, Colbourne JK (1998) The unusually long small-subunit ribosomal RNA of the crustacean, *Daphnia pulex:* Sequence and predicted secondary structure. J Mol Evol 46:307–313

DePriest PT, Been MD (1992) Numerous group I introns with variable distributions in the ribosomal DNA of a Lichen Fungus. J Mol Biol 228:315–321

De Wachter R, Neefs J-M, Goris A, Van de Peer Y (1992) The gene coding for small subunit ribosomal RNA in the basidiomycete Ustilago maydis contains a group I intron. Nucleic Acids Res 20: 1251–1257

Gerbi S (1986) The evolution of eukaryotic ribosomal DNA. Biosystems 19:247–258

Gilbert DG (1990) Dot plot sequence comparisons on Macintosh computers. Comput Appl Biosci 6:117

Gustincich S, Manfioletti G, Sar GD, Schneider C (1991) A fast method for high-quality genomic DNA extraction from whole human blood. BioTechniques 11:298–301

Hancock JM (1995) The contribution DNA slippage to eukaryotic nuclear 18S rRNA evolution. J Mol Evol 40:629–639

Hancock JM, Armstrong JS (1994) SIMPLE34: An improved and enhanced implementation for VAX and Sun computers of the SIMPLE algorithm for analysis of clustered repetitive motifs in nucleotide sequences. Comput Appl Biosci 10:67–70

Hancock JM, Dover GA (1988) Molecular co-evolution amongst cryptically simple expansion segments of eukaryotic 26S/28S rRNAs. Mol Biol Evol 5:377–391

Hancock JM, Dover GA (1990) Compensatory slippage in the evolution of ribosomal RNA expansion segments. Nucleic Acids Res 18:5949–5954

Hancock JM, Vogler AP (1998) Modelling secondary structures of hypervariable RNA regions evolving by slippage: The example of the tiger beetle 18S rRNA variable region V4. Nucleic Acids Res 26:1689–1699

Hinkel G, Leipe DD, Nerad TA, Sogin ML (1994) The unusually long small subunit ribosomal RNA of *Phreatamoeba balamuthi.* Nucleic Acids Res 22:465–469

Hwang UW, Ree HI, Kim W. Evolution of hypervariable regions, V4 and V7, of insect 18S rRNA and their phylogenetic implications. Zool Sci (in press)

Kwon O-Y, Ogino K, Ishikawa H (1991) The longest 18S ribosomal RNA ever known. Eur J Biochem 202:827–833

Neefs J.-M, Van de Peer Y, De Rijk P, Chapelle S, De Wachter R. (1993) Compilation of small subunit ribosomal RNA structures. Nucleic Acids Res 21:3025–3049

Nelles L, Fang BL, Volckaert G, Vandenberghe A, De Wachter R (1984) Nucleotide sequcence of a crustacean 18S ribosomal RNA gene and secondary structure of eukaryotic small subunit ribosomal RNAs. Nucleic Acids Res 12:8749–8768

Nunn GB, Theisen BF, Christensen B, Arctander P (1996) Simplicity-correlated size grow of the nuclear 28S ribosomal RNA D3 expansion segment in the Crustacean order Isopoda. J Mol Evol 42:211–223

Sogin ML, Edman JC (1989) A self-splicing intron in the samll subunit rRNA of *Pneumocystis carinii.* Nucleic Acids Res 17:5349–5359

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

Van de Peer Y, Caers A, De Wachter R. (1998) Database on the structure of small subunit ribosomal RNA. Nucleic Acids Res 26: 179–182

Vogler AP, Welsh A, Hancock JM (1997) Phylogenetic analysis of slippage-like sequence variation in the V4 rRNA expansion segment in tiger beetles (Cicindelidae). Mol Biol Evol 14:6–19

Wilcox LW, Louise LA, Fuerst PA, Floyd GL (1992) Group I introns within the nuclear-encoded small-subunit rRNA gene of three green algae. Mol Biol Evol 9:1103–1118