생명과학회지
Korean J. Life Science, 8(1), 32-39, 1998

Nucleotide Sequences of an Aphid Ribosomal RNA Unit

Tae-Young Kwon*, Seung-Lak An**, Cheol Song***, Jong-Kyun Park****,
Young-Sub Kim****, Jae-Sam Hwang***** and O-Yu Kwon†

Dept. of Anatomy, Chungnam National University
**Kyungpook Provincial Rural Develop Administration*
***Dept. of National History, National Science Museum*
****Screening Division, Korea Research Institute of Chemical Technology*
*****Dept. of Sericulture and Entomology Resource, Sangju National Polytechnic University*
******National Entomology and Sericulture Research Institute*

Abstract

The length and G/C content of regions of an aphid rDNA unit that spans 13,061bp with 59% G/C content. Following below are the those results, 5′ ETS is 843bp in length with 69% G/C content, 18S is 2,469bp in length with 59% G/C content, ITS I is 229bp in length with 70% G/C content, 5.8S is 160bp in length with 63% G/C content, ITS II is 325bp in length with 70% G/C content, 28S is 4,147bp in length with 60% G/C content, IGS is 4,888bp in length with 55% G/C content.

Key words : *aphid, rDNA(ribosomal RNA gene), ETS (External Transcribed Spacer), 18S, ITS (Internal Transcribed Spacer) I & II, IGS (InterGenic transcribed Spacer), 28S*

Introduction

Aphids are an extremely prosperous group in the insect in the present day. This should be based upon the several characteristic unique to these animals. First, aphids are distinguished from other groups in the Homoptera in that the females, of at least a few generations, do not require fertilization for the development of their embryos, and also in that these asexual females produce live offspring. No doubt, the parthenogenesis and viviparity are quite beneficial to these insects. Second, all aphids expect the Phylloxiridae have special cells, the mycetocytes which harbor endosymbiotic microorganisms. The symbionts are assumed to be particularly important to the life of aphids. Third, even at the molecular level, aphids are unique enough. it is ribosomal RNAs, In general in most eukaryotic organisms, ribosomal RNA genes (rDNA) are encoded in multiple copies and the 18S, 5.8S and 28S rRNA-coding regions are transcribed as a larger precursor (pre-rRNA) by RNA polymerase I, which undergoes a series of post-transcriptional processing steps and then ITS I and ITS II removed, finally yielding the 18S, 5.8S and 28S rRNA found in ribosomes[1], Three exceptional molecular properties of the aphid rRNAs[2,3] are, First, the aphid 28S rRNA does not contain the hidden break that is share by the 28S rRNAs from most other protostomes. Second, the aphid 18S rRNA is much larger in molecular mass than those of other insects. Third, the aphid 28S rRNA is significantly larger than those of other insects.

† Corresponding author

What causes is involved these exceptional molecular properties for the aphid, which may give some clues to understand parthenogenetic reproduction, normal rRNA transcriptional mechanism and function of rRNA in the ribosome.

All these characteristics and others yet unknown together will support the extraordinary prosperity of this group. Yet to be known is causal relation among these characteristics. When this is answered, we can fully understand the true basis for the aphid's prosperity and also the evolutionary stage of this group. Also, undoubtedly studies on unusual organisms like aphids, as is the case with the studies on the genetical mutants, will provide a short cut leading to an understanding of usual ones in general.

Materials and Methods

Along-established parthenogenetic clone of the pea aphid, *Acyrthosiphon pisum* (Harris) was reared on young broad-bean plants, *Vicia faba L,* at 15℃ in a short-day regime with a 12-h photo period. The collected aphids were stored at −80℃ until use. High-molecular-mass genomic DNA was isolated from the aphids by the method described[4]. *EcoRI* partial digests of aphid DNA were ligated into EMBL4 at *EcoRI* sites using T4 ligase, packaged as described[5] and introduced into *Escherichia coli* LE392. The genomic library was screened[6] by the plaque-hybridization method using fragments from *Bombyx mori* ribosomal DNA as probes[7]. Isolation of bacteriophage DNA, subcloning of the fragment into plasmid Bluescript SK^+ and KS^+, transformation into *E. coli* XL1 Blue, plasmid growth, isolation of plasmid DNA and restriction enzyme digestions were performed according to the standard methods described[5]. Plasmid subclones of deletion series were constructed and sequenced by the dideoxy chain termination method[8] using the Sequenase version 2.0 kit (U.S. Biochem. Co.).

A secondary structure model for the aphid rRNAs were constructed according to the consensus model for eukaryotic rRNAs established previously[9-12]. Alternative regions of the model have been identified directly using the DNASIS V 6.00 program (Hitachi Software Engineering Co., Japan). Arrangement of sequences obtained from the subcloned plasmids and secondary structure construction were performed with the aid of a personal computer program described above. Another program of Genetyx (Software Development Co.) was used for the GC content analysis. Total aphid RNA was prepared from the whole tissue by the guanidinium-CsCl isopycnic purification method[5]. For S1 mapping, uniformly labeled single-strand probes were used[5]. For dot-blot hybridization, total RNA from aphids was spotted on nitrocellulose membrane and probed with the fragments from aphid rDNA. For primer extension analysis, 5°-end labeled synthetic DNA primer was hybridized with total RNA from the aphid tissue and extended with reverse transcriptase[5].

Results and Discussion

ETS(External Transcribed Spacer)[13,14]

The *cis*-regulatory sequences of transcription initiation of rDNA by RNA polymerase I is apparently species specific, however, due to the lack of data of nucleotide sequences recognized by *trans-regulatory* factors, ETS (external transcribed spacer) sequence, few data have only accumulated for transcriptional mechanism by RNA I. Here to understand normal rRNA transcription, aphid ETS, transcribed and removed for final rRNAs, connected with unique rRNAs was sequenced and compared with those of other species determined so far studied data. The 5° ETS of aphid rDNA consists of 843 nucleotides with a G/C content of 69%, far higher than that of any other known 5°ETS for insects, *B. mori* (36.5%), *D. melanogaster* (24%) and *Glossina morsitans morsitans* (28%). However, the size of the ETS of the aphid rDNA resembled those of other insects, *B. mori*

(909 nucleotides), *D. melanogaster* (864 nucleotides) and *G. morsitans morsitans* (924 nucleotides), which are considerably shorter than mammalian, mouse (4,006 nucleotides) and human (3,658 nucleotides). A very low degree of similarity was found when the aphid ETS sequence was compared with that of other insects so far studied (Fig 1). This suggested that the function of ETS for transcription by RNA polymerase I is depend on species. The aphid transcription initiation site is C reside located 37bp downstream of the internal *Xho* I site in IGS. And four palindromes consisting strong GC-stem without loop are observed, which may be *trans*-elements recognized by *cis*-factors during synthesis of rRNAs. More detail functional information for the these palindromes are need interaction studies both palindromes and *cis*-factors in vitro.

18S rRNA[15,16]

The small subunit RNA (18S rRNA), which are encoded by nuclear genomes, have lengths close to 1.8-2kbp and share significant nucleotide sequence identity with similar secondary structures. Especially, those sequences have been extensively researched and used as molecular clocks for evolutional studies. Previously, we suggested that the aphid 18S rRNA has larger about 500kDa than those of other eukaryotic 18S rRNA, suggesting that this exceptional mass may provide a clue to understand the general functions of the rRNA in the eukaryotic ribosomes. And then aphid 18S rRNA gene was cloned and sequenced. Determination of the complete sequence of the aphid 18S rRNA revealed that it is 2,469bp with a G/C content of 59%. It was the longest and has the highest G/C content among the 18S rRNA examined

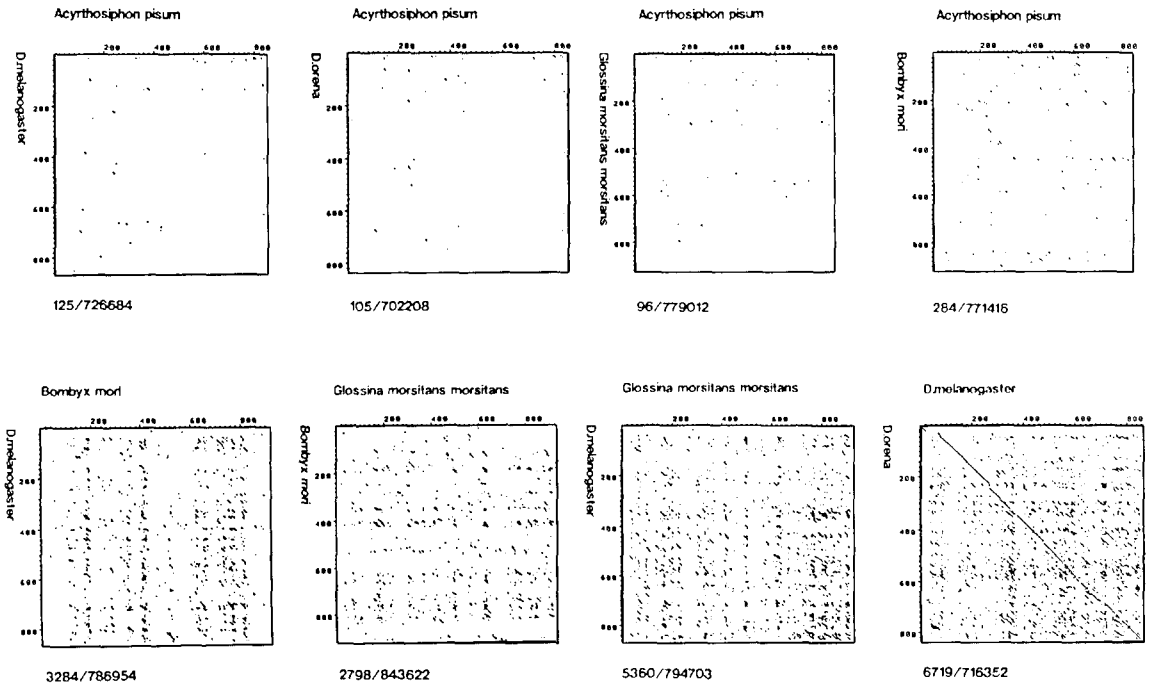

Fig. 1. Dot-matrix comparisons of ETS sequences between selected pairs of insects.
All comparisons were made by comparing 17 bases at a time, requiring 11 bases to identical to register a dot. The scores (total number of matched dots/total number of search dots) are displayed beneath the plots.

so far. However recent research of whitefly's has levealed the longest among the 18S rDNA in lengh by Campbell. From the result of construction of presumed secondary its structure, the eukaryote-specific E21 (Fig. 2) and 41 region are supposed to be longer and more complex than the counterparts of other 18S rRNA together exceptional X region, about 500bp. These additional nucleotide stretches account for most of the extraordinary length of this 18S rRNA. According to Hancock and Dover, such a length variation is attributed to the action of DNA slippage-like mechanisms. Furthermore, as an evolutionary process, 'compensatory slippage' has been proposed to explain the maintenance of secondary structure that accompanies slippage in such regions. The motifs of CGC and GCC found highly in the E21, 41 and X region (Fig. 3) which are taken to suggest that DNA slippage has taken place in the recent evolutionary history of this molecular species because no detected this like motifs in the other 18S rRNA exceptionally aphid 18S rRNA.

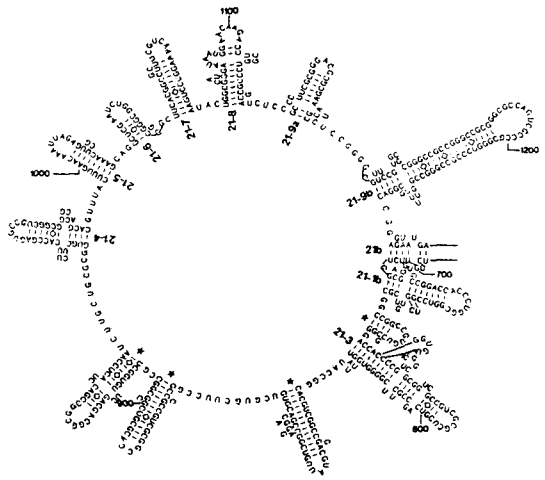

Fig. 2. A putative secondary structure for the E21 region of the aphid 18S rRNA.
The four asterisked portions are additional helices not present in the common model.

ITS I and II(Internal Transcribed Spacer I, II)[17,18]

What is called, ITS I and ITS II are located between 18S rRNA and 5.8S, 5.8S rRNA and 28S rRNA in a long precursor rRNA, which are removed during a series of posttranscriptional processing and those are not involved for the final rRNAs. While the nucleotide sequence of mature rRNAs has been strongly conserved throughout the evolution of prokaryotes and eukaryotes, both ITS I and ITS II sequences of distant organisms show dramatic differences in size and G/C content. ITS I and ITS II of rDNA of the aphid consisted of 229 and 280 nucleotides, whose G/C contents were 70% and 74%, respectively. Secondary structure models constructed for the ITS I and ITS II suggested that certain structural

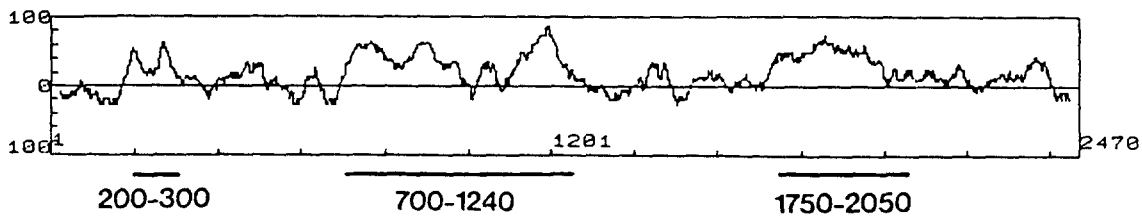

Fig. 3. Local G/C ratios of the aphid 18S rRNA.
G/C percentage was calculated for each 10 consecutive bases along the aphid 18S rRNA and the result was represented by a graph. G/C-rich regions are represented by the drawings above the zero line while A/U-rich regions are represented below the line. Eukaryotic specific regions (E10, E21 and 41) were indicated by horizontal lines below the graph.

motifs have been converted in these regions despite extensive divergence in nucleotide sequence due to species. The aphid ITS sequences were quite exceptional in that their G/C content was very high and comparable to that of the ITSs of vertebrates (Fig. 4). The model for ITS I contains four stem-loop structures. The proposed stems occur at nucleotide positions 4-68 (I), 73-100 and 185-210 (II), 104-114 (III) and 115-185 (IV). Free energy for the proposed secondary structure of the ITS I is -123.3kcal/mol. The model for ITS II contains six stem-loop structures. The proposed stems occurred at nucleotide positions 3-43 and 236-276 (I), 217-234 (II), 47-77 and 187-216 (III), 78-133 (IV), 138-153 (V) and 154-186 (VI). All of these stems, except stem V, are stable and rich in G/C pairs. Free energy for the proposed secondary structure of the ITS II is -207.7kcal/mol.

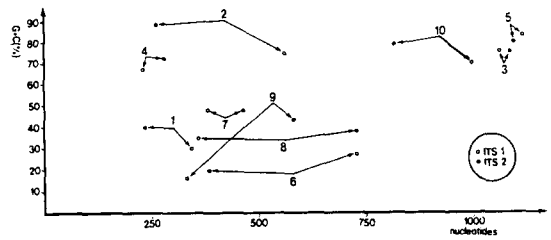

Fig. 4. Size and G/C content of eukaryotic ITSs.
1. Yeast, 2. *Xenopus*, 3. Mouse, 4. Aphid (this study), 5. Human, 6. *Drosophila*, 7. *C. elegans*, 8. *Dictyostelium*, 9. *Trypanosoma*, 10. Rat.

5.8S rRNA[6]

Since 5.8S rDNA is comparatively small comparing those of other rRNAs encoded in the nuclear, in general 5.8S rRNA genes from various species were at early time of the molecular biology sequenced and used as molecular index. Aphid 5.8S rRNA is 160 nucleotides in length with 63% G/C content. Sequence comparison suggested that the *A. pisum* 5.8S rRNA differs from that of *A. magnoliae* in only four bases. However, in view of the close relationship between the two aphid species, the 97% match between the two rRNAs is rather lower than expected because in mammals the 5.8S rRNA sequence is substantially the same. Even among vertebrates the sequence homology is around 95%. Even among vertebrates the sequence homology is around 95%. It is likely that divergence of the two aphid species is an event as old as that of most vertebrates. Alternatively, in insects molecular evolution may be faster than in vertebrates.

28S rRNA[19-21]

The aphid 28S rRNA is also significantly larger about 500 kDa than those of other insects. The hidden break of 28S rRNA is generated by excising a fixed length of polynucleotide from pre-rRNA posttranscriptionally and make two equally size subunits, α and ∨, which are hydrogen-bonded to each other at the regions close to the cleavage site *in vivo*. Here interesting point of aphid 28S rRNA is that hidden break is not generated in the middle part of aphid 28S rRNA. While in most protostomes share hidden break, in aphid, like in deuterostomes, such an excision does not take place in the 28S rRNA. We have completed the sequencing of aphid 28S rRNA, which has enable us to further characterize this unique rDNA from a molecular evolutionary point of view. Determination of the entire nucleotide sequence of the aphid 28S rDNA revealed that it is 4,147bp in length with a G/C content of 60.3%. The G/C content of the aphid's variable regions was much higher than that of the entire sequence of the 28S rRNA, which formed a sticking contrast to those of Drosophila with the G/C content (39%) much lower than the entire 28S molecule. The dot-matrix analysis suggested that the aphid 28S rRNA, just as other 28S rRNAs so far sequenced, has mosaic character, composed of two different types of sequence tracts. One is core segments whose structure is mostly conserved irrespective of species, and the other is variable regions, or expansion segments, whose

structure is extremely variable across species. On comparison of nucleotide sequences, it was shown that the aphid 28S rRNA has two types of variable regions. One is the variable regions that are comparatively similar to those found in 28S rRNAs of vertebrates, but not found in *Drosophila*. The other is those that seem to be unique to the aphid 28S rRNA. On the other hand, aphid species form only one exceptional group of insects that have been studies so far in that like those of deuterostomes their 28S rRNAs lack the hidden break. Large subunit of aphid rRNA does not contains hidden break in the its middle region, for comparison's sake, its possible secondary structure was constructed with other species. The stem-loop structure of aphid differs from the others not only in that it is smaller and more GC-rich but in that does not contain the UAAU tract (Fig. 5). The

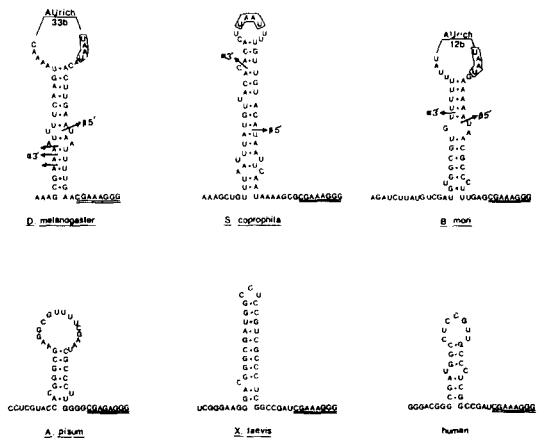

Fig. 5. Comparison of secondary structure models for the expansion segment 5 of the 28S rRNAs of several organisms.
Arrows indicate the processing sites (α3°, the 3° end of 28Sα ; β5° end of 28Sβ). The conserved UAAU tract in the loop is boxed. The highly conserved sequence found downstream of the expansion segment 5 in all of the 28S rRNAs is doubly underlined.

UAAU tract has been suggested to be a signal for excision of the gap region to create the hidden break because it is unique to the rDNA transcript leading to the 28S rRNA with the hidden break. Unlike those of other insects, those of deuterostomes in that it is GC-rich and without any UAAU tract.

IGS(InterGenic transcribed Spacer)[13]

In most rDNA studies, the IGS (intergenic transcribed spacer) contains a tandem array of small 'subrepeats', which are thought to be involved in the control of rDNA transcription. Although this region is not transcribed unlike ETS, however which is very important for transcriptional regulation to make normal rRNAs and to occurs chromosomal unequal crossing over in the same species. When the aphid transcription-initiation site sequence was determined and compared with those of other species so far studied, no significant similarities was observed between them. From the result of IGS sequence of the aphid, its full length is 4,888bp with 55 % G/C content, in which contains a tandem array subrepeats was detected, which region digested with Kpn I by identical 247bp tandemly repeating 11 times. Coen and Dover[15] have suggested that the use of the promotor duplications within subrepeats in the regulation of transcription may have evolved to overcome the original generation of such repeats by unequal crossing over. We have found such a promoter duplication at positions 431-439. IGS of the aphid shown length heterogeneity, which probably results from unequal exchange between the repeating units. And unequal crossing over is probably the mechanism responsible for the spread of minor and major mutational differences throughout the gene family and species. The size and sequence heterogeneity in thess repeating sequences appears to be a common feature of eukaryotes. In this regard, the repeating sequences in aphid IGS are quite exceptional in that they are homogeneous in both size and sequence. Such a high degree of homogeneity may have been generated by a high rate of unequal crossing over under the influence of molecular drive. It is also believed that in the course

of their diploid parthenogenesis, no recombination of DNA take place.

Discussion

The aphid rRNAs have three exceptional molecular properties, both 18S rRNA and 28S rRNA have about 500 kDa more bigger than other's, respectively. Although aphid is included in the protostomes, its 28S rRNA does not contain a hidden break gap in the its middle part. This like exceptional molecular properties in the rRNAs may give some ideas to known that transcription mechanism by RNA polymerase I and parthenogenetic reproduction of aphids. Actually aphid's 18S rRNA is shown that it is the longest and has complex eukaryotic specific regions, E21, 41 and X region, by compensatory slippage mechanism. the molecular properrty of aphid 28S rRNA is generally similar to its eukaryote, and unlike those of other insects it contains only GC-rich stem without any UAAU tract in the hidden break region. In IGS, perfectly identical 247bp is repeated 11 times by result of unequal crossing over exchange between the subrepeats, which suggested aphids have propagated themselves parthenogenetically during most of their evolution. One plausible explanation might be that unequal crossing over at high-enough frequency during phases of bisexual reproduction to 'spring clean' the rDNA and restore homogeneity of the IGS subrepeats. The model of the secondary structure for the pea aphid ITS I and ITS II are surprisingly similar to general structures that have been reported for ITSs of other organisms. They share common helical structure divergence, which play any role in the recognition mechanism of ribosomal RNA processing.

References

1. Perry, R. P. : Processing of RNA, *Ann. Rev. Biochem.*, **45**, 605-629(1976).
2. Ishikawa, H. : Arthropod ribosomes : integrity of ribosomal ribonucleic acids from aphids and waterfleas, *Biochem. Biophys. Acta.*, **435**, 258-268(1976).
3. Ishikawa, H. : Evolution of ribosomal RNA, Com. *Biochem. Physiol.*, **58B**, 1-7(1977).
4. Herrmann, B. H., Frischuf, A. M. : Isolation of genomic DNA, *Methods Enzymol.*, **152**, 346-358(1987).
5. Sambrook, J., Fritsch, E. F., Maniatis, T. : Molecular cloning : a laboratory manual, *Cold Spring Harber Laboratory, CSH,* NY(1989).
6. Ogino, K., Eda-Fujiwara, H., Fujiwara, H., Ishikawa, H. : What causes the aphid 28S rRNA to lack the hidden break ? , *J. Mol. Evol.*, **30**, 509-513(1990).
7. Fujiwara, H., Ishikawa, H. : Molecular mechanism of introduction of the hidden break into the 28S rRNA of insects : implication based on structural studies. *Nucl. Acids Res.*, **13**, 3581-3597(1985).
8. Sanger, F., Nicklen, S., Coulson, A. R. : DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA.*, **74**, 5463-5467(1977).
9. Chan, Y. L., Gutell, R., Noller, H. F., Wool, I. G. : *J. Biol. Chem.*, **259**, 224-230(1984).
10. Ellis, R. E,. Sulston, J. E,. Coulson, A. R. : The rDNA *C. elegans* : sequence and structure. *Nucl. Acids Res.*, **14**, 2345-2364(1986).
11. Subramanyam, C. S., cassidy, B., Busch, H., Rothblum, L. I. : Nucleotide sequence of the region between the 18S rRNA sequence and the 28S rRNA sequence of the rat ribosomal DNA, *Nucl. Acids Res.*, **10**, 3667-3680(1982).
12. Schnare, M. N., Cook, J. R., Gray, M. W. : Fourteen internal transcribed spacers in the circular ribosomal DNA of Euglena gracilis, *J. Mol. Biol.*, **215**, 85-91 (1990).
13. Kwon, O. -Y., Ishikawa, H. : Unique structure in the intergenic and 5° external transcribed spacer of the ribosomal RNA gene from the pea aphid *Acyrthosiphon pisum, Eur. J. Biochem.*, **206**, 935-940(1992).
14. Kwon, O. -Y., Lee, D. H., Kwon, T. Y. : Detection of the specific DNA-binding proteins for the aphid rDNA, *Korean J. Appl. Entomol.*, **34**, 100-105(1995).
15. Kwon, O. -Y., Ogino, K., Ishikawa, H. : The longest 18S ribosomal RNA ever known ; Nucleotide sequence and presumed secondary structure of the 18 S rRNA of the pea aphid, *Acyrthosiphon pisum, Eur. J. Biochem.*, **202**, 827-833(1991).

16. Campbell, B. C., Steffen-Campbell, Gill. R. J. : Evolutionary origin of whiteflies inferred from 18S rDNA sequences, *Insect Mol. Biol.*, 3, 73-88(1994).
17. Kwon, O. -Y., Ishikawa, H. : Nucleotide sequence and presumed secondary structure of the internal transcribed spacers of rDNA of the pea aphid, *Acyrthosiphon pisum., Com. Biochem. Physiol.*, **103B**, 651-655(1992).
18. Yeh, L. -C., Thweatt, R., Lee, J. C. : Internal transcribed spacer I of the yeast precursor ribosomal RNA : higher order structure and common structural motif, *Biochemistry*, **29**, 5911-5918(1990).
19. Amako, D., Kwon, O. -Y., Ishikawa, H. : Nucleotide sequence and presumed secondary structure of the 28S rRNA of pea aphid : Implication for diversification of insects rRNA, *J. Mol. Evol.*, **43**, 469-475(1996).
20. Linares, A. R., Hancock, J. M., Dover, G. A. : Secondary structure constraints of the evolution of *Drosophila* 28S ribosomal RNA expansion segments, *J. Mol. Biol.*, **219**, 381-390(1991).
21. Ware, V. C., Renkawitz, R., Clark, C. G. : rRNA processing : removal of only nineteen bases at the gap between 28Sα and β rRNAs in Sciara coprophila, *Nucl. Acids Res.*, **13**, 3581-3597(1985).

초록 : 진딧물의 전 ribosomal RNA 염기배열

권태영*, 안승락**, 송철***, 박종균****, 김영섭****, 황재삼*****, 권오유†
(경상북도 농촌진흥원*, 국립 중앙과학관 자연사연구실**, 한국 화학연구소 스크리닝부***, 상주산업대 잠사곤충자원학과****, 잠사곤충연구소 유전육종과*****, 충남대학교 해부학교실)

진딧물의 하나의 ribosomal RNA 유전자(rDNA)단위는 총 길이가 13,061bp이며 총 G/C비율은 59%이다. 그것을 구성하고있는 각 영역의 길이와 G/C비율은 다음과 같다. 5' ETS는 G/C비율이 69%이고 843bp이다. 18S rRNA는 2, 469bp이며 G/C비율은 59%이다. ITS I의 길이는 229bp이며 70%의 G/C비율이다. 5.8S rRNA는 160bp이며 63%의 G/C비율이다. ITS II는 325bp이며 70%의 G/C비율이다. 28S rRNA는 4,147bp이고 60%의 G/C비율이다. IGS는 4, 888bp로 55%의 G/C비율이다.