한국동물분류학회지 특간제 3 호
(金熏洙 敎授 古稀 紀念號)
The Korean Journal of Systematic Zoology
Special Issue No. 3: 139-146 (October 30, 1992)

A Study on the Nucleotide Analysis of 18S rRNA and the Molecular Evolution of the Korean Decapods (II)

Kim, Won, Min, Gi Sik and Kim, Sang Hee
(Department of Molecular Biology, College of Natural Sciences, Seoul National University, Seoul 151-742, Korea)

한국산 십각류의 18S 리보솜 RNA의 염기분석과 분자진화에 관한 연구(II)

김 원·민 기 식·김 상 희
(서울대학교 자연과학대학 분자생물학과)

적 요

중합효소연쇄반응(PCR)을 이용한 클로닝 기법과 Taq 염기서열분석법을 사용하여 갑각류에 속하는 뿔물맞이게(*Pugettia quadridens*)(십각 목, 범배 아목, 게 하목)에 대한 18S 리보솜 RNA 유전자의 1차염기서열을 밝혔다. 본 종의 18S 리보솜 RNA 유전자는 십각류에 속하는 또 다른 종인 두드러기어리게(*Oedignathus inermis*)보다 46개가 짧은 1837개의 염기로 이루어져있었다. 염기가 삽입되거나 결실된 부분을 고려하지 않았을때에 두 종간에 염기서열 유사도는 90.8%이었다. 염기서열이 가장 보존적인 부위는 1137−1206(70개) 부위로 이 부위에서는 두 종이 완전히 동일한 염기서열을 보이고 있었고, 변이가 연속적으로 가장 큰 부위는 46−55 부위였고 399−407 부위가 그 다음으로 많은 연속적 변이를 가지고 있었다. 18S 리보솜 RNA 유전자의 1차구조에 있어서 염기서열의 변이는 이 유전자 전체를 통해 고르게 분포하지 않았다.

Key words: *Pugettia quadridens*, 18S rRNA, Decapoda, PCR cloning.

This paper is dedicated to Professor Hoon Soo Kim on the occasion of his 70th birthday. This work was supported by research grant from the Ministry of Education of Korea (Institute for Molecular Biology and Genetics), 1991.

INTRODUCTION

Because of the conservative nature, the nucleotide sequences of 18S rRNA have been widely adopted in the construction of molecular phylogenies among the remotely related eukaryotes in recent years (Field *et al.*, 1988; Stock and Whitt, 1992). However, most studies used the partial sequences without complete understanding of the sequence variability of each species representing certain taxonomic categorical rank. Even though the results from those studies are congruent to those from the morphological data in many cases, the 18S rRNA shows sequence variability, such as conservative, variable, and highly variable, across the molecule. Therefore the nucleotide analysis of the complete sequences of many species will provide the further understanding of the evolutionary relationships among the organismal groups, especially when the groups being compared belong to the lower categorical rank.

As a part of study in examining sequence variability and general patterns of nucleotide substitution of the 18S rRNA gene with the complete sequences among crustaceans, the 18S rRNA gene from one decapod species, *Pugettia quadridens,* was cloned and sequenced. In crustacean decapods, there is only one species, *Oedignathus inermis,* in which the complete nucleotide sequences of 18S rRNA gene were known (Kim *et al.*, 1992). Therefore we compared the nucleotide sequences of *P. quadridens* with that of *O. inermis.*

MATERIALS AND METHODS

In this study, the 18S rRNA gene of *Pugettia quadridens* was sequenced by PCR cloning and Taq sequencing as Kim et al. (1992). This species belongs to the infraorder Brachyura, suborder Pleocyemata, order Decapoda in Crustacea (Bowman and Abele, 1982). The sequence datum of *O. inermis* is from Kim *et al.*(1992) and this species belongs to infraorder Anomura, order Decapoda. We aligned the nucleotide sequences of *O. inermis* and *P. quadridens* using FASTA program (Pearson and Lipman, 1988). The numbers indicating nucleotide position in the following text are those of *O. inermis* when the sequences are numbered without the insertion and/or deletion.

RESULTS

The total length of the 18S rRNA gene of *P. quadridens* is 1837 bases long, and 46 bases shorter than *O. inermis* sequences (Fig. 1). The occurrence of each nucleotide in the two species ranges in 23.3-26.9%. The G+C content is around 50% in each species (Table 1).

The total number of different, identical, and null nucleotides between two species are 167, 1655, and 76, respectively. The similarity between two species is 90.8% when the insertion and/or deletion sites were excluded. The number of different nucleotides in a row varies 1 to 10 and the one nucleotide difference is the majority (86 out of 167). If the insertion and/or deletion sites are included and these sites are considered consider as the unweighted nucleotide differences between two species, the similarity is 87.2% (Table 2). The ratio of transition/transversion is 0.72 in the present two species. There are 34 insertion regions

```
                10        20        30        40        50        60
Oed  TACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGATTAAGCCATGCATGTGT
Pug  TACCTGGTTGATCCTGCCAGTAGTCATATGCTTGTCTCGTTGTCTCGTTGCCTGCTGTCT

                70        80        90       100       110       120
Oed  AAGTACAAGCCGATGTAAGGTGAAACCGCGAATGGCTCATTAAATCAGCTATGTTTCATT
Pug  AAGTACAAGCCGATTCAAGGCGAAACCGCGAATGGCTCATTAAATCAGCTATGATTCATT

               130       140       150       160       170       180
Oed  GGAACTGTACCCCCACTTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCACAC
Pug  GGATCTGTACCCACATTTACTTGGATAACTGTGGTAATTCTAGAGCTAATACATGCCCA-

               190       200       210       220       230
Oed  AGAGTCCCCGGACCGCAGGGAGGGGCGCTTTTATTAGTTCAAAACCGGTCGGGCCTC--G
Pug  -GAGTCTCTG-ACCGCAAGGAAGAGCGCTTTTATTAGTTCAAAACCGGTCGGGCCTTCGG

     240       250       260       270       280       290
Oed  GTCCGTAACCACC-TGTGGTGAATCTGAATAACTTTGACTGAGCGCACGGTCTCCGCACT
Pug  GTCCGTCACCCTGGTGTGTTGAATCTGAATAACTTTT-CTCA-CGCACGGTCTCCGCGAG

     300       310       320       330       340       350
Oed  GGCGCCGCATCTTTCAAGTGTCTGCCTTATCAGCTTTCGATTGTA-GGTTATT-CGCCTT
Pug  GGC--CGCCTCTTTCAAGTGT-TGCCTTATCAGCTTTCGATTGTAAGGTTACTACGCCTT

     360       370       380       390       400       410
Oed  CCATGGCTATTACGGGTTACGGGGAATTCAGGGTTCGATTCCGGAGAGGGAGCCTTAGGA
Pug  CAATGGCTATTACGG-TTACGGGTAAT-CAGGGTTTTATTCCGTTTTCCCCCCCTTGGGA

     420       430       440       450       460       470
Oed  ACGGCTACCACATCTAAGGAAGGCAGGAGGGACGTAAATTACCCATTTCCAGACCGGGGA
Pug  ACGGGTACCACATCTAAGGAAGGCCGGAGGGACGGCAATTACCCAAC-CCGGCACGGGGA

     480       490       500       510       520       530
Oed  GGTAGTGACGGAAAATAACGATGCGAGACTCATCCGAGGCCTTGCAATCGGAATGAGTAC
Pug  GGTAGTGACCGAAAATAACGATGCGAGACTCATCCGAG-CCTCGCAATCGGAATGAGTAC

     540       550       560       570       580       590
Oed  ACTTTAAATCCTTTAACGAGGACCCATTGGAGGGGCAAGTCTGGTGCCAGCAGCCGCGGT
Pug  ACTTTAAATGGTTTAACGAGGATCCATTGGAGGG-CAAGTCTGGTGCCAGCAGCCGCGGT

     600       610       620       630       640       650
Oed  TAATTCCAGCTCCAATA-CGTATATTAAAGTTGTTGCGGTTAAAAAG-TCGTAGTTGGAT
Pug  -AATTCCAGCTCCAATAGCGTATATTAAAGTTGTTGCGGTTAAAAAGCTCGTAGTTGGAT

     660       670       680       690       700       710
Oed  TACAGTTCCGGACTGACGGTTACCGCCCGGTGCTTACTGTCACGCTCCGAACAGCCTGAA
Pug  TTCAGTTCTGGACTGACGGTTACCGCCCGGTGATCACTGTCACGCTCCGAACACTTTGAC

     720       730       740       750       760       770
Oed  ACATGGGCCCGGCGGCTCGCCGGGGTGCTCTTTACCGAGTGTCCCGAGTGGCCGGCATGT
Pug  -CAT----CCGCTGGCNCAC-GGGGTGCCCTTTTCCCAATTTCCCCTG-GCCCGGAGAGT

     780       790       800       810       820       830
Oed  TTACTTTGAAAAAATTAGAGTGCTCAGAGCAGGCTACATGAATTGGCCTGAATGTCTATG
Pug  TTACTTTGAAAAAATTAGAGTTCTCAAAGCAGGCTACACTGAC-GGCCTGAATGCCTATG

     840       850       860       870       880       890
Oed  CATGGAATAATGGAATAGGACCTCGGTTCTATTTTGTCGGTTTTCTTCGGAACCCGAGGT
Pug  CATGGAATAATGGAATAGGACCTCGGTTCTATTTTGTCGGTTTTCT---GAACCCGAGGT

     900       910       920       930       940       950
Oed  AATGACTAATAGAAACAAGGCGGGGGTATTAGTACTCGAGCCAGAGGTGAAATTCTTGGA
Pug  AATGACTAATGGAA--CAGGCGGGGTCATTC-TATTGACGCTAGAGGT-AAATTCTTGGA

     960       970       980       990      1000      1010
Oed  CCGTCGCAAGACTTACAACTGCAAAGTATTTACCAAGGATGTTTTCATTAATCAAGAACG
Pug  CCGTCGCAAGACGAACTACTCGAAATCATTTGCCAAGGATGTTTTCATTAATCAAGAACG
```

```
          1020      1030      1040      1050      1060      1070
Oed  AAAGTTAGAGGTTCGAAGGCGATCAGATACCGCCCTAGTTCTAACCATAAACGATGCTGA
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  AAAGTTAGAGGTTCGAAG-CGATCAGATACCGCCCTAGTTCTAACCATAAACGATTCTGA

          1080      1090      1100      1110      1120      1130
Oed  CTACGATCCGCCGGCGTTATTCCCATGACCGGCGGGCAGCTTCCGGGAAACCAAAGTCTT
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  CCAC-ATCCGCCGGAGTTATTCCCATGACCGGCGGGGAGCTTCCGGGAAACCAAAGTCTT

          1140      1150      1160      1170      1180      1190
Oed  TGAGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCAC
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  TGGGTTCCGGGGGAAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCAC

          1200      1210      1220      1230      1240      1250
Oed  CACCAGGAGTGGCATGCGGCTTAATTTGACTCAACACGGGAAACCTCACCAGGCCCAGAC
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  CACCAGGAGTGGC-TGCGGCTTAATTTGACTCAACACGGGGAACCTCACCAGGCCCAGAC

          1260      1270      1280      1290      1300      1310
Oed  ACTGGAAGGATTGACAGATTGAGAGCTCTTTCTCGATTCAGTGGGTGGTGGTGCATGGCC
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  ACCGGAAGGATTGACAGATTGAGAGCTCTTTCTTGATTCGGTGGGTGGTGGTGCATGGCC

          1320      1330      1340      1350      1360      1370
Oed  GTTCTTAGTTGGTGGAGCGATTTGTCTGGTTAATTCCGATAACGAACGAGACTCTAGCCT
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  GTTCTTAGTTGGTGGAACGATTTGTCTGGTTAATTCCGATAACGAACGAGACCCTAGCCT

          1380      1390      1400      1410      1420      1430
Oed  ATTAACTAGTCGACGGATCTCCAGCAATTGGTGTCCAGTCGCAACTTCTTCTTAGAGGGA
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  ACTAACTAGTCGACGGATCTCCAGCAATTGGTGTCCAGTCGCAGCTTCTTCTTACAGGGA

          1440      1450      1460      1470      1480      1490
Oed  TAACGGCAACTCTAAGCCGCACGAGAATTGAGCAATAACAGGTCTGTGATGCCCTTAGAT
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  AACGGGCAATTCTA-GCCGCACGAGA-TTGAGCAATAACAGGCTTGTGATGCCCTTAGAT

          1500      1510      1520      1530      1540      1550
Oed  GTTTTGGCCGAACCC-CGCTACACTGAAGGGATCAGCGTGTTGTCCCCCTCCGAGAGGAG
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  GTTCTGGGCGCACGCGCGCTACACTGAAGGGATCAACGTGTCCTCCCC-TCCGAGAGGAG

          1560      1570      1580      1590      1600      1610
Oed  CGGGTAACCCTATGAAACCCCTTCATGATAGGGATTGGGGCTTGCAATTGTT-TCCCATG
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  CGGGTAACCCG-TGAAATCCTTTCATGATAGGGATTGGGGTTTGCAATTGTTCTCCCATG

          1620      1630      1640      1650      1660      1670
Oed  AACGAGGAATTCCCAGTAAG-CGCAAGTCATCAGCTTGCGTTGATTACCGTCCTGCCCTT
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  AACGAGGAATTCCCAGTAAGGCGCAAGTCATGGGATTGCGTTGATTACG--CCTGCCCTT

          1680      1690      1700      1710      1720
Oed  -GTACACACCCCCG-TCGCTACTACCGATTGAATGATTTAGTGAG-CTTCGGACTG-CGC
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  CGTACACACGCCCGGTCGCTACTACCGATTGAATGATTTAGTGAGGCTTCGGATTGGCGC

          1730      1740      1750      1760      1770      1780
Oed  TCTTGGATGTCCGGCCGTTCCCGCCGTCTTCTCTCTGAGGGGGCGGTGGTCGCGGGTTCC
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  TCTTGGATGCCTGGC--------CCGCCTTC------------CGGTGG--GCTTTTTAG

          1790      1800      1810      1820      1830      1840
Oed  GGCCCTCGGGCTGACGGAAAGATGTCCAAACTTGATCATTTAGAGGAAGTAAAAGTCGTA
     ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Pug  GCGCCTCGAGCTGACT-AAAGATGTCCAAACTTGATCATTTAGAGGAAGTAAAAGTCGTA

          1850      1860      1870      1880
Oed  ACAAGGTTTCCGTAGGTGAAC-TGCGGAAGGATCATTA
     ::::::::::::::::::::::::::::::::::::::
Pug  ACAAGGTTTCCGTAGGTGAACCTGCGGAAGGATCATTA
```

Fig. 1. The nucleotide sequences of 18S rRNA gene of *O. inermis* and *P. quadridens* aligned by FASTA program. Oed = *O. inermis* ; Pug = *P. quadridens*.

Table 1. The total length, nucleotide compositon, and G+C content of the 18S rRNA gene in *O. inermis* and *P. quadridens*.

Species	Total length(base)	No. of each nucleotide				G+C content
		A	C	G	T	
*O. inermis**	1883	468 (24.8%)	439 (23.3%)	507 (26.9%)	469 (24.9%)	50.20%
P. quadridens	1837	444 (24.2%)	437 (23.8%)	481 (26.2%)	474 (25.8%)	49.97%

* Data from Kim *et al.* (1992)

Table 2. Pairwise comparison of nucleotide sequences in Art/Oed* and Oed/Pug. Art = *A. salina*; Oed = *O. inermis*; Pug = *P. quadridens*.

	Art/Oed	Oed/Pug
No. of nucleotide		
— identical	1552	1655
— different	237	167
— null (insertion and/or deletion)	115	76
Similarity(%)		
— with null sites	81.5	87.2
— without null sites	86.8	90.8
Transition / transversion (ratio)	1.26	0.72

* Data from Kim *et al.* (1992)

in *O. inermis* and 14 in *P. quadridens*. The number of nucleotide in each insertion region varies between 1 and 12. The longest insertion region is found at the postion of 1758-1769 and the second most at 1742-1749 (Fig. 1). The sequence variation in the primary structure of 18S rRNA gene are not evenly distributed throughout the molecule as indicated by Kim and Abele (1990).

DISCUSSION

Kim *et al.*(1992) compared the nucleotide sequences of *A. salina* (class Branchiopoda) and *O. inermis* to see the sequence variability between class levels and the present study compared the nucleotide sequences between infraorder levels (between Anomura and Brachyura) in Decapoda. Result from the present study show that the molecular data are congruent to the morphological data in terms of the higher similarity between infraorder levels than between class levels (Table 2).

It has been known that the nucleotide sequences of the 3′-terminal region is well conserved among many species. In the present two decapod species, the 3′-terminal region (position of 1804-1883) also shows

only one nucleotide insertion and/or deletion among total 81 nucleotides. Therefore this region appears to be quite conservative, while the most conservative (identical) region locates at the position of 1137-1206 and it is 70 bases long. The most variable nucleotide differences in a row occur at the position between 46-55 and the second most between 399-407. The nucleotide sequences of these regions are identical between *A. salina* (class Branchiopoda) and *O. inermis*. Thus the differences between *O. inermis* and *P. quadridens* in these regions are due to *P. quadridens* sequences. Because *A. salina* and *O. inermis* are considered to be more primitive than *P. quadridens*, the different sequences of *P. quadridens* in these regions seems to be the nucleotide substitutions occurred in the lineage leading *P. quadridens*.

We also compared the above three species together to see the insertion and/or deletion pattern between class levels by using Multialign programs (Corpet, 1988; sequence datum of *A. salina* from Nelles et al. 1984). The results of the multialignment show that there are 54 insertion regions where two decapod species show nucleotide sequences but no sequences in branchiopod species. Thus these regions may be important in recognizing the difference of class levels of Crustacea.

Our results indicate that many additional complete sequences are necessary to determine the sequence variability of 18S rRNA gene of each species among different taxonomic categorical groups. These sequence variability will give more elucidation to construct the molecular phylogenies and to see the molecular evolutionary patterns among the groups of other eukaryotic organisms as well as crustaceans.

ABSTRACT

The primary sequence of the 18S rRNA gene of a crustacean *Pugettia quadridens* (Decapoda: Pleocyemata: Brachyura) was determined by the PCR cloning and Taq sequencing. The 18S rRNA gene of this species is 1837 bases long, and 46 bases shorter than that of another crustacean decapod *Oedignathus inermis*. The similarity between two species is 90.8% when the insertion and/or deletion sites were excluded. Within the molecule, the most conservative (identical) region locates at the position of 1137-1206 and it is 70 bases long. The most long consecutive nucleotide differences occur at the position between 46-55 and the second most between 399-407. The sequence variation in the primary structure of 18S rRNA gene are not evenly distributed throughout the molecule.

REFERENCES

Bowman, T. E. and L. G. Abele, 1982. Classification of the recent Crustacea. In: The biology of the Crustacea, vol. 1, Systematics, the fossil record and biogeography (Abele, L.G. ed.). Academic Press, 1-27.

Corpet F., 1988. Multiple sequence alignment with hierarchial clustering. Nucl. Acids Res., **16:** 10881-10890.

Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin, E. C. Raff, N. R. Pace and R. A. Raff, 1988. Molecular phylogeny of the animal kingdom. Science, **239:** 748-753.

Kim, W. and L. G. Abele, 1990. Molecular phylogeny of selected decapod crustaceans based on 18S rRNA nucleotide sequences. J. Crustacean Biol., **10:** 1-13.

Kim, W., G. S. Min and S. H. Kim, 1992. The 18S ribosomal RNA gene of a crustacean decapod *Oedignathus inermis:*

A comparison with *Artemia salina* gene. Nucl. Acids Res., **20:** 4658.

Nelles, L., B. L. Fang, G. Volckaert, A. Vandenberghe, and R. De Wachter. 1984. Nucleotide sequence of a crustacean 18S ribosomal RNA gene and secondary structure of eukaryotic small subunit ribosomal RNAs. Nucl. Acids Res., **12:** 8749-8768.

Pearson, W. R. and D. J. Lipman, 1988. Improved tools for biological sequence comparison. Proc. Nat. Acad. Sci. Washington, **85:** 2444-2448.

Stock, D. W. and G. S. Whitt, 1992. Evidence from 18S ribosomal RNA sequences that lampreys and hagfishes form a natural group. Science, **257:** 787-789.

RECEIVED: 5 OCTOBER 1992
ACCEPTED: 10 OCTOBER 1992