Korean J. Zool. 35:80-86, 1992

Nucleotide Analysis of 18S rRNA and Molecular Phylogeny of the Korean Decapods

Won Kim and Geum Jong Bae

The Department of Molecular Biology, College of Natural Sciences, Seoul National University, Seoul 151-742, Korea

The nucleotide sequences of 18S rRNAs of the five Korean decapods were partially determined by the direct sequencing method using the reverse transcriptase. The average GC content of five species was 51.1% which is higher than that of yeast (45.0%) and lower than those of frog (53.0%) and rat (55.6%). This result follows the general patterns of the GC content in the nucleotides of the nucleic acid shown among the various phylogenetic groups. The average ratio of transitional/transversional nucleotide substitution of pairwise comparison among six species (including *Artemia salina*) was 1.200 ± 0.310 when whole region was examined. However, the ratio showed some differences when the conservative regions and variable regions were separately examined. The molecular phylogenies of the five species were constructed by using two different tree making methods. In general the results support the previously reported molecular phylogeny of the decapod crustaceans. However, our results indicate that, in the analysis of the sequence data, the UPGMA clustering method of the distance matrix method should be carefully employed after considering the rate of nucleotide substitution in the different regions of the molecule.

KEY WORDS: 18S rRNA, Decapods, Molecular phylogeny

Recently the nucleotide sequences of small-subunit (5S, 16S, 18S, etc.) rRNAs have been extensively used in the study of molecular evolution, especially in the study of molecular phylogeny because these molecules bear several characteristics suitable for these studies (Olsen *et al.*, 1986; Woese, 1987; Field *et al.*, 1988; Sogin *et al.*, 1989). After the publication of the rapid sequencing method (Lane *et al.*, 1985), the application of nucleotide sequences of small-subunit rRNAs to the construction of the phylogenetic trees have been widespreading among various taxonomic groups (Abele *et al.*, 1989; Kelly-borges *et al.*, 1991; Turbeville *et al.*, 1991). In the case of decapod crustaceans, Kim and Abele (1990) tested ideas on the phylogenetic relationships among selected groups based on the necleotide sequences of 18S rRNA and they revealed that 18S rRNAs have the different rate of nucleotide substitution across the molecule. Using same 18S rRNA sequencing method we attempted in the present study to elucidate the following aspects. Firstly, the general features such as the GC content and transitional/transversional nucleotide substitution among five Korean decapods representing suborder or infraorder were examined. Secondly, whether the two different tree making methods show the same results regardless of the regional variability of the nucleotide substitution.

This work was supported by research grant from the Ministry of Education of Korea (Institute for Molecular Biology and Genetics), 1990.

We also compared our data with the previously published molecular phylogeny of the decapod crustaceans.

Materials and Methods

Materials

The five Korean decapod species were collected. These are Dendrobranchiata, Caridea, Astacidea and Brachiura representing suborder or infraorder groups of the decapods and listed in Table 1.

Table 1. A list of species sequenced for this study.

Taxa	Number of nucleotides
Crustacea	
Class Malacostraca	
Order Decapoda	
Suborder Dendrobranchiata	
Family Penaeidae	
Penaeus japonicus (Bate) 보리새우	1,071
Suborder Pleocyemata	
Family Atyidae	
Caridina denticulata denticulata De Haan 새뱅이	1,001
Infraorder Astacidae	
Family Cambaridae	
Cambaroides similis (Koelbel) 가재	1,424
Infraorder Brachyura	
Family Portunidae	
Portunus (*Portunus*) *trituberculatus* (Miers) 꽃게	1,093
Family Xanthidae	
Macromedaeus distinguendus (De Haan) 꽃부채게	1,181

Nucleotide sequencing

The 18S rRNA nucleotide sequences were determined by the direct sequencing method using the reverse transcriptase. The experimental procedures used in this work have been previously described in detail (Kim and Abele, 1990).

Data analysis

There is no single accepted algorithm of multiple alignment. In the present study we used the FASTA program (Pearson, 1990) and Multialign program (Corpet, 1988) in part. Since the complete sequences were reported for *Artemia salina* (Nelles *et al.*, 1984), we aligned the nucleotide sequences of each of five species against that of *A. salina* first, then realigned all the sequences derived from five species together. From these total aligned set of nucleotides the variable regions, based on the secondary structure of the molecule (V_1-V_7 regions according to Nelles *et al.*, 1984), and the conservative regions (Hasegawa *et al.*, 1985) were determined. The GC content was examined from each species. The ratio of transitional /transversional nucleotide substitution was examined from pairwise comparison.

The molecular phylogenetic trees were constructed by two different tree making methods, distance matrix and parsimony methods. In the distance matrix methods, the proportion of different nucleotide (p) and estimates of the number of nucleotide substitution per site (d) between pairs of species were calculated by the method of Jukes and Cantor (1969). For the construction of the tree, the UPGMA clustering method was employed using d values. The NTSYS-pc software (version 1.50) was used for this method. In the parsimony methods, the ALLTREES option and the method of invariant/operator metrics of PAUP program were used to construct the phylogenetic trees.

Results

The total number of nucleotides from the aligned set of the six species (including *A. salina*), after deleting the unreadable and gap regions, was 732. The average GC value among five species was 51.1%. The GC content of each of five decapod species is shown in Table 2. Except for *Cambaroides similis*, the GC content was slightly higher than AU content.

Table 3 shows the raw differences with transitions over transversions between species. The average ratio of transition/transversion between species was 1.200 ± 0.310 when the whole regions were examined. However, in the separate examination, we obtained the different ratio be-

Table 2. Number and percentage of each base (A, C, G, U) in nucleotide sequences of 18S rRNA for five taxa studied.

( ): %

	A	C	G	U	CG/AU	total
Cambaroides	357 (25)	340 (24)	370 (26)	357 (25)	0.994	1,424
Caridina	267 (27)	230 (23)	275 (27)	229 (23)	1.018	1,001
Macromedaeus	300 (25)	294 (25)	313 (27)	274 (23)	1.057	1,181
Penaeus	261 (24)	268 (25)	307 (29)	235 (22)	1.159	1,071
Portunus	279 (26)	252 (23)	295 (27)	267 (24)	1.002	1,093

Table 3. Pairwise comparison of the taxa used in this study. Lower lefts are raw differences with transitions over transversions. Upper rights are raw distances (*p*) with distances corrected (*d*) in parentheses by the method of Jukes and Cantor (1969). N = number of nucleotides.

	Artemia	*Cambaroides*	*Caridina*	*Macromedaeus*	*Penaeus*	*Portunus*
Artemia	—	0.122 (0.133)	0.123 (0.134)	0.107 (0.115)	0.157 (0.176)	0.119 (0.130)
Cambaroides	47/42	—	0.087 (0.092)	0.025 (0.025)	0.101 (0.108)	0.026 (0.026)
Caridina	48/42	31/33	—	0.082 (0.087)	0.135 (0.149)	0.111 (0.120)
Macromedaeus	48/30	8/10	29/31	—	0.107 (0.115)	0.034 (0.035)
Penaeus	58/57	44/30	58/41	49/29	—	0.115 (0.125)
Portunus	49/38	11/8	44/37	9/16	49/35	—

tween the conservative regions (1.520 ± 0.826) and the variable regions (0.974 ± 0.421).

Figs. 1, 2, and 3 show the phylogenetic trees among the six species when the UPGMA clustering method was employed to the sequence data obtained from the whole regions, conservative regions, and variable regions, respectively. The calculated *d* values are shown in Tables 3, 4, and 5. The difference among three trees appeared to be in the position of *A. salina*. It was branched off

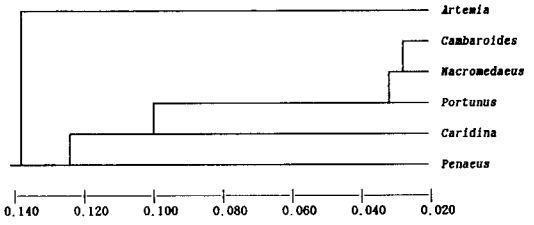

Fig. 1. Phylogenetic tree reconstructed by UPGMA method from the distance matrix in Table 3.

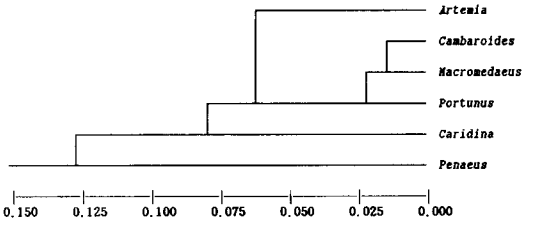

Fig. 2. Phylogenetic tree reconstructed by UPGMA method from the distance matrix in Table 4.

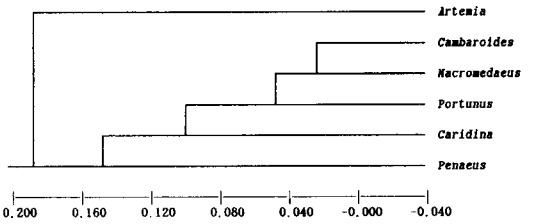

Fig. 3. Phylogenetic tree reconstructed by UPGMA method from the distance matrix in Table 5.

Table 4. Pairwise comparison of the taxa used in this study from the conservative regions of the 18S rRNA macromolecule (Hasegawa *et al.*, 1985). Lower lefts are raw differences with transitions over transversions. Upper rights are raw distances (*p*) with distances corrected (*d*) in parentheses by the method of Jukes and Cantor (1969). N = number of nucleotides.

	Artemia	*Cambaroides*	*Caridina*	*Macromedaeus*	*Penaeus*	*Portunus*
Artemia	—	0.060 (0.063)	0.086 (0.091)	0.050 (0.052)	0.129 (0.142)	0.067 (0.070)
Cambaroides	19/6	—	0.072 (0.076)	0.012 (0.012)	0.100 (0.107)	0.021 (0.021)
Caridina	22/14	18/12	—	0.062 (0.065)	0.143 (0.159)	0.074 (0.078)
Macromedaeus	16/5	3/2	16/10	—	0.098 (0.105)	0.019 (0.019)
Penaeus	34/20	26/16	34/26	26/15	—	0.112 (0.121)
Portunus	16/12	3/6	15/16	0/8	25/22	—

Table 5. Pairwise comparison of the taxa used in this study from the variable regions (V_1-V_7) of the 18S rRNA macromolecule (Nelles *et al.*, 1984). Lower lefts are raw differences with transitions over transversion. Upper rights are raw distances (*p*) with distances corrected (*d*) in parentheses by the method of Jukes and Cantor (1969). N = number of nucleotides.

	Artemia	*Cambaroides*	*Caridina*	*Macromedaeus*	*Penaeus*	*Portunus*
Artemia	—	0.140 (0.155)	0.157 (0.176)	0.149 (0.166)	0.190 (0.219)	0.174 (0.198)
Cambaroides	17/17	—	0.083 (0.088)	0.017 (0.017)	0.112 (0.121)	0.033 (0.034)
Caridina	18/20	8/12	—	0.075 (0.079)	0.154 (0.172)	0.112 (0.121)
Macromedaeus	20/16	1/3	8/10	—	0.112 (0.121)	0.050 (0.052)
Penaeus	22/24	16/11	18/19	18/9	—	0.141 (0.156)
Portunus	21/41	4/4	11/'6	5/7	20/14	—

first in the trees of Fig. 1 and Fig. 3 but, in contrast, branched off one leading to *Portunus-Macromedaeus-Cambaroides* in Fig. 2. Figs. 4 and 5 show the phylogenetic trees resulting from the parsimony method using ALLTREES option and evolutionary parsimony method (invariant/operator metrics), respectively.

The two trees showed the same topology each other but some differences from Figs 1, and 3 in the positions of *Penaeus/Caridina* and *Macromedaeus/Portunus*.

Discussion

There seems to be a tendency that the GC content in 18S rRNA molecule increases as the organisms evolve to the complex forms. Torczynski *et al.* (1983) showed that the GC contents of yeast, frog, and rat are 45.0%, 53.0%, and 55.6%. Salim and Maden (1981) also suggested that the point mutation from AT to GC occurs more frequently than mutation from GC to AT and that the insertion of GC takes place more often than AT does. Our results of the average GC content from

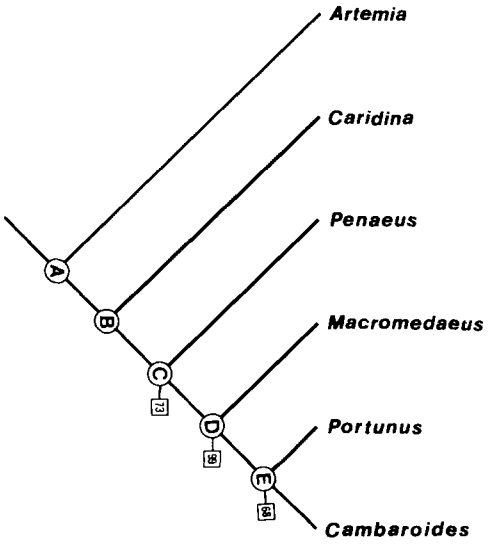

Fig. 4. Relationships among the taxa considered as estimated by PAUP using the ALLTREES option. Number = An estimate of the confidence intervals of the tree by the bootstrap method based on 100 replicates.

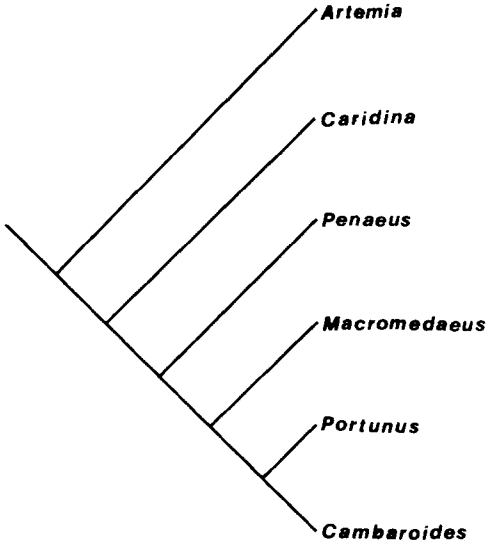

Fig. 5. Relationship among the taxa considered as estimated by the method of invariants/operator metrics.

five decapod species used here was 51.1%. This value is surely higher than that of yeast and lower than those of frog and rat. Therefore, the present result shows the consistency with the general evolutionary tendency in the GC content mentioned above.

The nucleotide substitutions consist of transition and, transversion and, theoretically, transversion is anticipated to be occurred twice more than transition if substitution occurs randomly among the four types of nucleotide (Nei, 1987). However, the rate of nucleotide substitution actually does not occur in random. Gojobori *et al.* (1982) showed that the substitution rate of G to A was 20.9 ± 3.1 and substitution rate of T to C was 3.7 ± 1.6 in the globin and ACTH (adrenocorticotrophic hormone) genes. This means that the patterns of the nucleotide substitutions are quite different among the four types of nucleotides. Hasegawa *et al.* (1985) also recognized that the transitional mutation occurs twice more than the transversional mutation in 18S rRNA molecules and tried to adjust this value to construct the phylogenetic trees. As Table 3 shows, the transitional substitution occurs more frequently than transversional substitution does in the decapod crustaceans when the whole regions were considered. However, analysis separately conducted with variable regions and conservative regions showed that the ratio of transitional substitution to transversional substitution is higher in conservative regions and lower in variable regions than in the whole sequence. Therefore, this results suggest that the rate of nucleotide substitution is not equal along the molecule, but quite different among the various regions of the molecule. However, most authors employed the partial sequences for the construction of trees without questioning whether the partial sequences are the representative data of the whole molecule.

As shown in the Figs. 1, 2, and 3, when it was employed in the data obtained from the different regions of the molecules, the UPGMA clustering method resulted in the different phylogenetic trees. We assume that these differences were derived from the different rate of nucleotide substitution. However, the UPGMA method treats with the equal weight of the characters. The present results therefore suggest that character-weighting should be used when the UPGMA method is employed for the whole sequences. If not, the variable region, neither whole sequences nor conservative regions, should be used in the construction of the phylogenetic tree at least within the

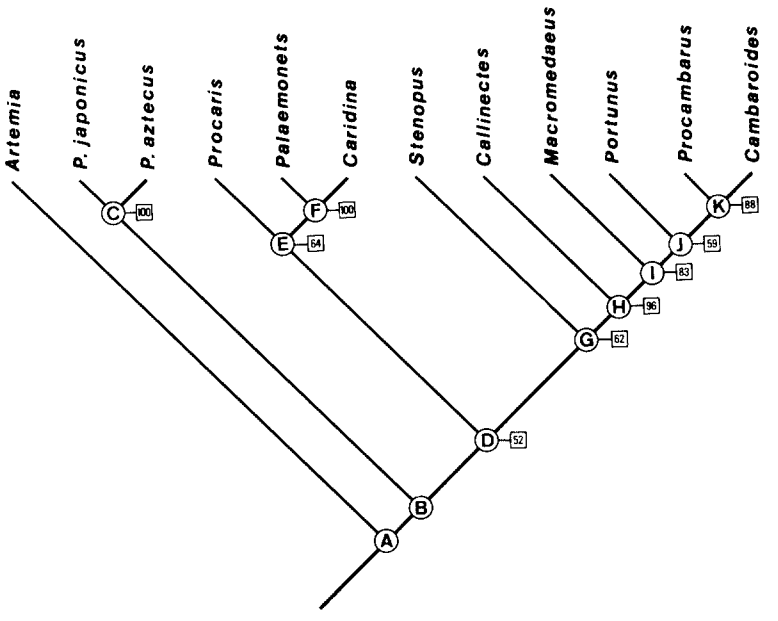

Fig. 6. Relationships among the decapods as estimated by PAUP using the ALLTREES option. Number = An estimate of the confidence intervals of the tree by the bootstrap method based on 100 replicates.

"order" level. The nucleotide sequences from the various organisms used in the construction of molecular phylogeny so far were obtained from the conservative regions and therefore these sequences should be used in the higher categorical rank, possibly between or above the "order" level.

Figs. 4 and 5 show the parsimony results and they look slightly different from those of Kim and Abele (1990). The reason may be due to the less number of species examined because the smaller number of species can affect the numbers of informative site and the transition and transversion. However, the overall topology of our results agrees well with the trees of Kim and Abele (1990).

In the study of the phylogeny of the decapod crustaceans, Kim and Abele (1990) concluded that there was general agreement between results obtained from morphological and molecular data, but also suggested that additional data would be necessary to determine in greater detail. We combined our data with those of Kim and Abele (1990) and reconstructed the phylogenetic trees by the ALLTREES option of PAUP program. As shown in Fig. 6, the resulting trees reinforce the previous morphological and molecular phylogenies of the decapod crustaceans and support the Kim and Abele's conclusion concerning the phylogeny among the major decapod groups.

In summary, the nucleotide sequence data from 18S rRNA provide the useful tool for the study of the molecular evolution, especially for the construction of the phylogenetic trees. However, the application of the partial sequences should be warranted in the use of the UPGMA clustering method because 18S rRNA molecule shows the different rate of nucleotide substitution across the molecule. Additional study concerning the analysis of the complete nucleotide sequences is neccessary to understand the nature of the 18S molecules and to construct more reliable phylogenetic trees with the proper method.

References

Abele, L. G., W. Kim, and B. E. Felgenhauer, 1989. Phylum Pentastomida: molecular evidence for inclusion in the crustacea. *Mol. Biol. Evol.* **6:** 685-691.

Corpet, F., 1988. Multiple sequence alignment with hierarchial clustering. *Nucleic 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.

Gojobori, T., K. Ishii, and M. Nei, 1982. Estimation of average number of nucleotide substitutions when the rate of substitution varies with nucleotide. *J. Mol. Evol.* **18:** 414-423.

Hasegawa, M., Y. Iida, T. Yano, F. Takaiwa, and M. Iwabuchi, 1985. Phylogenetic relationships among eukaryotic kingdoms inferred from ribosomal RNA sequences. *J. Mol. Evol.* **22:** 32-38.

Jukes, T. H., and R. C. Cantor, 1969. Evolution of protein molecules. *In* : H. N. Munro (ed.), Mammalian protein metabolism. pp. 21-133. Academic Press, New York.

Kelly-Borges, M., P. R. Bergquist, and P. L. Bergquist, 1991. Phylogenetic relationships within the order Hadromerida (Porifera, Demospongiae, Tetractinomorjpha) as indicated by ribosomal RNA sequence comparisons. *Biochem. Syst. Ecol.* **19:** 117-125.

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

Lane, D. J., B. Pace, G. J. Olsen, O. A. Stahl, M. L. Sogin, and N. R. Pace, 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analysis. *Proc. Nat. Acad. Sci. U.S.A.* **82:** 6955-6959.

Nei, M., 1987. Molecular Evolutionary Genetics. Columbia University Press, New York. 512 pages.

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. *Nucleic Acids Res.* **12:** 8749-8768.

Olsen, G. J., D. J. Lane, S. J. Giovannoni, N. R. Pace, and D. A. Stahl, 1986. Microbial ecology and evolution: a ribosomal RNA approach. *Ann. Rev. Microbio.* **40:** 337-365.

Pearson, W. R., 1990. Rapid and sensitive sequence comparison with FASTP and FASTA. *In* : R. F. Doolittle (ed.), Molecular evolution: computer analysis of protein and nucleic acid sequences. Adademic Press, Inc. pp. 1-736.

Salim, M., and B. E. H. Maden, 1981. Nucleotide sequence of *Xenopus laevis* 18S ribosomal RNA inferred from gene sequence. *Nature* **291:** 205-208.

Sogin, M. L., U. Edman, and H. Elwood, 1989. A single kingdom of eukaryotes. *In* : Fernholm B., Bremer K., H. Jörnvall (Eds.), The Hierarchy of Life. Amsterdam: Elsevier, pp. 133-143.

Torczynski, R. M., A. P. Bollon, and M. Fuke, 1983. The complete nucleotide sequence of the rat 18S ribosomal RNA gene and comparison with the respective yeast and frog genes. *Nucleic Acids Res.* **11:** 4879-4890.

Turbeville, J. M., D. M. Pfeifer, K. G. Field, and R. A. Raff, 1991. The phylogenetic status of Arthropods, as inferred from 18S rRNA sequences. *Mol. Biol. Evol.* **8:** 669-686.

Woese, C. R., 1987. Macroevolution in the microscopic world. pp. 177-202. *In* : C. Patterson (ed.), Molecules and Morphology in Evolution: Conflict or Compromise? Cambridge University Press, Cambridge.

(Accepted November 30, 1991)

한국산 십각류의 18S 리보솜 RNA의 염기분석과 분자계통에 관한 연구

김 원 · 배금종(서울대학교 자연과학대학 분자생물학과)

한국산 십각류 5종에 대한 18S 리보솜 RNA 염기서열을 역전사효소를 이용한 염기서열 결정법을 사용하여 얻었다. 5종의 평균 GC값은 51.1%로 효모(45.0%)보다는 높고, 개구리(53.0%)와 쥐(55.6%)보다는 낮았다. 이 결과는 여러 계통간 생물들 사이에서 나타나는 핵산 염기의 GC값의 일반적 경향과 일치한다. *Artemia salina*를 포함한 6종들 사이의 분자 전체를 조사했을 때 2종간의 transitional/transversional 평균 염기 치환율은 1.200 ± 0.310이었다. 그러나 변이가 심한 부위와 변이가 적은 부위를 각각 조사하였을 경우에는 이와 약간의 차이를 보였다. 또한 계통수를 추성하는 두 종류의 방법을 사용하여 5종에 대한 분자계통을 구성하였다. 본 결과는 일반적으로 이미 보고된 십각 갑각류의 분자계통과 일치하고 있다. 그러나 우리의 결과는 염기서열의 분석을 위해 distance matrix 방법의 UPGMA 집괴방법을 사용할 때에는 부위에 따른 염기치환율까지 고려하여 보다 신중히 사용해야 할 것을 시사하고 있다.