J Mol Evol (1999) 49:806–809

JOURNAL OF MOLECULAR EVOLUTION

Ribosomal DNA Intergenic Spacer of the Swimming Crab, *Charybdis japonica*

Seong Ho Ryu, Yoon Kyung Do, Ui Wook Hwang, Chong Pyo Choe, Won Kim

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

Received: 13 April 1999 / Accepted: 2 August 1999

Abstract. We have determined the full sequence of the ribosomal DNA intergenic spacer (IGS) of the swimming crab, *Charybdis japonica*, by long PCR for the first time in crustacean decapods. The IGS is 5376 bp long and contains two nonrepetitive regions separated by one long repetitive region, which is composed mainly of four subrepeats (subrepeats I, II, III, and IV). Subrepeat I contains nine copies of a 60-bp repeat unit, in which two similar repeat types (60 bp-a and 60 bp-b) occur alternatively. Subrepeat II consists of nine successive repeat units with a consensus sequence length of 142 bp. Subrepeat III consists of seven copies of another 60-bp repeat unit (60 bp-c) whose sequence is complementary to that of subrepeat I. Immediately downstream of subrepeat III is subrepeat IV, consisting of three copies of a 391-bp repeat unit. Based on comparative analysis among the subrepeats and repeat units, a possible evolutionary process responsible for the formation of the repetitive region is inferred, which involves the duplication of a 60-bp subrepeat unit (60 bp-c) as a prototype.

Key words: *Charybdis japonica* — Ribosomal DNA — Intergenic spacer — Subrepeat — Repeat unit

The nuclear ribosomal RNA gene (rDNA) of higher eukaryotes is a highly repetitive invariant gene. It is composed of transcription units containing the 18S, 5.8S, and 28S rRNA coding regions and the intergenic spacer (IGS) separating the 28S and 18S rRNA coding regions. The IGS contains some regulatory elements and is typically composed of reiterated subrepeats (Mandal 1984; Hemleben et al. 1987). It has been reported that differences in the number and sequence of these subrepeats account for most of the length variation between rDNA repeat units within closely related species, among populations, and even between individuals (Gerbi 1985; Murtif and Rae 1985; Rogers and Bendich 1987; Black et al. 1989). Despite the enormous variability in length and primary sequence, the IGS region of higher eukaryotes has broadly conserved structural features such as the existence of several kinds of repeating elements (subrepeats), repetitive enhancer elements, duplicated promoters, and conserved secondary structures (Hemleben et al. 1987; Soller-Webb and Tower 1986; Reeder 1989; Linares et al. 1991; Baldridge et al. 1992). These structural features may contain some information useful for molecular evolutionary studies in general. However, a comparative analysis of the IGS sequences to dig out the evolutionary information buried in these structural features has not been conducted, because the full sequences of IGS have been reported in only a few species. In Crustacea, no full sequence of IGS has been reported in Decapoda, the group considered to be most advanced within Crustacea. In the present study, we obtained the complete sequence of the IGS from the swimming crab, *Charybdis japonica*, and analyzed this sequence to investigate (1) whether the IGS of the swimming crab shows the typical repetitive structural features found in other eukaryotes and (2) which plausible evolutionary process is responsible for the formation of the repetitive region of the IGS.

The IGS was amplified from total genomic DNA of *C. japonica* using the Long and Accurate PCR (LA-

Correspondence to: Dr. Won Kim; *e-mail:* wonkim@plaza.snu.ac.kr

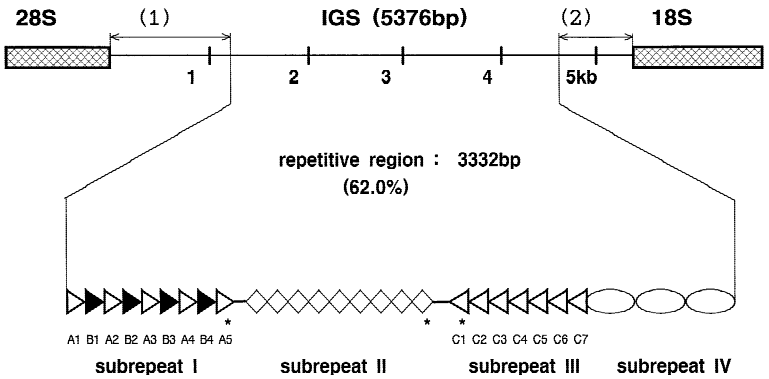

Fig. 1. Structural organization of the *C. japonica* rDNA repeat. The 1195-bp nonrepetititive region 1 and the 813-bp nonrepetitive region 2 are marked by the *numbers in parentheses* at the *top.* The 3332-bp repetitive region is composed of four subrepeats. The tandem array of nine 60-bp repeat units is represented by *arrowheads: open arrowheads,* 60 bp-a; *filled arrowheads,* 60 bp-b. The *letters above* mark each type of 60-bp repeat unit. The array of nine *open diamonds* represents the 142-bp repeat units. The *short vertical lines* located between the subrepeats represent flanked sequences. The open arrowheads reversed in relation to subrepeat I represent seven 60 bp-c repeat units. As indicated by the direction of the heads, the 60-bp repeat units of subrepeats I and III are complementary to each other. Without flanked sequences, the array of the three 391-bp repeat units is represented by *open ovals.* The *asterisks* indicate the truncated repeat unit.

PCR) kit (Takara Co.) following the manufacturer's recommendations. Since LA-PCR has a prolonged annealing time and requires long amplification primers of ≥30-mers, long primers were selected from the 3′ terminus of the 28S rDNA and the 5′-terminus of the 18S rDNA (the 28S rDNA end primer, 5′-tagggaacgtgagctgggtttagaccgtcgtga-3′ and the 18S rDNA end primer, 5′-gagacaagcatatgctactggcaggatcaacc-3′). The amplified products were extracted with the Qiagen PCR purification kit (Qiagene, Santa Clarita, CA). Ligation of the purified PCR products was performed with T4 ligase using pT7blue T-vector (Novagen Co.) and transformed to the DH5-α cell line. The clone, p15HRD, was digested with several restriction enzymes and the resulting fragments were subcloned into pGEM 3Z(+). Overlapping clones were sequenced with an automatic sequencer (ALF Express Co.) and the Sequenase Version 2.0 kit (USB Biochemicals Co.). Sequences were analyzed using the MacDNASIS Version 3.0 program (Hitachi Software Engineering Co.) and aligned by the FASTA (Pearson 1990) and the Clustal V (Higgins et al. 1992) programs.

Long PCR amplification of the IGS region resulted in one band on the PAGE gel, suggesting that *C. japonica* has no length variation of the IGS. The total length of the IGS was revealed as 5376 bp when the 3′ end of the 28S rDNA coding region and the 5′ end of the 18S rDNA coding region were determined by alignment with the sequences of *Daphnia pulex, Drosophilia melanogaster,* and the calanoid copepod, *Calamus finmarchicus* (Crease 1993; Tautz et al. 1988; Drouin et al. 1987; Nelles et al. 1984). It is the longest among the IGS known in arthropods. The overall primary structure of the IGS is organized into two nonrepetitive regions separated by one repetitive region as shown in Fig. 1. Downstream from the 3′ end of the 28S rDNA, a 1195-bp nonrepetitive region occurs, which is followed by a repetitive region and then another nonrepetitive region of 813 bp. The repetitive region is 3332 bp long and is composed of four subrepeats (termed subrepeats I, II, III, and IV, in the 5′–3′ direction). In subrepeat I, nine duplicated 60-bp repeat units were identified and two types of repeat units, termed 60 bp-a (A1–A5) and 60 bp-b (B1–B4) based on sequence similarity, were recognized. These two types appear alternatively and the last one (A5) is truncated. Eleven base pairs downstream from subrepeat I is subrepeat II, which contains nine successive repeat units, the first eight units being 142 bp long and the last one shorter by 29 bp. Subrepeat III consists of seven copies of another 60-bp repeat unit. However, the sequences of this repeat unit, termed 60 bp-c, are complementary to the 60-bp repeat units of subrepeat I. The first 60 bp-c repeat unit is truncated by about 30 bp nucleotides at the 5′ end. Immediately downstream from subrepeat II is subrepeat IV, which includes three copies of 391-bp repeat units. The last 30-bp sequence toward the 3′ end of each of the 391-bp repeat units is exactly the same as that of 60 bp-c.

The complementary primary structures of subrepeats I and III suggest the potential secondary structure of the IGS. Recent computer-based analyses of the IGS in eukaryotes proposed conservation of the potential secondary structure. Moreover, a few of them revealed extensive regions of self-complementarity that could generate an extensive secondary structure in the IGS (Linares et al. 1991; Hancock and Dover 1990; Baldridge et al. 1992). We could deduce a stem–loop structure from the sequences of subrepeats I, II, and III, where subrepeat I and III are responsible for the stem and subrepeat II for the loop. Consequently, the observation that a similar portion of the last unit (A5) of subrepeat I and the first unit (C1) of subrepeat III was deleted may reflect the involvement of these motifs in the same evolutionary events. Additionally, taking into account the truncated 142-bp repeat unit, it is interesting to note that the terminal repeats are less homogenized than the internal repeats. This fact was also observed in the IGS of the mosquito, *A. albopictus* (Baldridge and Fallon 1992).

The structural proportion of the four subrepeats is about 62% of the entire IGS sequence, which is almost twice that of *Daphnia*, 34.3% (Crease 1993). Sequence similarities among the 60-bp repeat units of subrepeat I (60 bp-a, 60 bp-b) and those of subrepeat III (60 bp-c) ranged from 85 to 87%. The high similarity values among the three kinds of 60-bp repeat units and their complementary sequences enabled us to consider the existence of one ancestral common prototype sequence in the formation of the repetitive region. We suggest the 60 bp-c repeat unit of subrepeat III as a prototype for the other 60-bp repeat units (60 bp-a and 60 bp-b), the 142-bp repeat unit of subrepeat II, and the 391-bp (repeat unit) of subrepeat IV, since the similarity value of 60 bp-c/60 bp-a or 60 bp-c/60 bp-b is higher than that of 60 bp-a/60 bp-b and the last 30-bp sequence toward the 3′ end of the 391-bp repeat unit is exactly the same as that of 60 bp-c. Furthermore, our assumption that all subrepeats of the IGS in the swimming crab, *C. japonica*, were derived from one prototype, 60 bp-c, gave us an idea about sequence alignments among several subrepeats of different lengths. The 142- and the 391-bp repeat units were split into three and eight parts, respectively, by referring to maximum alignment with tandem c3–60-bp repeat units. The overall similarity values between subrepeat II and three copies of 60 bp-c and between subrepeat IV and eight copies of 60 bp-c are 52.2 and 53.2% respectively. These values can be obtained from averaging the similarity values of each alignment set since each set value does not fluctuate severely. Moreover, these alignment pairs suggest the construction of the 142- and 391-bp repeat units from several copies of 60 bp-c after each 60 bp-c unit mutated independently (Fig. 2). The sequences of 60 bp-a, 60 bp-b, and 60 bp-c were probably conserved because of their functional importance, such as the formation of the secondary structure.

Based on the observations of the sequence similarities among the subrepeat units, we roughly infer the overall evolutionary process for the formation of the repetitive region of the IGS in the swimming carb, *C. japonica* (Fig. 2). At first, the chromosome containing the prototype 60 bp-c (C) went through inversion (C′). Then unequal crossover produced a new strand containing C′ and C (steps 1–3). After the first tandem repeat of the 60-bp repeat units (C′C′ and CC; step 4), the 60-bp repeat units (C′C′) toward the 5′ end were changed into types A and B (step 5), and then each unit was duplicated, to result in a tandem array of repeat structure (step 6). Finally, three and eight copies of the 60-bp repeat unit, of which the sequence was already mutated (step 7), were joined together to construct the 142- and the 391-bp repeat units, respectively (step 8). Finally, the third tandem repeat produced nine copies of the 142-bp repeat unit and three copies of the 391-bp repeat unit (step 9).

Since the present study is the first report of the full sequence of the IGS in Decapoda, the evolutionary process proposed above should be reinforced with the determination of more full sequences of various species both closely and remotely related. Such data will allow the discovery of other IGS structural features concerning the evolutionary relationship among repeat elements and shed more light on understanding the evolutionary mechanism of the formation of the repetitive region.

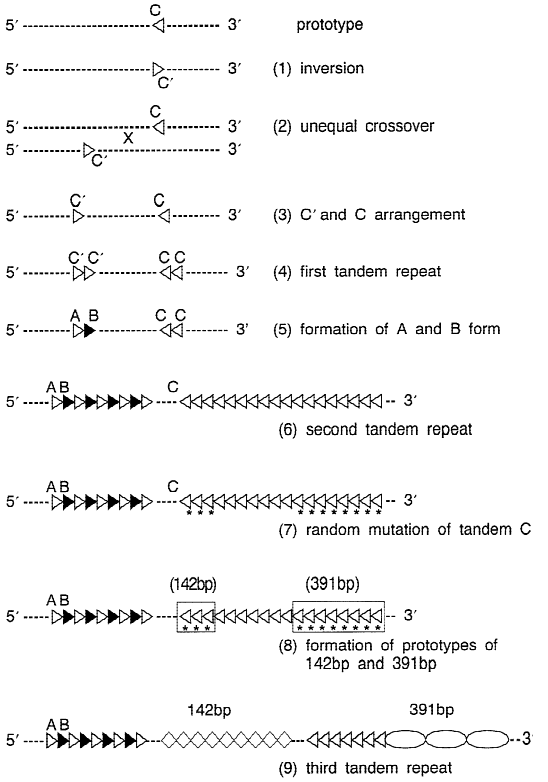

Fig. 2. A model demonstrating how the subrepeats in the IGS region of the swimming crab could be derived from a common ancestor. The *numbers in parentheses* indicate each step for completing the structure. The X in step 2 indicates a crossover between two strands; C′ represents a complementary sequence to C. The *asterisks* in step 7 represent random mutations and the *rectangles* in step 8, which encompass three and eight copies of C, represent prototypes for the 142- and the 391-bp subrepeats respectively. For the complete IGS structure, symbols, and numbers of the elements, see Fig. 1.

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

References

Baldridge GD, Fallon AM (1992) Primary structure of the ribosomal DNA intergenic spacer from the mosquito, *Ades albopicus*. DNA Cell Biol 11:51–59

Baldridge GD, Dalton M, Fallon AM (1992) Is higher-order structure conserved in eukaryotic ribosomal DNA intergenic spacers? J Mol Evol 35:514–523

Black WC, McLain DK, Rai KS (1989) Patterns of variation in the rDNA cistron within and among world populations of a mosquito, *Aedes albopictus* (Skuse). Genetics 121:539–550

Crease TJ (1993) Sequence of the intergenic spacer between the 28S and 18S rDNA-encoding genes of the crustacean, *Daphnia pulex*. Gene 134:245–249

Drouin G, Hofman JD, Dolittle WF (1987) Unusual ribosomal RNA gene organization in copepods of the genus *Calanus*. J Mol Biol 196:943–946

Gerbi SA (1985) Evolution of ribosomal DNA. In: MacIntyre RJ (ed) Molecular evolutionary genetics. Plenum Press, New York, pp 419–517

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

Hemleben V, Ganal M, Gerstner J, Scheibd K, Torres R (1987) Organization and length heterogeneity of plant ribosomal RNA genes. In: Kahl G (ed) Architecture of eukaryotic genes. VHC, Weinheim, pp 371–383

Higgins DG, Bleasby AJ, Fuchs R (1992) Clustal V: Improved software for multiple sequence alignment. CABIOS 8:189–191

Linares AR, Hancock JM, Dover GA (1991) Secondary structure constraints on the evolution of Drosophila 28S ribosomal RNA expansion segments. J Mol Evol 35:381–390

Mandal R (1984) The organization and transcription of eukaryotic ribosomal RNA genes. Prog Nucleic Acids Res Mol Biol 31:115–160

Murtif VL, Rae PMM (1985) *In vivo* transcription of rDNA spacers in *Drosophila*. Nucleic Acids Res 13:3221–3239

Nelles L, Fang B-L, Volckaert G, Vaderberghe A, De Wachter R (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.

Pearson WR (1990) Rapid and sensitive sequence comparison with FASTP and FASTA. In: Doolittle RF (ed) Molecular evolution: Computer analysis of protein and nucleotide sequences. Academic Press, San Diego, CA, pp 63–98

Reeder RH (1989) Regulatory elements of the generic ribosomal gene. Curr Opin Cell Biol 1:466–474

Rogers SO, Bendich AJ (1987) Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Mol Biol 9:509–520

Soller-Webb B, Tower J (1986) Transcription of cloned ribosomal RNA genes. Annu Rev Biochem 55:810–830

Tautz D, Hancock JM, Webb DA, Tautz C, Dover GA (1988) Complete sequences of the rDNA genes of *Drosophila melanogaster*. Mol Biol Evol 5:366–376