Journal of Heredity 2008:99(4):432–437
doi:10.1093/jhered/esn024
Advance Access publication May 13, 2008

The Organization of the Mitochondrial Control Region in 2 Brachyuran Crustaceans: *Ucides cordatus* (Ocypodidae) and *Cardisoma guanhumi* (Gecarcinidae)

MARCIO R. PIE, JOSÉ F. OLIVEIRA-NETO, WALTER A. BOEGER, ANTONIO OSTRENSKY, AND RAFAEL A. BAGGIO

From the Departamento de Zoologia, Universidade Federal do Paraná, Caixa Postal 19073, Curitiba, PR 81531-990, Brazil (Pie, Oliveira-Neto, Boeger, and Baggio); the Grupo Integrado de Aqüicultura e Estudos Ambientais, Universidade Federal do Paraná, Curitiba, Parana, Brazil (Pie, Oliveira-Neto, Boeger, Baggio, and Ostrensky); and the Departamento de Zootecnia, Universidade Federal do Paraná, Curitiba, Parana, Brazil (Ostrensky).

Address correspondence to M. R. Pie at the address above, or e-mail: pie@ufpr.br.

The control region (CR) is the largest noncoding segment of the mitochondrial DNA and includes the major regulatory elements for its replication and expression. In addition, the high level of intraspecific genetic variability found in the CR favors its use in phylogeographical and population genetic studies of a variety of organisms. However, most of the work on the structure of the CR has focused on vertebrates and insects, and little is known about the evolution of the CR in other taxa. In this study, we sequenced the entire CR of several individuals of 2 crab species: *Ucides cordatus* (Ocypodidae) and *Cardisoma guanhumi* (Gecarcinidae). There were neither large conserved regions in the CR of either species nor any similarity among species at the nucleotide level. However, the spatial pattern of genetic variability on the CR was similar among species. In addition, interesting similarities were found in the formation of stable secondary structures and in the position of regulatory elements. These results indicate that the evolution of CR in crustaceans is a remarkably dynamic process, with most homology among species being found at the secondary level.

The control region (CR) of the mtDNA, often called D-loop in vertebrates and "A + T"–rich region in invertebrates, contains the major regulatory elements for the replication and expression of the mitochondrial genome (Shadel and Clayton 1997). This region is characterized by an extraordinarily dynamic evolution. For instance, CR size in insects can range from 0.35 kb in butterflies (Taylor et al. 1993) to 13 kb in bark weevils (Boyce et al. 1989). The structure of the CR is also variable among animal groups. In mammals and birds, the CR is organized into 3 major regions, or domains, including the extended terminal associated sequence (ETAS), central, and conserved sequence block domains (e.g., Sbisà et al. 1997; Randi and Lucchini 1998; Matson and Baker 2001). However, such an organization is not shared by all vertebrate groups (e.g., Brehm et al. 2003). In insects, on the other hand, there seem to be 2 main types of CRs (Taylor et al. 1993; Zhang et al. 1995; Zhang and Hewitt 1996; Vila and Björklund 2004): Group 1, where a conserved domain is followed by a variable domain, is found in fruit flies, and Group 2, found in grasshoppers, locusts, butterflies, and mosquitoes, is characterized by a lack of distinct conserved regions.

Surprisingly, little is known about the structure of the CR in crustaceans. Grabowski and Stuck (1998) described the CR of the shrimp *Farfantepenaeus duorarum* with respect to its size, base composition, and the presence of 7–12 short repetitive sequences. Also, Diniz et al. (2005) studied the variability pattern and the base composition of the hypervariable region of the CR of the spiny lobster (*Panulirus argus*) to investigate its usefulness in phylogeographical studies. Finally, Kilpert and Podsiadlowski (2006) identified 2 sections with repetitive sequences in the isopod *Ligia oceanica*. The first consists of a series of 4 completely matching sequences of 10 bp extending into the adjacent tRNA, whereas the second section is formed by a consecutive triplicate 64-bp segment. No similarities were found between these sequences and any other mitochondrial gene. In addition, the position of the regulatory elements in *L. oceanica* indicates that the CR might have been inverted during the evolution of isopods.

In this study, we sequenced the entire CR of several individuals of 2 crab species: *Ucides cordatus* (Ocypodidae) and *Cardisoma guanhumi* (Gecarcinidae). Intra- and interspecific comparisons were used to describe the organization of the CR in these species as well as to search for possible structural similarities between them.

Materials and Methods

Samples of *U. cordatus* were collected in the Guaratuba bay, State of Paraná, Southern Brazil (25°50′14″S, 48°35′20″W), and samples of *C. guanhumi* were obtained in a local market in Aracaju, State of Sergipe, Northeastern Brazil (10°59′06″S, 37°04′24″W). Muscle tissue from one of the pereiopods of each specimen was removed, preserved in ethylenediaminetetraacetic acid–dimethyl sulfoxide buffer (Seutin et al. 1991), and maintained at −20 °C. Genomic DNA was extracted using the DNeasy kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. The primers 12SUCAF3 (5′-CCA GTA NRC CTA CTA TGT TAC GAC TTA T-3′) and ILEUCAR3 (5′-GCT AYC CTT TTA AAT CAG GCA C-3′) were used for the amplification of a ≈1.6-kb fragment including the complete CR (Oliveira-Neto et al. forthcoming). Each 25-μl polymerase chain reaction (PCR) included the following final concentrations: 6 mM of $MgCl_2$, 0.25 mM of each dNTP, 0.1 U/μl of *Taq* polymerase, 1× de buffer, 2 μM of each primer, and 1.2 ng/μl of template DNA. Thermocycling conditions included an initial denaturation at 95 °C for 2 min, followed by 35 cycles of 95 °C for 20 s, 55 °C for 30 s, and 72 °C for 90 s, and a final extension at 72 °C for 2 min. A 2-μl aliquot of each PCR product was electrophoresed in a 1.5% agarose gel, stained with ethidium bromide and visualized under ultraviolet light. Successfully amplified products were purified using a MinElute kit (Qiagen). Cycle sequencing in 10-μl solutions included the following final concentrations: 5 ng/μl of template DNA, 0.16 μM of primer, 0.15× of reaction buffer, and 0.5 μl of BigDye (Applied Biosystems, Foster City, CA). The final product was purified using Sephadex G50 and processed on an ABI3130 automatic sequencer. Forward and reverse strands were reconciled using the Staden package (Staden 1996). Five and 10 individuals were sequenced for *U. cordatus* and *C. guanhumi*, respectively. Sequences were aligned using ClustalX (Thompson et al. 1997), followed by visual inspection of the resulting alignments. All sequences were deposited in GenBank (accession numbers EU573697-EU573701, EU573687-EU573696). The limits of the CR were determined based on the genome of *Portunus trituberculatus*, the most closely related crustacean for which the complete mitochondrial genome has been characterized (Yamauchi et al. 2003).

Variation in the level of conservation along each studied alignment was obtained as an entropy function of nucleotide variation using the following equation: $\mathrm{Var} = -\sum_i = a, c, t, g \frac{n_i}{N} \ln \frac{n_i}{N}$, where n_i = the numbers of each nucleotide (G, A, C, T, or U) in a column of the alignment and N = total number of sequences analyzed, as implemented in the software SWAN (Proutski and Holmes 1998). The entropy function was calculated in a 10-bp sliding window along the studied fragment. The used window size is arbitrary, but the qualitative results are robust, even if different window sizes are used (data not shown). The most appropriate model of molecular evolution for the CR of each species was estimated using the software Modeltest 3.7, followed by hierarchical comparisons using the Akaike Information Criterion (Posada and Crandall 1998). Tandem repeat sequences, which might indicate the presence of regulatory elements, were searched using the software MREPS (Kolpakov et al. 2003). In addition, secondary structures and folding energies were determined using the software Mfold (Zucker 2003). Finally, potential promoter elements were searched using Proscan version 1.7 (Prestidge 1995).

Results and Discussion

Alignments of the obtained CR sequences of *U. cordatus* and *C. guanhumi* are shown in Figures 1 and 2, and a description of their basic features is shown in Table 1. There is a bias against G in both species, as commonly found in vertebrates (Wolstenholme 1992). Comparisons of the likelihood scores of alternative models using Modeltest indicated the need for fairly complex models to describe the evolution of the CR in either species. The best model for *C. guanhumi* was TIM + I + Γ using the following parameters: base = (0.4300, 0.1449, 0.0854), Nst = 6, Rmat = (1.0000, 18.8797, 0.2125, 0.2125, 8.4991), rates = gamma, shape = 0.5148, pinvar = 0.6374, whereas the best model for *U. cordatus* was TIM + Γ using the following parameters: base = (0.3968, 0.1617, 0.0694), Nst = 6, Rmat = (1.0000, 11.8967, 0.1320, 0.1320, 6.0735), rates = gamma, shape = 0.1245, pinvar = 0. Average distances among CR haplotypes using those models were 0.049 ± 0.013 and 0.074 ± 0.019 (mean ± standard deviation) for *C. guanhumi* and *U. cordatus*, respectively. These levels are more than 60% higher than those estimated from uncorrected average pairwise distances (0.031 and 0.044, respectively) or using a simple model of sequence evolution such as the K2P (0.032 and 0.046, respectively). Thus, the evolution of the CR cannot be described by such a simple model of sequence evolution, as commonly observed in phylogeographical studies, at the risk of severely underestimating molecular distances.

There was considerable nucleotide variation along the CR in both species (Figure 3), yet without forming a distinct large conserved region as observed in the Group 1—type of CR found in fruit flies, where a conserved domain is followed by a variable domain. However, a few smaller conserved regions could be seen throughout the alignment (Figures 1 and 2). Interestingly, there was considerable concordance in the spatial pattern of sequence variability between *U. cordatus* and *C. guanhumi* (Figure 3), even though an alignment of the conserved sequences of both species failed to detect any significant similarity between them, even when only the regions that are conserved intraspecifically were compared. This conclusion is supported by a Spearman's rank-order

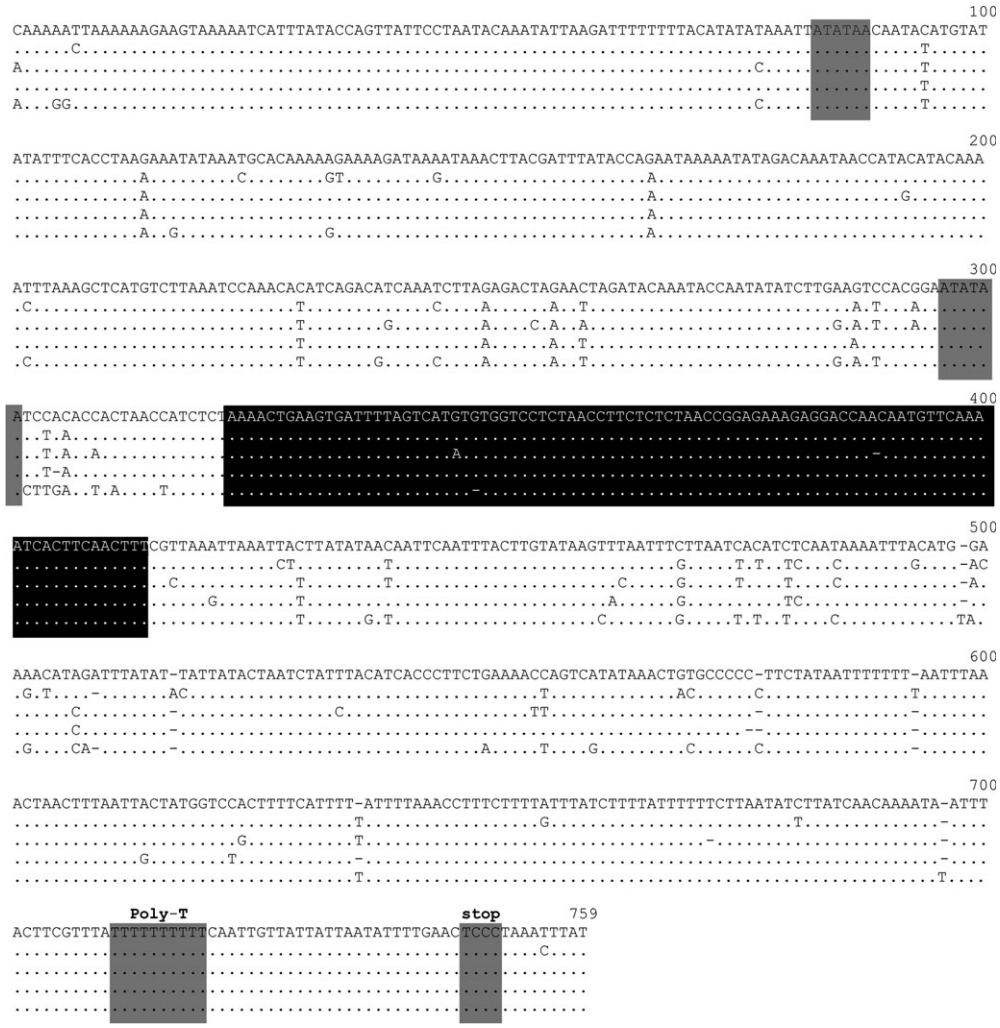

Figure 1. Aligned nucleotide sequences of the CR of *Ucides cordatus*. Dots indicate identification of the corresponding nucleotides to the reference (top) sequence; dashes indicate indels. Stem and loop regions are surrounded by a black box, and motifs are shown in boldface and surrounded by a gray box.

correlation between the variability levels between both species (as measured by the entropy function) and found it to be highly significant ($R_s = 0.427$, $P < 0.001$). This result indicates that *U. cordatus* and *C. guanhumi* might share similar CR organization at the level of its secondary structure despite little correspondence at the nucleotide level.

Several candidate common regulatory motifs were found in both species. These include a polythymine stretch near the $tRNA^{Ile}$ gene, which is often associated with DNA replication origins and transcriptional activators (see Campbell 1986; Delucia et al. 1986). The CR of *C. guanhumi* also included a $(TA)_5$, which is absent from *U. cordatus* sequences. Finally, a common modular element for most promoters, the ATATAA box, is repeated 2 times in *U. cordatus* and 3 times in *C. guanhumi*, with 2 such motifs being present before the conserved region.

There is a TCCC termination motif within the large hairpin of *C. guanhumi* (see below), mapping at nt 345–348. This motif is common in vertebrate CRs (e.g., Randi and Lucchini 1998) and has been associated experimentally with the termination of H strands (Dufresne et al. 1996). Given that it is the only occurrence of this motif in the CR of *C. guanhumi*, it might indeed play that role in this species. However, this motif is absent from the corresponding position in *U. cordatus*; rather, it maps at nt 747–750 at the end of the CR in that species, downstream of the polythymine stretch. Therefore, the interpretation of the functional role of TCCC motif in the studied species is still uncertain. On the other hand, several conserved motifs that are widespread among vertebrates were absent from either studied species. These include GYRCAT, commonly found in mammalian and bird ETAS1 (Randi and Lucchini 1998; Brehm et al. 2003) and tandem repeats at the end of the 3' end of the CR (Brehm et al. 2003). An important caveat is that the identification of candidate motifs is inherently tentative until experimental studies are carried out on the

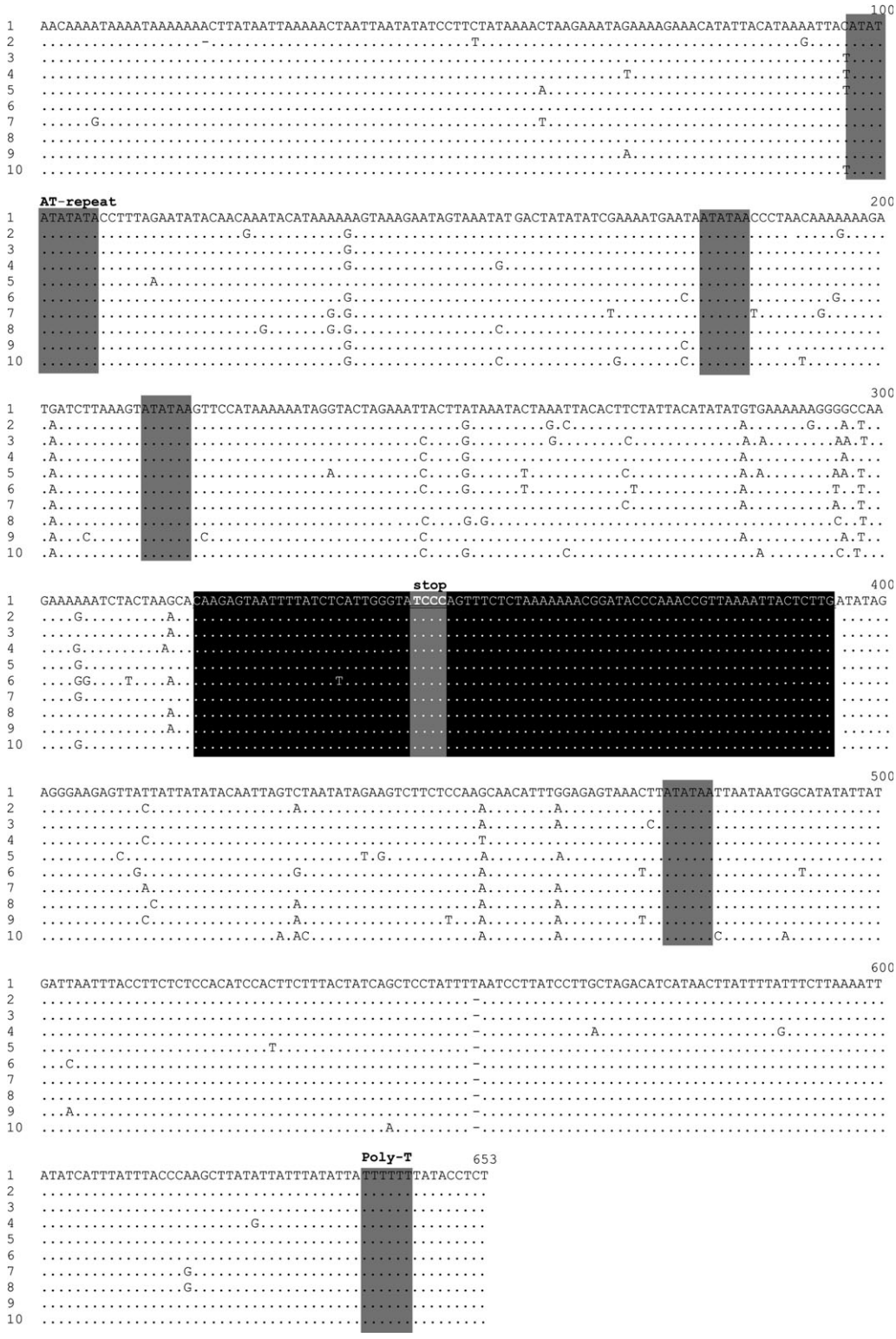

Figure 2. Aligned nucleotide sequences of the CR of *Cardisoma guanhumi*. Dots indicate identification of the corresponding nucleotides to the reference (top) sequence; dashes indicate indels. Stem and loop regions are surrounded by a black box, and motifs are surrounded by a gray box.

Table 1. Basic features of the studied sequences

	Ucides cordatus	Cardisoma guanhumi
A	39.7 (39.2–40.4)	43.5 (42.8–44.5)
G	6.9 (6.6–7.3)	8.3 (7.5–8.9)
C	16.0 (15.5–16.7)	14.6 (14.0–15.0)
T	37.4 (16.8–38)	33.7 (33.1–34.4)
Size (bp)	753.4 (751–756)	652.0 (651–653)

Values are shown as averages (%), followed by their respective ranges (%) in parentheses.

CR function in brachyuran crustaceans, particularly because the candidate motifs are different among the studied species.

The lack of CR sequence conservation between *C. guanhumi* and *U. cordatus* might raise concerns over whether the studied fragment is in fact the result of a cytonuclear transfer of mitochondrial DNA and not of mitochondrial origin. There are several reasons to believe that such artifact is not the case in our study. First, if mutations accumulated at such a high rate as to obscure interspecific similarities in CR, they should also have eliminated both the concordant variability patterns along the studied fragments and the similar secondary structures (see below). Moreover, the flanking 12S region showed minimal sequence divergence among individuals of the same species, suggesting that this region is indeed functional. Finally, a fragment of the CR of both species has been sequenced for more than 200 specimens in a comprehensive study on their comparative phylogeography and evolutionary demography along the Brazilian coast, providing results that were biologically meaningful (Oliveira-Neto JF, Boeger WO, Pie MR, unpublished results). This combined evidence strongly suggests that the studied fragments are indeed the CR of the studied species.

A conserved stem and loop (hairpin) structure was identified at the central region of the CR, with similar morphologies and folding energies (Figure 4). The central region of the CR is conserved intraspecifically in both species (Figure 3), although there is little interspecific correspondence between the nucleotide sequences in those regions. These results indicate that variation in the sequence level can be compensated by specific CR configurations or that novel nucleotide sequences (or protein factors) can provide the same function in different species (Shadel and Clayton 1997).

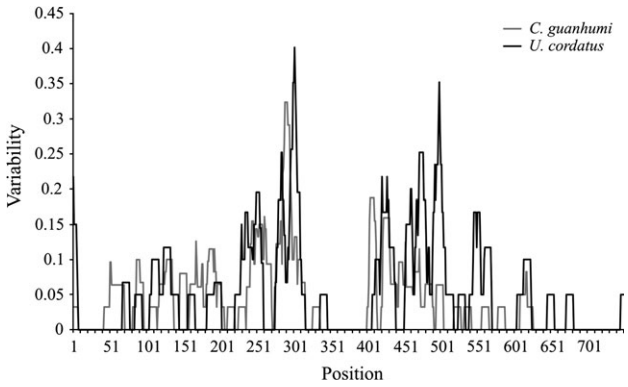

Figure 3. Pattern of sequence variability in the studied alignments of *Ucides cordatus* and *Cardisoma guanhumi* based on a 10-bp sliding window.

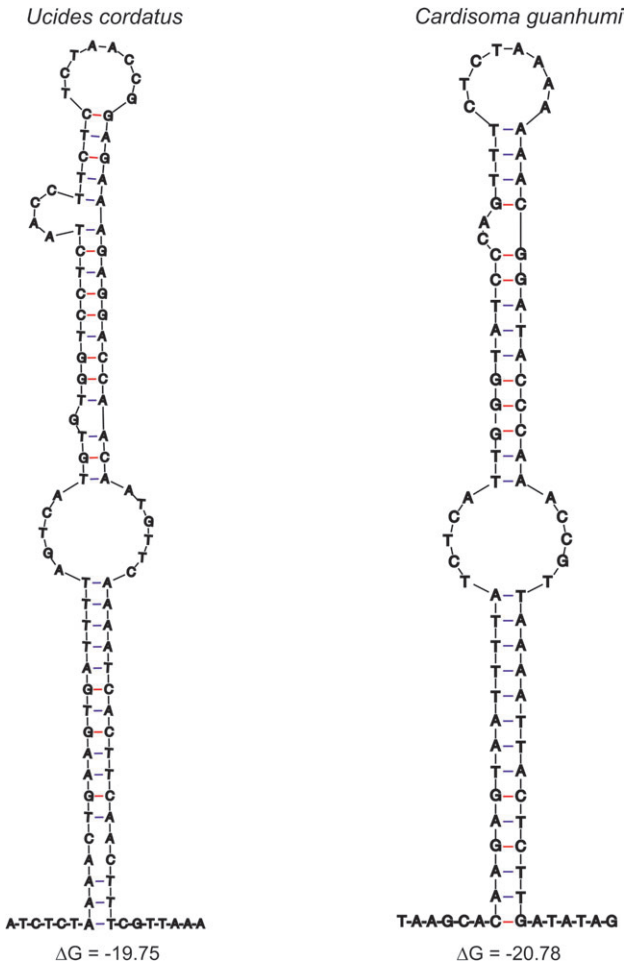

Figure 4. Putative secondary structures associated with conserved region segments. See text for details.

Funding

Unidade Gestora do Fundo Paraná de Ciência e Tecnologia, State of Paraná (through the Fundação da Universidade Federal do Paraná); Serviço Brasileiro de Apoio às Micro e Pequenas Empresas, State of Sergipe, Brazil.

Acknowledgments

The Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) provided research and scholar assistantships to W.A.B. and J.F.O.-N., respectively. L. Patella provided laboratory support.

References

Boyce TM, Zwick ME, Aquadro CF. 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics. 123:825–836.

Brehm A, Harris DJ, Alves C, Jesus J, Thomarat F, Vicente L. 2003. Structure and evolution of the mitochondrial DNA complete control region in the lizard *Lacerta dugesii* (Lacertidae, Sauria). J Mol Evol. 56:46–53.

Campbell JL. 1986. Eukaryotic DNA replication. Annu Rev Biochem. 55:733–737.

Delucia AL, Sumitra D, Partin K, Tegtmeyer P. 1986. Functional interactions of the simian virus 40 core origin of replication with flanking regulatory sequences. J Virol. 57:138–144.

Diniz FM, Maclean N, Ogawa M, Cintra IHA, Bentzen P. 2005. The hypervariable domain of the mitochondrial control region in Atlantic spiny lobsters and its potential as a marker for investigating phylogeographic structuring. Mar Biotechnol. 7:462–473.

Dufresne C, Mignotte F, Guéride M. 1996. The presence of tandem repeats and the initiation of replication in rabbit mitochondrial DNA. Eur J Biochem. 235:593–600.

Grabowski M, Stuck KC. 1999. Structure and intraspecific variability of the control region mtDNA in the pink shrimp, *Farfantepenaeus duorarum* (Decapoda, Penaeidae). In: Schram FR, Vaupel Klein JC, editors. Crustaceans and the biodiversity crisis, vol. I. Leiden (The Netherlands): Brill Academic Publishers. p. 333–344.

Kilpert F, Podsiadlowski L. 2006. The complete mitochondrial genome of the common sea slater, *Ligia oceanica* (Crustacea, Isopoda) bears a novel gene order and unusual control region features. BMC Genomics. 7:241.

Kolpakov R, Bana G, Kucherov G. 2003. MREPS: efficient and flexible detection of tandem repeats in DNA. Nucleic Acids Res. 31:3672–3678.

Matson CW, Baker RJ. 2001. DNA sequence variation in the mitochondrial control region of red-baked voles (Clethrionomys). Mol Biol Evol. 18:1494–1501.

Oliveira-Neto JF, Boeger WA, Pie MR, Ostrensky A, Hungria DB. Forthcoming 2008. Genetic structure of populations of the mangrove crab *Ucides cordatus* (Decapoda: Ocypodidae) at local and regional scales. Hydrobiologia.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.

Prestidge DS. 1995. Predicting Pol II promoter sequences using transcription factor binding sites. J Mol Biol. 23:923–932.

Proutski V, Holmes EC. 1998. SWAN: a new Macintosh application for the sliding window analysis of nucleotide sequence variability. Comput Appl Biosci. 14:467–468.

Randi E, Lucchini V. 1998. Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. J Mol Evol. 47:449–462.

Sbisà E, Tanzariello F, Reyes A, Pesole G, Saccone C. 1997. Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene. 205:125–140.

Seutin G, White BN, Boag PT. 1991. Preservation of avian blood and tissue samples for DNA analysis. Can J Zool. 69:82–90.

Shadel GS, Clayton DA. 1997. Mitochondrial DNA maintenance in vertebrates. Annu Rev Biochem. 66:409–435.

Staden R. 1996. The Staden sequence analysis package. Mol Biotechnol. 5:233–241.

Taylor MFJ, McKechnie SW, Pierce N, Kreitman M. 1993. The lepidopteran mitochondrial control region: structure and evolution. Mol Biol Evol. 10:1259–1272.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25: 4876–4882.

Vila M, Björklund M. 2004. The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). J Mol Evol. 58:280–290.

Wolstenholme DR. 1992. Animal mitochondrial DNA: structure and evolution. In: Wolstenholme DR, Jeon KW, editors. Mitochondrial genomes. International review of cytology. Vol. 141. San Diego (CA): Academic Press. p. 173–216.

Yamauchi MM, Miya MU, Nishida M. 2003. Complete mitochondrial DNA sequence of the swimming crab, *Portunus trituberculatus* (Crustacea: Decapoda: Brachyura). Gene. 311:129–135.

Zhang D-X, Hewitt GM. 1996. Insect mitochondrial control region: a review of its structure, evolution, and usefulness in evolutionary studies. Biochem Syst Ecol. 25:99–120.

Zhang D-X, Szymura JM, Hewitt GM. 1995. Evolution, and structural conservation of the control region of insect mitochondrial DNA. J Mol Evol. 40:382–391.

Zucker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31:3406–3415.

Received May 25, 2007
Accepted March 10, 2008

Corresponding Editor: Steve Karl