J Mol Evol (2000) 55:1–13
DOI: 10.1007/s00239-001-0075-0

JOURNAL OF MOLECULAR EVOLUTION

The Mitochondrial Genome of *Acropora tenuis* (Cnidaria; Scleractinia) Contains a Large Group I Intron and a Candidate Control Region

Madeleine J.H. van Oppen,[1]* Julian Catmull,[1] Brenda J. McDonald,[1] Nikki R. Hislop,[2] Paul J. Hagerman,[2] David J. Miller[1]**

[1] Biochemistry and Molecular Biology, James Cook University, Townsville, Queensland 4811, Australia
[2] Department of Biological Chemistry, School of Medicine, University of California, Davis, CA 95616, USA

Received: 21 May 2001 / Accepted: 24 October 2001

Abstract. The complete nucleotide sequence of the mitochondrial genome of the coral *Acropora tenuis* has been determined. The 18,338 bp *A. tenuis* mitochondrial genome contains the standard metazoan complement of 13 protein-coding and two rRNA genes, but only the same two tRNA genes (*trnM* and *trnW*) as are present in the mtDNA of the sea anemone, *Metridium senile*. The *A. tenuis nad5* gene is interrupted by a large group I intron which contains ten protein-coding genes and *rns*; *M. senile* has an intron at the same position but this contains only two protein-coding genes. Despite the large distance (about 11.5 kb) between the 5′-exon and 3′-exon boundaries, the *A. tenuis nad5* gene is functional, as we were able to RT-PCR across the predicted intron splice site using total RNA from *A. tenuis*. As in *M. senile*, all of the genes in the *A. tenuis* mt genome have the same orientation, but their organization is completely different in these two zoantharians: The only common gene boundaries are those at each end of the group I intron and between *trnM* and *rnl*. Finally, we provide evidence that the *rns-cox3* intergenic region in *A. tenuis* may correspond to the mitochondrial control region of higher animals. This region contains repetitive elements, and has the potential to form secondary structures of the type characteristic of vertebrate D-loops. Comparisons between a wide range of *Acropora* species showed that a long hairpin predicted in *rns-cox3* is phylogenetically conserved, and allowed the tentative identification of conserved sequence blocks.

Key words: *Acropora* — Cnidaria — mtDNA Control region — Group I intron

Introduction

The mitochondrial genomes of bilateral metazoans are remarkably uniform, and are distinguished from those of fungi, plants, and protists by a number of unique characteristics, including size, gene complement, and the absence of introns (reviewed in Wolstenholme 1992). However, in several respects, the mitochondrial genomes of cnidarians are intermediate between those of bilateral animals and other eukaryotes.

Of the four cnidarian classes, only the Anthozoa are thought to have circular mitochondrial genomes (Bridge et al. 1992), implying that the Anthozoa are the basal class, and that a linearization event occurred after the split with the lineage leading to the other classes (see Fig. 1A). Complete mitochondrial genome sequences are available for only two cnidarians, both of which are anthozoans: the soft coral *Sarcophyton glaucum* (Beaton et al. 1998) and the sea

**Present address:* Australian Institute of Marine Science, PMB No. 3, Townsville MC, Qld 4810, Australia
***Present address:* Tropical Environment Studies and Geography, James Cook University, Townsville 4811, Australia
Correspondence to: D.J. Miller; *email:* david.miller@jcu.edu.au

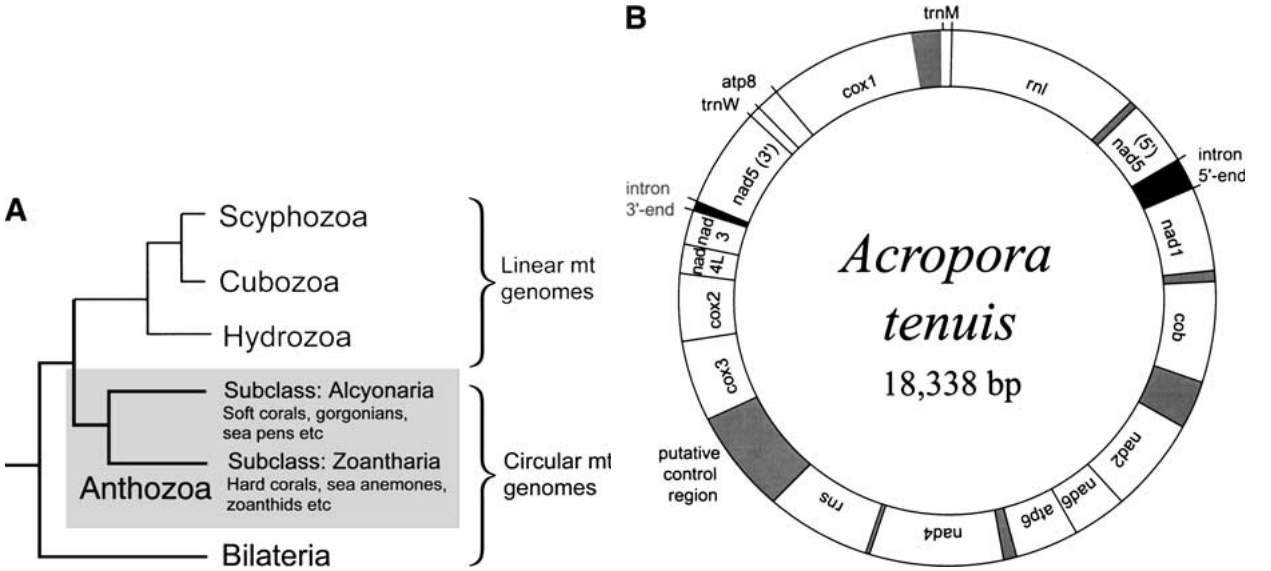

Fig. 1. **A** Phylogenetic position of the Anthozoa in relation to mitochondrial genome structure. Members of the cnidarian class Anthozoa and bilateral animals have circular mitochondrial genomes, whereas in representatives of the other three cnidarian classes these are linear. Note that mitochondrial gene order is exactly the same in two distantly-related members of the anthozoan subclass Alcyonaria (the soft coral *Sarcophyton* and the sea pen *Renilla*), whereas it differs completely between *Acropora* and *Metridium*, members of the subclass Zoantharia. **B** Gene map of the *Acropora tenuis* mitochondrial genome. Substantial intergenic spacers are shown shaded, and the 3′- and 5′-ends of the group I intron are filled. All of the genes are transcribed in a clockwise direction.

anemone *Metridium senile* (Beagley et al. 1998). These represent the two major subclasses of the Anthozoa, the Alcyonaria (= Octocorallia), and the Zoantharia (= Hexacorallia), respectively. For a second member of the Alcyonaria, the sea pen *Renilla kolikeri*, the complete sequence has been determined (Beagley et al. 1995), but this is not in the databases or otherwise available. For the hydrozoan *Hydra attenuata*, 3231 nt of one end of one of the two linear mitochondrial DNA molecules are now available (Pont-Kingdon et al. 2000).

Whereas the mitochondrial genomes of anthozoans are of a comparable size to those of higher animals (*M. senile* = 17,443 bp; *R. kolikeri* = 18,911 bp; *S. glaucum* = 18,453 bp), the genes are not as tightly packed. In the mitochondrial genomes of higher animals, the genes are usually immediately adjacent or overlap, whereas in cnidarians this is generally not the case—often there are extensive non-coding sequences between genes in the Cnidaria. The only exceptions to this pattern are that *S. glaucum nad5* overlaps (by 13 and 19 nt respectively) with *nad2* and *nad4*, *R. kolikeri nad5* overlaps (by 13 nt) with *nad2* and *H. attenuate atp8* overlaps with *atp6* (by 1 nt). Together with the fact that complete stop codons are present in every protein-coding gene, this suggests major differences between the Cnidaria and higher animals with respect to mitochondrial transcription mechanisms. Whereas all other animal mitochondrial genomes are devoid of introns, group I introns are present in both *cox1* and *nad5* of *M. senile* (Beagley et al. 1996). Only two tRNA genes (*trnM* and *trnW*) are present in *M. senile* (Beagley et al. 1998), and only *trnM* is present in the two octocorals. An ORF encoding a putative mismatch repair protein is present in the mitochondrial genomes of *R. kolikeri* and *S. glaucum* (Beagley et al. 1995; Pont-Kingdon et al. 1995). Extremely low mutation rates in the scleractinian coral *Acropora* suggest that mitochondrial mismatch repair occurs (van Oppen et al. 1999a), and therefore that (unlike the situation in higher animals) this property may be retained throughout the Cnidaria.

To provide additional perspectives on the evolution of the metazoan mitochondrial genome, we determined the complete nucleotide sequence of the mitochondrial genome of the scleractinian coral *Acropora tenuis*, a second representative of the anthozoan subclass Zoantharia. Gene order in *A. tenuis* differs markedly from that in *M. senile*, despite the fact that both are zoantharians: This contrasts with the situation in the Alcyonaria, where representatives of two evolutionarily distinct groups have identical mitochondrial genome organization (Beagley et al. 1995; Beaton et al. 1998). As in *M. senile*, the *A. tenuis* mitochondrial genome encodes only two tRNAs (*trnM* and *trnW*), hence the majority of the mitochondrial tRNAs must be imported; however, analysis of codon use implies that the import process is likely to be selective (van Oppen et al. 1999b). The *A. tenuis nad5* gene contains a group I intron that is likely to be homologous with that in *nad5* of *M. senile*. However, whereas in *M. senile* the intron contains only *nad1* and *nad3*, in the case of *A. tenuis* over half

of the mitochondrial genome is nested within the intron. Despite the size of intron, *A. tenuis nad5* appears to be functional, as RT-PCR detected a transcript from which the intron had been excised. Finally, we provide evidence to support the hypothesis that the *rns-cox3* intergenic spacer in *A. tenuis* corresponds to the mitochondrial control region of higher animals.

Materials and Methods

Sample Collection and DNA Extraction

Acropora tenuis colonies were collected on the anticipated day of spawning in the summer of 1998 and kept in a bucket under continuous aeration and water flow. After spawning, bundles of eggs and sperm were collected using a wide-bore Pasteur pipette. Approximately 10 ml of gamete bundles were placed in sterile plastic tubes and the tubes were inverted several times, causing the bundles to burst and the eggs to float to the surface. The tubes were left undisturbed for 5 min and the unfertilized eggs were collected from the surface, washed with filtered seawater to remove remaining sperm, and frozen in liquid nitrogen, after which the colony was returned to its collection site. Total DNA was extracted from 500 µl of eggs ground in liquid nitrogen following McMillan et al. (1988) and stored at 4°C. The partial mitochondrial DNA sequence published earlier (van Oppen et al. 1999b) was obtained from a different individual than the remaining sequence presented here, but both were collected in the same year and from the same reef. Note that we have no evidence for mitochondrial heteroplasmy in *Acropora tenuis* individuals at this time.

Restriction Digestion, Cloning, and Sequencing

Approximately 10 kb of the *A. tenuis* mitochondrial genome have previously been obtained by a combination of deletion cloning and primer walking of cloned mtDNA restriction fragments (van Oppen et al. 1999b; GenBank accession AF152244). Repeated attempts to clone the remaining fragments failed. We therefore PCR amplified the remaining part of the mtDNA using the high fidelity PCR supermix (Gibco BRL) and the primer pair AtCytbRP5e (5′-CGGAATTCTGTCCCACCGATGCAAAAGCCAAG-3′) and AtOIIFP3e (5′-CGGAATTCTATTATTAGGGGGTTGGGGAGTGG-3′), which contain an *Eco*RI restriction site for sticky-end cloning. The amplicon was approximately 9 kb in length. Repeated attempts to clone the fragment into pBluescript and pGEM-T (Promega) were unsuccessful. Subcloning of *Bgl*II digested fragments of the original PCR product also failed. Finally attempts were made to shotgun clone the 9 kb PCR product using *Taq*I and *Sau*3A. Although we obtained many true positives, part of the 9 kb fragment appeared unclonable and we used direct sequencing of PCR products to close these gaps in the sequence. The complete mitochondrial DNA sequence of *A. tenuis* has been submitted to GenBank (accession AF338425). With the exception of a short segment (nt 15156–15985 in the sequence, corresponding to part of the 3′ region of *nad5*), both strands have been sequenced with an overall redundancy of approximately fourfold.

For sequence comparison of the putative control region from a range of different coral colonies and species, DNA was extracted from approximately 1 cm^3 of coral branch as described in van Oppen et al. (1999a) and sequences determined using the primers and protocols described in van Oppen et al. (2001).

RT-PCR Analysis

To examine whether the *A. tenuis nad5* gene was functional, RT-PCR was conducted across the predicted group I intron splice site. Total RNA was prepared from approximately 1 cm^3 of *A. tenuis* eggs using the method of Chomczynski and Sacchi (1987). The RNA was treated with Gibco-BRL amplification grade DNAse I prior to its use in first strand cDNA synthesis from the ND5RP2 primer (5′-TGACTACAAGTCGAGTAAGC-3′) with an Amersham Pharmacia cDNA synthesis kit. Aliquots of first-strand cDNA were used in PCR (reagents as above) using the ND5RP2 and ND5FP1 (5′-CCGGATGCWATGGARGGT-3′) primers. The PCR program was as follows: 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 50°C, and 30 s at 72°C. Finally, the mixture was incubated at 72°C for 5 min. The PCR product was subjected to electrophoresis on 1.0% TAE-agarose, excised, and recovered by spinning through Whatman paper prior to spectrophotometric quantitation and sequencing using the ND5FP1 primer.

Data Analysis

Open Reading Frames (ORFs) of significant length were translated in MacVector 4.1.4 (Kaufman et al. 1994) using the *Metridium senile* genetic code (Pont-Kingdon et al. 1994) and compared with the database using the BLAST programs (Altschul et al. 1990).

The DNA Mfold server (http://bioinfo.math.rpi.edu/~mfold/dna/; SantaLucia 1998; Zuker 2000) was used to examine potential secondary structure of the intergenic regions and the putative control region. Sequences were folded at 28°C. The intergenic sequence was examined for the presence of canonical tRNAs using tRNAscan-SE (http://www.genetics.wustl.edu/eddy/tRNAscan-SE/; Lowe and Eddy 1997) and was also examined for the presence of non-canonical tRNAs as follows: sequence elements were searched for strings of 17 nt that are capable of forming 5 bp stems with central loops of 7 nt, without restrictions as to loop sequence. Stems were allowed to have 0 or 1 mismatch. Sequences passing this filter were investigated for the ability to form an acceptor stem of 7 nt, with 0 or 1 mismatch, and with the single-stranded regions connecting the anticodon and acceptor stems allowed to vary from 2–20 nt. Finally, sequences passing this second filter were examined for their ability to form at least one stem/loop (T or D or both).

Results and Discussion

Gene Content and Organization

The mitochondrial genome of *Acropora tenuis* is 18,338 nt and, in general, the closest matches at both the nucleotide and the amino acid levels were with the corresponding mitochondrial genes of *Metridium senile*.

The search strategy employed identified the complete cytochrome b (*cob*), NADH dehydrogenase subunit 2 (*nad2*), NADH dehydrogenase subunit 6 (*nad6*), ATP synthase subunit 6 (*atp6*), NADH dehydrogenase subunit 4 (*nad4*), small subunit ribosomal RNA (*rns*), cytochrome oxidase subunit III (*cox3*), cytochrome oxidase subunit II (*cox2*), NADH dehydrogenase subunit 4L (*nad4L*), NADH dehydrogenase subunit 3 (*nad3*), NADH dehydrogenase subunit 5 (*nad5*), ATP synthase subunit 8 (*atp8*), cytochrome oxidase subunit I (*cox1*), large sub-

Table 1. Position and size of all coding and non-coding regions and initiation/termination codons of all protein-coding genes of the *A. tenuis* mitochondrial genome

Position	Region	Start codon	Stop codon	Length (in bp)
1–71	*trnM*			71
72–2332	*rnl*			2261
2333–2434	igr1			102
2435–3154	*nad5* 5′	GTG		720
3155–3477	group I intron 5′			323
3478–4461	*nad1*	GTG	TAA	984
4462–4569	igr2			108
4570–5724	*cob*	ATG	TAG	1155
5725–6245	igr3			521
6246–7343	*nad2*	ATG	TAA	1098
7344–7375	igr4			32
7376–7969	*nad6*	ATA	TAA	594
7970–8037	igr5			68
8038–8736	*atp6*	ATG	TAG	699
8737–8887	igr6			151
8888–10363	*nad4*	GTG	TAG	1476
10364–10415	igr7			52
10416–11591	*ms*			1176
11592–12677	putative control region			1086
12678–13466	*cox3*	GTG	TAG	789
13467–13522	igr8			56
13523–14266	*cox2*	ATG	TAA	744
14267–14298 (or 14295)	igr9			32 (or 29)*
14299 (or 14296)–14598	*nad4L*	ATA/GTG	TAA	300 (or 303)*
14599–14630	igr10			32
14631–14987	*nad3*	GTG	TAG	357
14988–15082	group I intron 3′			95
15083–16198	*nad5* 3′		TAG	1116
16199–16228	igr11			30
16229–16298	*trnW*			70
16299–16330	igr12			32
16331–16549	*atp8*	ATG		219
16531–18132	*cox1*	ATG	TAA	1602
18133–18338	igr13			206

igr = intergenic region.
*Size depends on which of the two codons (ATA or GTG) is the true start codon of *nad4L*.
Numbering used here corresponds to that used in the database entry (AF338425).

unit ribosomal RNA (*rnl*), as well as two tRNA genes (*trnM* and *trnW*) in the sequence (Table 1, Figs. 1B and 2). These correspond to the two tRNAs encoded by the *M. senile* mitochondrial genome (Beagley et al. 1998), whereas the mitochondrial genomes of the alcyonarians *S. glaucum* (Beaton et al. 1998; Pont-Kingdon et al. 1998) and *R. kolikeri* (Beagley et al. 1995) encode only a *trnM*. As in *M. senile*, the major conserved features of standard prokaryotic and eukaryotic nuclear- and chloroplast-encoded tRNAs are present in the *A. tenuis trnM*: R_9, G_{10}, Y_{11}, A_{14}, R_{15}, G_{18}, G_{19}, A_{21}, Y_{25}, R_{26}, Y_{32}, U_{33}, R_{37}, G_{53}, U_{54}, U_{55}, C_{56}, R_{57}, A_{58}, and C_{61}, and the *A. tenuis trnW*: U_8, R_9, G_{10}, Y_{11}, A_{14}, R_{15}, G_{19}, A_{21}, Y_{25}, R_{26}, Y_{32}, U_{33}, R_{37}, G_{53}, U_{54}, C_{56}, R_{57}, A_{58}, and C_{61}. The proposed secondary structures of *trnM* and *trnW* are very similar between *A. tenuis* and *M. senile*.

The *A. tenuis* mitochondrial gene order is as described for *A. nasuta* in a preliminary mapping study (Fukami et al. 2000), although the *nad5* group I intron present in *A. tenuis* (see below) was not reported for *A. nasuta*. The intron is present in *A. nasuta* from the Great Barrier Reef, as we have PCR amplified across both the 5′ and 3′ intron segments in this and a range of other *Acropora* species (see below). Despite the fact that *M. senile* is in the same anthozoan suborder (the Zoantharia), its gene order differs dramatically from that in *Acropora* (van Oppen et al. 1999b); the only common gene boundaries are those at each end of the group I intron (see below) and between *trnM* and *rnl*. The two representatives of the Alcyonaria for which data are available have identical gene organization, but there are few similarities with that seen in the zoantharians. *nad1* and *cob* are organized the same way in both *A. tenuis* and the alcyonarians. Although *nad4L* and *nad3* are adjacent in both *A. tenuis* and the alcyonarians, the order is reversed. The reasons why major rearrangements in gene order have occurred within one anthozoan sub-

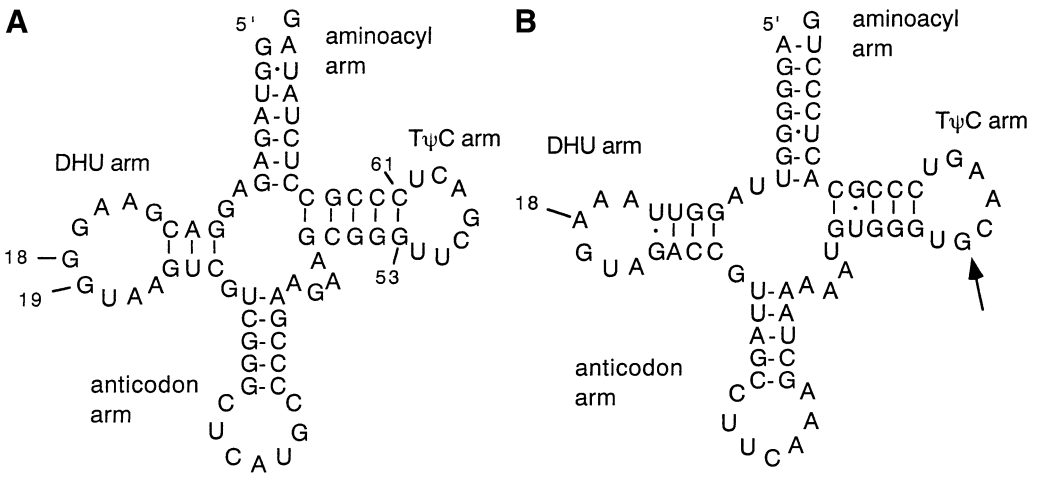

Fig. 2. Predicted secondary structures of tRNAs encoded by the *Acropora tenuis* mitochondrial genome. Numbering follows that of the *Metridium senile* tRNAs (Pont-Kingdon et al. 1994; Beagley et al. 1998). **A** *trnM*; **B** *trnW*.

class (the Zoantharia), but apparently not in another (the Alcyonaria), are unclear.

Base Composition and Codon Usage in the Sense Strand

The *A. tenuis* mtDNA consists of 62.1% A + T, which is at the low end of the range for invertebrates—insect mtDNA, for example, is very A + T-rich (>75%, e.g. Clary and Wolstenholme 1985; Crozier and Crozier 1993; Mitchell et al. 1993; Flook et al. 1995)—but similar to *M. senile* and *S. glaucum*. Cytosine is the least used and thymine the most used nucleotide in the *A. tenuis* mtDNA (25.1% A, 13.7% C, 24.2% G, and 37.0% T). A slight bias towards A and G (26.6% A and 25.5% G) and away from C and T (12.7% C and 35.2% T) was found in the noncoding regions.

In higher metazoans, mitochondrial protein-coding genes frequently lack complete termination codons—in these cases, after precise cleavage of the individual coding sequences from a polycistronic primary transcript, TAA codons are generated by polyadenylation. However, in all cases only full termination codons are used in the Cnidaria. In *A. tenuis*, as in other Cnidaria, both TAG and TAA are used as termination codons (Table 1), TGA coding for tryptophan (Beagley et al. 1998). The majority of cnidarian mitochondrial genes use ATG as start codon, whereas both ATA and GTG serve as translation initiation codons in some *A. tenuis* genes (Table 1). Use of GTG as a start codon has not been documented for other cnidarians; the only known deviations from the use of ATG are *nad1*, *nad5*, and *atp8* of *S. glaucum*, which are initiated by ATA (Beaton et al. 1998).

All 62 amino acid codons are used in the *A. tenuis* coding sequences, but only two tRNA genes are present—*trnM* and *trnW*. For some amino acids, codon use is strongly biased, and there are differences in codon use between mitochondrial and nuclear genes in *Acropora*; These patterns do not differ significantly from data reported earlier for a subset of mitochondrial genes (van Oppen et al. 1999b).

nad5 *Group I Intron*

Group I and group II introns are often present in the mitochondrial genomes of fungi and protists but are generally absent from those of animals, the sole exception to this pattern being that group I introns are present in *cox1* and *nad5* of *M. senile* (Beagley et al. 1996). The demonstration that group I introns are also likely to be present in a number of other sea anemones, but not in other cnidarians (including two scleractinians), led to the suggestion that the introns may have been acquired after the Actiniaria (sea anemones) branched off from other zoantharian orders (Beagley et al. 1996). Our results show that this is unlikely to be the case for the *nad5* intron, since we have found what is likely to be an orthologous intron (see Fig. 3) at the same insertion site in a scleractinian coral.

Several lines of evidence support the orthology of the *A. tenuis* and *M. senile nad5* introns. The *A. tenuis nad5* group I intron is at the same insertion site as those of *M. senile* (Beagley et al. 1996), the fungus *Neurospora crassa* (Nelson and Macino 1987) and the liverwort *Marchantia polymorpha* (Oda et al. 1992), while all other *nad5* group I introns have different insertion sites. A phylogenetic analysis of these four introns based on the conserved regions P, Q, R, and S plus the three 3′-end bases (following the alignment in Cech 1988) strongly supported a close relationship between the *A. tenuis* and *M. senile* introns (not

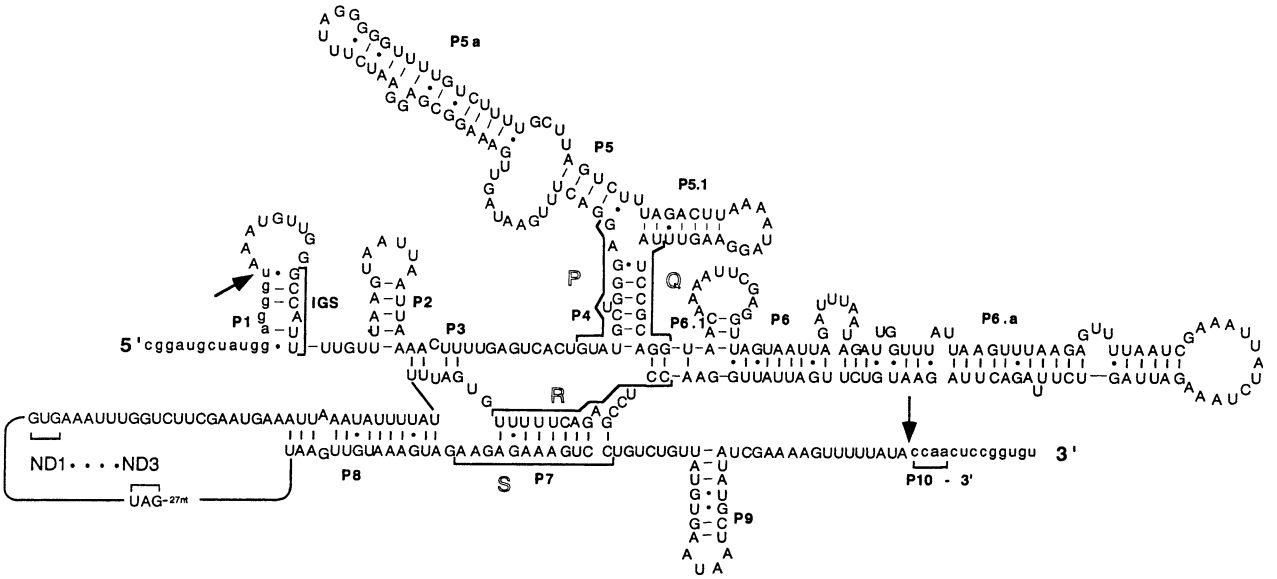

Fig. 3. Predicted secondary structure of the transcript of the *A. tenuis nad5* group I intron, showing the conserved core elements "P," "Q," "R," and "S" and the helical elements P1–P9. Intron and exon sequences are shown as uppercase and lowercase, respectively. Arrows indicate predicted splice sites. Helix P1 is predicted to form by base pairing between the last six nucleotides of the 5′ exon and the internal guide sequence (IGS). The loop of helix P8 contains *nad1*, *cob*, *nad2*, *nad6*, *atp6*, *nad4*, *rns*, the putative control region, *cox3*, *cox2*, *nad4L*, and *nad3*. The positions of the start and stop codons of *nad*1 and *nad*3 respectively are indicated. The 27 nucleotides of the group I intron that immediately follow *nad3*, cannot be folded into the typical group I intron secondary structure and are not shown.

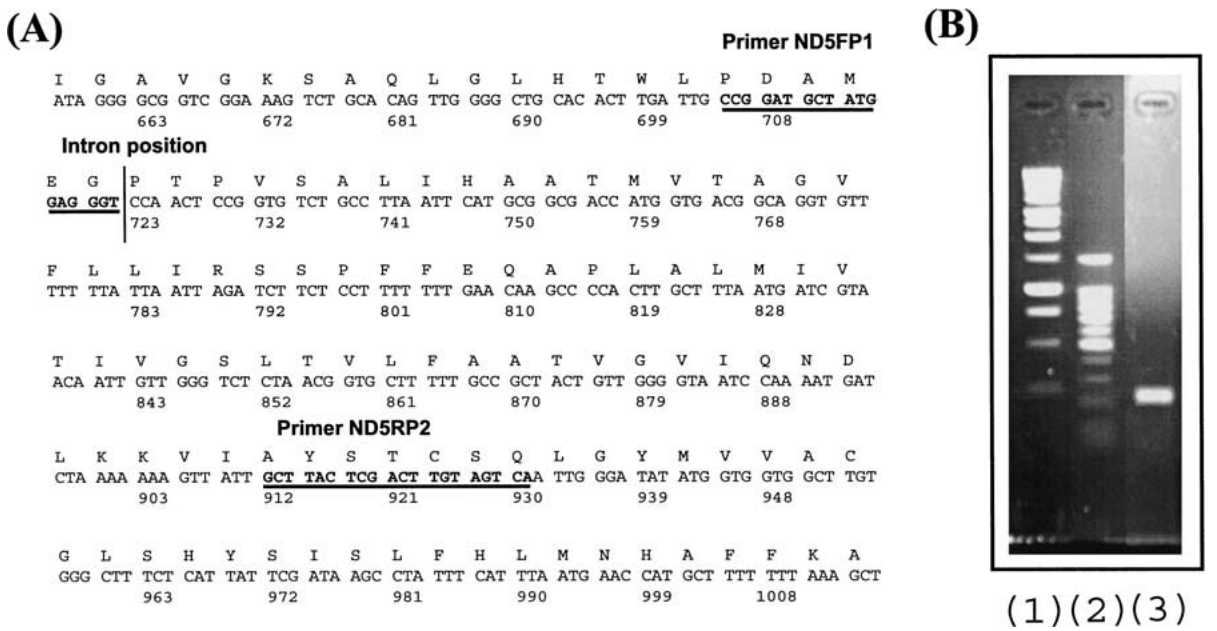

Fig. 4. The *A. tenuis* mitochondrial *nad5* gene is functional. **A** Nucleotide and predicted amino acid sequence of that part of the *A. tenuis nad5* gene proximal to the group I intron insertion site. The positions of the intron and those corresponding to primers used in the RT-PCR experiment are indicated. Numbering refers to nucleotide position in the *nad5* coding sequence. **B** Agarose gel electrophoresis analysis of the RT-PCR product spanning the *A. tenuis nad5* intron splice site. (1) High molecular weight standards (Promega 1 kb ladder). (2) Low molecular weight standards (Promega 100 bp ladder). (3) RT-PCR product. The size of the product (~220 bp) estimated by comparison with the standards corresponds well with that predicted from the sequence (227 bp). The product runs at a point intermediate between the smallest band visible in lane (1) and the second band in lane (2); the sizes of these standards are 250 bp and 200 bp, respectively.

shown). Note that the relatively small number of informative sites prohibited the inclusion of larger numbers of introns in phylogenetic analyses. Moreover, one very distinct feature—the presence of an A at the 3′ end of the intron—is shared only between the *A. tenuis* and *M. senile* introns, while all other group I introns end with G (Cech 1988; Michel and Westhof 1990). Finally, although the introns differ a great deal

in size—only *nad1* and *nad3* are within the *nad5* intron of *M. senile*, while approximately 11.5 kb of the mitochondrial genome are nested within the *A. tenuis nad5* intron—the genes 5′ and 3′ of the inserted region of the *A. tenuis* intron (*nad1* and *nad3*, respectively) correspond to those within the *M. senile* intron.

Only a single copy of *nad5* is present in the *A. tenuis* mitochondrial genome, implying that the group I intron is spliced correctly despite the large distance (approximately 11.5 kb) between the 5′ and 3′ exon boundaries. This was confirmed by the fact that we were able to RT-PCR across the predicted *nad5* intron splice site using as template total RNA prepared from *A. tenuis* eggs (see Fig. 4). The identity of the RT-PCR product was confirmed by direct sequencing after gel purification as above. Note that all *Acropora* colonies (>50) so far screened contained the *nad5* group I intron (data not shown).

To test the possibility that the nested segment was acquired (together with the group I intron) by lateral transfer, we compared base composition and codon use of the regions within and outside of the intron. A small but statistically significant ($p < 0.001$, Chi-square = 26.72, d.f. = 3) difference in base composition is present between the two regions (23.8% A, 13.8% C, 24.3% G, and 38.1% T in the nested sequence; 26.8% A, 13.9% C, 24.4% G, and 34.9% T in the sequence outside the intron). However, no significant differences in codon use were found ($p > 0.995$, Chi-square = 22.36, d.f. = 63), suggesting that the group I intron was acquired without any nested genes. Processes causing gene rearrangements must have been responsible for the transposition of genes into the intron. These results also imply that the putative (and vertebrate-like) control region of *Acropora* (see below), which is positioned within the intron, has not been acquired via lateral gene transfer, but has evolved independently in *Acropora* (or perhaps the Scleractinia) or has been lost in non-scleractinian cnidarians. Base composition of the group I intron itself differs considerably from that of the segment nested within it ($p < 0.001$) and that of the remaining mtDNA segment ($p < 0.01$). It has a bias away from G and C (20.3% and 8.9%, respectively) and toward A (32.8%). Since non-coding and gene regions of the *A. tenuis* mtDNA are similar in base composition, these data confirm that the intron is of foreign origin, while the segment nested within the intron is not.

Our data imply that the *nad5* group I intron was acquired by a common ancestor of the Scleractinia and Actiniaria and is therefore much older than previously assumed (Beagley et al. 1996). It is not clear whether the introns in the same position in *nad5* of *N. crassa* and *M. polymorpha* are homologous, or whether this position is a preferred site for intron insertion (Beagley et al. 1996). The absence of group I introns in the mitochondrial genomes of *S. glaucum* (Pont-Kingdon et al. 1998; Beaton et al. 1998) and *R. kolikeri* (Beagley et al. 1995) suggests that the *nad5* intron may have been acquired after the Zoantharia/Alcyonaria split. However, as yet data are available for relatively few cnidarians, and the interpretation of these is complicated by the fact that intron loss can occur over evolutionary time.

Intergenic Spacers

As in other cnidarian mitochondrial genomes, the noncoding regions between genes in *A. tenuis* are generally large by comparison with most other animals. Only in the case of *atp8* and *cox1* do genes overlap (by 19 nt) in *A. tenuis*; precedents for this in the Cnidaria are described above (Introduction). In general, the intergenic regions in *A. tenuis* are the largest of the cnidarians for which data are available (Table 1). The three longest intergenic regions (between *cob* and *nad2*, between *rns* and *cox3* and between *cox1* and *trnM*) contain open reading frames (ORFs) of maximally 342 nt, 777 nt, and 123 nt, respectively. However, no significant matches were detected between the predicted ORF product and known protein sequences in the databases. Two non-standard proteins are encoded by cnidarian mitochondrial genomes: *S. glaucum* and *R. kolikeri* encode a putative MutS ortholog (Pont-Kingdon et al. 1995; Pont-Kingdon et al. 1998), and a homing endonuclease of the LAGLI-DADG sub-class is encoded by an ORF within the *cox1* group I intron in *M. senile* (Beagley et al. 1996). However, we found no evidence that ORFs present in the *A. tenuis* mitochondrial genome encode proteins of these types; no significant matches with motifs characteristic of homing endonucleases (Lambowiz and Belfort 1993; Saguez et al. 2000) were detected, nor were there with MutS-related sequences (Eisen 1998; Culligan et al. 2000).

The intergenic regions in *A. tenuis* do not have significant matches at the nucleotide level with *M. senile* or *S. glaucum* mitochondrial sequences, and at present the function of these remains unclear. However, the largest intergenic region (that between *rns* and *cox3*) in *A. tenuis* has characteristics of a control region (see below). Short ORFs are present in some of the other intergenic regions, but again the translation products do not correspond to known proteins.

The rns-cox3 *Intergenic Spacer Has Features Typical of Control Regions*

We have previously suggested that the large intergenic space between *rns* and *cox3* may correspond to the control region of the *A. tenuis* mitochondrial genome (van Oppen et al. 1999b). Note that, although

very few control regions have been characterized to the extent that the origins of replication and transcription are known, several features are thought to be characteristic of control regions in higher animals. They typically contain repetitive sequences (e.g. Solignac et al. 1986; Rand and Harrison 1989; Hoelzel et al. 1993; Broughton and Dowling 1994; Flook et al. 1995; Prager et al. 1996; Nesbø et al. 1998; Cook

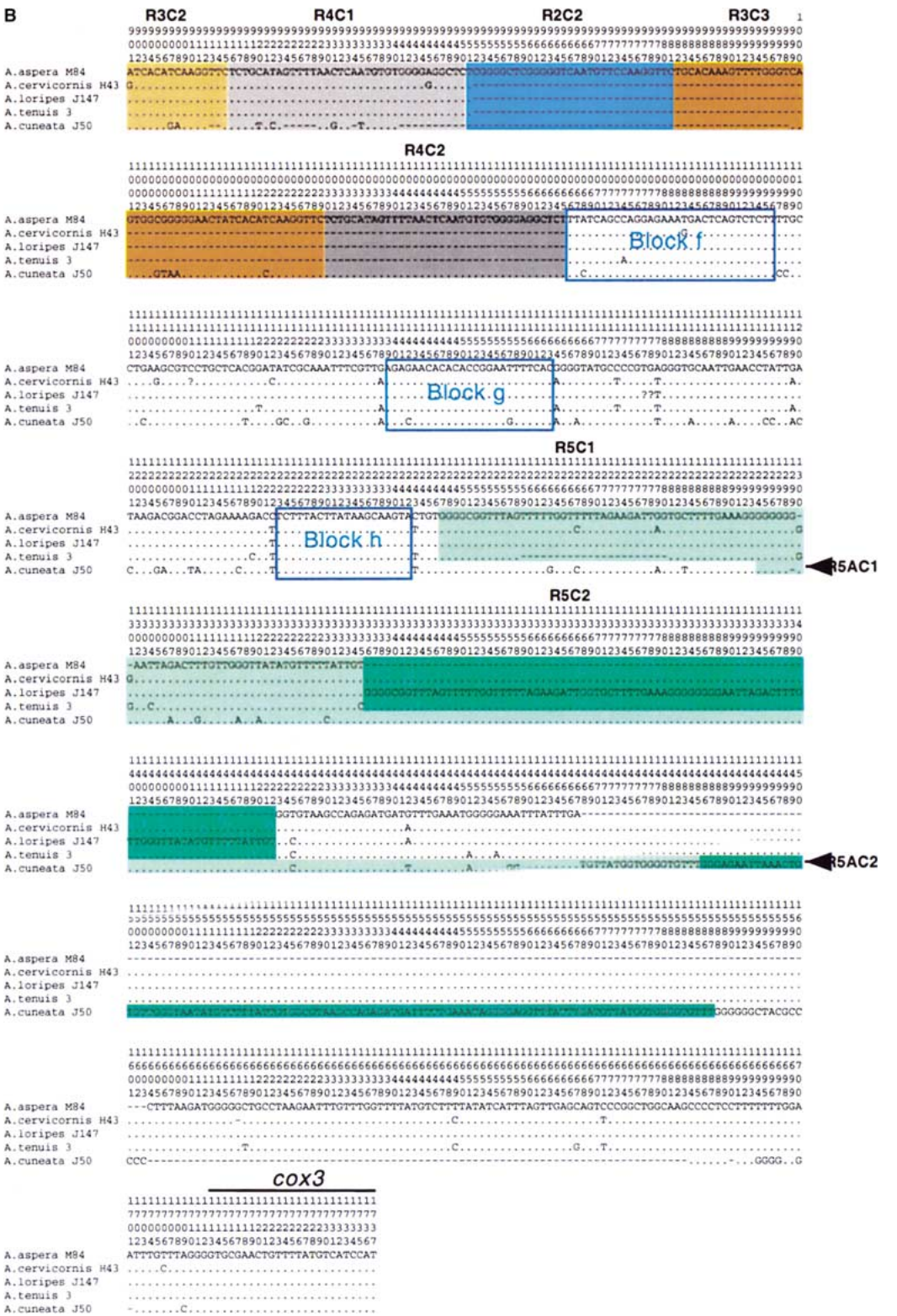

Fig. 5. **A** Schematic overview of repeat regions in the *Acropora* putative control region. Repeats are color-coded (different copies of the same repeat unit are marked by different shades of the same color). **B** Alignment of the putative control region plus flanking sequences of five *Acropora* species representing the two subgenera *Acropora Acropora* (*A. aspera*, *A. cervicornis*, *A. loripes*, and *A. tenuis*) and *Acropora Isopora* (*A. cuneata*). Dots indicate sequence identity with the top line (*A. aspera* M84). Color coding is the same as in **A.** Note that both copies of repeat 2 comprise two sub-repeats and that the repeat structure is different for *A. cuneata* as compared to the *A. Acropora* species. The first repeat extends further into the *rns* at the 5′ end in *A. cuneata* and the last repeat (R5/R5A) is completely different. Blocks A–H = conserved sequence blocks, R = repeat and C = copy (e.g. R1C1 = repeat 1, first copy). The accession numbers of the sequences shown are AY026418 (*A. aspera* M84), AY026428 (*A. cervicornis* H43), AY026447 (*A. loripes* J147), AY026459 (*A. tenuis* 3), and AY026429 (*A. cuneata* J50).

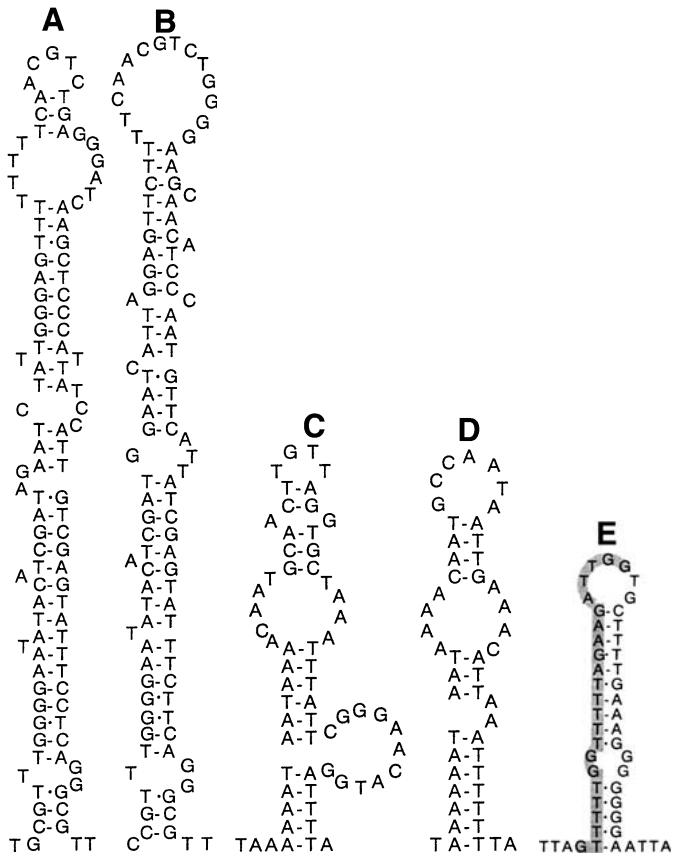

Fig. 6. Putative secondary structures in the *Acropora* candidate control region. **A** Hairpin structure at position 604–706 in *A*. (*A*.) *aspera* M84 (see Fig. 5B). **B** Hairpin structure at position 604–706 in *A*. (*I*.) *cuneata* J50 (see Fig. 5B). **C** Hairpin structure at position 52–114 in *A*. (*A*.) *aspera* M84 (see Fig. 5B). **D** Hairpin structure at position 52–99 in *A*. (*I*.) *cuneata* J50 (see Fig. 5B). **E** Hairpin structure at position 1255–1306 in *A*. (*A*.) *aspera* M84. The gray area is missing in several taxa (see Fig. 5B).

et al. 1999), conserved sequence blocks (CSBs) and secondary structures associated with the initiation of H-strand (heavy-strand) replication (Hixson et al. 1986; Wong and Clayton 1986; Monnerot et al. 1990; Saccone et al. 1991; Monforte et al. 1993; Stanton et al. 1994; Zhang et al. 1995; Lavrov et al. 2000). Accordingly, the putative *A. tenuis* control region was examined for each of these features.

Comparison of sequences from a wide range of *Acropora* species allowed the identification of conserved motifs and repeated elements in the *rns–cox3* intergenic region. *Acropora* is a large and complex genus; two subgenera are recognized—*A. Acropora* consists of over a hundred species, whereas *A. Isopora* contains only four (Wallace 1999). *rns–cox3* sequences were determined for 41 individuals representing 20 *A. Acropora* species from the Indo-Pacific and Caribbean and two individuals representing one *A. Isopora* species (van Oppen et al. 2001; accession numbers AY026418–AY026460). Alignment of the sequences allowed the recognition of five different (imperfect) repetitive regions, the second having distinct sub-repetitive structure (Fig. 5). Figure 5B shows an alignment of the five sequences showing the largest variation in copy number of the various repeats. With the exception of the first and fifth repeat units (repeats 1 and 5/5A), the repeat structure is the same in the two subgenera. The first repeat has a 16 bp long extension at the 5′ end in *A. Isopora* as compared to all *A. Acropora* sequences. The boundaries of repeats 5 and 5A are different at both ends and the repeats are very divergent in sequence between *A. Isopora* and *A. Acropora*. These results suggest that repeats 1 and 5/5A have originated after the *A. Acropora* and *A. Isopora* lineages diverged. Repeats observed at the 5′ end of control regions tend to be composed of long units (e.g. Bentzen et al. 1988; Brown et al. 1992), and this is also true for *Acropora*. Moreover, it is not unusual for duplicated segments to comprise part of a flanking gene region (e.g. Broughton and Dowling 1994; Zhang et al. 1995), as in the first repeat region of the *Acropora rns–cox3* intergenic region which includes the 3′ end of *rns* (Fig. 5). Remarkably, no repetitive regions are present in any of the intergenic regions of other cnidarians for which mitochondrial data are available. A DNA sequence comparison between the putative control regions of *M. senile* and *A. tenuis* also revealed no significant matches (not shown).

Comparisons across the subgenera allowed the tentative identification of *Acropora* CSBs: blocks of sequence that appear to be conserved and may be functionally significant (Fig. 5). Note that these have no significant sequence similarity with the CSBs of vertebrates (Broughton and Dowling 1994; Foran et al. 1988) or insects (Zhang et al. 1995), and that the relatively small number of substitutions observed between the *Acropora* subgenera mean that these results should be regarded as preliminary.

Replication of mtDNA has been shown to initiate within or close to hairpin structures (reviewed in Clayton 2000), and secondary structures are also believed to be involved in D-loop duplication events (e.g. Buroker et al. 1990; Stanton et al. 1994). We examined the *Acropora rns–cox3* intergenic region for the potential to form hairpin structures as follows. The complete *rns–cox3* intergenic regions of two divergent sequences—*A.* (*A.*) *aspera*_M84 and *A.* (*I.*) *cuneata*_J50—were folded at 28°C using the DNA Mfold server at http://bioinfo.math.rpi.edu/~mfold/dna/ (SantaLucia 1998). This identified several potential hairpin structures, a few of which were very similar in the two divergent sequences. In particular, these analyses predicted in both *Acropora Acropora* and *Acropora Isopora* a stable hairpin with a long stem corresponding to positions 604–706 in the alignment (Figs. 5B, 6A, and B).

Strikingly similar hairpin structures are predicted for all of the *Acropora* sequences examined in our laboratory to date (~70). To examine the biological significance of this potential hairpin, we compared the stabilities of the structures predicted from the real data with those predicted from randomized versions of the sequences. The *A. aspera*_M84 sequence corresponding to the predicted hairpin was randomized using the Shuffle program in the Wisconsin Genetics Computer Group (GCG) package. None of the ten randomized sequences had the potential to fold into a large hairpin, and the secondary structures predicted were significantly less stable than that predicted for the real data. The free energies of 41 potential secondary structures obtained by folding ten randomized sequences were −7.1 to −12.7 kcal/mole (average: −10.1 kcal/mole), whereas that of the original sequence was −12.8 kcal/mole. Because secondary structures may also be important in initiating duplication events, each of the repeat regions in the *A. aspera*_M84 sequence was separately examined for the potential for secondary structure formation. Repeats 1 and 5/5A both have the potential to form hairpins with relatively long stems (Figs. 6C, D, and E). The former of these may be significant, but latter is not phylogenetically conserved, implying that it may not be functional: *A. tenuis*, *A. intermedia*, and some *A. longicyathus* colonies have only one copy of this segment carrying a 22 bp deletion.

In summary, the *A. tenuis rns–cox3* intergenic region consists of repetitive and non-repetitive regions, contains several conserved blocks of sequence, and has the potential to form secondary structures like those characteristic of control regions in higher animals. However, note that regions with these features are not present in the mitochondrial genomes of the other cnidarians for which data are currently available. In these cases, the intergenic regions do not contain repetitive regions, and are significantly shorter than in *A. tenuis*—in the case of *S. glaucum*, the largest is 111 bp. In *M. senile* the largest intergenic regions (143 and 324 bp) have no significant sequence similarity with the *A. tenuis rns–cox3* region (not shown). The limited available data thus hint at possible heterogeneity in the mechanisms of mitochondrial transcription and replication within the Cnidaria, and highlight the need for more comparative data.

Acknowledgments. The authors wish to thank Danielle de Jong for performing the RT-PCR analysis and Elsi Vacano for her help with the analysis of the mtRNA database. This work was supported by grants from the Australian Research Council, James Cook University and the National Institutes of Health (GM35305 to PJH).

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Evol 215:403–410

Beagley CT, Macfarlane JL, Pont-Kingdon G, Okimoto R, Okada NA, Wolstenholme DR (1995) Mitochondrial genomes of Anthozoa (Cnidaria). In: Palmieri F, Papa S, Saccone C, Gadaleta N (eds) Progress in cell research—symposium on thirty years of progress in mitochondrial bioenergetics and molecular biology. Elsevier, Amsterdam, pp 149–153

Beagley CT, Okada NA, Wolstenholme DR (1996) Two mitochondrial group I introns in a metazoan, the sea anemone *Metridium senile*: One intron contains genes for subunits 1 and 3 of NADH dehydrogenase. Proc Natl Acad Sci USA 93:5619–5623

Beagley CT, Okimoto R, Wolstenholme DR (1998) The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): Introns, a paucity of tRNA genes, and a near-standard genetic code. Genetics 148:1091–1108

Beaton MJ, Roger AJ, Cavalier-Smith T (1998) Sequence analysis of the mitochondrial genome of *Sarcophyton glaucum:* Conserved gene order among octocorals. J Mol Evol 47:697–708

Bentzen P, Wright JM, Bryden LT, Sargent M, Zwanenburg KC (1998) Tandem repeat polymorphism and heteroplasmy in the mitochondrial control region of redfishes (Sebastes: Scorpaenidae). J Hered 89:1–7

Bridge D, Cunningham CW, Schierwater B, DeSalle R, Buss LW (1992) Class-level relationships in the phylum Cnidaria: Evidence from mitochondrial genome structure. Proc Natl Acad Sci USA 89:8750–8753

Broughton RE, Dowling TE (1994) Length variation in mitochondrial DNA of the minnow *Cyprinella spiloptera*. Genetics 138:179–190

Brown JR, Beckenbach AT, Smith MJ (1992) Mitochondrial DNA length variation and heteroplasmy in populations of white sturgeon (*Acipenser transmontanus*). Genetics 132:221–228

Buroker NE, Brown JR, Gilbert TA, O'Hara PJ, Beckenbach AT, Thomas WK, Smith MJ (1990) Length heteroplasmy of sturgeon mitochondrial DNA: An illegitimate elongation model. Genetics 124:157–163

Cech TR (1988) Conserved sequences and structures of group I introns: Building an active site for RNA catalysis—a review. Gene 73:259–271

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156–159

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. J Mol Evol 22:252–271

Clayton DA (2000) Transcription and replication of mitochondrial DNA. Hum Reprod 15(Suppl 2):11–17

Cook CE, Wang Y, Sensabaugh G (1999) A mitochondrial control region and cytochrome b phylogeny of sika deer (*Cervus nippon*) and report of tandem repeats in the control region. Mol Phylogen Evol 12:47–56

Crozier RH, Crozier YC (1993) The mitochondrial genome of the honey bee *Apis melifera*: Complete sequence and genome organization. Genetics 133:97–117

Culligan KM, Meyer-Gauen G, Lyons-Weiler J, Hays JB (2000) Evolutionary origin, diversification and specialization of eukaryotic MutS homolog mismatch repair proteins. Nucleic Acids Res 28:463–471

Eisen JA (1998) A phylogenomic study of the MutS family of proteins. Nucleic Acids Res 26:4291–4300

Flook PK, Rowell CH, Gellissen G (1995) The sequence, organization, and evolution of the *Locusta migratoria* mitochondrial genome. J Mol Evol 41:928–941

Foran DR, Hixson JE, Brown WM (1988) Comparison of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis. Nucleic Acids Res 16:5841–5861

Fukami H, Omori M, Hatta M (2000) Phylogenetic relationships in the coral family Acroporidae, reassessed by inference from mitochondrial genes. Zool Sci 17:689–696

Fumagalli L, Taberlet P, Favre L, Hausser J (1996) Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. Mol Biol Evol 13:31–46

Hixson JE, Wong TW, Clayton DA (1986) Both the conserved stemloop and divergent 5′-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J Biol Chem 261:2384–2390

Hoelzel AR, Hancock JM, Dover GA (1993) Generation of VNTRs and heteroplasmy by sequence turnover in the mitochondrial control region of 2 elephant seal species. J Mol Evol 37:190–197

Jaeger JA, Turner DH, Zuker M (1989) Predicting optimal and suboptimal secondary structure for RNA. Meth Enzymol 183:281–306

Kaufman J, Olson S, Panagrossi J (1994) MacVector (4.1.4). KODAK Scientific Imaging System, New Haven, CT

Lavrov DV, Boore JL, Brown WM (2000) The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Mol Biol Evol 17:813–824

Lambowiz AM, Belfort M (1993) Introns as mobile genetic elements. Annu Rev Biochem 62:587–622

Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964

McMillan J, Yellowlees D, Heyward A, Harrison P, Miller DJ (1988) Preparation of high molecular weight DNA from hermatypic corals and its use for DNA hybridization and cloning. Mar Biol 98:271–276

Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J Mol Biol 216:585–610

Mitchell S, Cockburn A, Seawright J (1993) The mitochondrial genome of *Anopheles quadrimaculatus* species A: Complete nucleotide sequence and gene organization. Genome 36:1058–1073

Monforte A, Barrio E, Latorre A (1993) Characterization of the length polymorphism in the A + T-rich region of the *Drosophila obscura* group species. J Mol Evol 36:214–223

Monnerot M, Solignac M, Wolstenholme DR (1990) Discrepancy in divergence of the mitochondrial and nuclear genomes of *Drosophila teissieri* and *Drosophila yakuba*. J Mol Evol 30:500–508

Nelson MA, Macino G (1987) Structure and expression of the overlapping ND4L and ND5 genes of *Neurospora crassa* mitochondria. Mol Gen Genet 206:307–317

Nesbø CL, Arab MO, Jakobsen KS (1998) Heteroplasmy, length and sequence variation in the mtDNA control regions of three percid fish species (*Perca fluviatilis, Acerina cernua, Stizostedion lucioperca*). Genetics 148:907–919

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. J Mol Biol 223:1–7

Pont-Kingdon G, Beagley CT, Okimoto R, Wolstenholme DR (1994) Mitochondrial DNA of the sea anemone, *Metridium senile* (Cnidaria): Prokaryote-like genes for $tRNA^{f\text{-}met}$ and small-subunit ribosomal RNA, and standard genetic code specificities for AGR and ATA codons. J Mol Evol 39:387–399

Pont-Kingdon G, Okada NA, Macfarlane JL, Beagley CT, Watkin-Sims CD, Cavalier-Smith T, Clark-Walker GD, Wolstenholme DR (1998) Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. J Mol Evol 46:419–431

Pont-Kingdon GA, Okada NA, Macfarlane JL, Beagley CT, Wolstenholme DR, Cavalier-Smith T, Clark-Walker GD (1995) A coral mitochondrial mutS gene. Nature 375:109–111

Prager EM, Tichy H, Sage RD (1996) Mitochondrial DNA sequence variation in the eastern house mouse, *Mus musculus*—Comparison with other house mice and report of a 75-bp tandem repeat. Genetics 143:427–446

Rand DM, Harrison RG (1989) Molecular population genetics of mtDNA size variation in crickets. Genetics 121:551–569

Saccone C, Pesole G, Sbisa E (1991) The main regulatory region of mammalian mitochondrial DNA: Structure-function model and evolutionary pattern. J Mol Evol 33:83–91

Saguez C, Lecellier G, Koll F (2000) Intronic GIY-YIG endonuclease gene in the mitochondrial genome of *Podospora curvicolla:* evidence for mobility. Nucleic Acids Res 28:1299–1306

SantaLucia J Jr (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. Proc Natl Acad Sci USA 95:1460–1465

Solignac M, Monnerot M, Mounolou JC (1986) Concerted evolution of sequence repeats in *Drosophila* mitochondrial DNA. J Mol Evol 24:53–60

Stanton D, Daehler LL, Moritz CC, Brown WM (1994) Sequences with the potential to form stem-and-loop structures are associated with coding-region duplications in animal mitochondrial DNA. Genetics 137:233–241

van Oppen MJH, Willis BL, Miller DJ (1999a) Atypically low rate of cytochrome *b* evolution in the scleractinian coral genus *Acropora*. Proc Roy Soc Lond B 266:179–183

van Oppen MJH, Hislop NR, Hagerman PJ, Miller DJ (1999b) Gene content and organization in a segment of the mitochondrial genome of the scleractinian coral *Acropora tenuis*: Major differences in gene order within the anthozoan subclass Zoantharia. Mol Biol Evol 16:1812–1815

van Oppen MJH, McDonald BJ, Willis BL, Miller DJ (2001) The evolutionary history of the coral genus *Acropora* (Scleractinia, Cnidaria) based on a mitochondrial and a nuclear marker: Reticulation, incomplete lineage sorting or morphological convergence? Mol Biol Evol 18:1315–1329

Wallace CC (1999) Staghorn corals of the world: A revision of the genus *Acropora*. CSIRO Publishing, Collingwood

Wilkinson GS, Mayer F, Kerth G, Petri B (1997) Evolution of repeated sequence arrays in the D-loop region of bat mitochondrial DNA. Genetics 146:1035–1048

Wilkinson GS, Chapman AM (1991) Length and sequence variation in evening bat D-loop mtDNA. Genetics 128:607–617

Wolstenholme DR (1992) Animal mitochondrial DNA: Structure and evolution. Int Rev Cytol 141:173–216

Wong TW, Clayton DA (1986) DNA primase of human mitochondria is associated with structural RNA that is essential for enzymatic activity. Cell 45:817–825

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

Zuker M (2000) Calculating nucleic acid secondary structure. Curr Opin Struct Biol 10:303–310