Evolution of monoblepharidalean fungi based on complete mitochondrial genome sequences

C. E. Bullerwell¹, L. Forget and B. F. Lang*

Program in Evolutionary Biology, Canadian Institute for Advanced Research, Département de Biochimie, Université de Montréal, 2900 Boulevard Edouard-Montpetit, Montréal, Québec H3T 1J4, Canada and
¹Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

ABSTRACT

We have determined the complete mitochondrial DNA (mtDNA) sequences of three chytridiomycete fungi, Monoblepharella15, Harpochytrium94, and Harpochytrium105. Our phylogenetic analysis based on concatenated mitochondrial protein sequences confirms the placement of Monoblepharella15 together with Harpochytrium spp. and Hyaloraphidium curvatum within the taxonomic order Monoblepharidales, with overwhelming support. These four mtDNA sequences encode the standard fungal mitochondrial gene complement and, like certain other chytridiomycete fungi, encode a reduced complement of 7–9 tRNAs, some of which require 5′-tRNA editing to be functional. Highly conserved sequence elements were identified upstream of almost all protein-coding genes in the mtDNAs of Monoblepharella15 and both Harpochytrium species. Finally, a guanosine residue is conserved upstream of the predicted ATG or GTG start codons of almost every protein-coding gene in these genomes. The appearance of this G residue correlates with the presence of a non-canonical cytosine residue at position 37 in the anticodon loop of the mitochondrial initiator tRNAs. Based on the unorthodox features in these four genomes, we propose that a 4 bp interaction between the CAUC anticodon of these tRNAs and GAUG/GUG codons is involved in translation initiation in monoblepharidalean mitochondria. Intriguingly, a similar interaction may also be involved in mitochondrial translation initiation in the sea anemone Metridium senile.

INTRODUCTION

Although members of the ‘higher’ fungi (ascomycetes and basidiomycetes) have received the vast majority of scientific study, the ‘lower’ fungi (chytridiomycetes and zygomycetes) contain most of the genetic diversity within the fungal kingdom. Unfortunately, few molecular, and even fewer genomic data, are available currently from lower fungi. In addition, the evolutionary relationships of lower fungi to each other and to the higher fungi are poorly understood.

The most widely used molecular data for inferring fungal phylogeny, rDNA sequence data, have not been able to resolve the fungal phylogeny with significant support (1–5), and thus cannot be used alone to confirm or refine the taxonomic classification of the fungi. In fact, much more sequence information than is contained in rRNA is required to recover the phylogenetic signal needed to resolve a phylogeny as deep as that of the fungi, which are believed to have diverged from a common ancestor with the animals ~1 billion years ago (6). Further, the taxon sampling in other data sets, such as tubulins (7), RPB1 [the largest subunit of RNA polymerase II (8)] and concatenated nuclear protein sequences (9) currently is too poor to be of use in the prediction of the global fungal phylogeny. In contrast, a large amount of complete mitochondrial genome data is available from a broad selection of species from all major lineages. Concatenated mitochondrial protein data presently are the best way of classifying and understanding the evolution of Chytridiomycota and other fungi (e.g. 10–12). A robust phylogenetic tree, such as that inferred using mitochondrial sequence data, can then serve as a framework for comparative mitochondrial genome analysis.

In recent years, considerable progress has been made in understanding mitochondrial genome structure and evolution in a broad selection of eukaryotes using a comparative genomics approach. For example, sequences from closely related organisms have been compared in order to identify promoters or other little conserved sequence motifs (12), and more substantial sequence conservation across larger evolutionary distances has been used to identify previously unrecognized genes [e.g. Rps3 (13) or the fungal mitochondrial RNase P gene rnpB (E.Seif and B.F.Lang, unpublished)]. In addition, sequence variation in gene homologs from organisms with well defined relationships has aided in RNA secondary structure prediction (e.g. the above-mentioned RNase P RNA). This approach has also allowed the time of emergence of traits to be pinpointed [e.g. tRNA editing in chytridiomycetes (14); M.-J.Laforest and B.F.Lang, unpublished results].
The Fungal Mitochondrial Genome Project (FMGP) has as its primary goal to sequence mitochondrial DNAs (mtDNAs) from representatives of all fungal lineages, with a particular focus on the lower fungi (15) (see also http://megasun.bch.umontreal.ca/People/lang/FMGP/progress.html). These new sequence data have been used to study genome structure and content in the fungi, trace the evolution of gene expression in these systems and reconstruct fungal phylogeny. To date, the FMGP has sequenced mtDNAs from members of Ascomycota, Basidiomycota and Zygomycota, as well as from members of four of the five known taxonomic orders of chytridiomycetes: Allomyces macroscopus (Blastocladiales), Rhizophydium136 (Chytridiales), Spizellomyces punctatus (Spizellomyces) and Hyaloraphidium curvatum (Monoblepharidales). Members of the Neocallimastigales, the remaining order in Chytridiomycota, have not been investigated as all known representatives are amitochondriate.

In continuation of the FMGP’s aims, we have determined the mitochondrial genome sequences of three chytridiomycetes, Monoblepharella15, Harpochytrium94 and Harpochytrium105, and compared these sequences with that of H.curvatum. In this study, we (i) present a phylogeny based on mitochondrial protein sequences which further clarifies relationships within the lower fungi; (ii) address issues of genome organization, gene content, gene order and gene expression in these mtDNAs; and (iii) propose that a 4 bp anticodon–codon interaction is involved in mitochondrial translation initiation in this fungal order.

MATERIALS AND METHODS

Strains and culture conditions

The strains Monoblepharella15, Harpochytrium94 and Harpochytrium105 were collected, identified and generously provided by M. R. N. Mollicone and J. E. Longcore (University of Maine). Cells were grown using TYG medium, consisting of 0.25% tryptone, 0.125% yeast extract and 0.3% glucose. Liquid cultures of 300 ml in 2 l Erlenmeyer flasks were inoculated with growing cells and placed at room temperature either with shaking (Harpochytrium spp., 70 r.p.m.) or without shaking (Monoblepharella15) for 2–4 weeks. Purification of mitochondria was performed from ~10 g (wet weight) of cells.

Purification of mtDNAs

Due to their differing growth forms, cells of Harpochytrium spp. and Monoblepharella15 were disrupted differently. Harpochytrium spp. cells ( unicellular) were harvested by centrifugation. After resuspension in a sorbitol buffer (0.6 M sorbitol, 5 mM EDTA, 50 mM Tris pH 8), the cells were broken by shaking in the presence of glass beads. Monoblepharella15 cells (filamentous) were harvested by filtration through cheesecloth, then ground using a pestle and mortar in the presence of glass beads. A sorbitol buffer was used to wash the beads after cell disruption in both cases.

Crude mitochondrial fractions were recovered by differential centrifugation, and these fractions were lysed in the presence of 1% SDS and 100 µg/ml proteinase K at 50°C for 1 h. SDS was eliminated from the lysate by addition of 1 M NaCl. After 1 h on ice, the SDS–protein complex was removed by centrifugation. Nucleic acids were fractionated on a CsCl gradient (1.1 g/ml, 40 000 r.p.m. for 48 h) in the presence of 10 µg/ml bis-benzimide (Hoechst dye 33258). The prominent upper band (AT-rich mtDNA) was extracted and recentrifuged in a second CsCl gradient.

Cloning, sequencing and sequence analysis of mtDNAs

MtDNA was physically sheared by nebulization (16), and a size fraction of 500–3000 bp was recovered after agarose gel electrophoresis. DNAs were incubated in the presence of dNTPs, the Klenow fragment of DNA polymerase I and T7 DNA polymerase to generate blunt ends, and cloned into the Smal site of a modified Bluescript II KS+ vector with a shortened multicloning site (pFBS; B.F.Lang, unpublished). Recombinant plasmids containing mtDNA inserts were identified by colony hybridization using mtDNA as a probe. DNA sequencing, sequence assembly and sequences analysis were performed as described (17).

The mtDNA sequences of Monoblepharella15, Harpochytrium94 and Harpochytrium105 have been deposited in GenBank (accession numbers AY182007, AY182005 and AY182006, respectively).

Purification of mtRNA

A crude mitochondrial pellet was recovered from cell lysates by differential centrifugation (see above). Then 8 M guanidinium chloride was added to lyse mitochondria. Ethanol-precipitated material from this lysate was incubated in the presence of 1% SDS and 100 µg/ml proteinase K at room temperature for 30 min. Following phenol extraction, SDS was eliminated from the lysate by addition of 1 M NaCl. After 20 min on ice, the SDS–protein complex was removed by centrifugation and 2 M LiCl was added for 1 h on ice. After centrifugation, the pellet was enriched in salt-insoluble large RNAs such as rRNAs and mRNAs, whereas the supernatant contained predominantly salt-soluble tRNAs and DNA.

Primer extension analysis

Oligonucleotides used in this study were: atp6, 5′-(G/A)-AA(5′)-(A/T)GGGACCAAGCACC(G/A)GC-3′; atp8, 5′-AAAGTATACGGCAGAAATGAG-3′; nad3, 5′-AAGAGCAGCATCACTATAAGG-3′; nad4L, 5′-ACACGAGATTCCAGACCAGCC-3′; nad5, 5′-GAAGTACAACGTCATGTAACC-3′; nad6, 5′-CCCCAGCTAGGTAAAGACC-3′; and cox3, 5′-CACCGGTTAATATCAGCGGAACCA-3′.

Oligonucleotides were phosphorylated by incubation in the presence of T4 polynucleotide kinase and [γ-32P]ATP. A fraction of the radiolabeled oligonucleotide was annealed to mRNA in the salt-insoluble mitochondrial RNA fraction. Reverse transcription was performed using AMV reverse transcriptase in the presence of dNTPs. The resulting 5′-radiolabeled cDNAs were subjected to electrophoresis in 4% polyacrylamide sequencing gels. The length of the reverse-transcribed cDNA was determined by comparison with a DNA sequencing ladder from a mtDNA clone containing the gene of interest, using the same oligonucleotide (5′-phosphorylated prior to use) as was used for the primer extension.
RT–PCR of tRNAs
Salt-soluble mitochondrial RNAs (enriched in tRNAs) were treated with DNase I. A DNA oligonucleotide (see below) was annealed to the tRNA$^{Met}$, and this mixture was incubated in the presence of AMV reverse transcriptase and dNTPs. cDNAs were amplified by PCR using Taq polymerase I and a second primer. Products were separated on 2% agarose gels. A negative control (DNase-treated tRNAs, no reverse transcription) did not produce an amplification product in the PCR, confirming the cDNA (and not the mitochondrial or nuclear DNA) origin of the PCR product. The PCR products were cloned and sequenced as described (see above).

Primers used were: first strand synthesis, 5′-GCAGAGAAGGATTCGAACC(T/C)C-3′; second strand synthesis, 5′-TAGAATGGAGTAAGGGTAACTCG-3′.

Phylogenetic inference
For phylogenetic analyses, we used a total of 2895 amino acid positions (termed ‘complete’ data set) from 13 concatenated, well conserved protein sequences (Cox1, 2 and 3, Cob, Atp6 and 9, and Nad1, 2, 3, 4, 4L, 5 and 6) that are encoded in most fungal mtDNAs. Multiple sequence alignments were performed with CLUSTALW (18), and regions of uncertain alignment were eliminated by comparing alignments inferred with various parameters, using SOAP (19), thus eliminating the need for further, disputable user interventions. The data used for the phylogenetic analysis will be made available on request.

Many fungal mtDNA sequences have a high rate of evolution. Therefore, certain artifacts of phylogenetic analyses, notably long branch attraction (20) and/or combinations of mutational oversaturation of sites, bias introduced by variation in nucleotide composition, strong codon preferences (10) and non-standard translation codes (14), may obscure true phylogenetic relationships. To minimize the effect of these artifacts, we have chosen to infer phylogenies with protein sequences and to base our conclusions on maximum likelihood (ML) methods, which are known to be least sensitive to potential artifacts due to mutational oversaturation of sites. Phylogenetic inferences employed either the ML method as implemented in PROML (21) or a distance approach using ML distance tables calculated with TREE-PUZZLE (22), and BIONJ (23) for tree construction. In all instances, a $\Gamma$-distribution model of site variation was used [$\alpha = 1.0$, calculated with PAML (24)]. To assess the level of confidence in tree selection, statistical tests were performed employing CONSEL (25), which provides the least biased and most rigorous statistical tests available to date (26,27). The standard AU and WKH tests (results not shown) confirm the topology of the monoblepharidalean lineage shown in Figure 1, i.e. they reject all alternative scenarios at a significance level of 0.05.

Taxon sampling was given particular consideration in these analyses. We excluded fungal taxa that completely lack mitochondrial nad genes [Saccharomyces cerevisiae, as well as the three available Schizosaccharomyces species (12)], limited the number of ascomycetes to the three species with the shortest relative branch lengths, and chose a variety of non-fungal species among protists, animals, plants and bacteria that have relatively low evolutionary rates. Finally, the number of taxa was limited to 30 due to the computational demands of the ML method when using evolutionary models that permit change of rate at all amino acid positions of the protein sequences, and when applying subsequent bootstrap analyses (28).

RESULTS AND DISCUSSION
Molecular phylogeny of the chytridiomycetes
Our molecular phylogenetic analyses (Fig. 1) using 13 concatenated mitochondrial protein sequences, and including an established member of the chytridiomycete order Monoblepharidales Monoblepharella15, support several features of fungal evolution with overwhelming support (95–100% bootstrap support both with ML and distance-based methods). In particular, these analyses support the specific relationship of fungi and animals [Opisthokonts; see for example Wainright et al. (29) and Paquin et al. (30)] and the deep divergence separating Chytridiomycota from Zygomycota, Basidiomycota and Ascomycota (but note that the phylogenetic affiliation of the blastocladialean Allomyces macrognys remains unresolved; see below). Within Chytridiomycota, S.punctatus (Spizellomyces in Fig. 1) and Rhizophydium136 (from the taxonomic orders Spizellomyctales and Chytridiales, respectively), branch separately from the Monoblepharidales group, as expected. Monoblepharella15 and the two Harpochytrium species are sister lineages within the Monoblepharidales to the exclusion of H.curvatum (Hyaloraphidium in Fig. 1). The close relationship of Monoblepharella15 and the two Harpochytrium species is also strongly supported by the high level of gene order conservation in these three genomes (Fig. 2). This grouping of Harpochytrium spp. within the Monoblepharidales is contrary to the assignment of this genus to the Chytridiales (31), but consistent with recent phylogenies using ribosomal sequence data (5). However, it should be noted that studies based on rRNA data have been unable to recover the Monoblepharidales within Chytridiomycota, with significant support (5,32). In contrast, our analyses not only reveal the branching order within the Monoblepharidales, but also confirm the specific affiliation of H.curvatum with this taxonomic order (Fig. 1) (10).

The branching position of A.macrogynys remains uncertain [Allomyces in Fig. 1; see also Leigh et al. (11)] despite the large array of fungal mitochondrial data currently available. In fact, it has proven difficult to confirm or disprove the hypothesis of chytridiomycete paraphyly (15). By using ML inference methods with among-site rate variation on our mitochondrial data set, A.macrogynys has migrated away from a previously well supported affiliation with the Zygomycota + Basidiomycota + Ascomycota clade to a basal position within Chytridiomycota. However, the bootstrap support for the monophyly of Chytridiomycota is only ~60% with both the ‘complete’ and ‘filtered’ data sets (see Materials and Methods). The inconsistency of results using distance and ML methods may be due to a long branch attraction...
phenomenon, which is known to be more prevalent in analyses employing distance methods. We are confident that addition of data from more chytridiomycetes will stabilize the position of A. macrognus within Chytridiomycota.

**Genome organization, gene content and gene order**

The mitochondrial genomes of the monoblepharidalean fungi Monoblepharella15, Harpochytrium94 and Harpochytrium105 were completely sequenced. In contrast to the linear conformation of the mtDNA of H. curvatum (10), the mtDNAs of Monoblepharella15, Harpochytrium94 and Harpochytrium105 map as circular molecules (Fig. 2), although they are probably multimeric concatemers in vivo as are most other mtDNAs (33). All four of these mtDNAs encode the basic fungal set of mitochondrial genes: the large and small subunit rRNAs (SSU rRNAs; rnl and rns, respectively), three subunits of the cytochrome oxidoreductase complex (cox1, 2 and 3), apocytochrome b (cob), three subunits of the ATP synthase complex (atp6, 8 and 9) and seven subunits of the NADH:ubiquinone oxidoreductase complex (nad1, 2, 3, 4, 4L, 5 and 6) (Table 1). As they are in the mtDNA of H. curvatum, genes are encoded on only one strand in the other three mtDNAs. The gene order in the two Harpochytrium species is identical, and highly similar to that in Monoblepharella15 (refer to color code in Fig. 2).

However, there is no gene order conservation between these genomes and that of H. curvatum. This agrees with our understanding of the phylogenetic distances between these organisms (Fig. 1).

The mtDNAs of H. curvatum, Harpochytrium94 and Harpochytrium105 are relatively compact, at 29.97, 19.473 and 24.169 kb, respectively. In contrast, the mtDNA of Monoblepharella15 is 60.440 kb. This is partly due to additional coding regions in the larger genome. Monoblepharella15 encodes five open reading frames (ORFs) with no significant similarity to known proteins, whereas H. curvatum encodes three, and both Harpochytrium species encode none; Monoblepharella15 encodes nine tRNAs whereas H. curvatum encodes seven and the Harpochytrium genomes each encode eight (see below); and Monoblepharella15 encodes eight introns [all of group I, seven of which encode ORFs with either LAGLIDADG or GIY-YIG motifs, characteristic of group I intron ORFs (34)], whereas H. curvatum encodes only one and the Harpochytrium genomes are devoid of introns. Intergenic (non-coding) regions account for the majority of mitochondrial genome size variation in Monoblepharidales.

G + C content is relatively high in these genomes compared with other fungi, at 39% in both H. curvatum and Monoblepharella15 and 36% in Harpochytrium spp. This is

Figure 1. Phylogenetic analysis based on concatenated mitochondrial proteins. The topology shown was inferred using ProML and the JTT model of protein evolution. It was constructed from the ‘complete’ data set of the unambiguously aligned portions of 13 concatenated protein sequences. Bootstrap support for this tree (%, upper number) was calculated from 100 replicates. A similar topology was obtained using ProtDist and BIONJ (%, second number; ‘±’ in cases of differences in topology compared with the ML tree). Bootstrap support for the distance tree was calculated from 1000 replicates. In order to explore the impact of oversaturated amino acid positions in the alignment, an alternative ProML analysis was performed based on the ‘filtered’ data set (see Materials and Methods). Bootstrap support for this ML tree, calculated from 100 replicates, is indicated only for the fungi (%, lower number). The postulated events marking the origin of quartet GAUG/GGUG initiation codons (I), fragmentation of the gene encoding the small subunit rRNA (F), the introduction of 5¢¢ tRNA editing (E), and the drastic reduction in mtDNA-encoded tRNAs (T) are indicated with arrows. Sequences obtained from GenBank: A. castellanii (NC_001637); Allomyces macrognus (U41288); Amoebidium parasiticum (AF538042-AF538052); Arabidopsis thaliana (NC_001284); Aspergillus nidulans (ODAS1, CAA33481, AAA99207, AAA31737, CA25707, AAA31736, CA23994, X15442, P15956, CA22395, CAA3316X00790, X15441, X06960, J01387, X01507); Chondrus crispus (NC_001677); Chrysodidymus synuroideus (NC_001274); Dictyostelium discoideum (NC_000895); H. curvatum (AF402142); Marchantia polymorpha (NC_001637); M. senile (AF538053); Mesorhizobium loti (NC_002678); Nephrolepis olivacea (AI110138); Phytophthora infestans (NC_002387); Podophora anserina (X55026); Porphyra purpurea (NC_002007); R. americana (AF007261); Rhizophyllum136 (NC_003053); Rickettisia prowazekii (NC_000963); Sarcophyton glaucum (AF064823, AF063191); Schizophyllum commune (AF402141); Sinorhizobium melliloti (NC_003047); Spizellomyces punctatus (AF402142); Yarrowia lipolytica (AJ307410). Protein sequences of R. stolonifer can be downloaded from http://megasun.bch.umontreal.ca/People/lang/FMGP/proteins/.
also true in intergenic regions, as Harpochytrium94, Harpochytrium105, H.curvatum and Monoblepharella15 have intergenic G + C contents of 32, 38, 43 and 53%, respectively. The remarkably high value in Monoblepharella15 intergenic regions is due predominantly to the presence of many G + C-rich repeats, possibly the vestiges of double hairpin elements [DHEs (35)]. The intergenic regions of Harpochytrium105 contain several canonical DHE structures (15), whereas those of Harpochytrium94 and H.curvatum do not contain any of these elements. The biology

---

### Table 1. Gene and intron content in fungal mtDNAs

<table>
<thead>
<tr>
<th>Genes</th>
<th>Monoblepharella15</th>
<th>Harpochytrium94</th>
<th>Harpochytrium105</th>
<th>H.curvatum</th>
<th>S.punctatus</th>
<th>A.macrogynus</th>
<th>S.pombe</th>
</tr>
</thead>
<tbody>
<tr>
<td>rnl, rns</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>atp6,8,9</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>cob</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>cox1,2,3</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>nad1-4</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>trnA-W</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>rpsB</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>rps3</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>ORFs&lt;sup&gt;b&lt;/sup&gt;</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Intron&lt;sup&gt;d&lt;/sup&gt;</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>26</td>
<td>2</td>
</tr>
<tr>
<td>Intron&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

<sup>a</sup>Filled squares indicate the presence, open circles the absence, of a gene or genes.

<sup>b</sup>Includes nad1, nad2, nad3, nad4, nad4L, nad5 and nad6.

<sup>c</sup>Only ORFs greater than 100 amino acids in length are listed.

<sup>d</sup>IntronI and IntronII denote introns of group I and group II, respectively.
of DHEs is not well understood, but they have been proposed to be mobile elements and contribute to mitochondrial genome plasticity (12,35).

**Conserved sequence elements upstream of protein-coding genes**

All genes are encoded on the same DNA strand in the four mtDNAs compared here. It is therefore possible that few, or even a single promoter is responsible for the transcription of all genes in each genome. However, nothing is known about promoter structure in chytridiomycete mtDNAs. Certain sequences in the *H. curvatum* mtDNA have been suggested as possible promoters based on their location (10), but this possibility has not been verified experimentally. Similar sequences to those in *H. curvatum* were not identified in the other three Monoblepharidales mtDNAs.

We identified a highly conserved sequence motif specific to the *Monoblepharella15, Harpochytrium94* and *Harpochytrium105* genomes (consensus: 5’-TTATAGGAAAT-3’; Fig. 3A) between 15 and 26 nt upstream of the start codons of the *atp6, atp8, nad3, nad4L, nad5* and *nad6* genes. Additional conservation (consensus: 5’-CTAGGT-3’; Fig. 3A) was identified immediately 3’ to this conserved sequence upstream of the *atp6* and *atp8* genes in all three species. Another conserved sequence (consensus: AGAGTGTANTNNAATT; Fig. 3B) was identified 8–13 nt upstream of the *cox3* gene in these three species. We speculated that some or all of these sequences might serve as promoters in these mtDNAs, like the nonanucleotide sequences which promote transcription in yeast mitochondria (36,37). If these sequences were truly functionally analogous to those in yeast, one would expect that RNA transcripts would begin near the last nucleotide in the conserved sequence block. To test this hypothesis, we performed primer extension experiments to map the 5’ ends of seven mitochondrial transcripts (not shown), which have conserved sequences upstream of their coding regions in the mtDNA of *Monoblepharella15*. In all cases, however, the 5’ end of the RNA mapped 1–6 nt upstream of the 5’ end of the conserved sequence (Fig. 3). These experiments indicate that
the conserved sequences are completely included within RNA transcripts (at least in *Monoblepharella15*), arguing against the role of these sequences as promoters.

The conservation of these sequence motifs close to the end of these RNAs may suggest their implication in 5′-RNA processing. However, we currently favor the alternative hypothesis that these sequences function in translation regulation. Poorly conserved structures and/or sequences [possibly a conserved octanucleotide (38)] in the 5′-untranslated regions of mitochondrial RNA transcripts have been implicated in translation regulation in yeast (39–41). The conserved sequences probably do not serve as Shine–Dalgarno-like sequences [as observed in the minimally derived mtDNA of *Reclinomonas americana* (42)] as no sequences can be identified at the 3′ end of monoblepharidalean SSU tRNAs with which they might interact.

### A reduced set of tRNAs, codon bias and tRNA editing

A reduced set of tRNA genes is found in these four monoblepharidalean mitochondrial genomes, insufficient to decode all codons (Table 2). This is also the case in the mtDNAs of two other chytridiomycete fungi, *S.punctatus* and *Rhizophydi um136*, which encode eight and seven tRNAs, respectively. *Monoblepharella15* encodes nine mitochondrial tRNAs (tRNAAsp, tRNAGlu, tRNAGly, tRNALys, tRNAfMet, tRNAPro, tRNAres, tRNATrp and tRNAYtr), both *Harpochytrium* species encode eight (the same set as in *Harpochytrium94* and *Harpochytrium105*), which encode eight and seven tRNAs, respectively. *Hcurvatum* encodes seven (the same set as in *Hcurvatum* spp. except that tRNALys is absent) and *H.curvatum* encodes seven (the same set as in *Harpochytrium* spp. except that tRNALys is absent). Additional tRNAs required for mitochondrial translation presumably are encoded in the nucleus and imported into mitochondria from the cytoplasm. Interestingly, tRNAres, tRNAPro, tRNAGln, tRNAtrp and tRNAtyr are found in all six of these chytridiomycete mtDNAs, testifying to a common history of tRNA gene loss in this lineage. Retention of these particular tRNAs perhaps is due to a difficulty in replacing features which are not found in their cytoplasmic tRNA counterparts. For example, tRNAMet must be recognized and formylated in order to serve as the initiator tRNA for mitochondrial protein synthesis. It might therefore be retained because the mitochondrial methionyl-tRNAfMet transformylase cannot recognize the cytoplasmic initiator tRNA. As the chytridiomycete *A.macrogynus* encodes a tRNA set capable of decoding all codons in its mtDNA (43), we assume that the majority of tRNAs were lost deep within the chytridiomycete lineage after the divergence of *A.macrogynus* (Fig. 1).

Several codons are not used in the standard protein-coding genes encoded in these genomes (i.e. not including intronic or free-standing ORFs), although the universal translation code is otherwise used. Of 64 possible codons, *H.curvatum* does not use seven (TGA, TCA, ATA, GCA, GCG, AGA and AGG). *Monoblepharella15* does not use 10 (TGA, CTG, TCA, TCG, ACA, ACG, CAG, AAG, GAG and CGG), and the two *Harpochytrium* species do not use 14 (the same as those in *Monoblepharella15*, as well as ATA, GCG, AGG and GGA). This marked non-use of codons is in contrast to other fungal mtDNAs, including those of *S.punctatus* and *Rhizophydi um136*, which use all but two and three codons, respectively. Interestingly, all of the unused codons in the four monoblepharidalean genomes have a purine residue in the third position. The significance of this observation is unknown.

Finally, the majority of tRNAs in these genomes contain mismatches in the first three base pairs of their acceptor stems, potentially rendering them non-functional (10). The acceptor stem mismatches in *Monoblepharel lla15*, *Harpochytrium94* and *Harpochytrium105* have been shown to be corrected at the RNA level to create standard Watson–Crick base pairs, by removal of the first 3 nt from the 5′-half of the acceptor stem followed by nucleotide addition (M.-J.Laforest and B.F.Lang, unpublished). This type of tRNA editing was first identified in the ameoboid protist *Acanthamoeba castellanii* (44), and has also been described in the chytridiomycete *S.punctatus* (14). The origin, evolution and mechanism of this form of RNA editing in chytridiomycete mitochondria will be discussed in detail elsewhere (M.-J.Laforest and B.F.Lang, unpublished).

### SSU rRNA in pieces

Fragmented and scrambled mitochondrial rRNA genes were first observed in *Tetrahymena pyriformis* (45). They also occur in the mitochondria of certain green algae (46–48) and in alveolates such as *Plasmodium falciparum* (49) and *Theileria parva* (50). In *Monoblepharella15* and both *Harpochytrium* species, the mitochondrial SSU rRNA is encoded in two pieces, exactly as it is in *H.curvatum* (10). The 3′-fragment is found upstream of the 5′-fragment in all four cases, separated by a gene (either *nad2* or *rnl*; see Fig. 2). The secondary structures of these rRNAs, derived from the gene sequences, indicate structures consistent with the eubacterial model. The inferred break point in the SSU rRNA is located within the same variable region in all four cases, corresponding to nucleotides 590–649 of the *Escherichia coli* SSU rRNA, and the two rRNA fragments have the potential to assemble by

### Table 2. Partially shared, reduced complement of tRNAs encoded by chytridiomycete mtDNAsab

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Monoblepharella15</em></td>
<td>●</td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td><em>Harpochytrium94</em></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td><em>Harpochytrium105</em></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td><em>H.curvatum</em></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td><em>S.punctatus</em></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td><em>Rhizophydi um136</em></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

a) tRNA species are identified by their corresponding amino acid (in one-letter code) and anticodon sequence.

b) Filled squares indicate the presence, open circles the absence, of a gene.

c) An unusual tRNA decodes UAG `stop' codons as leucine in *S.punctatus* and *Rhizophydi um136*. 

Nucleic Acids Research, 2003, Vol. 31, No. 6
intermolecular base pairing. It should be noted that there may be break points in other variable regions of these genes, and determination of rRNA ends would be necessary to establish the true fragmentation pattern. The rns gene is not fragmented in the mtDNAs of the three other characterized chytridomycetes, A. macrogyrus, S. punctatus and Rhizophyllum136. This suggests that fragmentation of the rns gene occurred within the Monoblephariales lineage prior to the divergence of H. curvatum (Fig. 1).

A 4 bp interaction is potentially involved in translation initiation

Comparison of the four Monoblephariales mitochondrial genome sequences revealed an unusual feature: a guanosine residue is found immediately 5' to the ATG and GTG start codons (inferred by protein sequence alignment) of almost all protein-coding genes. The only protein-coding genes for which a G does not precede the start codon are the cox1 genes of Monoblepharella15 and Harpochytrium spp. (whereas it is present in the cox1 gene of H. curvatum) and the nad1 gene of H. curvatum (whereas it is present in the nad1 genes of Monoblepharella15, Harpochytrium94 and Harpochytrium105). No alternative in-frame GATG or GGTT codons are present in either of these two cases. Further analysis revealed that the initiator tRNAsMet encoded by these mtDNAs also contain unusual features (Fig. 4). Most tRNAs adhere to a well defined secondary structure, which includes a 5 bp anticodon stem and a 7 nt anticodon loop. The tRNAsMet of Monoblepharella15, Harpochytrium94 and Harpochytrium105, on the other hand, have mismatches at the base of the anticodon stem (CxC in Monoblepharella15, UxU in Harpochytrium spp.). In fact, non-Watson–Crick base pairs are observed at this position in all four Monoblephariales tRNAsMet (a U–G pair is present at this position in H. curvatum). Further, the nucleotide at position 37 (immediately 3' to the anticodon) of all four of these tRNAs is occupied by a cytidine residue (Fig. 4). This feature is particularly surprising, as a purine residue, usually modified, is found in the corresponding position of the vast majority of tRNAs.

As editing of tRNA acceptor stems is observed in the mitochondria of these organisms, it was formally possible that another type of tRNA editing altered the sequence of the anticodon stem and/or loop. Therefore, we reverse-transcribed and PCR-amplified the anticodon stem–loop region of tRNAMet from both Monoblepharella15 and Harpochytrium105. In both cases, cDNA sequencing demonstrated that these unusual tRNAs are expressed, and that the RNA sequence of the amplified regions was identical to that encoded by the mtDNA (not shown). Therefore, we assume that these are functional mitochondrial tRNAs.

The unorthodox features of the initiator tRNAs encoded by these mtDNAs, in conjunction with those of initiation codons, strongly suggest that a 4 bp interaction between the CAUC anticodon and quartet GAUG/GGUG initiation codons is involved in translation initiation in Monoblephariales mitochondria (Fig. 4), in contrast to the standard 3 bp interaction. Extended anticodon–codon interactions occur in other contexts; in fact, selection experiments have shown that the most efficient engineered suppressor tRNAs for frameshifting at four-base codons have Watson–Crick complementarity at all four bases of the codon, supporting direct interaction of a four-base anticodon and codon (51). More recently, evidence has been presented supporting an extended interaction in translation initiation in Chlamydomonas reinhardtii chloroplasts (52). In this case, additional base pairing is between the unmodified adenosine at position 37 of tRNAMet and the uridine residue immediately 5' to an initiation codon (uridine is found at this position in 75% of initiation codons in this system). The absence of a modified purine residue at position 37 of monoblephariales mitochondrial tRNAsMet (as well as C. reinhardtii chloroplast tRNA^{Met}) may be important for this type of interaction to occur, as the presence of a modified purine residue would be expected to impede interaction with the mRNA during translation (53–55). Finally, the lack of a Watson–Crick base pair at the base of the anticodon stem of Monoblephariales tRNAsMet may be required for increased flexibility of the anticodon loop, allowing the presentation of the extended anticodon in a conformation which makes it possible for the interaction to occur. It is of great interest to understand the function of this unorthodox feature of mitochondrial translation. One attractive possibility is that...
the additional anticodon–codon complementarity allows more accurate selection of true initiation codons.

Intriguingly, the mitochondrial initiator tRNA in the sea anemone Metridium senile also encodes a cytosine at position 37 (56). Inspection of the M. senile mtDNA reveals that 5 of the 13 protein-coding genes in this genome (cox1, cox2, atp6, nad2 and nad3) have a guanosine residue immediately preceding the ATG initiation codon predicted by the authors (56). Three other genes have in-frame GATG codons in close proximity to the predicted initiation codon (one codon downstream in atp6; seven codons downstream in nad1; 12 codons downstream in cox2). The remaining five genes, cox3, nad4, nad4L, nad5 and nad6, do not have a GATG codon in close proximity to the predicted initiation codon, although an in-frame GGTTG codon is located five codons upstream of the predicted initiation codon of the nad4L gene. There is therefore a strong tendency for a G immediately 5’ to the likely ATG initiation codon (8 out of 13), whereas the overall G + C content of this mtDNA is only 37.5%. Based on the correlation of these features of initiation codons and tRNA-Mt, it is probable that a 4 bp interaction is involved in translation initiation in M. senile mitochondria, as we propose here for monoblepharidalean mitochondria. As there is no evidence to support a specific relationship between Monoblepharidae and the cnidarian M. senile (see Fig. 1), this extended anticodon–codon interaction must have evolved independently in each lineage. Furthermore, it does not appear to be widespread even among cnidarians, as the mitochondrial initiator tRNAs of two other distant members of this group, Acropora tenuis and Sarcophyton glaucum, have purine residues at position 37 (57,58). It is possible nevertheless that evidence of further 4 bp anticodon–codon interactions involved in translation initiation will come to light as mitochondrial genome sequences from close relatives of M. senile become available.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank J. E. Longcore and M. R. N. Mollicone (both of the University of Maine) for identification and supply of lower fungal strains. We also acknowledge the contributions of H. Hamel and M.-J. Laforest in DNA sequencing, and the excellent technical assistance of Z. Wang. This research was supported by the Canadian Institutes of Health Research (CIHR grants MT-14028 and MOP24176 to B.F.L.), an academic equipment grant by SUN Microsystems (Palo Alto, CA, to B.F.L.), the donation of an automatic sequencer by Licor (Lincoln, NB, to B.F.L.) and a Master’s Research Scholarship from Fonds F.C.A.R. (C.E.B.). Salary and interaction support from the Canadian Institute for Advanced Research is gratefully acknowledged.

REFERENCES


