648–654 *Nucleic Acids Research, 1996, Vol. 24, No. 4*

Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage

Nicolas Cermakian[1], Tatsuya M. Ikeda[2], Robert Cedergren[1] and Michael W. Gray[2,*]

Program in Evolutionary Biology, Canadian Institute for Advanced Research, [1]Département de Biochimie, Université de Montréal, Montréal, Québec H3C 3J7, Canada and [2]Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

Received October 30, 1995; Revised and Accepted January 5, 1996

GenBank accession nos U34283–U34286 and U34402–U34406

ABSTRACT

Although mitochondria and chloroplasts are considered to be descendants of eubacteria-like endosymbionts, the mitochondrial RNA polymerase of yeast is a nucleus-encoded, single-subunit enzyme homologous to bacteriophage T3 and T7 RNA polymerases, rather than a multi-component, eubacterial-type $\alpha_2\beta\beta'$ enzyme, as encoded in chloroplast DNA. To broaden our knowledge of the mitochondrial transcriptional apparatus, we have used a polymerase chain reaction (PCR) approach designed to amplify an internal portion of phage T3/T7-like RNA polymerase genes. Using this strategy, we have recovered sequences homologous to yeast mitochondrial and phage T3/T7 RNA polymerases from a phylogenetically broad range of multicellular and unicellular eukaryotes. These organisms display diverse patterns of mitochondrial genome organization and expression, and include species that separated from the main eukaryotic line early in the evolution of this lineage. In certain cases, we can deduce that PCR-amplified sequences, some of which contain small introns, are localized in nuclear DNA. We infer that the T3/T7-like RNA polymerase sequences reported here are likely derived from genes encoding the mitochondrial RNA polymerase in the organisms in which they occur, suggesting that a phage T3/T7-like RNA polymerase was recruited to act in transcription in the mitochondrion at an early stage in the evolution of this organelle.

INTRODUCTION

The evolutionary origin of the mitochondrial (mt) transcription system is puzzling. It is generally accepted that both mitochondria and chloroplasts arose from eubacteria-like endosymbionts, closely related to contemporary α-Proteobacteria and Cyanobacteria, respectively (1–3). Consistent with this view, chloroplast DNA is known to encode components of a eubacteria-like $\alpha_2\beta\beta'$ RNA polymerase (RNAP) (4). No such genes have been found in any of the mitochondrial genomes sequenced to date; instead, all of the genes for the mitochondrial transcriptional machinery appear to be encoded by nuclear DNA, with the protein products being imported into the organelle (5).

In yeast (*Saccharomyces cerevisiae*), the mitochondrial RNA polymerase (mtRNAP) is encoded by a nuclear gene, *RPO41* (6,7), and is a homolog of the single-polypeptide RNAPs of bacteriophages T3 and T7 (8). In view of the similar endosymbiotic origins of mitochondria and chloroplasts and the discovery of eubacteria-like RNAP genes in chloroplast DNA, this is a surprising finding. Moreover, this observation raises questions about the nature of the mitochondrial transcription system in other eukaryotes, and the evolutionary origin(s) of this system.

The identification of a second T3/T7-like mtRNAP sequence in *Neurospora crassa* (9), coupled with the recent appearance of homologous sequences in various expressed sequence tag (EST) databases (human, rice, *Caenorhabditis elegans*; Fig. 1), prompted us to devise a polymerase chain reaction (PCR) amplification strategy to search more widely within the eukaryotic lineage for phage T3/T7-like RNAP sequences. Information about the types and phylogenetic distribution of mtRNAPs, and their structural similarity to one another, is necessary to determine whether the mtRNAPs of different eukaryotes all arose from a single common ancestor and, if so, what the evolutionary source of this enzyme might have been. In particular, because the multicellular eukaryotes (animals, fungi, plants) represent relatively late radiations in the eukaryotic lineage (10), it is important to explore a phylogenetically broad range of earlier diverging unicellular eukaryotes (protists), in order to address the question of whether a T3/T7-like mtRNAP was acquired early or at a relatively late stage in the evolution of the mitochondrial transcription system. Comparative information about mtRNAPs and other transcriptional components in different eukaryotes is also essential for defining and understanding species-specific peculiarities in the biochemical mechanism of expression of mitochondrial genomes that can vary tremendously in size, base composition and organization.

MATERIALS AND METHODS

DNA and RNA preparations

Nuclear DNA from *Pycnococcus provasolii*, *Thraustochytrium aureum* and *Isochrysis* sp. Tahiti was isolated by B.F. Lang. The

* To whom correspondence should be addressed

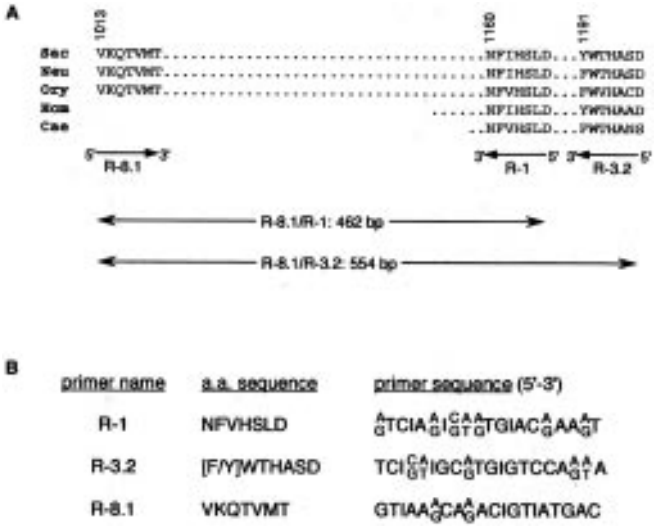

Figure 1. Design of PCR primers and strategy employed for amplification of putative mtRNAP sequences. (**A**) The amino acid sequences of established or putative mtRNAPs from five organisms were aligned, and oligonucleotide primers were designed against three highly conserved sequence motifs, as indicated. Abbreviations of organism names and GenBank accession numbers of the sequences used (square brackets) are: Sac, yeast (*Saccharomyces cerevisiae*) [M17539]; Neu, *Neurospora crassa* [L25087]; Ory, rice (*Oryza sativa*) [D24565, D23514]; Hom, human (*Homo sapiens*) [T97038, T93942, R31174, H03471, H03562]; Cae, *Caenorhabditis elegans* [D34229, D32914]. The three numbers above the aligned sequences (1013, 1160 and 1191) are the residue positions in the Sac sequence (8); the corresponding coordinates in the T7 RNAP sequence are 630, 781 and 814, respectively. Additional rice sequence beyond that in the database was generated as described in the text. (**B**) Designation and sequence of degenerate primers based on the alignment shown in (A) (I = inosine).

organisms were cultured axenically, with *Isochrysis* and *Pycnococcus* grown in sterilized seawater with additions (F/2 medium; for further details, consult WWW site URL http://megasun.bch.umontreal.ca/People/lang/FMGP/methods.html). Cells were broken mechanically by shaking with glass beads (11), DNA was solubilized in 0.5% SDS, and proteins were hydrolyzed in the presence of 100 µg/ml proteinase K. After removal of detergent by salt precipitation (1 M NaCl, 1 h on ice), the isolated total cellular DNA was further purified by equilibrium centrifugation in CsCl density gradients (1.1 g/ml CsCl, 10 µg/ml Hoechst dye 33258 (Serva); 40 000 r.p.m., 48 h). The main, lower band in the gradient, representing the nuclear DNA fraction in all three species, was used.

Nuclear DNA from *Acanthamoeba castellanii* (strain Neff; axenic culture) was prepared by K.M. Lonergan from isolated nuclei and further purified by two rounds of CsCl density-gradient centrifugation after RNase A treatment (12).

Cells from an axenic culture of *Cryptomonas* Φ were supplied by M.A. Ragan, and DNA and cellular RNA were isolated from these by D.F. Spencer. Cells were lysed in 1% SDS containing 1 M sodium perchlorate, following which the lysate was extracted with chloroform/isoamyl alcohol and nucleic acids precipitated with isopropanol. The pellet was redissolved and extracted several times with phenol–cresol, after which the nucleic acids were twice precipitated from ethanol before further purification by CsCl density-gradient centrifugation in the presence of Hoechst dye 33258. The main-band (nuclear) DNA fraction and RNA pellet were recovered.

DNA was prepared from isolated nuclei of an axenic culture of *Naegleria fowleri* (strain LEE) and further purified by CsCl density-gradient centrifugation (13). The *N.fowleri* DNA sample was obtained from R. N. Band via A. Roger and P. Keeling (Department of Biochemistry, Dalhousie University).

Cells of *Tetrahymena pyriformis* (axenic culture), from which total cellular DNA was prepared by phenol extraction followed by RNase treatment, were grown by J. Edqvist.

A wheat (*Triticum aestivum*) cDNA clone bank was graciously provided by B.G. Lane. The library was constructed with mRNA that had been isolated as described (14), using λgt11 as vector in a Y1090 host (B.G. Lane, personal communication).

Total cellular RNA from rice (*Oryza sativa* cv. Lacassine) was kindly provided by P. Gros. The RNA sample was prepared from leaves homogenized in the presence of guanidinium hydrochloride, followed by phenol extraction.

Oligonucleotide primers

Degenerate primers used in this study are listed in Figure 1. Primers specific for λgt11 DNA were used to pre-amplify cDNA inserts, as follows: forward, 5′-ACTCCTGGAGCCCGTCAGTA-3′; reverse, 5′-CAGACCAACTGGTAATGGTA-3′. For amplification of rice cDNA, the following were used: primer 1, 5′-AACAACCGGAAGACCAACG-3′; primer 2, 5′-CGTCCATTTGACAGGGTGG-3′.

PCR amplification

With *Pycnococcus*, *Thraustochytrium* and *Isochrysis* DNAs, PCR amplification was performed in a total volume of 50 µl containing PCR reaction buffer (Pharmacia) and 100 ng DNA, 0.2 mM each dNTP, 1 µM each primer (R-3.2 and R-8.1) and 100 µg/ml BSA. Reactions were carried out in a Perkin Elmer GeneAmp PCR System 9600, according to the following protocol: denaturation at 95°C for 4 min; addition of 1 U *Taq* DNA polymerase at 75°C; 3 cycles of 95°C for 1 min, 40°C for 1 min, slow ramp to 72°C for 1 min, and hold at 72°C for 3 min; 30 cycles of 95°C for 1 min, 50°C for 1 min, slow ramp to 72°C for 1 min, hold at 72°C for 3 min; hold at 72°C for 4 min. PCR products were resolved by agarose gel electrophoresis and extracted using QIAquick columns (QIAGEN).

With *Acanthamoeba*, *Cryptomonas*, *Naegleria*, *Tetrahymena* and *Triticum* (wheat) DNAs, PCR was performed in a total volume of 50 µl containing reaction buffer [25 mM Tricine (pH 8.5), 16 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$], 100 ng DNA, 0.1 mM each dNTP, 2 µM each primer (R-3.2 or R-1 and R-8.1), 1 U *Taq* DNA polymerase (Gibco-BRL) and 0.01 U *Pfu* DNA polymerase (Stratagene). Reactions were performed in a Perkin Elmer GeneAmp PCR System 2400, according to the following protocol: denaturation at 94°C for 3 min; 35 cycles of 94°C for 30 s, 50°C for 1 min, 55°C for 20 s, 60°C for 10 s and 72°C for 2–3 min, followed by an extension at 72°C for 10–15 min. PCR products were isolated after electrophoresis in low-melting-point agarose gels, as above. All PCR fragments were cloned into pT7Blue (Novagen) and sequenced using Sequenase version 2.0 (USB).

From a wheat cDNA library, inserts were pre-amplified in a total volume of 50 µl containing reaction buffer [25 mM Tricine (pH 8.5), 16 mM $(NH_4)_2SO_4$, 2 mM $MgCl_2$], 100 ng of DNA, 0.1 mM each dNTP, 0.2 µM each primer specific for λgt11 DNA (forward and reverse), 1 U *Taq* DNA polymerase and 0.01 U *Pfu* DNA polymerase. The reaction was performed according to the

following protocol: denaturation at 94°C for 3 min; 25 cycles of 94°C for 30 s, 55°C for 30 s and 72°C for 2 min, followed by an extension at 72°C for 5 min. A 100-fold dilution of amplified cDNAs was used for the PCR reaction following the same method as used for *Acanthamoeba*, *Cryptomonas*, *Naegleria* and *Tetrahymena* genomic DNAs.

Additional rice sequence was obtained by rapid amplification of cDNA ends (RACE) with construction of specific primers based on a rice EST (accession no. D23514). Rice total cellular RNA (5 µg) was used as template for reverse transcription by MuMLV (NEB) in the presence of 1 mM dNTP (Pharmacia), 15 U RNAguard (Pharmacia) and 50 ng of primer 1 in 20 µl containing 50 mM Tris–HCl (pH 8.3), 8 mM $MgCl_2$, 10 mM dithiothreitol for 1 h at 37°C, 30 min at 42°C and 15 min at 52°C. The resulting cDNAs were separated from primer using a Centricon-30 concentrator (Amicon) and extended at the 3′-end by terminal deoxynucleotidyl transferase (Gibco-BRL) in the presence of dATP. A 200-fold dilution of polyadenylated cDNAs was taken for two nested PCR runs, using primer 1 in the first run and primer 2 in the second. The conditions of the two runs and the sequences of the non-specific primers used were as described (15).

RESULTS

Identification of sequence homologs of T3/T7 RNA polymerase genes in eukaryotic DNA

From available protein sequences, we designed a set of degenerate oligonucleotide primers targeted to DNA sequences encoding regions highly conserved among known or putative mtRNAPs (all homologous to T3 and T7 RNAPs). The latter include rice, human and *C.elegans* expressed sequence tags (ESTs) that have recently appeared in public domain databases (Fig. 1). Primer combinations were tested in PCR amplification experiments with total cellular DNA preparations from a phylogenetically broad range of eukaryotes. The characteristics of positive PCR amplification products are summarized in Table 1. In a number of cases (including all red algae tested and several early diverging, amitochondriate eukaryotes), negative results were obtained with cellular DNA samples: i.e. either no discrete amplification products were obtained, or ones that were generated proved to be unrelated to T3/T7 RNAPs. Aside from the possibility that the gene in question may actually be absent in these negative cases, failure to generate a positive PCR product could result from any number of factors, including sub-optimal concentration of the target sequence in selected DNA preparations (high genomic complexity), spurious presence of competing sequences that sequester PCR primers, presence of introns that interfere with PCR amplification, and sequence divergence at primer binding sites.

An alignment of T3/T7-like RNAP sequences amplified from eight eukaryotes (seven of them protists) is shown in Figure 2. Also included are known bacteriophage (T7, K11, SP6) and fungal mitochondrial (yeast, *Neurospora crassa*) RNAP sequences, as well as human, rice and *C.elegans* EST homologs. The alignment excludes several blocks (the sizes of which are indicated in Fig. 2) that display pronounced sequence and length variation. Within the remaining alignable blocks (I–III), a high degree of amino acid sequence identity is evident, with a number of positions universally conserved or almost so among the eukaryotic sequences, and to a lesser extent between the eukaryotic and phage ones. Despite the phylogenetic breadth of the source organisms, the extent and degree of positional identity leaves little doubt that all of these sequences are related by descent from a common ancestral sequence (i.e. are homologous).

The amplified region (encompassing T7 positions 637–813) comprises part of the palm and fingers domain of the recently determined, hand-shaped T7 RNAP crystal structure (16). The most highly conserved portions of the alignment correspond mainly to regions facing the template-binding cleft of the enzyme. All of the sequences shown in Figure 2 contain the invariant and catalytically essential aspartate residue equivalent to D812 in T7 RNAP (17), as well as the adjacent and catalytically significant H811 (17). It should be noted that these two functionally important residues, as well as the equivalents of T7 RNAP residues Y639 and G440, are also found in DNA polymerases (18,19); however, overall sequence similarity clearly identifies the non-phage sequences in Figure 2 as RNAP, not DNAP, homologs.

Table 1. Characteristics of PCR amplification products containing T3/T7 RNAP-like sequences[a]

Organism	Abbreviation	Phylum	Type	Product (bp)	Cristal type	Introns
Triticum aestivum	Tri	Angiospermatophyta	land plant	500	flattened	?[b]
Pycnococcus provasolii	Pyc	Chlorophyta	green alga (prasinophyte)	692	flattened	0
Acanthamoeba castellanii	Aca	Rhizopoda	amoeboid protozoon	820	tubular	3
Isochrysis sp. Tahiti	Iso	Haptophyta	haptomonad alga	524	tubular	0
Thraustochytrium aureum	Thr	Heterokonta	slime net	581	tubular	0
Cryptomonas Φ	Cry	Cryptista	cryptomonad alga	824	flattened	6
Tetrahymena pyriformis	Tet	Ciliophora	ciliate protozoon	728	tubular	3
Naegleria fowleri	Nae	Percolozoa	amoeboid protozoon	411	discoidal	0

[a]All of these PCR products were obtained using the primer combination R-8.1/R-3.2 (see Fig. 1) except for the *N.fowleri* one, which amplified only with the R-8.1/R-1 combination. Negative results were obtained with *Dennstaedtia punctilobula* (hay-scented fern); *Chlamydomonas reinhardtii* (green alga); *Palmaria palmaria*, *Chondrus crispus*, *Mastocarpus stellatus* and *Gracilaria sordida* (red algae); *Cyanophora paradoxa* (glaucocystophyte); *Goniomonas truncata* Stein (heterotrophic nanoflagellate); *Crypthecodinium cohnii* (dinoflagellate); *Crithidia fasciculata* (kinetoplastid flagellate); *Euglena gracilis* (euglenoid flagellate); *Reclinomonas americana* and *Jakoba libera* (jakobid flagellates); and two amitochondriate eukaryotes (*Giardia lamblia*, *Trichomonas vaginalis*). Negative results were also obtained with a soybean cDNA library (Clontech).

[b]Sequence determined from cDNA clone.

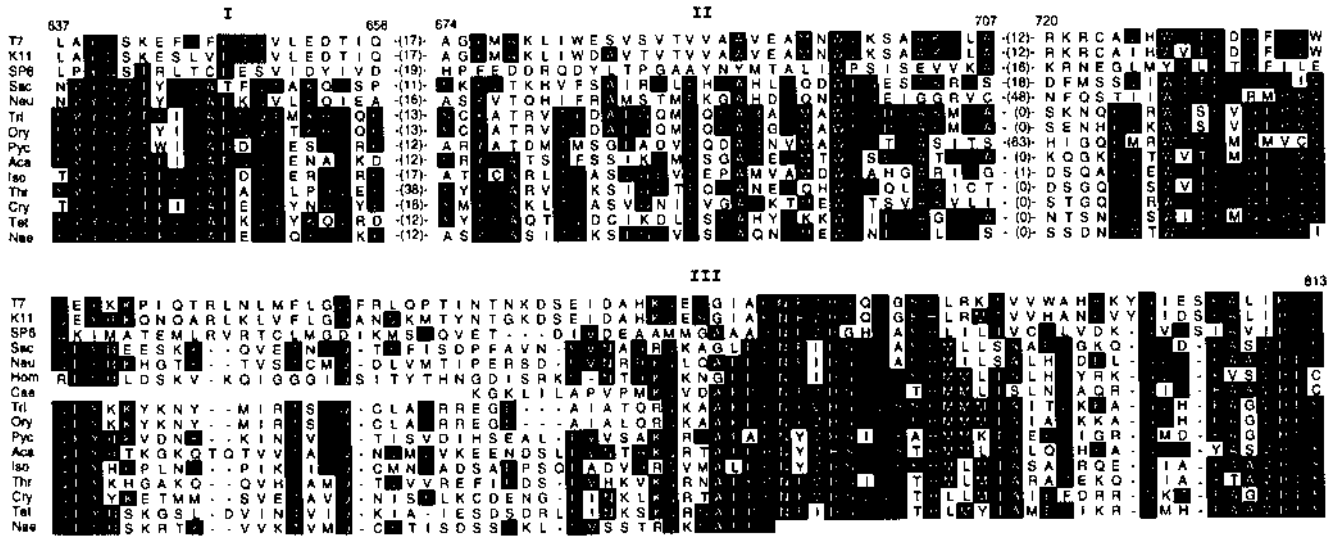

Figure 2. Alignment of RNAP sequences (sequences and abbreviations of organism names are listed in Figure 1 and Table 1). Sequences determined in the present study have been deposited in GenBank (accession numbers U34283–86 and U34402–06). Accession numbers for sequences taken from the database are listed in Figure 1, with additional ones as follows: T7 [M38308], K11 [X53238], SP6 [Y00105]. Note that EST databases contain several human and *C.elegans* sequences homologous to the yeast mtRNAP (see Fig. 1), only one of which is included here in each case (Hom, T97038; Cae, D34229). The alignment was constructed using the Multalin program (41) and refined manually; residue numbering refers to the bacteriophage T7 RNAP sequence. Only blocks of residues that can be unambiguously aligned (I, II and III, corresponding to positions 637–656, 674–707 and 720–813 in the T7 mtRNAP) are shown; numbers in parentheses indicate the sizes of less well conserved regions that vary considerably in length and sequence. Amino acids that are positionally identical in at least five of the sequences are highlighted (white on black). The complete alignment is available from the authors upon request (contact M.W.G.; Email address: mgray@ac.dal.ca).

At several positions (e.g. 638, 641, 646, 689, 738 and 810), the yeast mtRNAP and its PCR-amplified homologs are distinguished as a group from the phage RNAP sequences. For example, at T7 residue 641, all of the non-phage sequences have Val rather than the Ser found in the three phage sequences. Interestingly, an S641A mutant of T7 RNAP displays substantial DNA polymerase activity (20). Although they lack Ser at position 641, all of the non-phage sequences in Figure 2 have another hydroxy amino acid (Thr) at the immediately adjacent position (642).

Intron sequences in PCR amplification products

In several cases, PCR amplification generated positive products that were substantially larger than expected (compare Fig. 1 and Table 1). In these instances (*Acanthamoeba castellanii, Cryptomonas* Φ and *Tetrahymena pyriformis*), we infer the presence of small intron sequences in the PCR products (Fig. 3 and Table 2). Exon/intron boundaries were assigned based on consideration of optimal amino acid sequence alignment and maintenance of open reading frames, as well as the assumption that introns begin with GT and end in AG. The deduced intron sequences display distinctive base compositions relative to their flanking exons, being particularly C+T-rich (61–69%) in *Acanthamoeba* and A+T-rich (81–85%) in *Tetrahymena* (Table 2); the latter feature is characteristic of spliceosomal-type nuclear introns in this organism (21). Intron junction sequences also correspond closely to known consensus sequences at intron splice sites in *Acanthamoeba* and *Tetrahymena* nuclear genes (Table 3). A well-defined consensus sequence is also evident at *Cryptomonas* intron junctions; in this case, however, no published nuclear intron sequences are available for comparison.

In the case of *Cryptomonas*, RT–PCR using total cellular RNA gave a product having a size expected for amplification from a spliced message; sequencing of this product confirmed the absence of putative intron sequences and verified the inferred splice junctions (data not shown). For the *Acanthamoeba* and *Tetrahymena* PCR products, the actual existence of the deduced introns and the validity of the assigned exon–intron boundaries remain to be confirmed by cDNA analysis. Although the positions of several of the introns are quite close to one another (e.g. the final intron in each case), none of the assigned insertion sites is precisely shared among the three organisms.

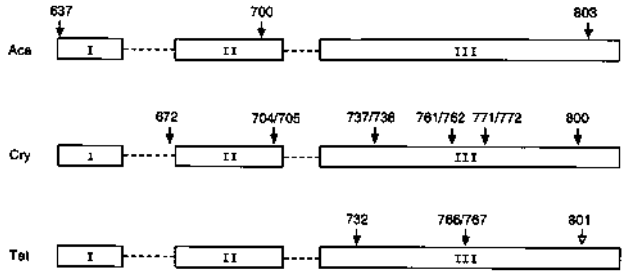

Figure 3. Positions of putative introns (arrows) in *Acanthamoeba* (Aca), *Cryptomonas* (Cry) and *Tetrahymena* (Tet) PCR products, deduced from considerations of amino acid sequence alignment, as described in the text. Intron positions are numbered with reference to the T7 RNAP sequence, with insertion sites falling either within (single numbers) or between (double numbers) the indicated codons. Introns that disrupt the RNAP reading frame are denoted by solid arrows, whereas one that is continuous with upstream and downstream exons and contains no termination codons is indicated by an open arrow. Rectangles labelled I, II and III refer to regions of the sequence specifying highly conserved blocks of amino acid residues; dashed lines refer to more variable regions (see Fig. 2).

Table 2. Characteristics of deduced introns in T3/T7 RNAP-like sequences isolated by PCR amplification of eukaryotic DNA

Organism	Intron position[a]	Size (bp)	G+C (%)	C+T (%)
A.castellanii	637	104	51.0	69.2
	700	97	47.4	67.0
	803	97	51.4	60.7
			(61.4)[b]	(49.3)[c]
Cryptomonas Φ	672	52	50.0	61.5
	704/705	47	53.2	63.8
	737/738	57	60.7	42.9
	761/762	49	51.0	69.4
	771/772	57	53.2	57.4
	800	52	65.4	55.8
			(52.2)[b]	(49.9)[c]
T.pyriformis	732	67	14.9	42.3
	766/767	62	19.4	54.8
	801	93	16.1	63.4
			(35.6)[b]	(51.9)[c]

[a]Numbering refers to the homologous codon(s) in the T7 RNAP sequence.
[b]G+C content (%) of flanking exons.
[c]C+T content (%) of flanking exons.

DISCUSSION

The T3/T7-like sequences reported here, which are also homologous at the amino acid level to the yeast mtRNAP sequence, represent a phylogenetically broad sampling within the eukaryotic lineage. In several cases (e.g. *Cryptomonas*, wheat), we have direct evidence that the PCR-amplified gene is expressed. In the remaining instances, additional work will be required to verify that the gene we have identified is expressed, and that it encodes a functional mitochondrial RNAP. Whether or not all of these sequences ultimately prove to be functional, their widespread occurrence throughout the eukaryotic lineage raises intriguing questions about their evolutionary origin, and their present or former function.

The species from which positive PCR products were obtained include members of a number of major protist phyla, representing all three mitochondrial cristal types (discoidal, tubular, flattened) (Table 1). The organisms branch widely within a eukaryotic phylogenetic tree [see e.g. Cavalier-Smith (22)]. Of particular interest is our finding of a mtRNAP homolog in the protozoon, *Naegleria fowleri. Naegleria* is thought to be one of the earliest diverging genera among mitochondria-containing eukaryotes (22). It is also striking that sequences homologous to yeast mtRNAP could be isolated from organisms displaying very diverse patterns of mitochondrial gene organization and expression, such variability being a hallmark of mitochondrial genomes (23). Allowing that the PCR-derived T3/T7-like sequences described here likely encode a portion of the mtRNAP in these organisms, this would suggest that transcriptional mechanisms

Table 3. Characteristics of inferred intron junction sequences in T3/T7-like RNAP amplification products[1]

Acanthamoeba castellanii														
consensus[2]				G100	T100	A89	C82	G75	T45 C39			C73	A100	G100
intron (637)				G	T	G	C	G	G			C	A	G
intron (700)				G	T	A	C	G	T			C	A	G
intron (803)				G	T	A	T	G	C			C	A	G
Tetrahymena pyriformis														
consensus[3]	A52 T33	A52 T33	G56	G100	T100	A93	A74			A67 T32	A64	T67	A100	G100
intron (732)	G	G	G	G	T	A	A			A	A	T	A	G
intron (766/767)	A	G	A	G	T	A	A			A	A	T	A	G
intron (801)	A	C	G	G	T	A	A			A	T	T	A	G
***Cryptomonas* Φ**														
consensus[4]		A/C	G	G	T						G/C	C	A	G
intron (672)		C	G	G	T	G	C			C	C	C	A	G
intron (704/705)		A	G	G	T	C	C			G	G	C	A	G
intron (737/738)		A	G	G	T	G	A			T	C	C	A	G
intron (761/762)		A	G	G	T	T	G			T	G	C	A	G
intron (771/772)		A	G	G	T	T	G			G	G	C	A	G
intron (800)		C	G	G	T	C	T			T	G	C	A	G

[1]Numbers to the right of each nucleotide indicate frequency of occurrence (%) at that position. Universal GT and AG dinucleotides at the beginning and end of each intron are underlined.
[2]Based on 44 *A.castellanii* nuclear introns (40).
[3]In *Tetrahymena thermophila* (21).
[4]Based on the six sequences shown.

and the transcriptional machinery itself (24) may have more features in common than the diversity of mitochondrial transcriptional patterns might suggest.

Phylogenetic trees (not shown) constructed using the aligned amino acid sequences in Figure 2 did not display robust branching patterns, a consequence of the limited information content in the sequence data currently available. Not surprisingly, the wheat and rice protein sequences are highly similar (142 identical residues over 153 positions), and associate strongly in the trees. The same is true of the *Neurospora* and yeast protein sequences, although branch lengths are much longer in this case, reflecting a greater degree of sequence divergence. Other than these two affiliations, the only other notable feature of this analysis is that, as a group, the putative mtRNAP sequences are more similar to one another than to the phage sequences, consistent with the idea that the mtRNAPs diverged from a more recent common ancestor than they shared with the phage RNAPs. The fact that none of the amplified sequences branches with the yeast-*Neurospora* clade makes fungal contamination of the DNA preparations used for PCR analysis an unlikely possibility.

Because PCR products were generated using total cellular DNA or partially purified nuclear DNA fractions as target, the genomic localization of each of the amplified sequences remains to be definitively established. In the case of *A.castellanii* and *T.pyriformis*, no T3/T7-like RNAP sequences are present in the completely sequenced mitochondrial genomes of these organisms (25; Burger,G., Zhu,Y., Littlejohn,T., Greenwood,S.J., Schnare,M.N. and Gray,M.W., in preparation), whereas the presence of small, splicesosomal-like intron sequences in the respective PCR products supports a nuclear location for the mtRNAP homologs. Moreover, in translating the *Tetrahymena* exon sequences for the alignment, it was necessary to use the modified genetic code (UAA and UAG decoded as Gln) that is known to be employed for nucleus-encoded mRNAs in this organism (26).

Introns in the *Cryptomonas* Φ PCR product also suggest that these RNAP sequences are localized in the nuclear genome. However, a complicating factor in this case is the additional presence of a nucleomorph genome that is the evolutionary remnant of the nuclear genome of an endosymbiotic alga (27–29). In Southern hybridization experiments with total DNA resolved by pulse-field gel electrophoresis, the *Cryptomonas* PCR product specifically hybridized with nuclear DNA (data not shown). In the case of wheat, the mtRNAP-homologous sequence could not be amplified from purified mtDNA; moreover, no sequences homologous to T3/T7 RNAPs have been found in several completely sequenced chloroplast DNAs, including those from rice (a monocotyledon closely related to wheat) (30) and tobacco (a dicotyledon) (31). These observations indicate that the wheat and rice homologs are also nuclear genes.

Fungal and plant mitochondria commonly contain linear plasmids encoding single-subunit RNAPs (32). These plasmid-encoded RNAP sequences form a distinct group that is only distantly related to the clade of phage T3 and T7 RNAPs and to the nucleus-encoded yeast mtRNAP (32), as well as to the PCR-derived RNAP sequences listed in Figure 2. In pairwise comparisons, the non-phage RNAP sequences listed in Figure 2 are clearly more closely related to one another than to either their phage- or plasmid-encoded RNAP homologs (data not shown). Moreover, as a consequence of sequence divergence at the binding sites against which PCR primers were constructed (Fig. 1), we would not expect that the equivalent region in mitochondrial plasmid-encoded RNAP genes would be amplified under the conditions employed in the current study (particularly with the R-8.1/R-3.2 primer combination used to recover all but the *N.fowleri* sequence; Table 1). For these reasons, we are confident that that sequences shown in Figure 2 do not originate from linear plasmids of the type characterized in fungal and plant mitochondria, the evolutionary origin of which is also obscure.

In photosynthetic organisms, there is evidence for the existence of a second chloroplast RNAP activity, not encoded in chloroplast DNA (33–35). In spinach chloroplasts, RNAP activity has been associated with a 110 kDa polypeptide that has some phage RNAP-like properties (36). Thus, it is possible that there are separate nuclear genes encoding distinct mitochondrial and chloroplast enzymes, each homologous to T3/T7 RNAPs, and that in certain cases, the gene for the chloroplast enzyme might be amplified preferentially over the gene for the mitochondrial enzyme under our conditions. In this regard, it may be significant that among the amplified sequences shown in Figure 2, the most divergent is the one from the prasinophyte *Pycnococcus provasolii*, a primitive green alga. On the other hand, in no case was more than one T3/T7 RNAP-homologous PCR product obtained from any of the DNA samples analyzed. In addition, a Southern hybridization experiment not only verified that the *Pycnococcus* amplification product comes from *Pycnococcus* DNA, but also indicated that the corresponding gene is single copy (data not shown). If a chloroplast homolog does exist, it is formally possible that both it and its mitochondrial counterpart are encoded by the same nuclear gene, with the protein products being targeted to the respective organelles. Precedent exists for targeting of the protein products of a single nuclear gene to different subcellular compartments (37). Additional experimentation will be required to establish the genomic location of each of the T3/T7-like RNAP genes identified in this study and the subcellular location and function of their encoded protein products.

The observations reported here support the thesis that a T3/T7-like RNAP was recruited to act as a mtRNAP at an early stage in the evolution of the mitochondrion. So far, however, we have few clues as to the evolutionary origin of the gene encoding this enzyme. In addition to negative results with amitochondriate eukaryotes, we were unable to amplify homologous sequences from DNAs of eubacteria or archaebacteria. Moreover, a search of the recently published complete genome sequences from *Haemophilus influenzae* (38) and *Mycoplasma genitalium* (39) failed to reveal any T3/T7-like RNAP sequences. Thus, it is not yet clear whether the gene for a T3/T7-like enzyme was acquired from a eubacteria-like symbiont or was provided by the ancestor of the nucleus-containing eukaryotic host cell. Other questions that warrant further investigation are whether mtRNAP genes of the phage T3/T7 type are present in any mitochondrial genomes not yet characterized, and whether such an enzyme is used as the mtRNAP in all eukaryotes. The results summarized here provide a useful entry to the study of mtRNAP evolution and function in a broad range of eukaryotes, particularly protists.

ACKNOWLEDGEMENTS

We are indebted to R. N. Band (Department of Zoology, Michigan State University), E. Denovan-Wright (Department of Biology, Dalhousie University), P. Gros (Department of Biochemistry, McGill University), B.F. Lang (Département de Biochimie,

Université de Montréal), and K.M. Lonergan and D.F. Spencer (Department of Biochemistry, Dalhousie University) for generous gifts of DNA and/or RNA; to M.A. Ragan (Institute for Marine Biosciences, National Research Council, Halifax) and J. Edqvist (Department of Biochemistry, Dalhousie University) for gifts of *Cryptomonas* Φ and *T. pyriformis* cells, respectively; to B.G. Lane (Department of Biochemistry, University of Toronto) for provision of a wheat cDNA library; and to G.I. McFadden (School of Botany, University of Melbourne) for a blot of *Cryptomonas* Φ chromosomes separated by pulse field gel electrophoresis. This work was supported by grants from MRC Canada (to M.W.G.; MT-4124) and from NSERC Canada (to R.C.), and by an NSERC 1967 Science and Engineering Scholarship (to N.C.). Salary and interactions support from the Canadian Institute for Advanced Research are gratefully acknowledged by M.W.G. and R.C., who are Fellows in the Program in Evolutionary Biology.

REFERENCES

1 Gray,M.W. (1989) *Trends Genet.* **5**, 294–299.
2 Gray,M.W. (1992) *Int. Rev. Cytol.* **141**, 233–357.
3 Gray,M.W. (1993) *Curr. Opin. Genet. Dev.* **3**, 884–890.
4 Bogorad,L. (1991) In Bogorad,L. and Vasil,I.K. (eds), *The Molecular Biology of Plastids.* Academic Press, Inc., San Diego, pp. 93–124.
5 Shadel,G.S. and Clayton,D.A. (1993) *J. Biol. Chem.* **268**, 16083–16086.
6 Greenleaf,A.L., Kelly,J.L. and Lehman,I.R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3391–3394.
7 Kelly,J.L., Greenleaf,A.L. and Lehman,I.R. (1986) *J. Biol. Chem.* **261**, 10348–10351.
8 Masters,B.S., Stohl,L.L. and Clayton,D.A. (1987) *Cell* **51**, 89–99.
9 Chen,B., Kubelik,A.R., Mohr,S. and Breitenberger,C.A. (1996) *J. Biol. Chem*, in press. L25087.
10 Patterson,D.J. and Sogin,M.L. (1992) In Hartman,H. and Matsuno,K. (eds), *The Origin and Evolution of the Cell.* World Scientific, Singapore, pp. 13–46.
11 Lang,B., Burger,G., Doxiadis,I., Thomas,D.Y., Bandlow,W. and Kaudewitz,F. (1977) *Anal. Biochem.* **77**, 110–121.
12 Lonergan,K.M. and Gray,M.W. (1993) *Science* **259**, 812–816.
13 Hu,W.-N., Kopachik,W. and Band,R.N. (1992) *Infect. Immunity* **60**, 2418–2424.
14 Haffner,M.H., Chin,M.B. and Lane,B.G. (1978) *Can. J. Biochem.* **56**, 729–733.
15 Frohman,M.A. (1994) *PCR Methods Applic.* **4**, S40-S58.
16 Sousa,R., Chung,Y.J., Rose,J.P. and Wang,B.-C. (1993) *Nature* **364**, 593–599.
17 Osumi-Davis,P.A., de Aguilera,M.C., Woody,R.W. and Woody,A-Y.M. (1992) *J. Mol. Biol.* **226**, 37–45.
18 Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) *Protein Engng.* **3**, 461–467.
19 Gross,L., Chen,W.-J. and McAllister,W.T. (1992) *J. Mol. Biol.* **228**, 488–505.
20 Kostyuk,D.A., Dragan,S.M., Lyakhov,D.L., Rechinsky,V.O., Tunitskaya,V.L., Chernov,B.K. and Kochetkov,S.N. (1995) *FEBS Lett.* **369**, 165–168.
21 Csank,C., Taylor,F.M. and Martindale,D.W. (1990) *Nucleic Acids Res.* **18**, 5133–5141.
22 Cavalier-Smith,T. (1993) *Microbiol. Rev.* **57**, 953–994.
23 Gray,M.W. (1989) *Annu. Rev. Cell Biol.* **5**, 25–50.
24 Tracy,R.L. and Stern,D.B. (1995) *Curr. Genet.* **28**, 205–216.
25 Burger,G., Plante,I., Lonergan,K.M. and Gray, M.W. (1995) *J. Mol. Biol.* **245**, 522–537.
26 Osawa,S., Jukes,T.H., Watanabe,K. and Muto,A. (1992) *Microbiol. Rev.* **56**, 229–264.
27 Douglas,S.E., Murphy,C.A., Spencer,D.F. and Gray,M.W. (1991) *Nature* **350**, 148–151.
28 McFadden,G.I., Gilson,P.R. and Douglas,S.E. (1994) *J. Cell Sci.* **107**, 649–657.
29 McFadden,G. and Gilson,P. (1995) *Trends Ecol. Evol.* **10**, 12–17.
30 Hiratsuka,J., Shimada,H., Whittier,R., Ishibashi,T., Sakamoto,M., Mori,M., Kondo,C., Honji,Y., Sun,C.-R., Meng,B.-Y., Li,Y.-Q., Kanno,A., Nishizawa,Y., Hirai, A., Shinozaki,K. and Sugiura,M. (1989) *Mol. Gen. Genet.* **217**, 185–194.
31 Shinozaki,K., Ohme,M., Tanaka,M., Wakasugi,T., Hayashida,N., Matsubayashi,T., Zaita,N., Chunwongse,J., Obokata,J., Yamaguchi-Shinozaki,K. *et al.* (1986) *EMBO J.* **5**, 2043–2049.
32 Kempken,F., Hermanns,J. and Osiewacz,H.D. (1992) *J. Mol. Evol.* **35**, 502–513.
33 Morden,C.W., Wolfe,K.H., dePamphilis,C.W. and Palmer,J.D. (1991) *EMBO J.* **10**, 3281–3288
34 Falk,J., Schmidt,A. and Krupinska,K. (1993) *J. Plant Physiol.* **141**, 176–181.
35 Hess,W.R., Prombona,A., Fieder,B., Subramanian,A.R. and Börner,T. (1993) *EMBO J.* **12**, 563–571.
36 Lerbs-Mache,S. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5509–5513.
37 Martin,N.C. and Hopper,A.K. (1994) *Biochimie* **76**, 1161–1167.
38 Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J.-F., Dougherty,B.A., Merrick,J.M. *et al.* (1995) *Science* **269**, 496–512.
39 Fraser,C.M., Gocayne,J.D., White,O., Adams,M.D., Clayton,R.A., Fleischmann,R.D., Bult,C.J. Kerlavage,A.R., Sutton,G., Kelley,J.M. *et al.* (1995) *Science* **270**, 397–403.
40 Wong,J.M., Liu,F. and Bateman,E. (1992) *Nucleic Acids Res.* **20**, 4817–4824.
41 Corpet,F. (1988) *Nucleic Acids Res.* **16**, 10881–10890.