J Mol Evol (1997) 45:671–681

JOURNAL OF MOLECULAR EVOLUTION

On the Evolution of the Single-Subunit RNA Polymerases

Nicolas Cermakian,[1] Tatsuya M. Ikeda,[2] Pedro Miramontes,[3] B. Franz Lang,[1] Michael W. Gray,[2] Robert Cedergren[1]

[1] Département de biochimie, Université de Montréal, C.P. 6128, succursale Centre-Ville, Montréal, QC H3C 3J7, Canada
[2] Department of Biochemistry, Dalhousie University, Sir Charles Tupper Building, Halifax, Nova Scotia B3H 4H7, Canada
[3] Departamento de Matemáticas, Facultad de Ciencias, Universidad Nacional Autónoma de México, Mexico

Received: 28 April 1997 / Accepted: 11 August 1997

Abstract. Many eukaryotic nuclear genomes as well as mitochondrial plasmids contain genes displaying evident sequence similarity to those encoding the single-subunit RNA polymerase (ssRNAP) of bacteriophage T7 and its relatives. We have collected and aligned these ssRNAP sequences and have constructed unrooted phylogenetic trees that demonstrate the separation of ssRNAPs into three well-defined and nonoverlapping clusters (phage-encoded, nucleus-encoded, and plasmid-encoded). Our analyses indicate that these three subfamiles of T7-like RNAPs shared a common ancestor; however, the order in which the groups diverged cannot be inferred from available data. On the basis of structural similarities and mutational data, we suggest that the ancestral ssRNAP gene may have arisen via duplication and divergence of a DNA polymerase or reverse transcriptase gene. Considering the current phylogenetic distribution of ssRNAP sequences, we further suggest that the origin of the ancestral ssRNAP gene closely paralleled in time the introduction of mitochondria into eukaryotic cells through a eubacterial endosymbiosis.

Key words: RNA polymerase — Evolution — Mitochondria — Bacteriophage — Linear plasmid — Phylogenetic tree

Introduction

Considering the pivotal role played by the synthesis of RNA (transcription) in the expression of genetic information, knowledge of the origin and evolution of this process could provide crucial insights into the origin of cellular life itself. Transcription is generally carried out by complex, multicomponent RNA polymerases (RNAPs) consisting of two highly conserved subunits and a number of smaller polypeptides (Palenik 1992). However, in bacteriophages related to the coliphage T7, most genes are transcribed by a single-subunit RNA polymerase (ssRNAP; McAllister 1993), encoded by the phage genome. This enzyme not only initiates transcription but also functions in elongation and termination during RNA synthesis. The origin of the ssRNAP is an evolutionary enigma, as is its relationship to the multisubunit RNAPs, with which it shares no obvious sequence similarity. In fact, ssRNAPs possess short sequence motifs (Delarue et al. 1990) and the ''hand-like'' three-dimensional structure (Sousa et al. 1993) of the Klenow fragment of *Escherichia coli* DNA polymerase (DNAP) I (Sousa 1996).

The mystery surrounding the origin of the ssRNAP gene deepened with the discovery that the mitochondrial RNAP (mtRNAP) from *Saccharomyces cerevisiae* displays evident sequence similarity to the phage T7 RNAP (Masters et al. 1987). To test the premise that other mitochondria-containing eukaryotes might harbor ssRNAP-like genes, we designed PCR primers to scan eukaryotic genomes for T7-like ssRNAP genes. We were able to

Abbreviations: RNAP, RNA polymerase; ssRNAP, single-subunit RNA polymerase; mtRNAP, mitochondrial RNA polymerase; DNAP, DNA polymerase; RT, reverse transcriptase

Correspondence to: R. Cedergren; e-mail ceder@medcn.umontreal.ca

show that such sequences are widespread within the eukaryotic lineage (Cermakian et al. 1996). Moreover, the mitochondria of certain plants and fungi contain linear plasmids that bear RNAP genes strongly resembling the one encoding the T7 ssRNAP (Kempken et al. 1992; Griffiths 1995). These plasmid-like DNAs are characterized by double-stranded structure, terminal inverted repeats, and open reading frames specifying both a DNAP and a ssRNAP (Meinhardt et al. 1990; Griffiths 1995). The two polymerases are encoded either on a single plasmid, as in various fungi (Kempken et al. 1992), or on distinct plasmids, as in the case of the S-1 and S-2 plasmids of maize (Kuzmin et al. 1988) and, presumably, the pLm9 and pLm10 plasmids of *Leptosphaeria maculans* (Lim and Howlett 1994). It has been proposed that mitochondrial linear plasmids are evolutionarily related to bacteriophages (Meinhardt et al. 1990; Kempken et al. 1992). Both their protein-terminated double-stranded DNA structure (Meinhardt et al. 1990) and B-type DNAPs (Braithwaite and Ito 1993) are features shared with several bacteriophages, including Φ29 and PRD1 (Salas 1988; Blanco and Salas 1996). Although phages Φ29 and PRD1 diverge greatly from each other (Ackermann and DuBow 1987), phylogenies based on DNAP sequences show a close relationship between them and linear plasmid DNAPs (Kempken et al. 1992; Rohe et al. 1992). In both of these phages, the terminal protein gene is located just upstream of the DNAP gene (Yoshikawa and Ito 1982; Grahn et al. 1994), whereas in the case of linear plasmids, it has been suggested that this role is assumed by the amino-terminal domain of the plasmid-encoded DNAP, consistent with the view that linear plasmids may be remnants of an ancestral bacteriophage.

The existence of two RNAP families (single- and multisubunit) raises fundamental questions about the origin of transcription. In particular, we would like to know whether the two families share a common history, and which (if either) was responsible for transcription at the early stages in the evolution of this process. A further issue relevant to the ssRNAPs is whether the apparent similarity of these gene sequences from diverse genetic sources (eukaryotic nuclei, mitochondrial plasmids, and bacteriophage genomes) belies a true evolutionary relationship. Here we address these questions through phylogenetic analysis of trees constructed from an alignment of the ssRNAP sequences and discuss these and other data on ssRNAPs with a view to considering how this critical protein family may have originated and subsequently evolved and how the known phylogenetic distribution of ssRNAP genes can be rationalized.

Methods

Rhizophydium sp. #136, an unnamed representative of the Chytridiales (chytridiomycete fungi), was isolated from garden soil by Dr. J. Longcore, University of Maine. Sporangia were grown on PmTG (Barr 1986) nutrient agar. Cells were broken with glass beads (Lang et al. 1977) and DNA was isolated as described in Cermakian et al. (1996). PCR experiments and subsequent cloning and sequencing procedures were also performed as described in Cermakian et al. (1996), using the PCR primers R-8.1 and R-3.2.

The single-letter code for amino acids is used throughout this discussion and residues are numbered with reference to their position in the bacteriophage T7 RNAP sequence. Alignments of the ssRNAP sequences were performed using the Multalin program (Corpet 1988). Because the ssRNAP database consists of a diverse collection of sequences, we first performed separate alignments of the nucleus-encoded, plasmid-encoded and phage-encoded enzymes before all sequences were assembled into a globally aligned collection. Partial sequences were added manually to the collection based on the alignment of Fig. 1 and that of Cermakian et al. (1996). Phylogenetic trees were constructed from this alignment using parsimony (PAUP; Swofford 1993) and distance methods (SEQBOOT, PROTDIST, FITCH, NEIGHBOR, and CONSENSE of the PHYLIP package; Felsenstein 1993). For all trees, 100 bootstrap resamplings of the data were performed. All alignments and trees are available on request from the authors.

A Monte Carlo analysis of the sequence data was made to verify the null hypothesis of randomness in the distance scores (Manly 1991). One hundred sets of 16 random sequences of the same amino acid content and length at the *S. cerevisiae* ssRNAP sequence shown in Fig. 1 were generated. These data sets were then used to generate 100 distance matrices using PROTDIST of PHYLIP package (with ''categories distance model'' included in this program). The statistics chosen to test the null hypothesis were d_{max}, the maximum of distances between elements in each set of 16 sequences, and the radius (r) of the distance matrix (d_{max}–d_{min}). The first parameter measures the worst possible case in a distance matrix and the second the degree of inhomogeneity among the sequences in a single distance matrix.

Results and Discussion

Alignments and Phylogenies of ssRNAP Sequences

The sequences used in this study (Table 1) include those employed previously (Kempken et al. 1992) as well as the nucleus-encoded mitochondrial ssRNAP sequences from *Neurospora crassa* (Chen et al. 1996), *Chenopodium album* (Weihe et al. 1997), and *Arabidopsis thaliana* (Acc. No. Y09432), the linear plasmid sequences from *Gelasinospora* sp. (Griffiths 1995) and *Claviceps purpurea* strain T5 (Oeser et al. 1993), and the bacteriophage sequences from *Klebsiella* phage K11 (Dietz et al. 1990) and *Salmonella* phage SP6 (Kotani et al. 1987). To these we have added all partial sequences previously published by our group (Cermakian et al. 1996) as well as the new Rhizophydium sequence reported here.

The most conserved blocks in the alignment (Fig. 1) are similar to the regions delineated by Masters et al. (1987) and Chan et al. (1991), with the exception that we have eliminated the N-terminal region from the alignment because it is not well conserved in our more extensive data set. We also use an additional block (F) from the carboxy-terminal region of the protein sequence.

One of the most striking aspects of the alignment is the considerable length variation in the N-terminal re-

gion of the sequence, which suggests that this region could be involved in species-specific processes (Ikeda and Richardson 1987; Ikeda et al. 1993). This length variation accounts for most of the overall length variation of ssRNAP sequences (Fig. 1). On the other hand, the higher degree of conservation in the C-terminal region together with experimental data supports the idea that this portion of the protein is involved in general RNAP functions such as template binding and nucleotide polymerization (Sousa 1996). Another obvious feature of this alignment is the existence of three distinct classes of ssRNAPs. This might have been anticipated from the fact that we had aligned each group separately prior to generating the global alignment of Fig. 1; however, intragroup similarities are markedly stronger than would be expected from such an alignment bias. For example, the penultimate amino acid (882 in T7 RNAP) is F in all phage and nuclear sequences, but not in the plasmid-borne polymerases. Only bacteriophage sequences have L at position 637, and only the nuclear sequences have P at position 738.

Phylogenetic trees were derived from the alignment of Fig. 1 and the different phylogenetic treeing techniques gave almost identical results. One such phylogeny is shown in Fig. 2. As indicated by low bootstrap values, the branching order within the plasmid-borne RNAP gene representatives is poorly defined; this may reflect looser structural and functional constraints on this class. This observation is underscored as well by the presence of several unusual amino acids at a number of positions that are otherwise highly conserved among ssRNAP sequences. The overall topology is robust with the three treeing methods: the phage-encoded, nucleus-encoded, and plasmid-encoded sequences form three well-separated phylogenetic clusters, confirming what had been observed in the course of tree building and in analyzing the alignment. Even in the case where there is a nucleus-encoded and a plasmid-encoded sequence, i.e., *N. crassa*, the two sequences branch separately, each being akin to sequences encoded by the same type of genome.

To examine the effect of adding a number of branches to the nuclear clade, phylogenetic trees were then generated from an alignment that included the eight partial RNAP sequences obtained by PCR amplification (Cermakian et al. 1996) as well as a new sequence from the chytridiomycete fungus *Rhizophydium* sp. #136 (Fig. 3). The use of PCR to generate these gene sequences limits this particular analysis to the amplified region. One of the trees derived from this abbreviated alignment is presented in Fig. 4. Although the SP6 sequence is not grouped with those from the other phages, the monophyly of phage sequences is evident when a longer sequence alignment is used in the phylogenetic analysis (see Fig. 2). The plasmid-encoded sequences are monophyletic, as are the nucleus-encoded ones; for these branches, the bootstrap values, although satisfactory, are not as high as in Fig. 2, most likely because the sequences are shorter. Finally, this topology indicates that the plasmid-encoded and nucleus-encoded ssRNAPs (both of which are found in eukaryotic cells) do not share a recent common ancestor, although rapid sequence divergence among the plasmid-borne genes makes it impossible to rule out other, less parsimonious conclusions.

Relationships Among ssRNAPs

The clear separation of ssRNAP sequences into three clusters raises the question of whether they actually form an homologous family, or instead represent three distinct origins with subsequent convergent evolution. Close examination of the sequences and the alignment in Fig. 1 highlights several important aspects of the ssRNAP sequence database. In concert with tertiary structure considerations, these sequence features strongly support the conclusion that the three ssRNAP classes are derived from a common ancestral sequence.

The prototype of this family, the bacteriophage T7 RNAP, has been extensively studied to identify amino acid residues important for catalysis. Comparison of the three-dimensional structure of the T7 ssRNAP (Sousa et al. 1993) with those of other RNAPs (Sousa 1996) and examination of polymerase alignments (Delarue et al. 1990) have directed structure/function studies to several highly conserved amino acids. Among these, D537 and D812 are involved in the binding of two divalent metal ions at the active site and are, therefore, essential for catalysis (Osumi-Davis et al. 1992; Woody et al. 1996). Residues K631 and H811 are implicated in phosphodiester bond formation, and the latter residue appears to be involved in nucleotide binding as well (Osumi-Davis et al. 1992). In light of these data, it is significant that these signature residues are found in all sequences compiled in Fig. 1. Only in the RNAP encoded by the kalilo plasmid of *N. intermedia* has D812 been substituted (by E). Although this is a conservative change, the fact that RNAP activity is decreased 3,400-fold in a D812E mutant of T7 RNAP (Bonner et al. 1992) suggests either that the kalilo-encoded enzyme is not expressed or that it has low RNAP activity. In this case, expression of the ssRNAP gene has been demonstrated (Vickery and Griffiths 1993).

Another mutation, Y639P, is located at a highly conserved site in these RNAPs (Delarue et al. 1990) and disturbs the dNTP/rNTP discrimination of the enzyme (Sousa and Padilla 1995), whereas a S641A mutation directs T7 RNAP to use dNTPs instead of rNTPs (Kostyuk et al. 1995). The double mutant Y639F/S641A can use both sets of nucleoside triphosphates (Kostyuk et al. 1995). Residue Y639 is present in all ssRNAPs, whereas S641 occurs exclusively in the phage enzymes; however, another hydroxy amino acid (T) is found at the adjacent position in all of the available mtRNAP sequences (Fig. 1; Cermakian et al. 1996).

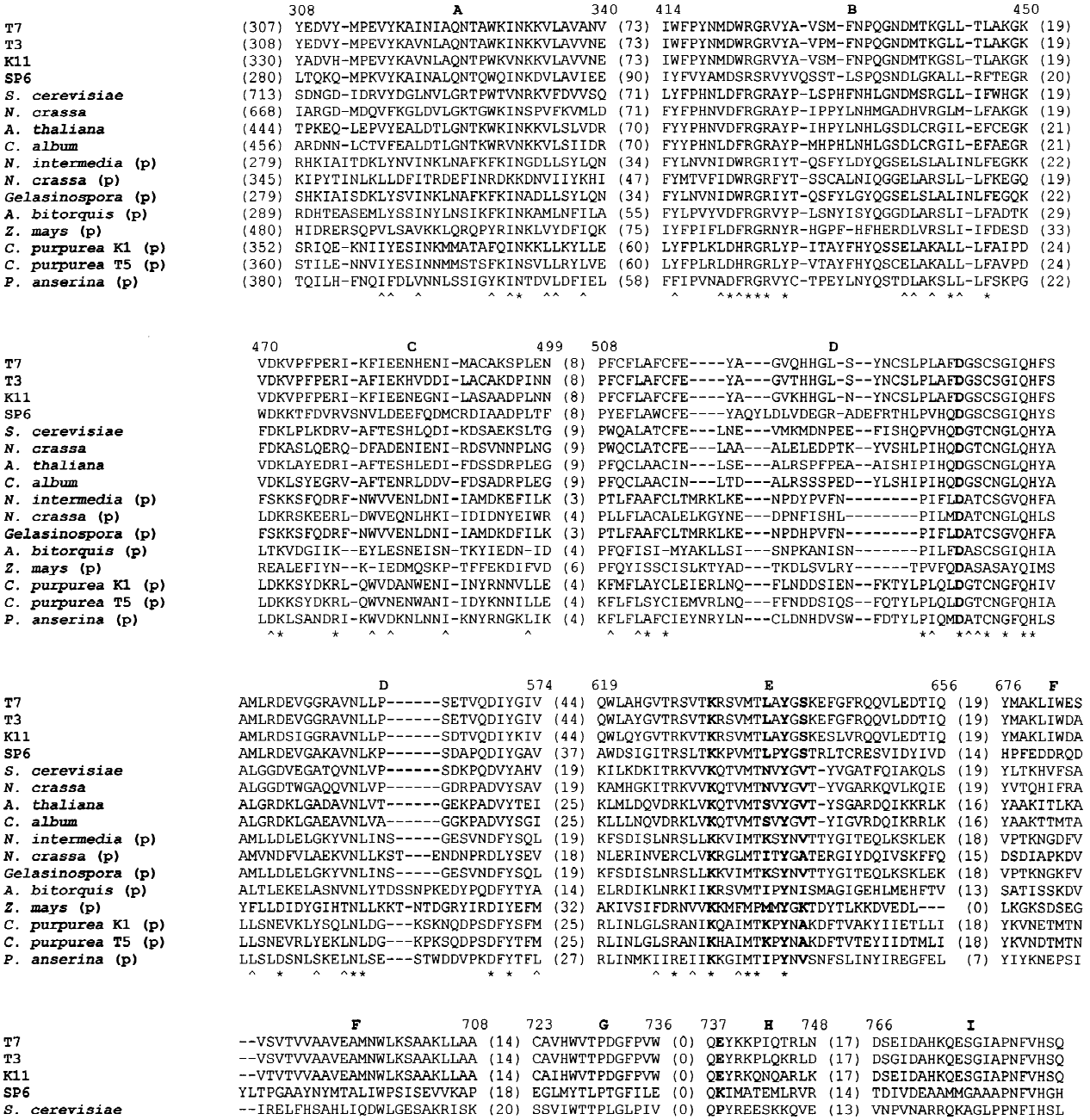

Fig. 1. Alignment of complete ssRNAP amino acid sequences. Blocks of conserved sequence are delineated, with the *number of amino acids* in the more variable regions between each block given in *parentheses*. Sequences from linear plasmids are indicated by a (*p*) beside the name of the organism. The number of amino acids upstream of block A in sequences from *A. bitorquis*, *Z. mays* and *C. purpurea* plasmids is tentative. In all cases, *block L* represents the 17 last amino acids of the protein, except in *Z. mays* RNAP, where there is an additional residue at the end (*Y*). *Numbers at the top* of each block refer to positions in the phage T7 RNAP sequence. Residues discussed in the text are set in *bold*. Gaps are denoted by *dashes* (-). Residues identical and similar in more than 80% of the sequences are indicated under the alignment by a * and a ^, respectively. The total number of amino acids in each sequence is given after block L.

```
                           I  794      804    J       818    819       K          837     867       L       883   Total
T7                   DGSHLRKT  (9)  IESFALIHDSFGTIP  (0)  ADAANLFKAVRETMVDTYE  (29)  KGNLNLRDILESDFAFA    883
T3                   DGSHLRMT  (9)  IESFALIHDSFGTIP  (0)  ADAGKLFKAVRETMVITYE  (29)  KGNLNLQDILKSDFAFA    884
K11                  DGSHLRMT  (9)  IDSFALIHDSSGTIP  (0)  ADAGNLFKAVRETMVKTYE  (29)  KGDLNLRDILESDFAFA    906
SP6                  DASHLILT  (8)  VTSIAVIHDSFGTHA  (0)  DNTLTLRVALKGQMVAMYI  (26)  QGEFDLNEIMDSEYVFA    874
S. cerevisiae        DASHMLLS  (7)  GLDFASVHDSYWTHA  (0)  SDIDTMNVVLREQFIKLHE (120)  KGDFDVTVLRNSQYFFS   1351
N. crassa            DASHMILS  (7)  GLTFAAVHDSFWTHA  (0)  SDIDSMNAVLRDAFIRIHS (201)  KGDFDVRSLKDSTYFFS   1422
A. thaliana          DGSHMMMT  (7)  GLSFAGVHDSFWTHA  (0)  CDVDVMNTILREKFVELYE  (25)  RGDFDLRKVLESTYFFN    976
C. album             DGSHMMMT  (7)  GMNFAGVHDSYWTHA  (0)  CDVDKMNQILREKFVELYE  (25)  RGDFDLREVLESPYFFN    988
N. intermedia (p)    DASHLTMI  (4)  DSYILPIHECFGTHP  (0)  NDMYKLAEQVRECFILLYS  (51)  GELNVEDIRDMGKYMIS    811
N. crassa (p)        DGSNIVLL  (9)  KINFASIHDCFATHA  (0)  NDTAWLSWYVKQSFIRIYS  (56)  KDNKIYKEILHSEYFIN    896
Gelasinospora (p)    DAAHLPML  (4)  DSYILPIHDCFGTHP  (0)  NDMFKLAEQVRECFILLYS  (51)  GELNVEDIRDMGKYMIS    811
A. bitorquis (p)     DA-SNVHL  (8)  NLPVYTVHDCFASTA  (4)  KLEK-LV---KNAFINIYF  (80)  NINEFVKGILNSKYFIG    874
Z. mays (p)          DAFTAIQL (11)  SIPIYAVHDNFITMP (15)  RMGHPLIIINKFLFDHILI (125)  GTQADSLDKGEDDYCIH   1159
C. purpurea K1 (p)   DAASLIML (11)  VVNFYSVHDCYGVTA  (0)  KYIDLLISQLRAVYIELYS  (55)  HIRKAYEELAKANMFIK    936
C. purpurea T5 (p)   DAASLILL (11)  PVNFYGVHDCYGVTA  (0)  KFIDLLISHLRAVYIELYS  (56)  LVHKAYEELSKANMFIK    945
P. anserina (p)      DAASLTLL (10)  VKNIYTIHDCFAVPA  (0)  NKMECLISLLKLTYIKLYS  (53)  PSDFDFNVLKKSSYILN    948
                     *^  ^ ^                ^** ^ ^               ^  ^^  ^ ^*                  ^    ^
```

Fig. 1. Continued.

Table 1. RNA polymerase sequences used in this study

Organism name	Gene name	Gene source	Reference	Acc. No.
Bacteriophage T7	Gene 1	Phage genome	Moffat et al. 1984	M38308
Bacteriophage T3	Gene 1	Phage genome	McGraw et al. 1985	X02981
Bacteriophage K11	Gene 1	Phage genome	Dietz et al. 1990	X53238
Bacteriophage SP6	Gene 1	Phage genome	Kotani et al. 1987	Y00105
Saccharomyces cerevisiae	RPO41	Nuclear genome	Masters et al. 1987	M17539
Neurospora crassa	*cyt-5*	Nuclear genome	Chen et al. 1996	L25087
Chenopodium album	—	Nuclear genome	Weihe et al. 1997	Y08067
Arabidopsis thaliana	—	Nuclear genome	Unpublished	Y09432
Triticum aestivum	—	Nuclear genome	Cermakian et al. 1996	U34402
Oryza sativa	—	Nuclear genome	Cermakian et al. 1996	U34283
Pycnococcus provasolii	—	Nuclear genome	Cermakian et al. 1996	U34286
Acanthamoeba castellanii	—	Nuclear genome	Cermakian et al. 1996	U34405
Isochrysis sp. Tahiti	—	Nuclear genome	Cermakian et al. 1996	U34284
Thraustochytrium aureum	—	Nuclear genome	Cermakian et al. 1996	U34285
Cryptomonas phi	—	Nuclear genome	Cermakian et al. 1996	U34404
Tetrahymena pyriformis	—	Nuclear genome	Cermakian et al. 1996	U34406
Naegleria fowleri	—	Nuclear genome	Cermakian et al. 1996	U34403
Rhizophydium sp.	—	Nuclear genome	this study	AF000226
Neurospora intermedia	ORF-1	Plasmid kalilo	Chan et al. 1991	X52106
Neurospora crassa	ORF-1	Plasmid maranhar	Court and Bertrand 1992	X55361
Gelasinospora sp.	ORF-1	Plasmid kalilo	Griffiths 1995	L40494
Agaricus bitorquis	ORF2	Plasmid pEM	Robison et al. 1991	X63075
			Levings and Sederoff 1983	
Zea mays	ORF1	Plasmid S-2	Kuzmin et al. 1988	J01426
Claviceps purpurea	ORF2	Plasmid pC1K1	Oeser and Tudzynski 1989	X15648
Claviceps purpurea	ORF2	Plasmid pC1T5	Oeser et al. 1993	X68490
Podospora anserina	ORF2	Plasmid pAL2-1	Hermanns and Osiewacz 1992	X60707

The penultimate amino acid, F882, is important for rNTP binding (Patra et al. 1992; Sousa et al. 1993) and is present in phage and nucleus-encoded polymerases (Fig. 1); however, it is replaced by another large hydrophobic amino acid (I or L) in all plasmid sequences. Other positions that are characteristic of plasmid-encoded genes again raise the possibility that these gene products might have little or no activity, although compensatory changes in other regions of the molecule could be operating in these cases. Moreover, at this stage, editing of mRNA to generate a required amino acid residue cannot be ruled out.

If it was the case that only those amino acids important for activity are conserved among the ssRNAP sequences, we might question the homology or monophyly of these sequences, suggesting as an alternative convergent evolution to a common structure and functionality. However, as can be seen in Fig. 1, sequence similarity extends well beyond these few catalytic motifs, and relatively large blocks of conserved sequence can be delineated in the alignment. Moreover, these blocks appear in all the sequences in the same linear order.

In order to support the homology of the sequences in our alignment, we performed a Monte Carlo analysis on these data. We calculated the basic statistics of the Monte Carlo randomization for each of the two parameters described in Methods. Our results reject the null hypothesis of randomness. The d_{max} value of the distance matrix for ssRNAP sequences (7.7) is 10 standard deviations (S_D) from the mean determined on the random sequences

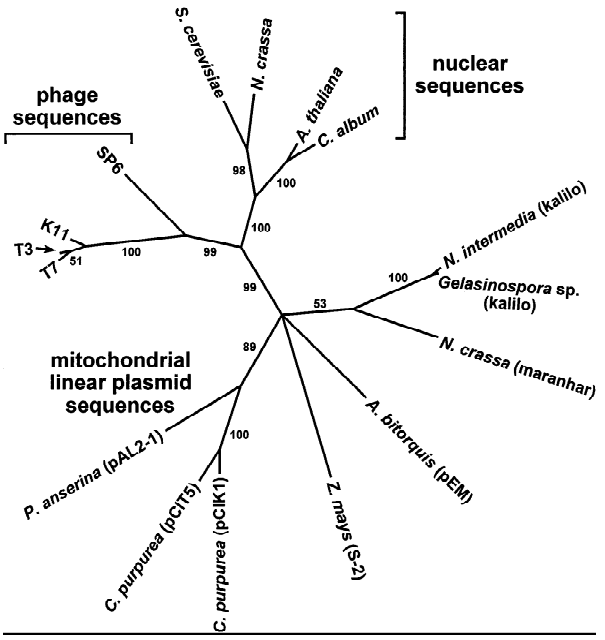

Fig. 2. Maximum parsimony phylogenetic tree derived from the alignment of Fig. 1. The bootstrap *values* (100 resamplings) are indicated on the branches. Branches with bootstrap values lower than 50 were collapsed. The three groups of ssRNAP sequences discussed in the text are indicated.

(108.8, with a S_D of 10.2) and none of the individual maxima is smaller than the d_{max} of real sequences. This d_{max} is even significantly smaller than the mean of d_{min} in random sequence matrices (14.9, with a S_D of 1.0). On the other hand, the radius of the matrix of ssRNAP sequences (7.6) is less than that of any of the randomized sequences matrices and significantly smaller than their mean (93.8, with a S_D of 10.3). These significant results show that similarity in the ssRNAP alignment is well above the noise level.

Even though these results support common ancestry of the ssRNAP sequences, the possibility remains, however, that they originated through a type of modular evolution (Li 1997). This possibility could be discounted if parts of the alignment of Fig. 1 showed topological congruency upon phylogenetic analysis. We thus divided the sequence alignment into four parts of approximately equal length (blocks A + B, C + D, E + F, G + H + I + J + K + L) and constructed phylogenetic trees from the four subsets, using parsimony and distance methods. All subsets yielded the same overall topology as in Figs. 2 and 4, with three well-defined clusters. The only deviation from congruency is the SP6 sequence, which did not group with the T7/T3/K11 clade in the C + D and E + F phylogenies (with distance method, but not parsimony), similar to its behavior in Fig. 4.

Thus, there is little doubt that these three classes of ssRNAP gene originated from a common ancestral sequence. However, because the phylogenetic trees are unrooted, the order of this divergence cannot be inferred from the available data. On another hand, even if blocks in our alignment share common ancestry, modular evolution still can be invoked and is even likely in some cases. For instance, fungal nuclear sequences (from *S. cerevisiae* and *N. crassa*) are much longer than the others, mainly due to a N-terminal extension and an insertion between blocks K and L. In fact, the three-dimensional structure of the T7 RNAP suggests that these two regions are outside of the polymerase core (Sousa et al. 1993; Sousa 1996).

Origin of the Ancestral ssRNAP Gene

Major questions in the evolution of the ssRNAPs are: What type of cell or genetic system could have given rise to their genes and how might this origin explain their present-day distribution pattern? Because the phylogenetic and structural data support a unique origin for ssRNAPs, it is useful to consider possible evolutionary scenarios.

In addition to the several sequence motifs discussed above (Delarue et al. 1990), the T7 RNAP possesses an intriguingly high structural similarity to the Klenow fragment of DNAP I (Sousa 1996; Sousa et al. 1993). Many secondary structure elements are in the same order in the sequence and can readily be aligned at the level of tertiary structure. Sequence motifs found in ssRNAPs and DNAPs do not exist in multisubunit RNAPs (Delarue et al. 1990). Moreover, as previously noted, single amino-acid changes are able to lower the ability of the ssRNAP of T7 phage to discriminate between rNTPs and dNTPs (Sousa and Padilla 1995) and can even change the substrate specificity from rNTPs to dNTPs (Kostyuk et al. 1995). More recently, Gao et al. (1997) have shown that mutation of a single residue in a reverse transcriptase (RT) results in a variant enzyme now capable of acting as an RNAP. These observations strongly suggest that ssRNAPs and DNAP I–like enzymes are homologous, i.e., that they arose by divergent evolution, in spite of the fact that homology is not clearly evident at the level of primary sequence similarity, which is limited to the few instances of sequence conservation within the catalytic pocket.

What can be said about the timing of the origin of the ancestral ssRNAP gene? To date, no member of this family has been found in any of the available complete prokaryotic genomic sequences (four eubacterial and one archaeal species; Fleischmann et al. 1995; Fraser et al. 1995; Bult et al. 1996; Kaneko et al. 1996; Blattner et al. 1997), nor has such a sequence emerged from the extensive database of partial eubacterial and archaeal genome sequences. Moreover, it is not clear what purpose a ssRNAP would have served even in the last common ancestor (cenancestor) of the three domains of life, considering that the transcription system in this entity almost certainly utilized a multicomponent RNAP (Klenk and Doolittle 1994). Among many prokaryotic viruses stud-

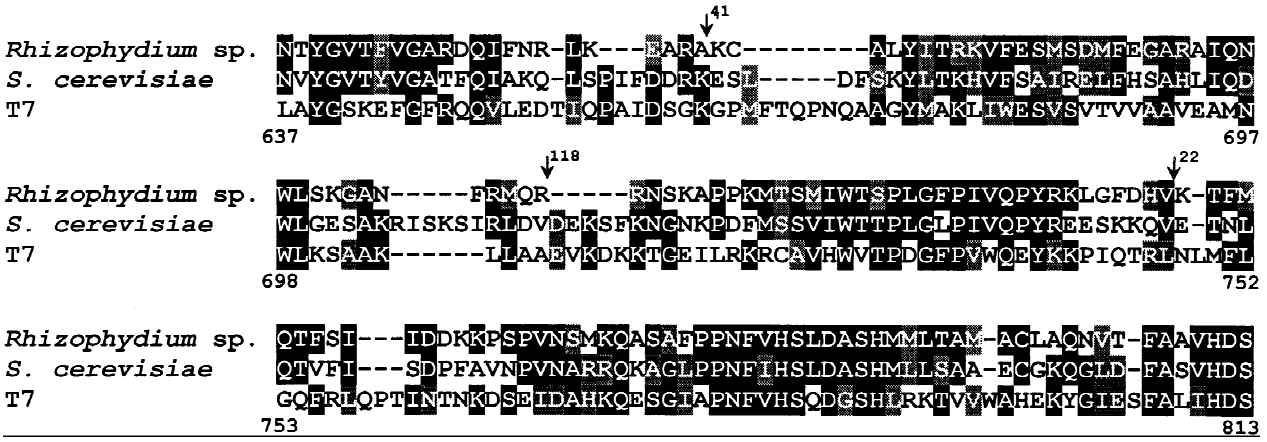

Fig. 3. Partial ssRNAP sequence obtained from the chyridiomycete fungus, *Rhizophydium* sp. #136. The protein sequence was deduced from the sequence of a PCR-amplified DNA product obtained using *Rhizophydium* DNA, as described in the text. The sequence is aligned with that of the corresponding region in *Saccharomyces cerevisiae* mtRNAP and bacteriophage T7 RNAP, and identical and similar amino acids are highlighted *white on black and grey*, respectively. *Numbers under the alignment* refer to positions in the phage T7 RNAP sequence. Three putative small introns were inferred, based on conservation of reading frame, optimization of sequence alignment, and identification of GU-AG intron boundaries. The positions of these introns are denoted by *arrows*, with the accompanying numbers indicating intron size in bp. The *Rhizophydium* sequence has been deposited in the GenBank database of NCBI (Acc. No. AF000226).

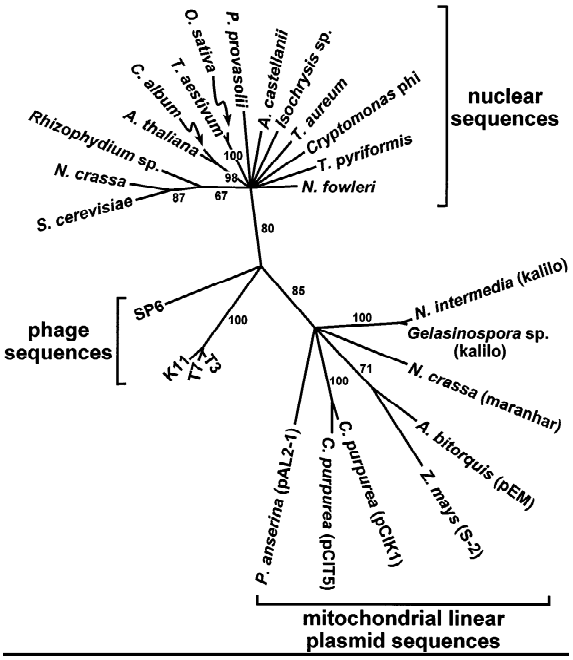

Fig. 4. Maximum parsimony phylogenetic tree generated as in Fig. 2, but using an alignment of partial protein sequences, corresponding to the region shown in Fig. 3. Branches with bootstrap values lower than 50 were collapsed. The *N. fowleri* sequence used in the alignment is shorter than the other sequences (31 fewer amino acids; see Cermakian et al. 1996).

ied to date, only four (phages T3, T7, K11, and SP6) are known to encode a ssRNAP; moreover, the host organisms for these phages (*E. coli, Klebsiella, Salmonella*) are all relatively closely related members of the same subgroup of γ-Proteobacteria. Otherwise, ssRNAP sequences are limited in occurrence to the eukaryotic nucleus (where they encode the mtRNAP) and to plasmid-like DNAs in the mitochondria of certain plants and fungi.

Taken at face value, the available data do not support the view that the ssRNAP gene originated early in evolutionary history, although the data do not rigorously exclude this view, either. A relatively late origin is more consistent with the known, phylogenetically limited distribution of ssRNAP sequences; conversely, an early origin better explains the extent of sequence divergence among the three classes of ssRNAP gene and between these and DNAP genes. An early origin scenario (prior to the separation of the three domains of life, or even early in eubacterial evolution) would demand numerous independent losses of an ssRNAP gene from prokaryotic genomes (assuming vertical descent of the nucleus-encoded gene from a homolog in a cenancestral genome) and/or the sequestration of this gene in a quasi-independent infectious entity, such as a bacterial virus, which as far as we know displays only a very narrow host range. At present, there is little we can say with confidence about the timing of the appearance of the ancestral ssRNAP gene and whether this origin occurred in a cellular or virus-like genome.

Possible Scenarios for the Origin of the Mitochondrial RNA Polymerase

Given the expectation that the proto-mitochondrial, endosymbiont genome should have encoded a multicomponent, $\alpha_2\beta\beta'$-type enzyme, as the chloroplast genome still does (Bogorad 1991; Reith 1995), it is rather surprising that a nucleus-encoded ssRNAP serves as the mitochondrial transcription enzyme. In this regard, it is notable that the mitochondrial transcription factor required by yeast mitochondria for accurate initiation of

transcription has some sequence similarities with bacterial σ factors (Jang and Jaehning 1991). However, of greater significance is the fact that four eubacteria-like *rpo* genes have recently been discovered (Lang et al. 1997) in the mtDNA of *Reclinomonas americana* (Flavin and Nerad 1993), an early diverging mitochondria-containing relative of certain amitochondriate eukaryotes (retortamonads) (O'Kelly 1993; Brugerolle and Mignot 1990). This finding in particular lends credence to the view that the mitochondrial transcription system initially employed a eubacteria-like, multisubunit RNAP but that this mtDNA-encoded system was subsequently lost upon recruitment of a nucleus-encoded ssRNAP to act as the mtRNAP. This replacement must have occurred early in the evolution of the eukaryotic lineage because a T7-like RNAP gene has been identified in *Naegleria fowleri* (Cermakian et al. 1996), one of the earliest diverging of mitochondria-containing eukaryotes (Cavalier-Smith 1993). A very recent study provides evidence for the existence of a phage-like ssRNAP in chloroplasts of *Arabidopsis thaliana*, whose chloroplast genome still encodes a bacterial-like RNAP (Hedtke et al. 1997). Such data suggest that partial replacement of the latter type of RNAP by a ssRNAP has also taken place in chloroplasts.

In considering the evolutionary source of the nucleus-encoded ssRNAP gene, several scenarios can be entertained.

1. *An ssRNAP gene was already present in the nuclear genome of the eukaryotic cell that served as host for the mitochondrial endosymbiosis.* The origin of the eukaryotic genome is currently controversial (Doolittle 1996), with scenarios ranging from direct descent from a common ancestor with the Archaea (Keeling and Doolittle 1995) to a fusion of two prokaryotic cells, one a gram-negative eubacterium and the other an archaeon (Gupta and Golding 1996). In either case, as noted above, there is currently no evidence of an ssRNAP homolog in either eubacterial or archaeal genomes, and our attempts to amplify such a gene by PCR from the DNA of early diverging eukaryotes (including amitochondriate ones) have so far been unsuccessful. Although there are many technical reasons why PCR amplification might fail in particular instances, the available data do suggest that an ssRNAP gene was introduced into the eukaryotic nuclear genome at some early stage in the evolution of this lineage, subsequent to the acquisition of mitochondria.
2. *An ssRNAP gene was present in the DNA of the eubacterial endosymbiont that contributed the mitochondrial genome and was subsequently transferred to the nucleus.* There is abundant evidence that mitochondria originated from the rickettsial subdivision of the α-Proteobacteria (see Gray and Spencer 1996), and mitochondrion-to-nucleus gene transfer is a well-established feature of eukaryotic cell evolution (Gray 1992). Transfer during a cryptic endosymbiosis, such as that suggested (Henze et al. 1995) for the cytosolic glyceraldehyde-3-phosphate dehydrogenase of some protists, is also a possibility. Relatively little is known about the rickettsial group of α-Proteobacteria and their genomes, but the arguments outlined in 1, above, can also be advanced here against the idea that an ssRNAP gene was selectively present in the genome of the protomitochondrial endosymbiont. In previous work (Cermakian et al. 1996), we were unable to amplify an ssRNAP sequence from several α-proteobacterial DNAs, including those of *Rickettsia*, *Ehrlichia*, and *Agrobacterium*. More significant perhaps is that in sequenced mtDNAs, including 11 complete protist mtDNA sequences recently determined by the Organelle Genome Megasequencing Program (see http://megasun.bch.umontreal.ca/ogmp/projects.html), there is no evidence of remnants of a protomitochondrial ssRNAP gene that might have served as the source of the nucleus-encoded mtRNAP gene. Although ssRNAP sequences have been identified in some fungal mtDNAs (Robison and Horgen 1996; Robison et al. 1997), these are clearly related to plasmid-encoded, not nucleus-encoded, ssRNAP sequences. Their presence in mtDNA is best rationalized as the result of incorporation of mitochondrial plasmid-like DNA sequences into mtDNA via a recombination event.
3. *An ssRNAP gene was contributed by a bacterial virus that accompanied the protomitochondrial endosymbiont.* It has previously been suggested that a T7-like phage may have brought the ssRNAP gene into the eukaryotic host cell, perhaps by virtue of being present within the α-proteobacterial endosymbiont that contributed the mitochondrial genome (Gray 1989; Schinkel and Tabak 1989). One could imagine that a phage genome might have served as a convenient vector for transfer of the ssRNAP gene to the nucleus. Arguing against this scenario is the fact that whereas T7-like phages are known to infect a select group of Proteobacteria (enterobacteria or other γ-Proteobacteria such as *Pseudomonas putida* and *Caulobacter crescentus*; Hausmann 1988), they are not known to be present in rickettsial-type α-Proteobacteria. A relationship with phage T7 has been suggested for the cyanobacterial virus LPP-1 on morphological and biochemical grounds (Sherman and Haselkorn 1970), raising the possibility that the range of T7 like phage hosts is in fact broader than currently supposed. However, at the moment there are no compelling molecular data showing that an ssRNAP is encoded by the genome of LPP-1 or any other member of the Podoviridae (the morphological group that includes phage T7), outside of T7 and its close relatives.

An additional consideration is that T7 and its rela-

tives are virulent phages, whose multiplication leads to destruction of the eubacterial host cell. This makes it unlikely that a eubacterial endosymbiont harboring such a phage could have persisted for the length of time required to effect its evolutionary conversion to a mitochondrion. In this regard, it would seem that a temperate phage accompanying the protomitochondrial endosymbiont would be a much more likely source of a nucleus-encoded mtRNAP gene than a virulent phage, and it is entirely possible that such an entity exists within the α-Proteobacteria. Finally, although the nucleus-encoded and phage-encoded ssRNAPs clearly derive from a common ancestor, our data provide no support for the view that these two groups are *specifically* related to the exclusion of the plasmid-encoded sequences.

4. *An ssRNAP gene was contributed by a plasmid-like agent that accompanied the protomitochondrial endosymbiont.* If mitochondrial linear plasmids are indeed of bacteriophage ancestry (see Introduction), it is possible that such an entity, rather than a phage per se, was present in the mitochondrial endosymbiont, and that a plasmid ssRNAP gene was the direct precursor of the nuclear gene. The plasmid-borne ssRNAP gene could have been transferred to the nucleus along with many other mitochondrial genes. Known mitochondrial plasmids are stably maintained within the organelle, with no obvious effect on mitochondrial function, and such long-term maintenance would presumably have favored plasmid-to-nucleus gene transfer during the early stages of the mitochondrial endosymbiosis. However, as in the case of phage-encoded ssRNAP sequences, the phylogenetic evidence does not support a *specific* common ancestry of plasmid-encoded and nucleus-encoded ssRNAP genes. Moreover, mitochondrial plasmids that contain an ssRNAP gene have so far been identified only in certain angiosperms and fungi. This limited phylogenetic distribution, coupled with evidence of horizontal transfer in some cases, makes it difficult to argue that mitochondrial linear plasmids represent a transitional form in the acquisition of an ssRNAP gene by the nuclear genome, although it is entirely possible that a different plasmid-like DNA element, not directly related to contemporary mitochondrial plasmids, brought an ssRNAP gene into eukaryotes.

Summary and Future Prospects

From the existing data, we can draw a number of inferences about the origin and evolutionary relationships of known ssRNAP sequences. We suggest that the most likely origin of the ssRNAP gene was via duplication and divergence of a DNAP or reverse transcriptase gene. Phylogenetic and structural analysis supports the existence of three distinct classes of ssRNAP (phage-encoded, plasmid-encoded, and nucleus-encoded) sharing a common ancestor; however, the order of divergence of these clades, and therefore their specific relationships to one another, cannot be ascertained at present. The nuclear ssRNAP (gene encoding the mitochondrial RNAP in most eukaryotes) is clearly of monophyletic origin and appears to have arisen early in the evolution of the eukaryotic cell, coincident with or shortly after the mitochondrial endosymbiosis. It is possible to propose scenarios in which the source of the mtRNAP gene was either a T7-like phage or a plasmid-like DNA, but the data are not strongly supportive of either of these possibilities. It is equally possible that either the phage-encoded or the plasmid-encoded ssRNAP gene, or both, are derived in evolution from the nuclear gene: Current data are simply not able to distinguish the directionality of possible gene transfers. It does seem unlikely that an ssRNAP gene was already present in the host nuclear genome prior to the mitochondrial endosymbiosis or that it was introduced into the eukaryotic cell via the genome of a eubacteria-like endosymbiont.

Current data on phylogenetic distribution are also not strongly supportive of the idea that the ssRNAP gene is the remnant of a primitive transcription system that predated and was the evolutionary precursor of the multicomponent RNAPs used in the transcription of eubacterial, archaeal, and eukaryotic nuclear genomes. Rather, it would appear that this gene emerged relatively late in evolution, around the time of the origin of mitochondria. This raises the possibility that the origin of the nucleus-encoded mtRNAP and the origin of the ssRNAP gene per se may be temporally related: i.e., that the postulated duplication and divergence of a DNAP or RT gene may have occurred either in the proto-mitochondrial genome (with subsequent transfer to the nucleus) or in the eukaryotic nuclear genome, perhaps after transfer of an initially mtDNA-encoded DNAP or RT gene to the nucleus. Subsequent horizontal transfer events might have led to incorporation of a copy of the nuclear ssRNAP gene into a select group of bacteriophage genomes (those of T7 and its relatives) and/or into linear mitochondrial plasmid-like DNAs, with subsequent sequence divergence (which would have had to have been both rapid and extensive) of the three classes of ssRNAP. In this regard, it is worth emphasizing that genes encoding single-subunit DNAPs are much more widely distributed among both bacteriophage and plasmid DNAs than are ssRNAP genes. In addition, DNAP genes of this type have now been found in several protist mitochondrial genomes (see http://megasun.bch.umontreal.ca/ogmp/projects.html), and reverse transcriptases are commonly encoded by mitochondrial group II introns and some fungal mitochondrial plasmids (Xiong and Eickbush 1990). Further exploration of a possible evolutionary link between RNAP and DNAP and/or RT genes would seem

worthwhile, as would further study of those α-Proteobacterial species (especially minimally derived, free-living ones) that tree closest to mitochondria.

Acknowledgments. We thank Alice B. Rae (Université de Montréal) for critical reading of this manuscript, and Hans-Wolfgang Ackermann (Université Laval, Québec) and David F. Spencer (Dalhousie University) for helpful discussions. This work was supported by grants from NSERC Canada (to R.C.) and MRC Canada (MT-4124, to M.W.G., and MT-14028, to B.F.L.) and by an NSERC 1967 Science and Engineering Scholarship and an FCAR Scholarship (to N.C.). R.C., M.W.G., and B.F.L., who are fellows in the Program in Evolutionary Biology of the Canadian Institute for Advanced Research, gratefully acknowledge salary and interactions support from CIAR.

References

Ackermann H-W, DuBow MS (1987) Viruses of prokaryotes, Volume 2, natural groups of bacteriophages. CRC Press, Boca Raton, pp 72–81, 174–178

Barr DJS (1986) *Allochytridium expandens* rediscovered: morphology, physiology and zoospore ultrastructure. Mycologia 78:439–448

Blanco L, Salas M (1996) Relating structure to function in Φ29 DNA polymerase. J Biol Chem 271:8509–8512

Blattner FR, Plunkett G III, Mayhew GF, Perna NT, Glasner FD (1997) National Center for Biotechnology Information; locus ECOLI; accession number U00096

Bogorad L (1991) Replication and transcription of plastid DNA. In: Bogorad L, Vasil IK (eds) The molecular biology of plastids. Academic Press, San Diego, CA, pp 93–124

Bonner G, Patra D, Lafer EM, Sousa R (1992) Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. EMBO J 11:3767–3775

Braithwaite DK, Ito J (1993) Compilation, alignment and phylogenetic relationships of DNA polymerases. Nucleic Acids Res 21:787–802

Brugerolle G, Mignot J-P (1990) Phylum zoomastigina, class Retortamonadida. In: Margulis L, Corliss JO, Melkonian M, Chapman DJ (eds) Handbook of Protoctista. Jones and Bartlett, Boston, MA, pp 259–265

Bult CJ et al. (1996) Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. Science 273:1058–1073

Cavalier-Smith T (1993) Kingdom Protozoa and its 18 phyla. Microbiol Rev 57:953–994

Cermakian N, Ikeda TM, Cedergren R, Gray MW (1996) Sequences homologous to yeast mitochondrial and bacteriophage T3 and T7 RNA polymerases are widespread throughout the eukaryotic lineage. Nucleic Acids Res 24:648–654

Chan BS, Court DA, Vierula PJ, Bertrand H (1991) The *kalilo* linear senescence-inducing plasmid of *Neurospora* is an invertron and encodes DNA and RNA polymerases. Curr Genet 20:225–237

Chen B, Kubelik AR, Mohr S, Breitenberger CA (1996) Cloning and characterization of the *Neurospora crassa cyt-5* gene: a nuclear-coded mitochondrial RNA polymerase with a polyglutamine repeat. J Biol Chem 271:6537–6544

Corpet F (1988) Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res 16:10881–10890

Court DA, Bertrand H (1992) Genetic organization and structural features of maranhar, a senescence-inducing linear mitochondrial plasmid of *Neurospora crassa*. Curr Genet 22:385–397

Delarue M, Poch O, Tordo N, Moras D, Argos P (1990) An attempt to unify the structure of polymerases. Protein Eng 3:461–467

Dietz A, Weisser HJ, Kossel H, Hausmann R (1990) The gene for *Klebsiella* bacteriophage K11 RNA polymerase: sequence and comparison with the homologous genes of phages T7, T3 and SP6. Mol Gen Genet 221:283–286

Doolittle WF (1996) Some aspects of the biology of cells and their possible evolutionary significance. In: Roberts DM, Sharp P, Alderson G, Collins M (eds) Evolution of microbial life. Cambridge University Press, Cambridge, UK, pp 1–21

Felsenstein J (1993) PHYLIP: phylogeny inference package, version 3.5c. University of Washington, Seattle, WA

Flavin M, Nerad TA (1993) *Reclinomonas americana* N. G., N. Sp., a new freshwater heterotrophic flagellate. J Euk Microbiol 40:172–179

Fleischmann RD et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenza* Rd. Science 269:496–512

Fraser CM et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270:397–403

Gao G, Orlova M, Georgiadis MM, Hendrickson WA, Goff SP (1997) Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcription controls substrate selection. Proc Natl Acad Sci USA 94:407–411

Grahn AM, Bamford JKH, O'Neill MC, Bamford DH (1994) Functional organization of the bacteriophage PRD1 genome. J Bacteriol 176:3062–3068

Gray MW (1989) The evolutionary origins of organelles. Trends Genet 5:294–299

Gray MW (1992) The endosymbiont hypothesis revisited. Int Rev Cytol 141:233–357

Gray MW, Spencer DF (1996) Organellar evolution. In: Roberts DM, Sharp P, Alderson G, Collins M (eds) Evolution of microbial life. Cambridge University Press, Cambridge, UK, pp 109–126

Griffiths AJF (1995) Natural plasmids of filamentous fungi. Microbiol Rev 59:673–685

Gupta RS, Golding GB (1996) The origin of the eukaryotic cell. Trends Biochem Sci 21:166–171

Hausmann R (1988) The T7 group. In: Calendar R (ed) The bacteriophages, vol 1. Plenum Press, New York pp 259–289

Hedtke B, Börner T, Weihe A (1997) Mitochondrial and chloroplast phage-type RNA polymerases in *Arabidopsis*. Science 277:809–811

Henze K, Badr A, Wettern M, Cerff R, Martin W (1995) A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc Natl Acad Sci USA 92: 9122–9126

Hermanns J, Osiewacz HD (1992) The linear mitochondrial plasmid pAL2-1 of a long-lived *Podospora anserina* mutant is an invertron encoding a DNA and RNA polymerase. Curr Genet 22:491–500

Ikeda RA, Richardson CC (1987) Enzymatic properties of a proteolytically nicked RNA polymerase of bacteriophage T7. J Biol Chem 262:3790–3799

Ikeda RA, Chang LL, Warshamana GS (1993) Selection and characterization of a mutant T7 RNA polymerase that recognizes an expanded range of T7 promoter-like sequences. Biochemistry 32: 9115–9124

Jang SH, Jaehning JA (1991) The yeast mitochondrial RNA polymerase specificity factor, MTF1, is similar to bacterial σ factors. J Biol Chem 266:22671–22677

Kaneko T et al. (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res 3:109–136

Keeling PJ, Doolittle WF (1995) Archaea: narrowing the gap between prokaryotes and eukaryotes. Proc Natl Acad Sci USA 92:5761–5764

Kempken F, Hermanns J, Osiewacz HD (1992) Evolution of linear plasmids. J Mol Evol 35:502–513

Klenk H-P, Doolittle WF (1994) Archaea and eukaryotes versus bacteria? Curr Biol 4:920–922

Kostyuk DA, Dragan SM, Lyakkov DL, Rechinsky VO, Tunitskaya VL, Chernov BK, Kochetkov SN (1995) Mutants of T7 RNA polymerase that are able to synthesize both RNA and DNA. FEBS Lett 369:165–168

Kotani H, Ishizaki Y, Hiraoka N, Obayashi A (1987) Nucleotide se-

quence and expression of the cloned gene of bacteriophage SP6 RNA polymerase. Nucleic Acids Res 15:2653–2664

Kuzmin EV, Levchenko IV, Zaitseva GN (1988) S2 plasmid from cms-S-maize mitochondria potentially encodes a specific RNA polymerase. Nucleic Acids Res 16:4177

Lang BF, Burger G, Doxiadis I, Thomas DY, Bandlow W, Kaudewitz F (1977) A simple method for the large-scale preparation of mitochondria from microorganisms. Anal Biochem 77:110–121

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature 387:493–497

Levings CS III, Sederoff RR (1983) Nucleotide sequence of the S-2 mitochondrial DNA from the S cytoplasm of maize. Proc Natl Acad Sci USA 80:4055–4059

Li W-H (1997) Molecular evolution. Sinauer, Sunderland, MA

Lim L, Howlett BJ (1994) Linear plasmids, pLm9 and pLm10, can be isolated from the phytopathogenic ascomycete *Leptosphaeria maculans* by pulsed-field gel electrophoresis. Curr Genet 26:276–280

Manly BFJ (1991) Randomization and Monte Carlo methods in biology. Chapman and Hall, London

Masters BS, Stohl LL, Clayton DA (1987) Yeast mitochondrial RNA polymerase is homologous to those encoded by bacteriophages T3 and T7. Cell 51:89–99

McAllister WT (1993) Structure and function of the bacteriophage T7 RNA polymerase (or, the virtues of simplicity). Cell Mol Biol Res 39:385–391

McGraw NJ, Bailey JN, Cleaves GR, Dembinski DR, Gocke CR, Joliffe LK, MacWright RS, McAllister WT (1985) Sequence and analysis of the gene for bacteriophage T3 RNA polymerase. Nucleic Acids Res 13:6753–6766

Meinhardt F, Kempken F, Kämper J, Esser K (1990) Linear plasmids among eukaryotes: fundamentals and application. Curr Genet 17: 89–95

Moffatt BA, Dunn JJ, Studier FW (1984) Nucleotide sequence of the gene for bacteriophage T7 RNA polymerase. J Mol Biol 173:265–269

Oeser B, Tudzynski P (1989) The linear mitochondrial plasmid pC1K1 of the phytopathogenic fungus *Claviceps purpurea* may code for a DNA polymerase and an RNA polymerase. Mol Gen Genet 217: 132–140

Oeser B, Gessner-Ulrich K, Deing P, Tudzynski P (1993) pC1K1 and pC1T5—two linear mitochondrial plasmids from unrelated *Claviceps purpurea* strains: a comparison. Plasmid 30:274–280

O'Kelly CJ (1993) The jakobid flagellates: structural features of *Jakoba*, *Reclinomonas* and *Histiona* and implications for the early diversification of eukaryotes. J Euk Microbiol 40:627–636

Osumi-Davis PA, de Aguilera MC, Woody RW, Woody A-YM (1992) Asp537, Asp812 are essential and Lys631, His811 are catalytically significant in bacteriophage T7 RNA polymerase activity. J Mol Biol 226:37–45

Palenik B (1992) Polymerase evolution and organism evolution. Curr Opin Genet Dev 2:931–936

Patra D, Lafer EM, Sousa R (1992) Isolation and characterization of mutant bacteriophage T7 RNA polymerases. J Mol Biol 224:307–318

Reith M (1995) Molecular biology of rhodophyte and chromophyte plastids. Annu Rev Plant Physiol Plant Mol Biol 46:549–575

Robison MM, Horgen PA (1996) Plasmid RNA-polymerase-like mitochondrial sequences in *Agaricus bitorquis*. Curr Genet 29:370–376

Robison MM, Royer JC, Horgen PA (1991) Homology between mitochondrial DNA of *Agaricus bisporus* and an internal portion of a linear mitochondrial plasmid of *Agaricus bitorquis*. Curr Genet 19:495–502

Robison MM, Kerrigan RW, Horgen PA (1997) Distribution of plasmids and a plasmid-like mitochondrial sequence in the genus *Agaricus*. Mycologia 89:43–47

Rohe M, Schründer J, Tudzynski P, Meinhardt F (1992) Phylogenetic relationships of linear, protein-primed replicating genomes. Curr Genet 21:173–176

Salas M (1988) Phages with protein attached to the DNA ends. In: Calendar R (ed) The bacteriophages, vol 1. Plenum Press, New York, pp 169–191

Schinkel AH, Tabak HF (1989) Mitochondrial RNA polymerase: dual role in transcription and replication. Trends Genet 5:149–154

Sherman LA, Haselkorn R (1970) LPP-1 infection of the blue-green alga *Plectonema boryanum*. J Virol 6:841–846

Sousa R (1996) Structural and mechanistic relationships between nucleic acid polymerases. Trends Biochem Sci 21:186–190

Sousa R, Padilla R (1995) A mutant T7 RNA polymerase as a DNA polymerase. EMBO J 14:4609–4621

Sousa R, Chung YJ, Rose JP, Wang B-C (1993) Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. Nature 364:593–599

Swofford DL (1993) PAUP: phylogenetic analysis using parsimony, version 3.1. Illinois Natural History Survey, Champaign, IL

Vickery DB, Griffiths AJ (1993) Transcription of the kalilo linear senescence plasmid from *Neurospora intermedia*. Plasmid 29:180–192

Weihe A, Hedtke B, Börner T (1997) Cloning and characterization of a cDNA encoding a bacteriophage-type RNA polymerase from the higher plant *Chenopodium album*. Nucleic Acids Res 25:2319–2325

Woody A-YM, Eaton SS, Osumi-Davis PA, Woody RW (1996) Asp537 and Asp812 in bacteriophage T7 RNA polymerase as metal ion-binding sites studied by EPR, flow-dialysis, and transcription. Biochemistry 35:144–152

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362

Yoshikawa H, Ito J (1982) Nucleotide sequence of the major early region of bacteriophage Φ29. Gene 17:323–335