

Cloning and characterization of a cDNA encoding a bacteriophage-type RNA polymerase from the higher plant *Chenopodium album*

Andreas Weihe*, Boris Hedtke and Thomas Börner

Institute of Biology, Humboldt University Berlin, Chausseestrasse 117, D-10115 Berlin, Germany

Received March 18, 1997; Revised and Accepted April 23, 1997

DBDJ/EMBL/GenBank accession nos Y08067, Y11599, Y11600

ABSTRACT

We have cloned a full-length cDNA from the higher plant *Chenopodium album* coding for a single subunit bacteriophage-type RNA polymerase. The cDNA isolated from an actively growing cell suspension culture recognized a 3.8 kb transcript on Northern blots. The open reading frame comprises 987 amino acids with a predicted molecular mass of 112 kDa. A comparison of the protein sequence with those of the two known fungal mitochondrial RNA polymerases, from *Saccharomyces cerevisiae* and *Neurospora crassa*, reveals extensive homology between the three enzymes, with complete conservation of all catalytically essential amino acids. The putative mitochondrial RNA polymerase from *C.album*, as well as homologous sequences from rice and barley, which have been partially cloned, lack two catalytically non-essential regions of up to 176 amino acids near the C-terminus present in the two fungal mitochondrial RNA polymerases. The extreme N-terminus of the cloned *C.album* RNA polymerase displays features of a potential mitochondrial transit sequence. In phylogenetic trees constructed to compare the evolutionary relationships between the different single subunit RNA polymerases the *C.album* sequence forms a subgroup together with the *S.cerevisiae* and the *N.crassa* mitochondrial RNA polymerases, well separating from both bacteriophage enzymes and plasmid-encoded RNA polymerases found in mitochondria of many fungi and some higher plants.

INTRODUCTION

A major component of the mitochondrial transcriptional machinery of higher plants, the RNA polymerase transcribing the mitochondrial genes, has neither been purified nor characterized in detail, nor has the corresponding gene been isolated. In contrast, at least two fungal mitochondrial RNA (mtRNA) polymerases, from *Saccharomyces cerevisiae* and *Neurospora crassa*, are relatively well characterized (1–3). The yeast mtRNA polymerase represents a nuclear encoded single subunit enzyme of 145 kDa which requires

a specificity factor for promoter recognition, also coded in the nucleus (4–5). The core polymerase displays strong homology with bacteriophage RNA polymerases (6). No eubacteria-like polymerase genes have been found in any of the mitochondrial genomes sequenced so far, including the first total sequence of a higher plant mitochondrial genome from *Arabidopsis thaliana* (7). It is likely that the similarity of mtRNA polymerases is not restricted to fungi. Thus, the purified *Xenopus laevis* mtRNA polymerase also consists of a 140 kDa core subunit with an associated specificity factor (8) and similar data indicate that the same situation is met in human and mouse mitochondria (reviewed in 9). Accepting that mitochondria arose from eubacteria-like endosymbionts (10) and in contrast to the presence of eubacteria-like RNA polymerase genes in plant chloroplasts (11), these data support the idea that the mtRNA polymerase of higher plants is encoded by the nucleus. For chloroplasts, the existence of a second bacteriophage-like RNA polymerase was postulated, being imported into the organelle and involved in transcription of the chloroplast genome, in addition to the organelle-encoded eubacteria-like multicomponent RNA polymerase (12–14). However, identification of the corresponding gene and detailed characterization of the enzyme has not yet been achieved.

In fungal as well as in some plant mitochondria, single subunit RNA polymerase genes have been identified on linear plasmids displaying significant homology to the phage polymerases (15). However, these genes, whose role is still obscure, do not occur ubiquitously throughout the plant kingdom and it is generally assumed that they do not code for an RNA polymerase playing the role of the major component of the mitochondrial transcriptional apparatus.

Recently, a number of cDNA sequences with significant homology to the yeast mtRNA polymerase gene have appeared in expressed sequence tag (EST) databases (human, rice and *Caenorhabditis elegans*). The identification of these ESTs might be an indication of a common T3/T7-like organellar RNA polymerase type in many, if not all, eukaryotes. Using a PCR approach, Cermakian *et al.* reported the partial recovery of sequences homologous to yeast mitochondrial and T3/T7 RNA polymerases from a phylogenetically broad range of unicellular and multicellular eukaryotes, including rice and wheat (16).

*To whom correspondence should be addressed. Tel: +49 30 2093 8153; Fax: +49 30 2093 8141; Email: andreas=weihe@biologie.hu-berlin.de

In the present paper we report, for the first time, the molecular cloning of a full-length cDNA from a higher plant, *Chenopodium album*, encoding a putative single subunit RNA polymerase with strong homology to the yeast mitochondrial and bacteriophage RNA polymerases. The sequence information available from homologous rice ESTs permitted the generation of probes hybridizing to regions relatively well conserved between all polymerases of the single subunit type. Screening of a cDNA library and various RT-PCR strategies enabled us to clone the full-length cDNA for the *C.album* enzyme and to isolate partial cDNAs from rice and barley. Here we report these novel gene sequences and give a first characterization of the putative organellar RNA polymerases coded by these cDNAs.

MATERIALS AND METHODS

The sequences reported were deposited in the EMBL database under accession nos Y08067, Y11599 and Y11600.

RNA preparation

Total cellular RNA was isolated from a cell suspension of *C.album* L. maintained in permanent culture since 1989 (17), from green leaves of a commercial cultivar of rice (*Oryza sativa*) and from white leaves of the plastid ribosome-deficient *albostrians* mutant of barley (*Hordeum vulgare*) (18) using monophasic guanidinium/phenol as described by Chomczynski (19).

Poly(A)⁺ RNA was purified from total RNA with oligo(dT) using the Message Maker kit from Life Technologies according to the recommendations of the manufacturer.

PCR cloning

Chenopodium album first-strand cDNA was synthesized from 1 µg poly(A)⁺ RNA primed by an oligo(dT) primer [Q_T, CCAGTGAGCAGAGTGACGAGGACTCGAGCTCAAGC(T)₁₇] (20) with Superscript II reverse transcriptase (Life Technologies) in a total volume of 20 µl, for 60 min at 42°C and 30 min at 52°C. PCR amplification of internal fragments using primers OS-P1 and OS-M1 derived from a rice EST sequence (see Fig. 1) was performed in 50 µl reaction volumes containing PCR reaction buffer (Eurogentec) and 1 µl of a 1:50 dilution of first-strand *C.album* cDNA, 0.2 mM each dNTP, 1 µM each primer and 1 U Goldstar Taq polymerase (Eurogentec). Reactions were carried out in a Biometra TRIO thermocycler: with an initial denaturation at 94°C for 2 min; 5 cycles of 94°C for 1 min, 45°C for 1 min, 72°C for 1 min; 35 cycles of 94°C for 1 min, 55°C for 1 min, 72°C for 1 min; a hold at 72°C for 5 min.

The 3'- and 5'-ends of the putative *C.album* mtRNA polymerase cDNA were obtained by a RACE strategy using double-stranded adapter-ligated cDNA ('Marathon' cDNA Amplification Kit, Clontech; 21). Briefly, cDNA was synthesized from 1 µg poly(A)⁺ RNA primed by a lock-docking oligo(dT) cDNA synthesis primer [TTCTAGAATTCAGCGGCCGC(T)₃₀ (G/A/C)N] with Superscript II reverse transcriptase. Second-strand synthesis was performed by the Gubler-Hoffmann method (22). After blunting with T4 DNA polymerase, an adaptor was ligated to both ends ('marathon' cDNA adaptor: 5'-CTAATACGACTCACTATAGG-GCTCGAGCGGCCGCCGGCAGGT-3'; 3'-H₂N-CCCGTCA-PO₄-5'). RACE-PCR reactions were performed using an adaptor primer, AP1 (CCATCCTAATACGACTCACTATAGG-

GC), and gene-specific primers (see Fig. 1) in a total volume of 50 µl containing KlenTaq reaction buffer (Clontech), 1 µl of a 1:50 dilution of the 'marathon' cDNA, 0.2 mM each dNTP, 1 µM each primer and 1 µl KlenTaq polymerase mix containing an anti-Taq antibody for 'hot start' PCR (Clontech). Reactions were performed according to the following protocol: initial denaturation at 94°C for 1 min; 35 cycles of 94°C for 30 s, 68°C for 2 (3'-RACE) or 3 min (5'-RACE); a hold at 72°C for 10 min.

The 'marathon' cDNA population was also employed to amplify a near full-length fragment spanning the complete coding region using primers CA-CDS-P1 and CA-CDS-M1 (see Fig. 1). The PCR reaction mixture was the same as for the RACE amplifications and the cycling protocol consisted of: 1 min initial denaturation at 94°C; 35 cycles of 94°C for 30 s, 65°C for 30 s, 68°C for 4 min; a hold at 72°C for 10 min.

The 3'-ends of rice cDNA homologous to mtRNA polymerase were obtained by RACE using single-stranded cDNA as template according to the protocol of Frohman (20). Rice poly(A)⁺ RNA (1 µg) was used to synthesize first-strand cDNA as described for *C.album*. A first round of PCR amplification was performed in a 50 µl reaction volume containing PCR reaction buffer (Eurogentec) and 1 µl of a 1:50 dilution of first-strand rice cDNA, 0.2 mM each dNTP, 1 µM each primer Q₀ (CCAGTGAGCAGAGTGACG), homologous to the 5'-extension of the oligo(dT) cDNA synthesis primer Q_T, and gene-specific primer OS-P0 (GGCAAAGGCA-GCTTTTCCC), derived from a rice EST (accession no. D24565), and 1 U Goldstar Taq polymerase. An aliquot of 1 µl of a 1:20 dilution of this first PCR reaction was used in a second round of PCR amplification using nested primers Q₁ (GAGGACTCGAGC-TCAAGC), overlapping with Q₀ by one nucleotide, and OS-P2 (TTGCAGGGGTGCATGACTC). Cycling conditions were as described (20).

PCR products were gel purified using Jetsorb silica matrix (Genomed) and cloned into pCRII vector (Invitrogen).

Screening of cDNA libraries

A *C.album* cDNA library in λZiPlox (Life Technologies) was graciously provided by K.Meissner. The library was constructed with mRNA from *C.album* suspension culture cells harvested in the logarithmic phase of growth (K.Meissner, personal communication). The library was screened with a 0.9 kb fragment obtained by 3'-RACE from *C.album* cDNA (CA-3R, see Fig. 1). Using standard methods (23), positive clones were identified by plaque hybridization and the inserts recovered by *in vivo* excision.

A 0.9 kb 3'-RACE fragment amplified from rice cDNA was used to screen a cDNA clone bank constructed with mRNA from white leaves of the *albostrians* mutant of barley using λgt22 as vector. The library was kindly provided by W.R.Hess. Positive clones were identified by hybridization and the inserts were recovered by PCR using λgt22-specific primers.

Sequencing

Cloned PCR products and recombinant plasmids excised from λ cDNA clones were sequenced by primer walking using dye terminators and AmpliTaq FS in a cycle sequencing protocol according to the recommendations of the manufacturer (Applied Biosystems) on an ABI373 automatic DNA sequencer. DNA inserts recovered from barley cDNA library clones by PCR were sequenced directly using dye terminators.

a.

Amplified fragment	Sense primer	Antisense primer
CA-int	OS-P1: ATGATGATGACTGCTATTGC	OS-M1: GAAATTGTAATGTTGGGAA
CA-3R	CA-P1: GCATGTGATGTTGACAAAATGAACC	AP1: CCATCCTAATACGACTCACTATAGGGC
CA-5R	AP1: CCATCCTAATACGACTCACTATAGGGC	CA-M1: CCAAACCCCTGCGGACCAAAGG
CA-CDS	CA-CDSP: CCCATTTC AATTTC AACGCC	CA-CDSM: CTCCATTATCACATGCGGG

b.

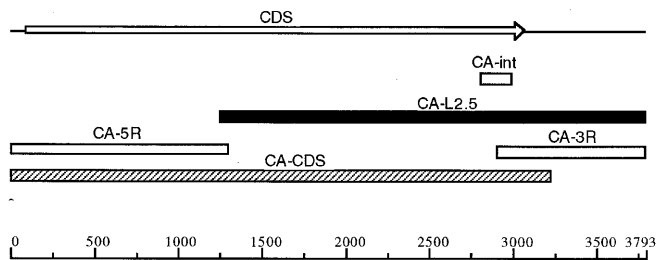


Figure 1. PCR primer design and strategy for amplification of putative mtRNA polymerase sequences from *C. album*. (a) Designation and sequence of primers. Primer sequences are given in 5'→3' orientation. Primers OS-P1 and OS-M1 were derived from an EST rice sequence homologous to mtRNA polymerase (GenBank accession no. D24565); base positions with a mismatch relative to the *C. album* sequence are indicated by italics. For RACE amplifications, the 'marathon' adaptor primer AP1 (Clontech) was used in combination with the gene-specific primers. (b) Schematic representation of RT-PCR and RACE fragments relative to the full-length *C. album* cDNA homologous to mtRNA polymerases. Internal RT-PCR and RACE fragments are represented by open bars. The clone CA-L2.5 isolated from a *C. album* cDNA library (using fragment CAS-3R as probe) is shown by a black bar, the near full-length PCR product (CA-CDS) spanning the whole coding region is hatched.

Northern blots and hybridization

Poly(A)⁺ and total RNA (10 and 40 µg/lane, respectively) from *C. album* were separated by electrophoresis on 1.2% formaldehyde-agarose gels and blotted onto Amersham Hybond-N membranes by capillary transfer with 20× SSC. The blots were probed with a ³²P-labeled riboprobe transcribed from a 2.5 kb fragment excised from a *C. album* λ cDNA library clone (CA-L2.5) containing the 3'-end of the putative mtRNA polymerase cDNA. Hybridization was carried out in 7% SDS, 0.25 M NaCl, 0.12 M sodium phosphate, pH 7.2, 50% formamide, at 55°C overnight. The blots were washed sequentially with 2× SSC, 1% SDS, 1× SSC, 1% SDS and 0.1× SSC, 0.1% SDS at 68°C for 15 min twice. The blots were exposed to phosphorimaging screens which were scanned on a Molecular Imager (BioRad, Hercules).

Computational sequence analyses

Database searches and general sequence comparisons relied on the BLAST world wide web server at the National Center for Biotechnology Information (24). Amino acid alignments were performed using the Multalin (25) and the ClustalW (26) programs. For phylogenetic analyses the PAUP program (27) was used.

RESULTS

Cloning of a *C. album* cDNA homologous to mitochondrial RNA polymerase

In a first attempt to isolate mitochondrial RNA polymerase sequences from higher plants we designed oligonucleotide primers targeted to a sequence region near the C-terminus of the polypeptides highly conserved among the fungal mitochondrial RNA polymerases and homologous ESTs available in databases. When using degenerate oligonucleotides for the amplification of

sequences from *C. album* cDNA, PCR products were obtained with an unacceptable high background of non-specific amplification products (data not shown). Therefore, in subsequent experiments we used non-degenerate primers derived from a rice EST sequence homologous to the two fungal mtRNA polymerase genes (see Fig. 1 for an overview of the oligonucleotide primers used in the present study). One primer combination permitted the amplification of a 196 bp fragment (CA-int) from *C. album* cDNA. The fragment was sequenced and used to design a new primer for 3'-RACE. Employing a PCR strategy based on adapter-ligated double-stranded cDNA as template ('marathon' RACE), 3'-ends of cDNAs were amplified, sequenced and identified as highly homologous to the mtRNA polymerase genes from *S. cerevisiae* and *N. crassa*. The *C. album* fragment (CA-3R) was used as a probe for screening of a *C. album* cDNA library. After screening of 600 000 recombinant phages, six positive clones were isolated and the DNA inserts recovered by *in vivo* excision. The largest cDNA clone, CA-L2.5, containing an insert of 2.5 kb, was fully sequenced by primer walking. The clone showed extended homology to the C-terminal part of the *Saccharomyces* and *Neurospora* genes. The cDNA sequence was, however, considered to be incomplete, since it lacked sequences related to the N-terminal part of the fungal mtRNA polymerases. 5'-RACE was performed to complete cloning of the *C. album* cDNA homologous to mtRNA polymerase. Primers were designed to amplify 5'-ends overlapping with the clones isolated from the cDNA library. Employing 'marathon' 5'-RACE, a fragment of the expected size was obtained (CA-5R), cloned into vector pCRII and sequenced. The composite sequence of the overlapping clones contained 3793 nt comprising an open reading frame of 987 amino acids with a predicted molecular mass of 112 kDa. The sequence contained 97 nt of a putative 5'-untranslated leader including one stop codon in-frame with the putative initiator ATG

and 717 nt of a 3'-untranslated trailer sequence. To obtain one contiguous clone, two primers were designed to amplify the region of the whole coding part of the cDNA and a PCR product of the expected size was amplified from *C.album* cDNA (CA-CDS, see Fig. 1). The fragment was cloned into vector pCRII and the insert sequenced. Comparison with the composite sequence (consisting of the recombinant phage clone and the 5'-RACE product) showed that both sequences were identical.

Cloning of 3'-ends of cDNAs from rice and barley homologous to mtRNA polymerase

Sense-oriented primers derived from a rice EST sequence were used to amplify 3'-ends of mtRNA polymerase-related cDNAs from rice. A PCR fragment of 913 nt was amplified, cloned into pCRII vector and fully sequenced. The sequence showed high homology to the 3'-end of the *C.album* cDNA and the two fungal mtRNA polymerase sequences. Downstream of the stop codon the sequence contained 649 nt of a 3'-untranslated trailer.

The rice 3'-RACE fragment was used as a probe to screen a cDNA library from white leaves of the *albostrians* mutant of barley, from which a positive clone with an insert of 756 nt could be isolated. Sequence analysis revealed high homology with the rice and the *C.album* cDNAs. No clones longer than that isolated could be found in the barley cDNA library after intensive screening of >500 000 recombinant clones.

Northern blot analysis

The mtRNA polymerase-related clone isolated from the *C.album* cDNA library was used to probe Northern blots of RNA preparations from *Chenopodium* suspension cells and leaves of rice and barley. When the formaldehyde-agarose gels were loaded with even high amounts of total RNA (40 µg) no signals were detected (data not shown), indicating that the cDNA represents a very low abundance transcript. Therefore, in subsequent experiments we used highly purified poly(A)⁺ RNA for Northern analysis. Finally, using 10 µg/lane mRNA and a riboprobe of high specific activity allowed us to detect a transcript in a Northern blot from *C.album* after 20 h exposure to a phosphorimaging screen (Fig. 2). The probe revealed a single mRNA species of 3.8 kb corresponding to the size of the composite full-length cDNA. We were not able to detect clearly discernible signals in blots from rice and barley. We suppose that the transcripts are present at an extremely low level in leaf cells, which corresponds with the low number of positive clones identified in the two cDNA libraries we screened.

Characterization of the putative plant mtRNA polymerase sequences

Sequence analysis of the *C.album* cDNA suggests that the encoded protein represents a functional RNA polymerase with strong homology to fungal mtRNA polymerases and bacteriophage polymerases. An alignment of the predicted full-length *C.album* amino acid sequence and the partial sequences from rice and barley with the two fungal mtRNA polymerases is shown in Figure 3. The overall degree of homology decreases from the C- to the N-terminus. All catalytically essential residues, as determined on the basis of the recently reported T7 RNA polymerase crystal structure (28), are strongly conserved. Extended blocks of homology as described for single subunit RNA polymerases (29)

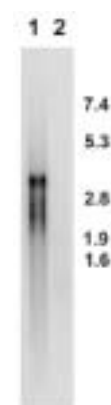


Figure 2. Northern blot of *C.album* mRNA (lane 1) and total RNA (lane 2). The putative mtRNA polymerase mRNA was detected by hybridization to a ³²P-labeled riboprobe generated from clone CA-L2.5 (see Materials and Methods). Relative mobilities of RNA standards (Boehringer Mannheim) are shown on the right.

are clearly discernible. The absence of two long stretches of 21 and 95 (relative to the *S.cerevisiae* enzyme) and 48 and 176 (relative to *N.crassa*) catalytically non-essential amino acids specifically present in the fungal polymerases but lacking in all three plant sequences should be noted.

The extreme N-terminus of the *C.album* sequence is rich in hydroxylated amino acids, mainly serine, interrupted by basic residues, a feature consistent with the general characteristics of mitochondrial transit peptides (30).

We have constructed phylogenetic trees to investigate the evolutionary relationships between the different RNA polymerase sequences. The *C.album* polymerase and the two fungal mtRNA polymerases clearly associate with each other, well separated from a branch containing both plant and fungal single subunit polymerases found on linear mitochondrial plasmids (Fig. 4). When the partial sequences from rice, barley, wheat and a number of lower eukaryotes (16) were included (in this case only the corresponding regions from the other 'full-length' polymerases were used for tree construction), all plant species, i.e. rice, barley, wheat and *C.album*, were found within one branch, together with the lower eukaryotes and the two fungal mtRNA polymerases (data not shown).

DISCUSSION

We used the sequence of a rice EST with strong homology to yeast mtRNA polymerase to clone a homologous, full-length cDNA from the higher plant *C.album* and the 3'-ends of homologous cDNAs from rice and barley.

Alignment of the three plant cDNA sequences with fungal mtRNA and bacteriophage polymerases reveals extensive sequence homology and suggests that the sequences all code for the same enzyme, a putative single subunit mtRNA polymerase.

Recently, the X-ray crystal structure of the T7 RNA polymerase has been determined, and the active site was found to be represented by a right hand-shaped 'palm and finger' domain (28). The template binding cleft forming the catalytic pocket in the palm of the T7 RNA polymerase comprises several invariant amino acid residues essential for catalytic activity (corresponding to D945, K1014, Y1022, G1023, H1188, D1189 and F1350 in the

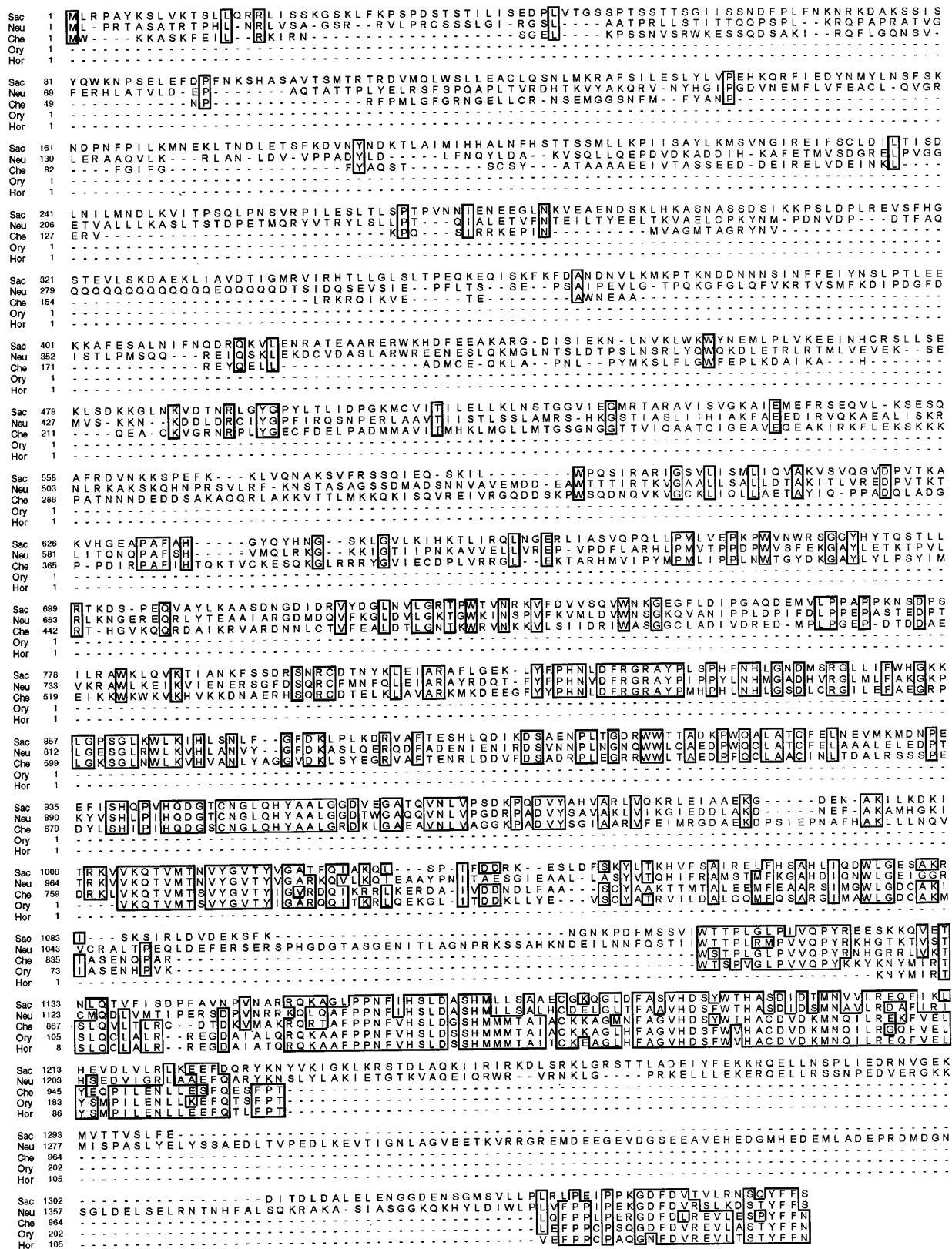


Figure 3. Alignment of single subunit bacteriophage-type RNA polymerases. Abbreviations of organism names (accession nos in square brackets): Sac, *Saccharomyces cerevisiae* (P13433); Neu, *Neurospora crassa* (P38671); Che, *Chenopodium album* (Y08067); Ory, *Oryza sativa* (rice), composite sequence of a partial cDNA clone from (16) (U34283) overlapping with the RACE fragment reported in the present study (Y11599); Hor, *Hordeum vulgare* (barley) (Y11600). The alignment was constructed using the Multalin (9) and ClustalW (10) programs and refined manually. Amino acids that are positionally identical in at least three of the sequences are boxed.

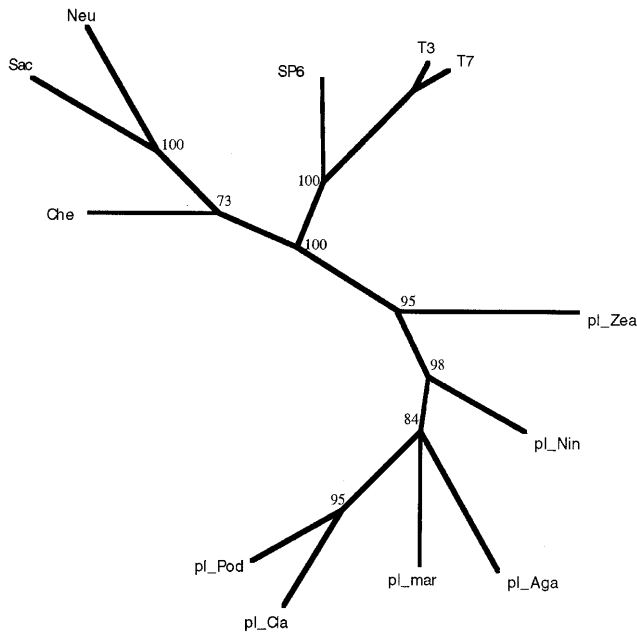


Figure 4. Evolutionary relatedness of bacteriophage-type RNA polymerases. A consensus minimum length tree based on the predicted amino acid sequences was constructed using the principle of maximum parsimony with the PAUP program (release 3.1.1). Bootstrap confidence values are shown at nodes. Branch lengths are proportional to relative sequence divergence. Abbreviations as in Figure 3, with additional ones as follows: T7, bacteriophage T7 (P00573); T3, bacteriophage T3 (P07659); SP6, bacteriophage SP6 (P06221); pl_mar, mitochondrial plasmid maranhar from *Neurospora crassa* (33540); pl_Aga, mitochondrial plasmid from *Agaricus bitorquis* (P33539); pl_Pod, mitochondrial plasmid pAL2-1 from *Podospora anserina* (505556); pl_kal, mitochondrial plasmid kalilo from *N.intermedia* (P33541); pl_Cla, mitochondrial plasmid pCLK1 from *Claviceps purpurea* (P22372); pl_Zea, mitochondrial plasmid S-2 from *Zea mays* (P10581).

yeast sequence). As we can show, all these residues are conserved between the plant, fungal and bacteriophage RNA polymerases (see Fig. 3). Remarkably, there is almost no sequence homology between the N-termini of the polymerases. It has been noted that the highly variable N-termini of the fungal mtRNA polymerases could represent extensions of a subdomain which in the T7 enzyme folds towards the right hand-shaped catalytic domain, without disrupting the active site (31). The extended N-termini in the fungal polymerases have been considered as candidates for binding sites of transcription factors (3). The N-terminus of the *C.album* polymerase is considerably shorter than in the case of the fungal enzymes, excluding the possibility that potential transcription factors bind to such a region in the plant RNA polymerase.

In the T7 RNA polymerase the extreme C-terminus is topologically adjacent to the catalytically active domain and phenylalanine F882, the penultimate residue, has been shown to be essential for interaction with the incoming rNTP (32). Consistent with these data, there is a high degree of conservation of the extreme C-terminus between all single subunit RNA polymerases, including the three plant species. Peculiar to the fungal mtRNA polymerases is the presence of two longer stretches of up to 176 amino acids near the C-terminus, which is missing in all plant sequences (as well as in the bacteriophage enzymes). No special function has been yet assigned to these residues. One could speculate, however, that they might be involved

in interaction with co-factors specific for the fungal mitochondrial transcriptional machinery.

The high degree of sequence similarity between the *C.album* RNA polymerase and the fungal enzymes suggests that the *C.album* cDNA encodes a plant mtRNA polymerase. Another indication that the cDNA cloned from *C.album* most likely encodes a mitochondrial enzyme is the structure of the extreme N-terminus of the predicted protein: it is rich in hydroxylated amino acids (mainly serine), interspersed with positively charged and non-polar residues and capable of forming an amphiphilic α -helix, features consistent with the general characteristics of mitochondrial transit peptides (30). Thus, considering the sequence features of the *C.album* cDNA, targeting of the predicted protein to mitochondria is most likely, although we cannot totally rule out that the protein is also, or alternatively, targeted to chloroplasts.

There are several lines of evidence suggesting that in plant cells there might exist, in addition to the chloroplast DNA-encoded enzyme, a second chloroplast RNA polymerase encoded in the nucleus (12–14) which might be associated with a 110 kDa enzyme displaying some bacteriophage-like features (13). We cannot therefore exclude the possibility that there are two bacteriophage-like single subunit RNA polymerases, being imported into the mitochondria and chloroplasts, respectively. On the other hand, the possibility exists that the same RNA polymerase (encoded by one nuclear gene) is targeted to both organelles. Thus, it has been demonstrated that the presequence of pea glutathione reductase is capable of co-targeting this enzyme or a foreign protein simultaneously to chloroplasts and mitochondria *in vivo* (33). However, analysis of the *C.album* sequence by a knowledge-based system for prediction of protein targeting, PSORT (<http://psort.nibb.ac.jp>), does not reveal any features of the N-terminus consistent with targeting to plastids. Furthermore, preliminary experimental data from *in vitro* import studies suggests that the N-terminus of the sequence is capable of targeting proteins to mitochondria (data not shown).

The low representation of the *C.album* cDNA in the phage library as well as the extremely low levels of transcript detected in Northern blots might be an indication of a very low expression level of the corresponding gene. On the other hand, from the multicopy nature of mitochondrial (or chloroplast) genes in the plant cell, one would expect that there is a requirement for a considerable amount of active organellar RNA polymerases. This discrepancy could probably be explained by high stability of the transcripts and/or their translation products.

Analysis of the phylogenetic relationships between the different single subunit RNA polymerases shows that the *C.album* polymerase is most closely related to the fungal mtRNA polymerases. The robust branching pattern excludes the possibility that the plant enzyme originates from RNA polymerases encoded on linear plasmids, which form a well-separated group only distantly related to the fungal mtRNA polymerases. We also constructed phylogenetic trees including a broader spectrum of RNA polymerases and related sequences (not shown). Due to the limited sequence data available for rice, barley, wheat and a number of lower eukaryotes, only a relatively small part of the sequences could be used for the construction of phylogenetic trees. Although this limitation influenced the robustness of the branching patterns, the analyses suggest a general scheme in which three main branches can be distinguished: (i) bacteriophage RNA polymerases; (ii) plasmid-encoded RNA polymerases (occurring in

many fungal and some plant mitochondria); (iii) a group comprising the mtRNA polymerases from *S.cerevisiae* and *N.crassa*, the putative mtRNA polymerase from the higher plant *C.album* and partial cDNA-derived sequences from rice, barley, wheat and from a number of lower eukaryotes. Although the limited sequence information does not allow final conclusions, we suppose that the third group represents organellar RNA polymerases.

While this manuscript was in preparation, a genomic DNA sequence from *A.thaliana* became available in the EMBL database (accession no. Y09432, Schuster *et al.*, unpublished) which carries a gene with extensive sequence homology to fungal mtRNA polymerases. This gene and its corresponding cDNA have also been isolated in our laboratory and deposited in the EMBL database as accession nos Y09006 and Y08137, respectively. *In vitro* import experiments suggest that it encodes the *Arabidopsis* mtRNA polymerase (Hedtke *et al.*, in preparation). The *Arabidopsis* sequence reveals 64% identical and 75% functionally similar amino acids compared with the *C.album* sequence reported here. This data further supports the idea that bacteriophage-like enzymes represent the common type of mtRNA polymerases in higher plants.

ACKNOWLEDGEMENTS

We are indebted to K.Meissner and W.R.Hess (Institute of Biology, Humboldt University Berlin) for provision of a *C.album* and a *albostrians* barley mutant cDNA library, respectively. The excellent technical assistance of C.Stock, J.Müller and I.Wagner is gratefully acknowledged. Part of the work was supported by grants of the Deutsche Forschungsgemeinschaft, Bonn, to A.W., and the Fond der Chemischen Industrie, Frankfurt am Main, to T.B.

REFERENCES

- Kelly,J.L. and Lehmann,I.R. (1986) *J. Biol. Chem.*, **261**, 10340–10347.
- Schinkel,A.H., Groot-Koerkamp,M.J.A. and Tabak,H.F. (1988) *EMBO J.*, **7**, 3255–3262.
- Chen,B., Kubelik,A.R., Mohr,S. and Breitenberger,C.A. (1996) *J. Biol. Chem.*, **271**, 6537–6544.
- Kelly,J.L., Greenleaf,A.L. and Lehmann,I.R. (1986) *J. Biol. Chem.*, **261**, 10348–10351.
- Lisowsky,T. and Michaelis,G. (1988) *Mol. Gen. Genet.*, **214**, 218–223.
- Masters,B.S., Stohl,L.L. and Clayton,D.A. (1987) *Cell*, **51**, 89–99.
- Unsold,M., Marienfeld,J.R., Brandt,P. and Brennicke,A. (1977) *Nature Genet.*, **15**, 57–61.
- Bogenghagen,D.F. and Insdorf,N.F. (1988) *Mol. Cell. Biol.*, **8**, 2910–2916.
- Clayton,D.A. (1991) *Annu. Rev. Cell Biol.*, **7**, 453–478.
- Gray,M.W. (1989) *Trends Genet.*, **5**, 294–299.
- Bogorad,L. (1991) In Bogorad,L. and Vasil,I.K. (eds), *The Molecular Biology of Plastids*. Academic Press, San Diego, CA, pp. 93–124.
- Hess,W.R., Prombona,A., Fieder,B., Subramanian,A.R. and Börner,T. (1993) *EMBO J.*, **12**, 563–571.
- Lerbs-Mache,S. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 5509–5513.
- Allison,L.A., Simon, L.D. and Maliga,P. (1996) *EMBO J.*, **15**, 2802–2809.
- Kemken,F., Hermanns,J. and Osiewacz,H.D. (1992) *J. Mol. Evol.*, **35**, 502–513.
- Cermakian,N., Ikeda,T.M., Cedergren,R. and Gray,M.W. (1996) *Nucleic Acids Res.*, **24**, 648–654.
- Dörfel,P., Weihe,A., Knösche,R. and Börner,T. (1989) *Curr. Genet.*, **16**, 375–380.
- Hagemann,R. and Scholz,F. (1962) *Der Züchter*, **32**, 50–59.
- Chomczynski,P. and Sacchi,N. (1986) *Anal. Biochem.*, **162**, 156–159.
- Frohman,M.A. (1994) In Mullis,K.B., Ferré,F. and Gibbs,R.A. (eds), *The Polymerase Chain Reaction*. Birkhäuser, Boston, MA, pp. 14–37.
- Siebert,P.D., Chenchik,A., Kellogg,D.E., Lukyanov,K.A. and Lukyanov,S.A. (1995) *Nucleic Acids Res.*, **23**, 1087–1088.
- Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
- Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) *J. Mol. Biol.*, **215**, 403–410.
- Corpet,F. (1988) *Nucleic Acids Res.*, **16**, 10881–10890.
- Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) *Nucleic Acids Res.*, **22**, 4673–4680.
- Swofford,D. and Olsen,G. (1990) In Hillis,D.M. and Moritz,C. (eds), *Molecular Systematics*. Sinauer, Sunderland, pp. 411–501.
- Sousa,R., Chung,Y.J., Rose,J.P. and Wang,B.-C. (1993) *Nature*, **364**, 593–599.
- DeLarue,M., Poch,O., Tordo,N., Morse,D. and Argos,P.J. (1990) *J. Protein Engng*, **10**, 461–467.
- Chaumont,F. and Boutry,M. (1995) In Levings,C.S. and Vasil,I.K. (eds), *The Molecular Biology of Plant Mitochondria*. Kluwer Academic, Dordrecht, The Netherlands, pp. 207–235.
- Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A. and Steitz,T.A. (1992) *Science*, **256**, 1783–1790.
- Patra,D., Lafer,E.M. and Sousa,R. (1992) *J. Mol. Biol.*, **224**, 307–318.
- Creissen,G., Reynolds,H., Xue,Y. and Mullineaux,P. (1995) *Plant J.*, **8**, 167–175.