J Mol Evol (1997) 45:301–310

JOURNAL OF MOLECULAR EVOLUTION

Phylogenetic Relationships Among Hypotrichous Ciliates Determined with the Macronuclear Gene Encoding the Large, Catalytic Subunit of DNA Polymerase α

David C. Hoffman, David M. Prescott

Department of Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, Campus Box 347, Boulder, CO 80309-0347, USA

Received: 7 January 1997 / Accepted: 7 April 1997

Abstract. The complete macronuclear DNA polymerase α gene, previously sequenced in *Oxytricha nova*, has been cloned from a genomic macronuclear library and sequenced for the hypotrich *O. trifallax*. Macronuclear DNA clones of DNA polymerase α encoding ~1000 amino acids, or approximately two-thirds of the open reading frame, have been obtained by PCR and sequenced for *Halteria grandinella*, *Holosticha* species, *Paraurostyla viridis*, *Pleurotricha lanceolata*, *Stylonychia lemnae* Teller, *Sty. mytilus*, *Uroleptus gallina*, and *Urostyla grandis*. Phylogenetic relationships inferred from DNA polymerase α amino acid sequences have been used to clarify taxonomic relationships previously determined by morphology of the cell cortex. Hypotrich phylogenies based on DNA polymerase α amino acid sequences are incongruent with morphological and other molecular phylogenies. Based upon these data, we assert that, contrary to morphological data, *O. nova* and *O. trifallax* are different species, and we propose that the oligotrich *Halteria grandinella* be reclassified as a hypotrich. This work also extends the available data base of eukaryotic DNA polymerase α sequences, and suggests new amino acid sequence targets for mutagenesis experiments to continue the functional dissection of DNA pol α biochemistry at the molecular level.

Key words: DNA polymerase α gene — Hypotrichous ciliates — Macronuclear genes — Phylogenetics

Correspondence to: D.M. Prescott

Introduction

Taxonomists have defined ciliate taxonomic groups using morphological criteria, primarily of the cell cortex. Different classification schemes emphasizing a variety of morphological features have identified from three to eight major classes (Corliss 1979; Lynn and Sogin 1988). More recently, the two primary schemes have been unified using a combination of classical phylogenetic analysis and molecular data from rRNA sequences isolated from a few species (Lynn and Corliss 1991). The molecular data include nucleotide sequences of the small subunit rRNA of *Euplotes aediculatus*, *Halteria grandinella*, *Stylonychia pustulata*, and *Oxytricha nova*, and nucleotide sequences of the large subunit rRNA of *E. aediculatus*, *H. grandinella*, *Paraurostyla* species, *S. lemnae*, *Urostyla* species, and *Uronychia* species. These sequences represent a beginning to the definition of phylogenetic relationships among members of the order Hypotrichida using molecular data (Elwood et al. 1985; Lynn and Sogin 1988; Schlegel et al. 1991; Baroin-Tourancheau et al. 1992). We have extended the analysis of phylogenetic relationships in 10 hypotrichs with the macronuclear gene that encodes DNA polymerase α (DNA pol α).

The eukaryotic DNA pol α holoenzyme is composed of a large catalytic subunit carrying the polymerase activity of ~180 kilodaltons (kDa) and three accessory subunits of ~70 kDa, ~60 kDa, and ~50 kDa (Lehman and Kaguni 1989; Kornberg and Baker 1992). DNA pol α has been classified as a family B DNA polymerase,

based upon amino acid sequence similarity to *Escherichia coli* DNA polymerase II (Braithwaite and Ito 1993). The two smallest polypeptides exhibit DNA primase activity (Kaguni et al. 1983; Plevani et al. 1985; Lucchini et al. 1987), and the ~70 kDa subunit is thought to have a regulatory function (Cotterill et al. 1987). The DNA pol α holoenzyme lacks a 3′-5′ proofreading exonuclease, and displays only moderate processivity, polymerizing from ~200 to ~2000 nucleotides per binding event (Kaguni et al. 1983; Plevani et al. 1985; Lucchini et al. 1987; Blow 1989; Kornberg and Baker 1992). Because of its associated DNA primase activity and moderate processivity, DNA pol α is thought to polymerize the lagging strand; it may be also required to prime leading strand synthesis (So and Downey 1988). However, additional work has demonstrated that DNA polymerase II, the yeast homolog of eukaryotic DNA pol ϵ, is required for chromosomal DNA replication in the budding yeast *Saccharomyces cerevisiae* (Araki et al. 1992). Temperature-sensitive mutants of the *S. cerevisiae* DNA pol ϵ have an elongated S phase and arrest with a phenotype characteristic of DNA replication mutants, strongly suggesting that DNA pol ϵ has a role in chromosomal replication in *S. cerevisiae* (Araki et al. 1992), raising the possibility that DNA pol α may be only required for primer synthesis during DNA replication.

Previous amino acid sequence comparisons of family B DNA polymerases from a variety of sources identified numerous putative functional domains based upon varying degrees of amino acid sequence conservation (from ~16 to ~88% identity; Wong et al. 1988; Wang et al. 1989; Braithwaite and Ito 1993; White et al. 1993). Subsequent functional analysis of these conserved regions by site-directed mutagenesis has shown that some of these amino acid sequences are involved in magnesium (Mg^{2+}), deoxynucleotide triphosphate (dNTP), or primer/template binding, as well as in the interaction of the polymerase subunit with the subunits carrying the DNA primase activity (Gibbs et al. 1988; Pizzagalli et al. 1988; Copeland and Wang 1993; Copeland et al. 1993; Dong et al. 1993a,b; Dong and Wang 1995).

In this article, we describe the isolation and characterization of a complete macronuclear DNA pol α gene sequence from *Oxytricha trifallax* and of partial macronuclear DNA pol α gene sequences from eight additional species of hypotrichous ciliates. The complete gene sequence of a tenth species, *O. nova*, was previously published (Mansour et al. 1994). Clones were obtained by either screening a genomic library (*O. trifallax*) or the polymerase chain reaction (PCR) using degenerate deoxyoligonucleotide primers (eight hypotrichs). Derived DNA pol α amino acid sequences were used to infer phylogenetic relationships among the hypotrichs. In addition, alignment of predicted amino acid sequences identified both functional domains associated with family B DNA polymerases in general (Wang et al. 1989; Braithwaite and Ito 1993), and with eukaryotic DNA polymerases α in particular (White et al. 1993). These sequence comparisons revealed 11 additional, unreported putative functional domains conserved among eukaryotes, as well as domains conserved only among the hypotrichs. Thus this work extends the available data base of eukaryotic DNA polymerase α sequences, and suggests new amino acid sequence targets for mutagenesis experiments to continue the functional dissection of DNA pol α biochemistry at the molecular level.

Materials and Methods

Cell Culture, Isolation of Nuclei, and Preparation of Genomic DNA. Cultivation of the various hypotrich species, preparation of macro- and micronuclei and isolation of macro- and micronuclear DNA were done as described elsewhere (Prescott and Greslin 1992).

Library Construction, Screening, and λ Phage Preparation. Genomic macronuclear libraries were prepared using λgt10 (Promega, Madison, WI) as described elsewhere (Hoffman 1996). Plaque lifts were done using Nitropure-supported nitrocellulose membranes (Micron Separations Inc., Westboro, MA) according to standard procedures, and hybridized to radiolabelled probe (prepared as described below) in NES (0.5 M Na_2HPO_4, pH 7.2/1 mM EDTA/7% SDS). Lifts were washed in SSC (1 × SSC = 0.15 M NaCl/0.015 M Na_3 · citrate) and exposed to Kodak X-Omat X ray film (Eastman-Kodak) at −70°C with a Dupont Lightning intensifying screen (DuPont/New England Nuclear). Purified λ clones were amplified and purified by equilibrium density gradient centrifugation in a linear CsCl gradient (Amersham Life Sciences, Arlington Heights, IL; Sambrook et al. 1989).

Southern Hybridization. All restriction enzymes were purchased from New England Biolabs (NEB). DNA was electrophoresed through agarose gels (FMC Bioproducts, Rockland, ME) of an appropriate percentage containing ethidium bromide at 0.5 μg/ml. DNA was transferred to Zetabind-supported nitrocellulose (Life Science Products Inc., Denver, CO) by capillary blotting with 0.4 M NaOH. Membranes were then neutralized in 1.5 M NaCl/0.5 M Tris-HCl (pH = 7.5) and cross-linked under UV light in a StrataLinker (Stratagene, La Jolla, CA). Probe fragments were electrophoresed through agarose and purified using the Qiaquick gel extraction kit (Qiagen, Chatsworth, CA), and labeled by incorporation of [α-^{32}P]dATP and [α-^{32}P]dTTP (DuPont/New England Nuclear, Wilmington, DE) using either random hexamers (Pharmacia Corp., Piscataway, NJ) or specific DNA primers (DNA International, Lake Oswego, OR or Integrated DNA Technologies Inc., Coralville, IA) and the Klenow fragment of *E. coli* DNA polymerase I (NEB). Hybridizations were performed using a Hybaid hybridization oven (National Labnet, Woodbridge, NJ) or a water bath in NES. Washes were generally done to a final stringency of 0.1× SSC/0.1% SDS at the hybridization temperature (52°–65°C).

Subcloning. Restriction fragments were excised from agarose gels and purified with the Qiaquick gel extraction kit (Qiagen). PCR products were purified with the Qiaquick PCR Cleanup Kit (Qiagen); those with restriction sites incorporated into the primers were digested with restriction enzymes and repurified with the PCR Cleanup Kit. Ligations were performed using T4 DNA Ligase (NEB), and ligation mixtures were used to transform competent *E. coli* (strain DH5α) prepared by the method of Hanahan (Hanahan 1983). Transformations were done by either the standard heat-shock method as described elsewhere (Hanahan 1983) or a high-efficiency method without heat-shock (Pope and Kent 1996).

Plasmid DNA Preparation and Sequencing. Small-scale preparations of plasmid DNA were obtained by the alkaline lysis method using the Qiaprep Spin Plasmid Miniprep Kit (Qiagen). Plasmid sequence was obtained by the dideoxynucleotide chain termination method using Sequenase DNA polymerase (Amersham Life Sciences) and [α-^{35}S]dATP (Dupont/New England Nuclear) incorporation. PCR product sequence was obtained by the same method using the Sequenase PCR Product Sequencing Kit (Amersham Life Sciences). Sequencing primers were obtained from either DNA International, Integrated DNA Technologies Inc., GIBCO Life Sciences (Grand Island, NY), or Only DNA (Midland, TX). All sequencing gels were prepared using Long Ranger acrylamide solution (FMC Bioproducts). Plasmid sequencing gels were 6% Long Ranger/7 M urea/1× TBE (89 mM Tris-borate/89 mM boric acid/2 mM EDTA, pH 8.0). PCR product sequencing gels were 6% Long Ranger/7 M urea/1× GTB (89 mM Tris-HCl/29 mM taurine/0.54 mM EDTA, pH 8.0). Sequence alignments and analysis were performed on an Apple Macintosh computer with the AssemblyLIGN/MacVector sequence analysis package (IBI-Kodak, Rochester, NY).

Degenerate PCR. DNA polymerase α clones were obtained from hypotrich species with a PCR strategy utilizing degenerate primers designed from the two most highly conserved sequences of α-type DNA polymerases:

pdR2-5′-GGG GAA TTC GAY TTY AAY WSI YTI TAY CC-3′

mdR1-5′-CCC CGG ATC CTC IGT RTC ICC RTA IAC IAC-3′

Primer pdR2 is 128-fold degenerate, contains an *EcoR* I restriction site, and amplifies all nucleotide sequences specifying the amino acid sequence DFNSLYP, from conserved region II (Wong et al. 1988). Primer mdR1 is fourfold degenerate, contains a *BamH* I restriction site, and amplifies all nucleotide sequences specifying the amino acid sequence VVYGDTD, from conserved region I (Wong et al. 1988). The PCR mixture was 60 mM Tris-SO_4 (pH 9.1 at 25°C), 18 mM $(NH_4)_2SO_4$, and 1–2 mM $MgSO_4$ containing all 4 dNTPs at 125 mM and 100 pmol of each primer. Typically ~10 ng of macronuclear DNA was used as template. After a 3- to 5-min hot start at 94°C, 1 μl of eLONGase enzyme mix (GIBCO-BRL) was added, and the thermal cycling was initiated. Thermal cycling was done in an Ericomp thermal reactor (San Diego, CA) according to the following protocol (modified from Asai and Criswell 1995): 5 cycles of 94°C/30 s, 50°C/1 min, 68°C/1 min and 40 cycles of 94°C/30 s, 42°C/1 min, 68°C/1 min.

Genomic walking toward the NH_3-terminus of the DNA polymerase α gene employed a similar strategy. Using a specific primer designed to amplify sequences within conserved region III (contained within the pdR2/mdR1 PCR product) and a third degenerate primer, PCR products containing approximately two-thirds of the DNA pol α open reading frame were amplified. The third degenerate primer was designed to amplify all nucleotide sequences specifying the amino acid sequence FIVDDDG, from conserved region A:

pdRA-5′-CGC GGA TCC TTY ATH GTI GAY GAY GAY GG-3′

Primer pdRA is 48-fold degenerate and contains a *BamH* I site. The PCR protocol was the same, with two exceptions: only 10 pmol of the specific primer were added per reaction, and the extension time was increased from 1 to 3 min.

Phylogenetic Analysis. Amino acid sequence alignments were generated with the Pileup function of the GCG sequence analysis package, using a gap creation penalty of 3.0 and a gap extension penalty of 0.1 (GCG 1994). Only previously reported amino acid sequences conserved in all 15 species were included in the alignment (regions I–IV, VI, A, B, D, and E). Eleven newly reported amino acid sequences (conserved in all 15 species) identified by this analysis were included in the alignment as well. The amino acid sequences FIVDDDG (region A) and VVYGDTD (region I) specified by the degenerate primers pdRA and mdR1 were not included in the alignment; a total of 284 amino acid characters were utilized in this analysis.

Using Felsenstein's PHYLIP program package (version 3.572), phylogenetic trees were constructed by the neighbor-joining method (Saitou and Nei 1987; Felsenstein 1989). Specifically, amino acid sequence alignments were converted to the PHYLIP input format using ClustalW v1.6 on an Apple Macintosh Performa 6300 (Thompson et al. 1994). One hundred bootstrap samples were generated using the SEQBOOT program. Using the PROTDIST program, the bootstrapped alignments were converted into distance matrices corrected for multiple substitutions by the Dayhoff-PAM method (Dayhoff 1979). Positions with gaps were excluded from the analysis. The resulting matrices were entered into the NEIGHBOR program and used to construct phylogenetic trees by the neighbor-joining method of Saitou and Nei (1987). Lastly, those 100 trees were reduced to the majority-rule consensus tree using the CONSENSE program. Phylogenetic trees constructed with PHYLIP were drawn on a Performa 6300 using the program TreeView (Page 1996).

Results and Discussion

The Macronuclear Gene Encoding the Catalytic Subunit of DNA Polymerase α in Hypotrichous Ciliates

The DNA pol α gene has been detected in all 13 hypotrich species examined thus far (Fig. 1; Mansour et al. 1994). Macronuclear gene-sized molecules encoding DNA pol α in various species range from ~4.8- to ~5.2 kb in length. The observed size variation is not unexpected because family B DNA polymerase polypeptides are generally composed of multiple conserved regions separated by variable length ''spacer'' sequences (Wong et al. 1988; Kornberg and Baker 1992; Braithwaite and Ito 1993). In addition, in hypotrich macronuclear genes the 5′ nontranslated leader or 3′ nontranslated trailer sequences may vary considerably in length among species (Hoffman et al. 1995).

The Macronuclear DNA pol α Gene in Two Oxytricha *Species*

The complete nucleotide sequence of the macronuclear DNA pol α gene has previously been determined for *Oxytricha nova* (Genbank accession number U02001; Mansour et al. 1994). The *O. trifallax* DNA pol α macronuclear clone was isolated from a genomic library using the *O. nova* gene as a probe. A single λ clone was purified by ultracentrifugation in a linear CsCl gradient, subcloned, sequenced, and shown to contain an insert of 4952 bp, not including 20 bp of telomeric repeats at the 5′ end and 28 bp at the 3′ end of the insert (data not shown). The presence of telomeres indicates that the λ clone contained an intact gene-sized molecule. The complete nucleotide sequence has been deposited in the GenBank sequence data base, accession number U59426 (Hoffman 1996).

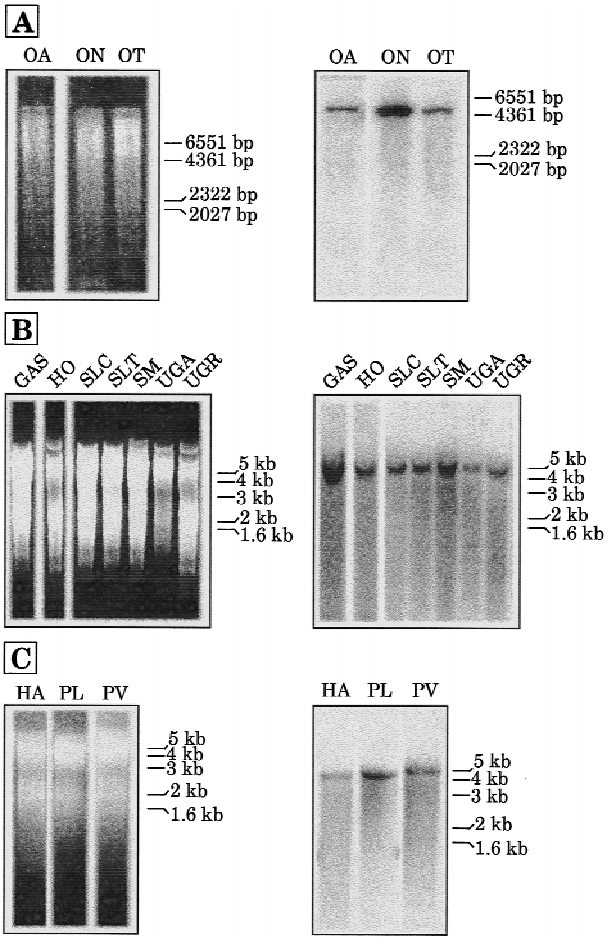

Fig. 1. **A** An ethidium-stained agarose gel of macronuclear DNA from three *Oxytricha* species and the corresponding Southern blot. The blot was probed with the full-length macronuclear DNA pol α gene from *O. nova,* and washed to a final stringency of 0.2× SSC/0.1% SDS at 45°C. OA = *O. sp. Aspen;* ON = *O. nova;* OTB = *O. trifallax.* **B** An ethidium-stained agarose gel of macronuclear DNA from seven hypotrich species and the corresponding Southern blot. The blot was probed with a 2.7-kb carboxy-terminal *EcoR* I restriction fragment of the macronuclear DNA pol α gene from *O. trifallax,* and washed to a final stringency of 0.1× SSC/0.1% SDS at 52°C. GAS = *Gastrostyla* sp.; HO = *Holosticha* sp.; SLC = *Stylonychia lemnae* CSH; SLT = *S. lemnae* Teller; UGA = *Uroleptus gallina;* UGA = *Urostyla grandis.* **C** An ethidium-stained agarose gel of macronuclear DNA from two hypotrich and a single oligotrich species and the corresponding Southern blot. The blot was probed with a 2.7-kb carboxy-terminal *EcoR* I restriction fragment of the macronuclear DNA pol α gene from *O. trifallax,* and washed to a final stringency of 0.1× SSC/0.1% SDS at 45°C. HA = *Halteria grandinella* (the oligotrich); PL = *Pleurotricha lanceolata;* PV = *Paraurostyla viridis.*

The gene-sized molecule in *O. trifallax* consists of a 5′ nontranslated leader (LDR) of 255 bp, an open reading frame (ORF) of 4542 bp (including the TGA codon), and a 3′ nontranslated trailer (TLR) of 162 bp. The 5′ LDR is 65% A+T, and lacks conventional eukaryotic promoter sequences; the 3′ TLR is 77% A+T, and lacks the eukaryotic poly(A) addition signal. The 4542-bp ORF is 64% A+T, contains no introns, and specifies a putative polypeptide of 1513 amino acids (≈173 kilodaltons), which contains the 12 conserved domains found in both prokaryotic and eukaryotic family B DNA polymerases (data not shown; Wong et al. 1988; Damagnez et al. 1991; Braithwaite and Ito 1993; White et al. 1993). The ORF contains 35 TAA and 16 TAG codons, which specify glutamine in most hypotrichs, rather than translation termination (reviewed in Hoffman et al. 1995).

A Comparison of the O. nova *and* O. trifallax *DNA pol* α *Polypeptides*

Overall, the amino acid sequences of the two proteins are ≈68% identical; the amino terminal ≈350 amino acids extending from the initiator methionine to conserved region E diverge considerably (≈46% identity) compared to the remaining ≈1150 amino acids (≈72% identity; data not shown). The 12 conserved domains are separated by ''spacers'' of variable sequence and length; some ''spacers'' are more variable than others. It is possible to define a core catalytic domain extending from region E through region V, in which the *O. nova* and *O. trifallax* polypeptides are 80% identical. This domain separates a highly variable amino-terminal domain (46% identical) and a less variable carboxy-terminal domain (63% identical).

Another feature conserved between the *O. nova* and *O. trifallax* DNA pol α proteins is a cysteine-rich domain near the carboxy-terminus. The domain encompasses amino acids 1314–1398 in *O. nova* and amino acids 1344–1428 in *O. trifallax* (Mansour et al. 1994). The spacing between cysteine residues is C-X_2-C-X_{23}-C-X_2-C-X_{30}-C-X_4-C-X_{12}-C-X_4-C, where X represents any amino acid except cysteine. This arrangement closely matches that of the *Sa. cerevisiae* (C-X_2-C-X_{23}-C-X_2-C-X_{30}-C-X_4-C-X_{13}-C-X_4-C) and the human (C-X_2-C-X_{23}-C-X_4-C-X_{32}-C-X_4-C-X_{17}-C-X_2-C) proteins (Wang et al. 1989). The C-X_2-C-X_{23}-C-X_2-C and C-X_4-C-X_{12}-C-X_4-C motifs are capable of forming a tetrahedral box structure with an extended protein loop characteristic of Zn-finger DNA binding proteins (Wang et al. 1989; Mansour et al. 1994).

Generation of Partial Macronuclear DNA pol α *Clones with PCR*

Degenerate oligonucleotide primers (Preston 1993) were used to amplify specific, conserved amino acid sequences by the polymerase chain reaction (PCR). Primers were synthesized to account for all possible permutations of nucleotide sequence that specify particular amino acid sequences, and were designed to amplify the two most highly conserved amino acid sequences in family B DNA polymerases (designated regions I and II in Wang et al. 1989; Braithwaite and Ito 1993). Primer pdR2 is 128-fold degenerate and amplifies all nucleotide sequences specifying the amino acid sequence DFNSLYP in conserved region II (Fig. 2; Wang et al.

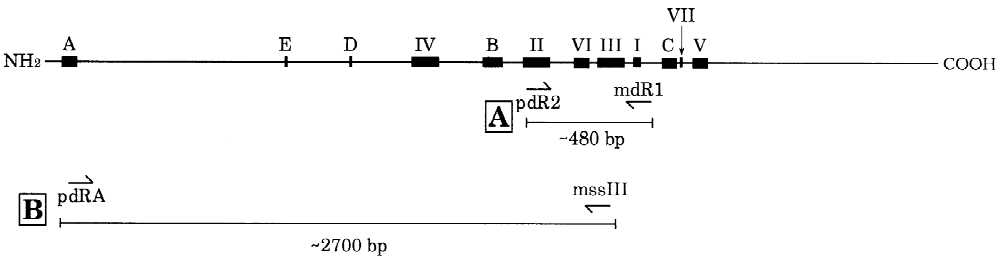

Fig. 2. Diagram of the DNA pol α polypeptide and the arrangement of conserved domains along the protein. Conserved domains are indicated by black rectangles. A degenerate PCR and genomic walking strategy was used to obtain partial DNA pol α clones from eight ciliate species. The PCR product labeled **A** was obtained first using degenerate primers designed to amplify all possible nucleotide sequences encoding the highly conserved amino acid sequences DFNSLYP (from region II) and VVYGDTD (from region I). After sequencing the short clones, a walking strategy was devised using a third degenerate primer designed to amplify all possible nucleotide sequences encoding the amino acid sequence FIVDDDG (from region **A**), and used with a specific nucleotide sequence primer designed to amplify sequences from region III (contained within PCR product **A**). This primer pair amplified the PCR product labeled **B** from total macronuclear DNA. pdR2 = plus strand degenerate primer for region II; mdR1 = minus strand degenerate primer for region I; pdRA = plus strand degenerate primer for region A; mssIII = minus strand specific sequence primer for region III.

1989). Primer mdR1 is fourfold degenerate and amplifies all nucleotide sequences specifying the amino acid sequence VVYGDTD in conserved region I (Fig. 2; Wang et al. 1989). The latter sequence is a portion of the DNA polymerase family B signature that has been observed in higher eukaryotic DNA polymerases α and δ, yeast DNA polymerases I, II, and III, *E. coli* DNA polymerase II, and herpes-, adeno-, and baculovirus DNA polymerases (Wong et al. 1988; Wang et al., 1989; Braithwaite and Ito 1993).

After sequencing PCR products from the macronuclear DNA pol α gene generated with the primer pair pdR2/mdR1, larger DNA pol α clones were obtained from eight species by genomic walking. Using a specific primer designed to amplify sequences from conserved region III (contained within the pdR2/mdR1 PCR product) and a third degenerate primer, PCR products containing approximately two-thirds of the DNA pol α open reading frame were amplified (Fig. 2). Primer pdRA (48-fold degenerate) was designed to amplify all nucleotide sequences specifying the amino acid sequence FIVDDDG in conserved region A (Damagnez et al. 1991; White et al. 1993). Figure 3a shows an ethidium-stained agarose gel of the long PCR product from eight hypotrich species, and Figure 3b shows the corresponding Southern blot probed with the *O. nova* macronuclear gene. Because of the large size of these PCR products, they were not sequenced directly, but were cloned first. Of the eight species for which large PCR clones were obtained, only the *Paraurostyla viridis* and *Urol. gallina* PCR products were cloned intact. PCR products for the remaining species were cloned in either two (*Halteria grandinella, Holosticha* sp., *Pleurotricha lanceolata, Sty. mytilus,* and *Uros. grandis*) or three (*Sty. lemnae* T) pieces. Specific sequence PCR primers were designed to amplify a product across the restriction sites used for cloning to test whether any small restriction fragments were missed during cloning. In every case, the short PCR products contained only a single restriction site for the appropriate enzyme (data not shown).

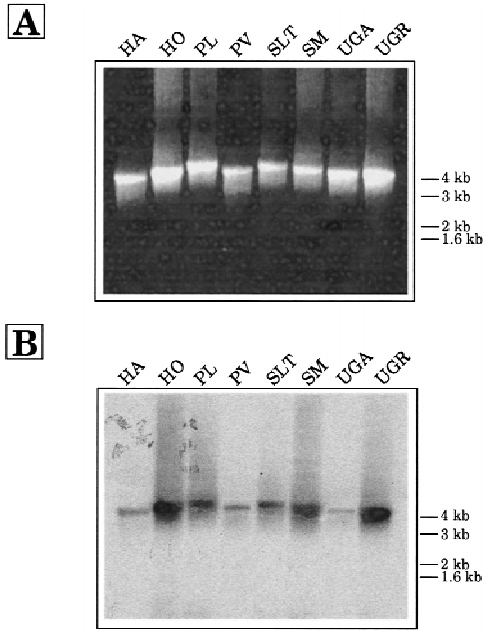

Fig. 3. **A** An ethidium-stained agarose gel with partial DNA pol α clones obtained using degenerate PCR as described in Figure 3. **B** Southern blot of the gel in **A** probed with a 2.7-kb, carboxy-terminal *EcoR* I restriction fragment of *O. trifallax* DNA pol α, and washed to a final stringency of 0.1 × SSC/0.1% SDS at 60°C. HA = *Halteria grandinella;* HO = *Holosticha* sp.; PL = *Pleurotricha lanceolata;* PV = *Paraurostyla viridis;* SLT = *Stylonychia lemnae* T; SM = *Stylonychia mytilus;* UGA = *Uroleptus gallina;* UGR = *Urostyla grandis.*

Figure 4 contains an alignment of conserved amino acid sequences present in the long PCR products (regions I–IV, VI, A, B, D, and E) obtained by the strategy outlined above and in Figure 2. Also included are 11 newly discovered conserved amino acid sequences identified in this analysis. Ten hypotrich species are included, as well as the corresponding sequences of DNA pol α proteins from human, *Mus musculus, Drosophila melanogaster, Sa. cerevisiae,* and *Sc. pombe.*

Region A is moderately conserved among all species, although the last four amino acids have diverged in the

```
                    A           E                                                            D
O. nova           FIVDDDGYGYRDHG VFGKI VSCSIKVN FVLP KYAFE GNTFECIFGANQSMLESFILKRKIR GPCWM
O. trifallax      .............. I.... ....M... YA.. ..... .........ST.....L........ .....
Pl. lanceolata    .............. I.... ........ .A.. ..... .........C...I..L.......K .....
Ha. grandinella   .......L...... I...V ....L... YA.. ..... .......L.....L..L.......K .....
Pa. viridis       .............. I.... ...AL... YA.. ..... ..S...L.........LL......K ....I
Sty. lemnae T     .......F...... I.... ....M... YA.. ..... .......L..C.....L.......K .....
Sty. mytilus      .......F...... I.... ....M... YA.. ..... ....D-VI.L......L......LK .....
Urol. gallina     .............F. I...V ...AL.I. .A.. ..... ................L.....R.K .....
Holosticha sp.    .............N. ....V ....M.I. YA.. ..... .......L..T.....L.......K .....
Uros. grandis     .......I..KEK. L...V ....L... YA.. ..... .S....L.NTQ.....LLL..Q..K ....I
human             W......I..VED. L...V ...CVM.K YF.. N.... .E..SHV..T.T.S..L.LMN...K ....L
M. musculus       W......I..VED. L...V ...CVM.K YF.. N.... .E..SHV..T.T.S..L.LMN...K ....L
D. melanogaster   W.-EE..T..AEDL L..R- K.VCLR.Q YL.. SF.HH Y.SIAH....TTNA..R.L.D...K ....L
Sa. cerevisiae    .V.....V..V.R. L...V ..AMVQI. .F.. ..S.. SD..YHV..G.SNIF...V.QNR.M ....L
Sc. pombe         .V...N.A..V.N. L...V ...FL..D YF.. ..... .SS.SHV..T.TALF.Q.V.SRRVM ....L

                                            IV
O. nova           PPLT IQFFQHERQLIEAFVAKIYQLDPDLMVAHNLCGGMFDLLLAR WSRIGRL TCGRL TAKELIRETNYDL
O. trifallax      .... ..K.EN...M....I..VFIV....V................ ....... ..... S....V...S...
Pl. lanceolata    .N.. .TY.MQ...M....IS.V..I....I........I....... ....... ..... S............
Ha. grandinella   .A.S VIY.PS.K......IN.VT.I....LI.......L.E..... ....... ..... .............
Pa. viridis       ...S L...TN...M.D..INRLTVV....L.........E...... ......M HAVD- .............
Sty. lemnae T     .... ..Y.TN...M.DL..TR.F.I....I........I....... ......F ..... S....V...S...
Sty. mytilus      .... ..Y.TN...M.DL..TR.F.I....I........I...K... ......F ..... S............
Urol. gallina     .... ..Y.TN...M....ISRLHLI....LI.......I....... ......F ..... S............
Holosticha sp.    .... ..Y..N...M....IS.LFLI....L................ ......F ..... S............
Uros. grandis     .... .FT..N.KA.L.N.ISRLHLI...MLI.......ILE..... ...L... S.... S........S...
human             ...V VEVAAT..T.LGF.L..VHKI...II.G..IY.FELEV..Q. ..K.... ....M S......CKS.H.
M. musculus       ...V VEIAAT..T..GF.L..VHKI...IL.G..I.SFELEV..Q. ..K.... ....M S.....HCKS.H.
D. melanogaster   .... VHKHDS..A.LSW.L.QYQKI.A..I.TFDSMDCQLNVITDQ ...M... FV..M S.E.C..ARS...
Sa. cerevisiae    .N.R VRL.NN.KAMLSC.C.MLKVE...VIIG.R.QNVYL.V.AH. F.S...R CS... GQSLTPKCQSW..
Sc. pombe         ..M. FCE-RS.VS.LNN.LN.VRTY...VYFG.DFE-MCYSV..S. ..S.... IA... GRSMI-KAQSWS.

                                        B
O. nova           PLTKQLTNI AGNLWFRSLQNARAERNEMLLLHEFKKKKFVLPDKK RKKAA
O. trifallax      ...L..... ..............................I..... .....
Pl. lanceolata    ......... ...M.....R....................L..... .....
Ha. grandinella   ......... ..............................I..... .....
Pa. viridis       ......... ..............................I..... ....Q
Sty. lemnae T     ..A...... ..............................I..... ....K
Sty. mytilus      ......... ..............................I..... ....K
Urol. gallina     ......... ...........VC.................I..... ....Q
Holosticha sp.    ......... .................EI...........I..... .....
Uros. grandis     .......S. .....Y...................YG..YLC..-. ....Q
human             ..AL.I... ...IMS.T.MGG.S....F....A.YENNYIV...Q ...G.
M. musculus       ..AL.I... ...IMS.T.MGG.S....F....A.YENNYIV...Q ....T
D. melanogaster   ..AL.I... C..TMT.T..GG.S....F....ASTE.NYIV...K K....
Sa. cerevisiae    T.......L ...A.AQT.GGT..G...YI.....SRNG.IV...E -...K
Sc. pombe         Q.S.N.... ...S.A.T.TGT......YI.......NGYIV...Q -..DK
```

Fig. 4. An alignment of conserved amino acid sequences present in the long PCR products (regions I–IV, VI, A, B, D, and E) obtained by the strategy described in Figure 2. Also included are 11 newly identified amino acid sequences identified in this analysis. Ten hypotrich species are included, as well as the corresponding sequences of DNA pol α proteins from human, *M. musculus, D. melanogaster, Sa. cerevisiae,* and *Sc. pombe.* Gaps introduced during sequence alignment are indicated with dashes (−); amino acids at any position that are the same as in the *O. nova* gene are indicated with dots (.). Complete sequences have been submitted to the GenBank sequence data base under the accession numbers U02001 (*O. nova*), U59426 (*O. trifallax*), U89699 (*Halteria grandinella*), U89700 (*Holosticha* species), U89701 (*Paraurostyla viridis*), U89702 (*Pleurotricha lanceolata*), U89703 (*Stylonychia lemnae*), U89704 (*Stylonychia mytilus*), U89705 (*Uroleptus gallina*), and U89706 (*Urostyla grandis*).

nonciliate species. Region E is five amino acids long in all species except for *D. melanogaster.* Based upon this alignment, it contains a core sequence of three amino acids (FGK) that is almost absolutely conserved, with *Drosophila* the lone exception. The first and last amino acids in region E can vary, although all observed substitutions are conservative: valine for isoleucine (and vice versa) and leucine for isoleucine. The first four positions of region D, including the essential G are exactly conserved in all 15 organisms in Figure 4. Two substitutions have occurred in the last position: isoleucine for leucine in *Pa. viridis,* and leucine for methionine in all five nonciliate species.

Region IV is 42 amino acids long and is centered around the nearly invariant sequence DPD, a motif resembling one absolutely required for DNA pol α activity in vitro which is present in conserved region I (DTD; Copeland and Wang 1993; Copeland et al. 1993). Al-

```
                                    II                                  VI
O. nova             YAGGLVIEPKAGFYDNIILLLDFNSLYPSIIQEYNLCFTTV LPD AVLPMVLRDLVQKRKAV
O. trifallax        ..................................... ... ....N.............
Pl. lanceolata      ..................................... ... ....N.........R..
Ha. grandinella     ..............S...................... ... ....N.............
Pa. viridis         ..................................... ... ....N.........R..
Sty. lemnae T       ................V..................... ... ....N.........R..
Sty. mytilus        ..................................... ... ....N.............
Urol. gallina       .S.....D...R..................VL..... ... ....N.........R..
Holosticha sp.      ..................................... ... ....N.............
Uros. grandis       .S............T...................... ... .I..N..KN......I.
human               ......LD..V....KF..............F.I..... ... GI..REI.K..ER..Q.
M. musculus         ......LD..V....KF..............F.I..... ... GI..REI.K..ER..Q.
D. melanogaster     ......L..MR.L.EKYV..M.L...........I..NP. ... GI..LQ.KR..ES..E.
Sa. cerevisiae      .Q....F..EK.LHKNYV.VM...........F.I..... V.P G...RL.AN..DR.RE.
Sc. pombe           .K....F..QK.LYETC..VM............I..... T.S GIF.RLIAN..ER.RQI

                                   III                                   I
O. nova             EIRQKAIKLTANSMYGCLGFGSSRFHA-QAIAALITRTGRETL VVYGDTDS
O. trifallax        ............................-........K...D.. ........
Pl. lanceolata      ................
Ha. grandinella     ..................S......-K.......K...N.. ....
Pa. viridis         ..................Q......-K.............. .....
Sty. lemnae T       .N................S..K...-S.......K...... ....
Sty. mytilus        ..................S......-S.
Urol. gallina       ........................-K.............. .....
Holosticha sp.      ..................S....Y.FRR.........S..
Uros. grandis       ..................S.......-K.......K...... ....
human               D.....L............SY...Y.-KPL...V.YK...I. .I......
M. musculus         D.....L............SY...Y.-KPL...V.YK...I. .I......
D. melanogaster     H...M.L............AH...F.-.HL...V.HK..D-. ........
Sa. cerevisiae      D...Q.L............YVN...Y.-KPL.M.V.NK...I. ........
Sc. pombe           D.Q.Q.L............YTK...Y.-RPL.V...YK...A. .I......
```

Fig. 4. Continued.

though the sequence is generally well conserved among the hypotrichs, the ciliate sequence bears little resemblance to that of the other organisms; only ≈12 positions (≈29%) are well conserved in both groups.

Region B is 36 amino acids long. It is highly conserved in the hypotrichs, although only short sections of the sequence are conserved between ciliates and other eukaryotes. Region II is 41 amino acids long and centered around the invariant core sequence DFNSLYPSII. The remaining sequence is also well conserved. Many of the substitutions are conservative: leucine for isoleucine, valine for alanine or leucine, phenylalanine for tyrosine, and isoleucine for leucine. Region II has been implicated in binding of dNTP and primer annealed to template (Dong et al. 1993a,b).

Region VI was identified previously as the region with the lowest amount of sequence conservation among eukaryotic, viral, and bacteriophage DNA pol α genes (Wong et al. 1988, Wang et al. 1989). However, among the hypotrichs region VI is highly conserved, with amino acid substitutions at only two of 16 positions, except for *Uros. grandis*, in which four positions have been altered. However, the sequence has diverged in the hypotrichs compared to the other eukaryotes, with substitutions occurring at 12 of 16 positions (75%). Region III, which is centered around the invariant core sequence of KLTANSMYGCLG is highly conserved in all species examined.

The alignment in Figure 4 clearly shows that the hy-

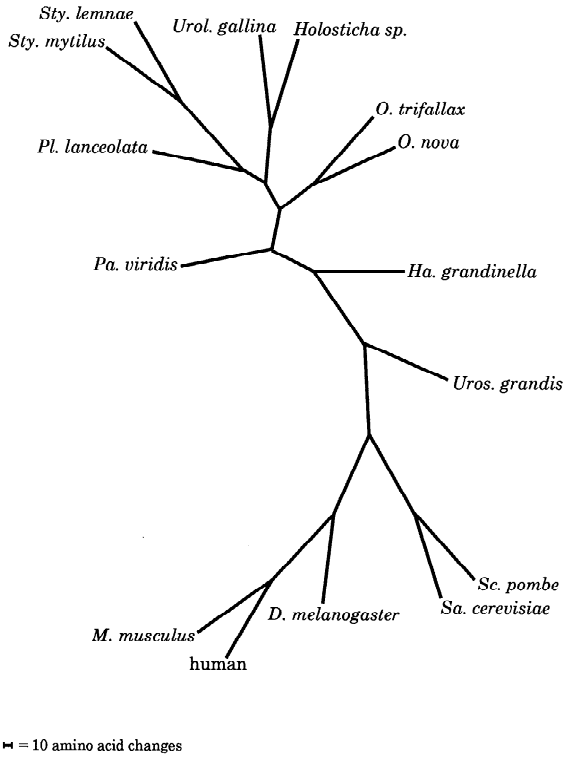

Fig. 5. Neighbor-joining tree inferred from the amino acid sequence alignment in Figure 4.

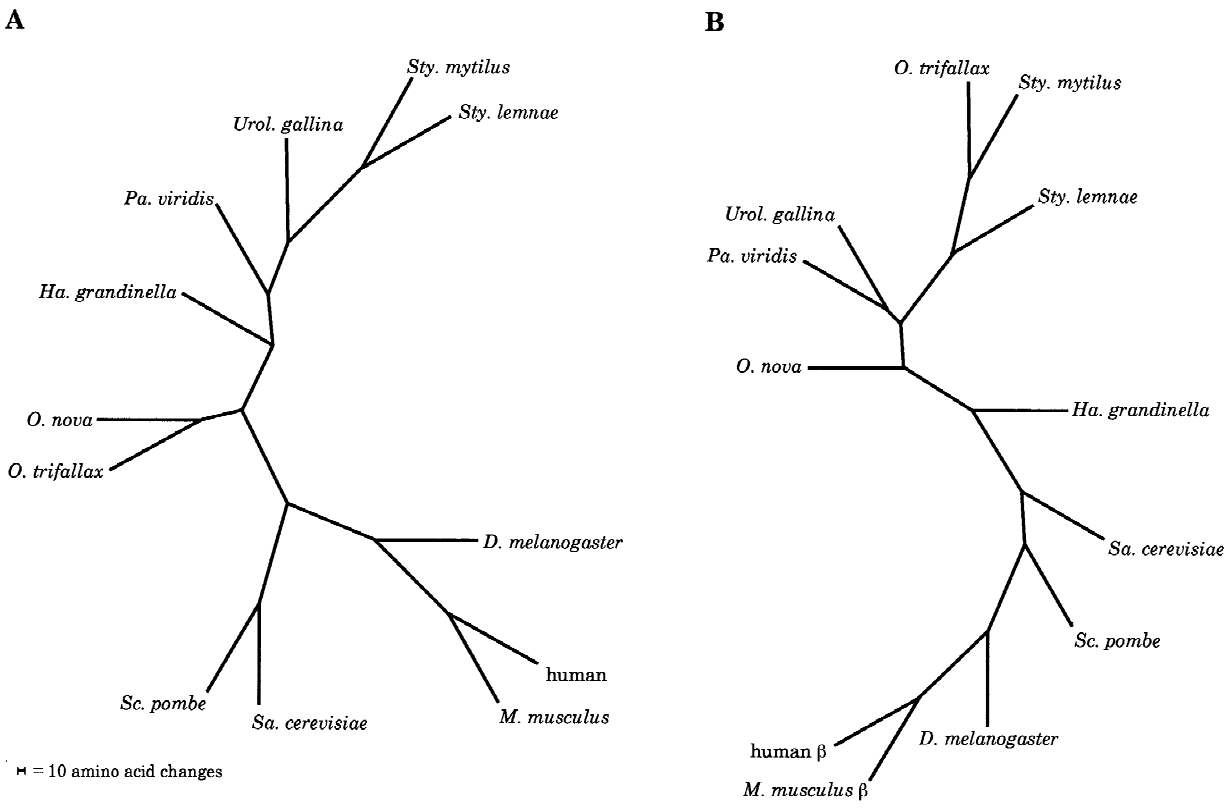

Fig. 6. **A** Neighbor-joining tree inferred from a subset of the DNA pol α sequences aligned in Figure 4, and including the same outgroups. **B** Neighbor-joining tree inferred from actin I sequences of the same species included in **A.**

potrich species analyzed are rather closely related, with the exception of *Uros. grandis.* Alignment of the full-length sequences identified several sequences conserved only among the hypotrichs (e.g., the ~130 amino acids between regions E and D, the ~110 amino acids between regions D and IV, and the ~120 amino acids between regions IV and D; data not shown). Although it is not clear what role these sequences play in DNA pol α function, given the fundamentally different organization of macronuclear DNA synthesis into replication bands, it is possible that these amino acid sequences are required to establish or maintain interactions between DNA pol α and other components of the replication band. Most of the protein constituents of the replication band have not been characterized biochemically, although antibody staining has demonstrated localization of both PCNA/cyclin and acetylated histone H4 in the replication band in *Euplotes eurystomus* (Olins et al. 1989; Olins and Olins 1994).

Phylogenetic Analysis

Alignment of amino acid sequences was done with the PILEUP function of the GCG sequence analysis package (GCG 1994). Pileup multiple sequence format amino acid alignments were then used to infer phylogenetic trees by the neighbor-joining method of Saitou and Nei (1987), using the NEIGHBOR program of the PHYLIP sequence analysis package (v. 3.572) for the Macintosh PowerPC (Felsenstein 1989). One hundred bootstrap data sets were converted to distance matrices corrected for multiple substitutions by the Dayhoff-PAM method using the PROTDIST program of PHYLIP (Dayhoff 1979; Felsenstein 1989). Trees were constructed using the NEIGHBOR program of PHYLIP, and the majority-rule consensus tree was identified using the CONSENSE program of PHYLIP (Felsenstein 1989).

Figure 5 contains a neighbor-joining tree inferred from the amino acid sequence alignment in Figure 4, including human, *M. musculus, D. melanogaster, Sa. cerevisiae,* and *Sc. pombe* as outgroups. Previous analyses have concluded that the major groups within the ciliates are widely separated based upon divergence of rRNA sequences, in accordance with classical systematics, and that each group is essentially monophyletic (Lynn and Sogin 1988; Baroin-Tourancheau et al. 1992). Comparison of DNA pol α amino acid sequences confirms this result, with the possible exception of *Urostyla grandis.*

An attempt to clarify taxonomic relationships among the hypotrichs with protein sequences was made previously with the actin I gene (DuBois 1995). Figure 6A contains a neighbor-joining tree inferred from DNA pol α amino acid sequences of seven of the hypotrich species

aligned in Figure 4 compared to a similar tree inferred from actin I amino acid sequences obtained from the same seven species (Figure 6B), using the same species as outgroups. DNA pol α sequences from *Holosticha* sp., *Pleurotricha lanceolata*, and *Urostyla grandis* were omitted from Figure 6A because actin I has not been sequenced in those species. Comparison of the two trees reveals two notable differences. First, based upon actin I sequences, *Halteria grandinella*, which is classified as an oligotrich by morphological criteria, has diverged widely from the hypotrichs, while DNA pol α sequences place it unambiguously with the hypotrichs. Second, actin I sequences indicate that *O. trifallax* is most closely related to *Sty. mytilus* within the main grouping that also contains *Sty. lemnae*, *Uroleptus*, and *Paraurostyla.* In contrast, DNA pol α sequences indicate that *O. trifallax* is most closely related to *O. nova*, and that both are separated from the two remaining groupings of *Uroleptus* and *Stylonychia*, and *Paraurostyla* and *Halteria.*

Phylogenetic comparison using molecular sequence data from DNA pol α has resolved two questions raised by classical systematic analyses of the hypotrichs. Based upon morphological criteria, *Oxytricha nova* and *O. trifallax* have been considered to be the same species (W. Foissner, personal communication). The molecular data from both actin I and DNA pol α contradict this assertion; clearly *O. nova* and *O. trifallax* are different species (this report and DuBois 1995). *Halteria grandinella* has been previously classified as an oligotrich based upon cell cortex morphology (Corliss 1979). We propose that *H. grandinella* be reclassified as a hypotrich based upon three criteria: (1) It contains macronuclear DNA as gene-sized molecules. (2) Macronuclear DNA replication occurs via replication bands. (3) Molecular data from DNA pol α, and small and large subunit rRNA sequences place it among the hypotrichs (Lynn and Sogin 1988).

Although DNA pol α sequences have enabled the resolution of two issues raised in the above paragraph, the differences between the relationships described by the actin I and DNA pol α sequence data emphasize the importance of analyzing multiple genes in order to establish evolutionary relationships with security. We are continuing the study of phylogenetic relationships among the hypotrichs using the genes encoding the α and β subunits of the telomere binding protein heterodimer.

Acknowledgments. The authors would like to thank Robin Gutell for advice on phylogenetic methods, and Don Burke for a critical reading of the manuscript. This work was supported by NSF grant no. MCB9506287 to D.M.P.

References

Araki H, Ropp PA, Johnson AL, Johnston LH, Morrison A, Sugino A (1992) DNA polymerase II, the probable homolog of mammalian DNA polymerase ϵ, replicated chromosomal DNA in the yeast *Saccharomyces cerevisiae.* EMBO J 11:733–740

Asai DJ, Criswell PG (1995) Identification of new dynein heavy-chain genes by RNA-directed polymerase chain reaction. Methods Cell Biol 47:579–585

Baroin-Tourancheau A, Delgado P, Perasso R, Adoutte A (1992) A broad molecular phylogeny of ciliates: identification of major evolutionary trends and radiations within the phylum. Proc Natl Acad Sci USA 89:9764–9768

Blow J (1989) Eukaryotic chromosome replication requires both α and δ DNA polymerases. Trends Genet 5:134–136

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

Copeland WC, Wang TS-F (1993) Mutational analysis of the human DNA polymerase α: the most conserved region in α-like DNA polymerases is involved in metal-specific catalysis. J Biol Chem 268:11028–11040

Copeland WC, Lam NK, Wang TS-F (1993) Fidelity studies of the human DNA polymerase α: the most conserved region among α-like DNA polymerases is responsible for metal-induced infidelity in DNA synthesis. J Biol Chem 268:11041–11049

Corliss JO (1979) The ciliated protozoa—characterization, classification, and guide to the literature. Pergamon Press, New York, New York

Cotterill SM, Reyland ME, Loeb LA, Lehman IR (1987) A cryptic proofreading 3′-5′ exonuclease associated with the polymerase subunit of the DNA polymerase-primase from *Drosophila melanogaster.* Proc Natl Acad Sci USA 84:5635–5639

Damagnez V, Tillit J, de Recondo A-M, Baldacci G (1991) The *POL1* gene from the fission yeast, *Schizosaccharomyces pombe*, shows conserved amino acid blocks specific for eukaryotic DNA polymerase α. Mol Gen Genet 226:182–189

Dayhoff MO (1979) Atlas of Protein Sequence and Structure, Volume 5, Supplement 3. National Biomedical Research Foundation, Washington, DC

Dong Q, Copeland WC, Wang TS-F (1993a) Mutational studies of human DNA polymerase α: identification of residues critical for deoxynucleotide binding and misinsertion fidelity. J Biol Chem 268:24163–24174

Dong Q, Copeland WC, Wang TS-F (1993b) Mutational studies of human DNA polymerase α: serine 867 in the second most conserved region among α-like DNA polymerases is involved in primer binding and mispair primer extension. J Biol Chem 268: 24175–24182

Dong Q, Wang TS-F (1995) Mutational studies of human DNA polymerase α: lysine 950 in the third most conserved region of α-like DNA polymerases is involved in binding the deoxynucleotide triphosphate. J Biol Chem 270:21563–21570

DuBois ML (1995) The unusual structure and evolution of the germline genome of hypotrichous ciliates. PhD thesis, University of Colorado, Boulder

Elwood HJ, Olsen GJ, Sogin ML (1985) The small-subunit ribosomal RNA gene sequences from the hypotrichous ciliates *Oxytricha nova* and *Stylonychia pustulata.* Mol Biol Evol 2:399–410

Felsenstein J (1989) PHYLIP—Phylogeny Inference Package (Version 3.2). Cladistics 5:164–166

GCG (1994) Program manual for the Wisconsin package, 8th edition, Genetics Computer Group, Madison, Wisconsin

Gibbs JS, Chiou HC, Bastow KF, Cheng Y-C, Coen DM (1988) Identification of amino acids in herpes simplex virus DNA polymerase involved in substrate and drug recognition. Proc Natl Acad Sci USA 85:6672–6676

Hanahan D (1983) Studies on transformation of *Escherichia coli* with plasmids. J Mol Biol 166:557–580

Hoffman DC (1996) The scrambled gene encoding the large, catalytic subunit of DNA polymerase α in the germline and somatic genomes of hypotrichous ciliates. PhD thesis, University of Colorado, Boulder

Hoffman DC, Anderson RC, DuBois ML, Prescott DM (1995) Mac-

ronuclear gene-sized molecules of hypotrichs. Nucl Acids Res 23: 1279–1283

Kaguni LS, Rossignol J-M, Conaway RC, Banks GR, Lehman IR (1983) Association of DNA primase with the β/γ subunits of DNA polymerase α from *Drosophila melanogaster* embryos. J Biol Chem 258:9037–9039

Kornberg A, Baker TA (1992) DNA replication. WH Freeman and Company, New York, New York, pp 197–223

Lehman IR, Kaguni LS (1989) DNA polymerase α. J Biol Chem 264: 4265–4268

Lucchini G, Francesconi S, Foiani M, Badaracco G, Plevani P (1987) Yeast DNA polymerase-DNA primase complex: cloning of PRI1, a single essential gene related to DNA primase activity. EMBO J 6:737–742

Lynn DH, Corliss JO (1991) Ciliophora. In: Harrison, FW (ed) Microscopic anatomy of invertebrates. John Wiley & Sons, New York, New York, pp 333–467

Lynn DH, Sogin ML (1988) Assessment of phylogenetic relationships among ciliated protists using partial ribosomal RNA sequences derived from reverse transcripts. Biosystems 21:249–254

Mansour SJ, Hoffman DC, Prescott DM (1994) A gene-sized molecule encoding the catalytic subunit of DNA polymerase α in the macronucleus of *Oxytricha nova.* Gene 144:155–161

Olins DE, Olins AL, Cacheiro LH, Tan EM (1989) Proliferating cell nuclear antigen/cyclin in the ciliate *Euplotes eurystomus:* localization in the replication band and in micronuclei. J Cell Biol 109: 1399–1410

Olins DE, Olins AL (1994) The replication band of ciliated protozoa. Int Rev Cytol 153:137–170

Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. Comp Appl Biosc, in press

Pizzagalli A, Valsasnini P, Plevani P, Lucchini G (1988) DNA polymerase I gene of *Saccharomyces cerevisiae:* nucleotide sequence, mapping of a temperature-sensitive mutation, and protein homology with other DNA polymerases. Proc Natl Acad Sci USA 85: 3772–3776

Plevani P, Foiani M, Valsasnini P, Badaracco G, Cheriathundam E, Chang LMS (1985) Polypeptide structure of DNA primase from a yeast DNA polymerase-primase complex. J Biol Chem 260:7102–7107

Pope B, Kent HM (1996) High efficiency 5 minute transformation of *Escherichia coli.* Nucl Acids Res 24:536–537

Prescott DM, Greslin AF (1992) The scrambled actin I gene in the micronucleus of *Oxytricha nova.* Dev Genet 13:66–74

Preston GM (1993) Use of degenerate oligonucleotide primers and the polymerase chain reaction to clone gene family members. In: White, BA (ed) Methods in molecular biology, Vol 15: PCR protocols: current methods and applications. Humana Press, Totowa, New Jersey, pp 317–337

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning—a laboratory manual, 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York

Schlegel M, Elwood HJ, Sogin ML (1991) Molecular evolution in hypotrichous ciliates: sequence of the small subunit ribosomal RNA genes from *Onychodromus quadricornatus* and *Oxytricha granulifera.* J Mol Evol 32:64–69

So AG, Downey KM (1988) Mammalian DNA polymerases α and δ: current status in DNA replication. Biochemistry 27:4591–4595

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Wang TS-F, Wong SW, Korn D (1989) Human DNA polymerase α: predicted functional domains and relationships with viral DNA polymerases. FASEB J 3:14–21

White JH, Kilbey BJ, de Vries E, Goman M, Alano P, Cheesman S, McAleese S, Ridley RG (1993) The gene encoding DNA polymerase α from *Plasmodium falciparum.* Nucleic Acids Res 21:3643–3646

Wong SW, Wahl AF, Yuan P-M, Arai N, Pearson BE, Arai KI, Korn D, Hunkapiller MW, Wang TS-F (1988) Human DNA polymerase α gene expression is cell proliferation dependent and its primary structure is similar to both prokaryotic and eukaryotic replicative DNA polymerases. EMBO J 7:37–47