JOURNAL OF BACTERIOLOGY, Apr. 1997, p. 2632–2640 Vol. 179, No. 8
0021-9193/97/$04.00+0

Gene Duplications in Evolution of Archaeal Family B DNA Polymerases

DAVID R. EDGELL,* HANS-PETER KLENK,† AND W. FORD DOOLITTLE

Department of Biochemistry, Canadian Institute for Advanced Research, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7

Received 23 October 1996/Accepted 11 February 1997

All archaeal DNA-dependent DNA polymerases sequenced to date are homologous to family B DNA polymerases from eukaryotes and eubacteria. Presently, representatives of the euryarchaeote division of archaea appear to have a single family B DNA polymerase, whereas two crenarchaeotes, *Pyrodictium occultum* and *Sulfolobus solfataricus*, each possess two family B DNA polymerases. We have found the gene for yet a third family B DNA polymerase, designated B3, in the crenarchaeote *S. solfataricus* P2. The encoded protein is highly divergent at the amino acid level from the previously characterized family B polymerases in *S. solfataricus* P2 and contains a number of nonconserved amino acid substitutions in catalytic domains. We have cloned and sequenced the ortholog of this gene from the closely related *Sulfolobus shibatae*. It is also highly divergent from other archaeal family B DNA polymerases and, surprisingly, from the *S. solfataricus* B3 ortholog. Phylogenetic analysis using all available archaeal family B DNA polymerases suggests that the *S. solfataricus* P2 B3 and *S. shibatae* B3 paralogs are related to one of the two DNA polymerases of *P. occultum*. These sequences are members of a group which includes all euryarchaeote family B homologs, while the remaining crenarchaeote sequences form another distinct group. Archaeal family B DNA polymerases together constitute a monophyletic subfamily whose evolution has been characterized by a number of gene duplication events.

Studies on the mechanisms of DNA replication in eubacteria and eukaryotes have led to the identification of numerous DNA-dependent DNA polymerases (31). These have been classified into families based on amino acid sequence similarity of the catalytic subunit to one of the three *Escherichia coli* DNA polymerases (5). Family A DNA polymerases include *E. coli* DNA polymerase I (polI), all eubacterial polI homologs, some eubacterial phage DNA polymerases, and mitochondrial DNA polymerases (often called γ polymerase). Family B DNA polymerases include *E. coli* DNA polymerase II, some eubacterial phage DNA polymerases, the eukaryotic nuclear replicative DNA polymerases (α, δ, and ε), and eukaryotic viral and plasmid-borne enzymes. Family C includes only eubacterial polIII homologs: there are no known phage, viral, archaeal, or eukaryotic family C DNA polymerases. An additional eukaryotic nuclear encoded DNA polymerase, β, which functions in repair, is assigned to family X. Members of family X have little amino acid sequence similarity with DNA polymerases but instead exhibit amino acid sequence similarity to terminal transferases.

Although our understanding of DNA replication in eubacteria and eukaryotes is quite advanced, comparatively little is known about DNA replication in archaea. Early studies on DNA replication showed that aphidicolin, a specific inhibitor of eukaryotic DNA replication, inhibited cell growth and DNA synthesis in halophilic archaea (21, 43). Aphidicolin-sensitive DNA polymerases were subsequently purified from halophilic, methanogenic, and some thermophilic archaea, suggesting that, like eukaryotes and unlike eubacteria, archaea use an aphidicolin-sensitive DNA polymerase for DNA replication (summarized in reference 20). However, not all archaea demonstrate a sensitivity to aphidicolin, and aphidicolin-resistant DNA polymerases have recently been isolated and biochemically characterized from archaea that had been shown to also contain aphidicolin-sensitive DNA enzymes (14, 25, 29, 35, 48, 49). Sequencing of the genes for several aphidicolin-sensitive and one aphidicolin-resistant DNA polymerase revealed that these DNA polymerases are all homologous to family B DNA polymerases from eubacteria and eukaryotes (34, 38, 48, 49) and more similar at the amino acid level to eukaryotic homologs. This similarity does not prove a specific (sister) relationship between archaea and eukaryotes (19), as it is impossible to root phylogenetic trees of family B DNA polymerases.

Eukaryotic replicative polymerases (α, δ, and ε) are more similar to each other at the amino acid level than to either archaeal or eubacterial family B homologs (13a). These three DNA polymerases must be the products of gene duplication events which occurred at or before the origin of the eukaryotic nucleus. Since eukaryotic nuclear genomes are likely derived from the genomes of an archaea-like ancestor (13), the family B DNA polymerase complement of members of the domain *Archaea* is of particular evolutionary interest. To date, euryarchaeote genomes appear to encode only a single B-type enzyme (7), while the crenarchaeotes *Pyrodictium occultum* and *Sulfolobus solfataricus* show two or (as we present here) three B-type polymerases (38, 40, 49).

The third DNA polymerase from *S. solfataricus* P2 described here is divergent from other archaeal family B homologs and is missing a number of invariant amino acids in catalytic motifs. We have cloned and sequenced the ortholog of this gene from *Sulfolobus shibatae*, a closely related member of the crenarchaeote order *Sulfolobales*. The *S. shibatae* ortholog is also highly divergent from other archaeal family B DNA polymerases and from the *S. solfataricus* P2 DNA polymerase. Our phylogenetic analysis splits the archaeal family B DNA polymerases into two paralogous groups. Group I includes all euryarchaeote DNA polymerases, one of the two *P. occultum* DNA polymerases, and the new DNA polymerases from *S. solfataricus* and *S. shibatae*. Group II encompasses the remain-

* Corresponding author. Phone: (902) 494-2968. Fax: (902) 494-1355. E-mail: dedgell@cs.dal.ca.

† Present address: The Institute for Genomic Research, Rockville, MD 20850.

ing crenarchaeote DNA polymerases. This is consistent with the last common ancestor of archaea possessing multiple family B DNA polymerases, one of which was lost in evolution of the euryarchaeote lineage. However, due to extremely rapid rates of molecular evolution of the *S. solfataricus* P2 paralogs, the relationship between the two groups of DNA polymerases is poorly resolved and we cannot rule out the competing hypothesis that the common ancestor of archaea possessed a single family B DNA polymerase and the multiple DNA polymerases of crenarchaeotes arose by lineage-specific gene duplications.

MATERIALS AND METHODS

Definitions and gene nomenclature. We use the term paralog to refer to homologous DNA polymerases related by gene duplication and the term ortholog to refer to homologous DNA polymerases related by speciation. We propose to name crenarchaeote family B DNA polymerases on the basis of their relationship to one of the three *S. solfataricus* P2 DNA polymerases. DNA polymerases have been sequenced from two *S. solfataricus* strains, MT4 and P2. The letter B refers to family B DNA polymerase, and a number after the letter B refers to one of the three family B DNA polymerases of *S. solfataricus* P2. For instance, one of the two *P. occultum* DNA polymerases (referred to as DNA polymerase A in reference 49) is an ortholog of the *S. solfataricus* P2 B1 DNA polymerase. We call this polymerase *P. occultum* B1. The other *P. occultum* DNA polymerase (B in reference 49) is likely an ortholog of the *S. solfataricus* P2 B3 DNA polymerase. We call this polymerase *P. occultum* B3.

Strains and genomic DNAs. *E. coli* DH5α and INVαF (Invitrogen) were used for cloning PCR products and preparation of plasmid DNA for sequencing. Genomic DNA from the crenarchaeotes *Acidianus ambivalens* (formerly *Desulfurolobus ambivalens*), *S. solfataricus* MT4, and *S. shibatae* was prepared as described previously (28). Genomic DNAs from *S. solfataricus* P2 and *S. acidocaldarius* were a gift from Margaret E. Schenk (Dalhousie University).

PCR and sequencing. Exact-match oligonucleotides surrounding the *S. solfataricus* P2 B3 DNA polymerase were designed by using DNA sequence data from the Sulfolobus Genome Project (44). Primers and their sequences (5′ to 3′) are as follows: ShibB3-1, TAGCCATGTTTATGTTC; ShibB3-2, TTGACTAGAG TATCTGG; UPS-1, AGAGGGCACATAGTCATAGC; DST-1, CGTTCTCTA TGATAATAATTGG. Conditions for amplification were 92°C denaturation for 2 min, annealing at various temperatures (42 to 50°C depending on which primer set was used), and extension at 72°C for 2 min. Approximately 50 to 100 ng of genomic DNA was used in each amplification reaction mixture, which consisted of 10 mM Tris HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% Triton X-100, 0.2 mg of bovine serum albumin per ml, 2 U of *Taq* polymerase (Gibco-BRL), and 5% acetamide (Sigma). PCR products of the correct molecular weight were purified from agarose gels (Bio-Rad) and ligated into a T-tailed vector, pCR2.1 (Invitrogen). Ligations were either electrotransformed into *E. coli* DH5α or heat shocked into *E. coli* INVαF (42). Clones were first manually sequenced to confirm the identity of the insert, and then two clones were completely sequenced on both strands by using ABI and Lycor automated sequencers.

Alignments and phylogenetic analysis. The sequences of the catalytic subunits of archaeal and eukaryotic family B DNA polymerases obtained from GenBank (release 94.0) are summarized in Table 1. Numbering of amino acids involved in 3′-5′ exonuclease and polymerase activity was as previously described (50). Two separate alignments were created by use of the PILEUP option of the Genetics Computer Group program with default values, one which contained all archaeal DNA polymerases, and another which contained the eukaryotic α and δ homologs. The two separate alignments were edited by hand and then combined into a final alignment that consisted of 168 amino acid characters. Only amino acids were used for phylogenetic analysis. Functional information on catalytic residues of the exonuclease domains of archaeal and other family B DNA polymerases was used to aid in the alignment of exonuclease domains. (Alignments are available from D. R. Edgell upon request.)

For phylogenetic analysis, the choice of outgroup sequences was based on BLASTP and BLASTX scores obtained by using numerous archaeal DNA polymerases as query sequences; in each case, the eukaryotic δ and α DNA polymerases were recovered by both BLASTP and BLASTX before eubacterial or other eukaryotic homologs. Parsimony analysis was performed by using PAUP 3.1.1 (47) with 100 random replicates with TBR branch swapping to search for the shortest tree. One hundred bootstrap replicates were performed with simple stepwise addition to determine confidence in the branching order. Distance analysis was performed with PHYLIP 3.57c (17). A PAM-corrected distance matrix was obtained with PROTDIST, and this matrix was used to calculate a tree by the neighbor-joining method (the NEIGHBOR option of PHYLIP). SEQBOOT was used for bootstrap analysis. Templeton tests were carried out with the PROTPARS option of PHYLIP 3.57c with user-defined trees. The standard error for parsimony trees was determined by dividing the numbers of steps by the standard deviation. Due to the time constraints of exhaustive maximum likelihood searches with many taxa, partially constrained trees based on optimal trees found by both parsimony and distance methods were used. An exhaustive search was then performed with PROTML (1).

Calculation of nucleotide substitution rates was done with the program MEGA (32). DNA alignments were created by first aligning the amino acid sequences of the *S. solfataricus* P2 B3 and *S. shibatae* B3 DNA polymerases and the *S. solfataricus* (accession number M34696) and *S. shibatae* (accession number L47841) β-galactosidase genes. The amino acid alignments were used as templates to align the DNA sequences. Nonsynonymous and synonymous substitutions per site were calculated by the method of Nei and Gojobori (37).

Nucleotide sequence accession numbers. The *Sulfolobus solfataricus* P2 and *Sulfolobus shibatae* DNA polymerase sequences described in this paper have been deposited in GenBank under accession numbers U92875 and U92874, respectively.

TABLE 1. Archaeal and eukaryotic family B DNA polymerases used for phylogenetic analyses

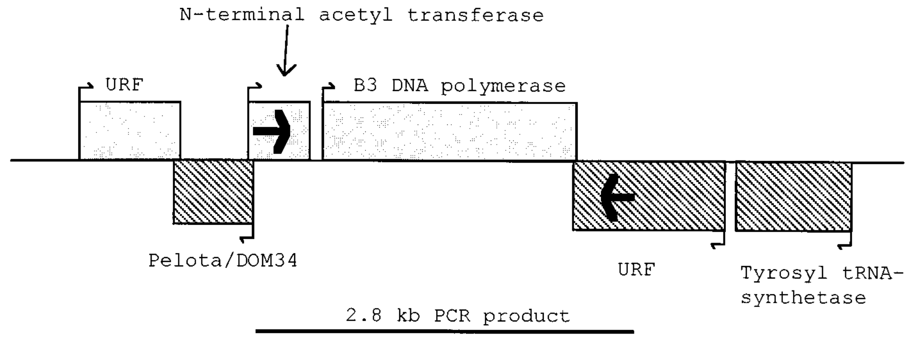
DNA polymerase	Accession no.
Pyrodictium occultum A (B1)	D12983
Pyrodictium occultum B (B2)	D12984
Sulfolobus solfataricus MT4 B1	X64466
Sulfolobus solfataricus P2 B1	U92875
Sulfolobus solfataricus P2 B2	X71597
Sulfolobus solfataricus P2 B3	Y08257
Sulfolobus shibatae B3	U92874
Sulfolobus acidocaldarius B1	U33846
Pyrococcus sp.	U00707
Pyrococcus sp. strain K0D1	D26971
Pyrococcus furiosus	D12983
Thermococcus sp. strain 9oN-7	U47108
Thermococcus litoralis	M47198
Methanococcus voltae	L33366
Methanococcus jannaschii	U67532
Homo sapiens δ and α	M80397 and X06745, respectively
Saccharomyces cerevisiae α and δ	J03268 and X15477, respectively
Schizbsaccharomyces pombe δ	X62423
Plasmodium falciparum α and δ	L18785 and X62423, respectively
Trypanosoma bruceii α	S71823
Bos tarus δ	M80395
Caenorhabditis elegans δ	Z81497
Mus musculus α and δ	D13543 and Z21848, respectively
Oxytricha nova α	U02001
Oxytricha fallax α	U59426

RESULTS

***S. solfataricus* P2 has three family B DNA polymerases.** Two *S. solfataricus* family B DNA polymerases are described in the literature, an enzyme which we designate B1 from *S. solfataricus* MT4, and a paralog, B2, from *S. solfataricus* P2 (38, 40). In the course of sequencing the genome of *S. solfataricus* P2, two open reading frames (ORFs) that were highly similar to archaeal DNA polymerases were found by BLASTX and BLASTP searches (44). One of the ORFs had 880 of 882 residues identical on the amino acid level to the *S. solfataricus* MT4 B1 polymerase (38). The second ORF, designated B3, was not specifically close at the amino acid level to any *S. solfataricus* DNA polymerase sequenced to date and represented an as-yet-undescribed family B DNA polymerase (44). Putative BoxA motifs, essential for transcription initiation, could be identified in the 5′ noncoding regions of all four DNA polymerase genes (see Fig. 2) (10, 24). *S. solfataricus* P2 is the first archaeon reported in which three family B DNA polymerases have been found, namely, the two DNA polymerases we report here and the previously sequenced B2 gene (40).

The catalytic subunits of family B DNA polymerases are difficult to align due to short, highly conserved exonuclease and polymerase domains separated by long stretches of low or no amino acid conservation. The new *S. solfataricus* P2 B3 DNA

A



B

```
S.solf P2 B3   MIKDFFILDFSYEIKGNTPLVYIWSVDDEGNSSVVIDNNFRPYFYIIYEGNENEIIENIKKNCEALQITKVKRKYLGNIV
S.shib B3      MIKDFFILDFSYEIKDNIPLIYIWSIDDEGNSCVVVERNFKPYFYVVYEGNGDEIIENIRKNCEVLLITKVKRKYLGNVV

S.solf P2 B3   ALLIQTSTPTQIKKCREKISELNNIKGIFDADIRYTMRYSLDFDLRPFTWFRAEVNEVKFDGFRTKKAYILDKILSHYEG
S.shib B3      ALLVQTFTPTQIKRCREKISRINGIKSIFDADIRFTMRYSIDFDLRPFTWFKAEVSEVKLEGFRAKKVYILDKILSHYEG

                                                                            Exonuclease
                 Exonuclease                                                 Domain II
                  Domain I                                 Polymerase Domain IV
S.solf P2 B3   NMPELRTIGVDFQIYSKYGSLNPRKDPIVVMSLWSKEGPMQFSLDEGIDDLKIIRRFVDYILNYDPDIIFVYDSDLLPWK
S.shib B3      KIPELRAIGIDFQIYSKYGSLNPRKDPIVVLSLWSKEGSMQFSLDESMDDLKIIRKFVDYILNYDPDIIYVFDVDVFHWK

S.solf P2 B3   YITERASSLGVKIDIGRKIGSEVSVGTYGHYSISGRLNVDLTGLLVNERSLGHVDLIDVSNYLGISPSRYSFKWYEISRY
S.shib B3      YITERANSLGVKIDIGRKIGSEVSQGTYGHYSISGRLNVDLVGLLMNERLTGHIDLIEVANYLGISPKRDSLNWYEISRY

                   Exonuclease
                    Domain III
S.solf P2 B3   WDNEKNRRIIREYSIENARSIYLLGNYLLSTYSELVKIVGLPLDKLSVASWGNRIETSLIRTATKSGELIPIRMDNPNRP
S.shib B3      WDDEKNRDLVKQYSLENAKSIYLLGNFLLSPYSELVKIIGLPLDKLSVASWGNRIEASLIRTAAKSEELIPIRMDNPNRS

                                                                          Polymerase
                    Polymerase Domain II                                   Domain VI
S.solf P2 B3   SKIKKNIIIQPKVGIYTDVYVLDISSVYSLVIRKFNIAPDTLVKEQCDDCYSSPISNYKFKREPSGLYKTFLDELSNVRD
S.shib B3      SKIKKTVI.EPKIGIYSDVYVLDISSVYLSVIRKFNISPDTLVKGQCDDCYVSTISNYKFKKEPSGLYKTFLEELSNIQD

                                                                 Polymerase
                  Polymerase Domain III                           Domain I
S.solf P2 B3   SNKIKVIEELISSFNDYVHWVNARWYSREIASAFDEFSNEIIRFIIDLIKSSGLDVILANDLLIFVTGGSRDKVNELITK
S.shib B3      TRKSKVIEELMSSFYDYIHWINSRWYSREIASAVDELSYEIGKLVIDLIKNSGFEVILANDFLVFVKGGSGDKLNELIFK

                          Polymerase       Polymerase
                          Domain VII        Domain V
S.solf P2 B3   INSLYNLDVKVKIFYKSLLVLDNNRYAGLSEGDKIDIARKGEEDMNLCELARNIKRKIIEEILISKDVKKAIKLVKSTVI
S.shib B3      INSLYDLNLKVRKIYRSLLILGNDRYAGLLEGDKIDIARIGKEDRDLCELVRNVKRKVVEEILISKDVKKAVKLVKSAVI

S.solf P2 B3   KLRRGEFDNEELITWAKIERDLNEYNNQLPFVTAARKAIQSGYLISKDSKIGYVIVKGLGPLNDRAEPFFLVKEKNRIDI
S.shib B3      KLRRGEFDIGELITWVHIEKDFSEYDKQLPFVVAARKAIQSGYLISKDSRIGYLIVKGHGSVHDRAEPFFFVKEKNRIDI

S.solf P2 B3   EYYVDQIFRETLKLLKPLGVNEESLKKTNITDILDLFGASKKK
S.shib B3      EYYVDQLLRESLKVLTPLGVSEESLKKTNITDILDMFGASKKK
```

FIG. 1. (A) Schematic representation (not to scale) of the genomic region of *S. solfataricus* P2 surrounding the B3 DNA polymerase ORF. Boxes indicate ORFs identified as potential coding regions. Names above or below boxes indicate ORFs which have significant matches in databases. URF, unidentified ORF. Direction of transcription is indicated by arrows. Solid arrows indicate the approximate location of direct-match PCR primers used to amplify the region from *S. shibatae*. (B) Amino acid alignment of the orthologous B3 DNA polymerase from *S. shibatae* (S. shib B3) with the *S. solfataricus* P2 B3 DNA polymerase (S. solf P2 B3). Conserved or nonconserved amino acid substitutions are highlighted in bold. Conserved functional regions are boxed. Polymerase domain IV and exonuclease domain II overlap; amino acids corresponding to the polymerase domain are shaded.

polymerase sequence could be aligned with other archaeal and eukaryotic family B homologs except in exonuclease domain III and polymerase domain VI; these two domains were excluded from phylogenetic analysis. Four additional domains not identified in previous analyses of archaeal DNA polymerases, and designated A through D in Fig. 6, were included in the alignment. Of these newly identified regions, only domain A was used in phylogenetic analysis since it alone could be confidently aligned with homologous domains from the eukaryotic α and δ DNA polymerases.

As noted previously, the *S. solfataricus* P2 B2 sequence contains a number of unusual amino acid substitutions in polymerase and exonuclease domains; this is also true for the *S. solfataricus* P2 B3 sequence (40). Neither DNA polymerase has the consensus Asp-Ile-Glu (DIE) motif found in the 3′-5′ exonuclease domain I of other archaeal DNA polymerases.

```
S. solfataricus MT4 B1  taaaaCTTATAgcgtatttctcagaaaataatAtAtgttagaaaATG
 S. solfataricus P2 B1  taaaaCTTATAgcgtatttctcagaaaataatAtAtgttagaaaATG
 S. solfataricus P2 B2  aaagaCTTAATttaccagaggagagAtgtaacAcATG
 S. solfataricus P2 B3  aagaaTTTATAttataaatatctggattaattgttAa[42 bps]atATG
        P. occultum B1  gaactGTTATCggaaatatcctcatctaggagAcgcg[51 bps]gcATG
        P. occultum B3  atacgATTATGtaggggcgggtggtggtagAttctccagggcagagccagcccATG
       Pyrococcus fur.  aaggtTTTATActccaaactgagttagtagAtAtgtggggagcAtaATG
        Pyrococcus sp.  gcgttCTTAAAggCTTAAAtacgtgaatttagcgtaaAttAttgagggattaagtATG
     Thermococcus lit.  gggggTTTAAAaatttggcggaacttttaTTTAATttgaactccagtttatatctggtggtAtttATG

       archaebacterial       t   t
             consensus        tta a
                             c   a
```

FIG. 2. Putative promoter motifs in the 5′ regions of archaeal family B DNA polymerases (22). Sequences shown in capitalized, underlined letters correspond to the consensus archaeal BoxA motif. Nucleotides which are in bold type indicate a pyrimidine-purine pair and the probable site of transcription initiation. Start codons of the DNA polymerase ORFs are capitalized.

These residues are critical for exonuclease activity, since introduction of Asp→Glu or Glu→Ala substitutions in exonuclease domain I of the *Thermococcus litoralis* family B DNA polymerase abolishes exonuclease activity (30). Mechanistic studies with other DNA polymerases indicate that these acidic amino acids play crucial roles in exonuclease activity and are responsible for coordination of divalent metal ions (3, 11). The absence of this exonuclease domain has been noted before in other family B DNA polymerases, notably all of the eukaryotic α DNA polymerases, but these homologs still retain polymerase function (5, 31). It is possible that the 3′-5′ exonuclease activity in *S. solfataricus* is performed by the *S. solfataricus* P2 B1 paralog, which does possess a consensus exonuclease domain I sequence.

Both the B2 and B3 sequences from *S. solfataricus* P2 exhibit a number of nonconserved substitutions in two metal-binding polymerase domains (I and II). The amino acid motif Asp-Thr-Asp (DTD), which is present in polymerase domain I of all other archaeal enzymes, is replaced by Ile-Ile-Asp and Asn-Asp-Leu in *S. solfataricus* P2 B2 and B3, respectively (Fig. 2). Copeland and Wang found that mutation of the Asp-Thr-Asp motif to Asn-Thr-Asp, Asp-Ser-Asp, or Asp-Thr-Asn in human DNA polymerase α drastically reduced DNA polymerase activity (9). Dong and Wang (12) also found that Lys-950 of human DNA polymerase α is essential for the binding of deoxynucleoside triphosphates. The *S. solfataricus* P2 B2 sequence possesses this amino acid in the homologous position but it is replaced by a Glu residue in the *S. solfataricus* P2 B3 and the *S. shibatae* B3 (see below) sequences.

***S. shibatae* possesses a rapidly evolving ortholog of the *S. solfataricus* P2 B3 DNA polymerase.** *S. solfataricus* P2 is the first archaeon reported to have three family B DNA polymerases. However, the extremely divergent amino acid sequence of the B3 DNA polymerase raises the question of whether this gene actually codes for a functional DNA polymerase. In an attempt to address this issue, we have cloned and sequenced an ortholog of the *S. solfataricus* P2 B3 DNA polymerase from *Sulfolobus shibatae*, a closely related member in the order *Sulfolobales.*

Using the *S. solfataricus* P2 genome sequence (44), we designed nondegenerate PCR primers flanking the B3 ORF (Fig. 1A) and attempted to amplify this genomic region from representatives of the *Sulfolobales*. We were consistently able to amplify a fragment of the expected size (2.8 kb) from *S. solfataricus* P1, *S. solfataricus* MT4, and *S. shibatae* but not from *Sulfolobus acidocaldarius* (data not shown). Of the organisms for which we could obtain amplification, *S. shibatae* is the most distant from *S. solfataricus* P2 on the basis of 16S rRNA phylogeny (23). We cloned and sequenced the *S. shibatae* PCR product and found it to be identical to *S. solfataricus* P2 in gene identity and order. However, the predicted amino acid sequence of the *S. shibatae* B3 DNA polymerase is only 79% identical to the *S. solfataricus* P2 B3 sequence (158 of 764 amino acids are different [Fig. 1B]).

This number of amino acid substitutions, both conserved and nonconserved, was surprising. Protein-coding genes evolving neutrally and under stabilizing selection for maintenance of a function will have high synonymous (no change of amino acid) substitution rates and low nonsynonymous (change of amino acid) rates (33). Protein-coding genes which are not evolving at neutral rates will have a higher rate of nonsynonymous substitutions and a lower rate of synonymous substitutions per site. We calculated both synonymous (K_S) and nonsynonymous (K_A) nucleotide substitution rates for the B3 DNA polymerases and for the only other protein-coding gene sequenced from both organisms, β-galactosidase (Table 2). The synonymous substitution rate of β-galactosidase is typical of protein-coding genes evolving neutrally (33). The substitution rates of the B3 DNA polymerase genes suggest that functional constraints have been relaxed, allowing these genes to accumulate nonsynonymous substitutions. However, the rate of nonsynonymous substitutions is not high enough to suggest that these proteins are under positive selection for a novel function (36).

Phylogenetic analysis of archaeal family B DNA polymerases is confounded by rapid rates of sequence evolution. To determine the relationship of the three *S. solfataricus* P2 family B DNA polymerases to other archaeal DNA poly-

TABLE 2. Comparison of rates of nucleotide substitution of *S. solfataricus* and *S. shibatae* protein-coding genes[a]

Gene coding for:	No. of nt	No. of differences		Sub rate		Nsy/Syn ratio
		Nsy	Syn	Nsy	Syn	
B3 DNA polymerase	2,292	183.5	201.5	0.111 ± 0.0083	0.572 ± 0.0468	0.192
β-Galactosidase	1,467	41.0	171.0	0.037 ± 0.058	0.886 ± 0.0898	0.042

[a] Abbreviations: nt, nucleotides; Nsy, nonsynonomous; Syn, synonomous; sub rate, number of substitutions per site.

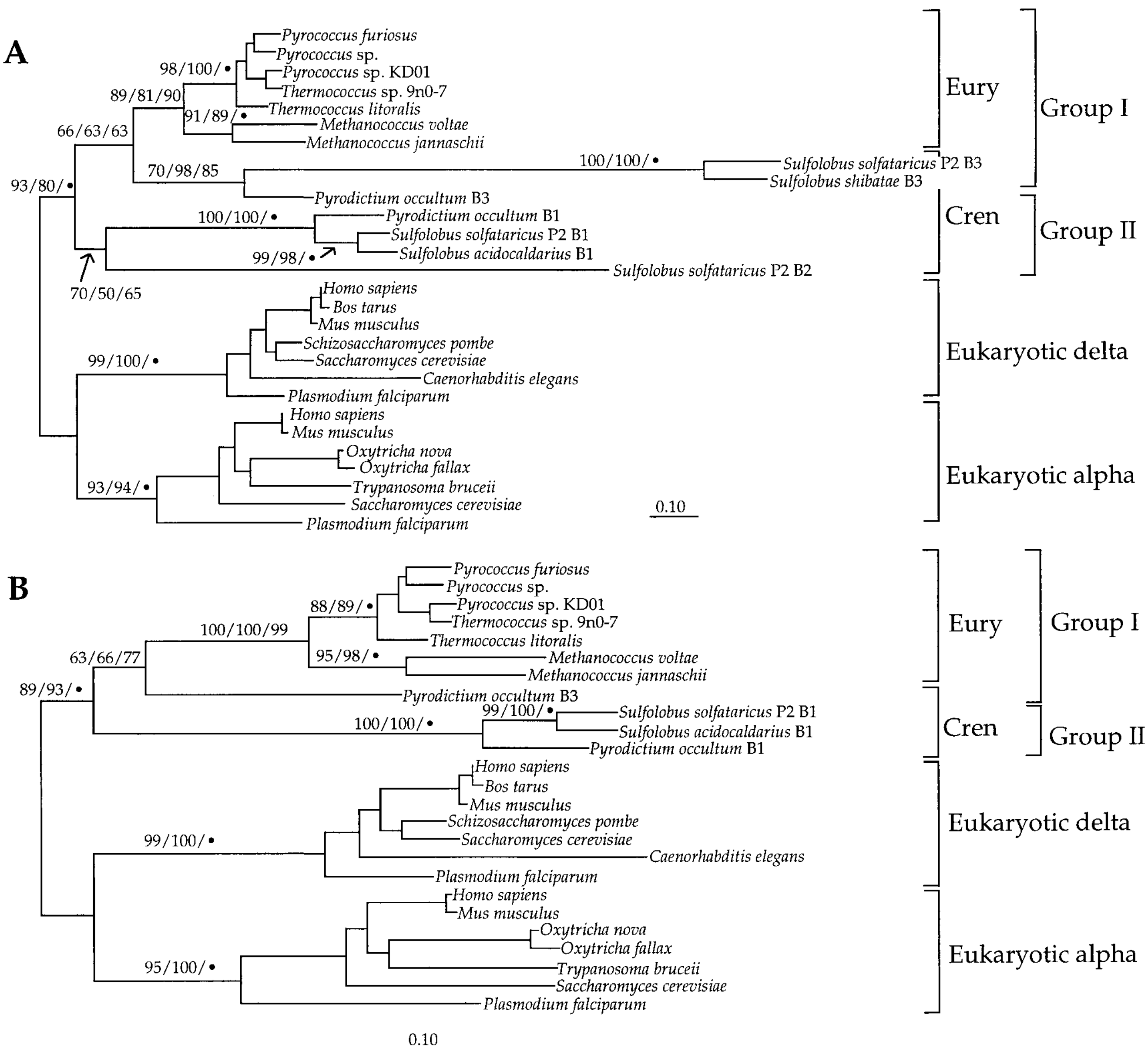
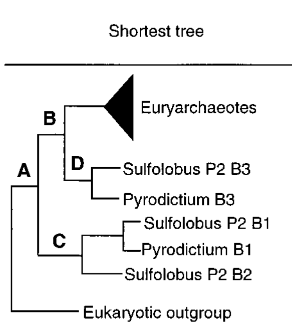


FIG. 3. Phylogenetic analysis of archaeal and eukaryotic family B DNA polymerases. (A) PROTDIST analysis with all taxa. An identical topology was found by using PAUP (100 random replicates of stepwise addition; the shortest tree was 748 steps [confidence interval = 0.741; HI = 0.259]). Based on the phylogeny obtained by PAUP and distance analyses, a partially constrained tree was used for a maximum likelihood search with PROTML. Nodes which were constrained in the maximum likelihood analysis are indicated by a small solid circle. Bootstrap values for nodes are indicated in the order parsimony/distance/maximum likelihood. EURY, euryarchaeote; CREN, crenarchaeote. (B) PROTDIST analysis with the rapidly evolving *S. solfataricus* P2 B2 and B3 sequences removed. An identical tree was found by use of parsimony and maximum likelihood methods. Bootstrap values are as described for panel A.

merases, phylogenetic analysis was performed on a data set which contained all available archaeal family B DNA polymerase sequences. These include family B paralogs from euryarchaeotes, the *S. solfataricus* P2 B1, B2, and B3 paralogs, a B1 DNA polymerase from *S. acidocaldarius*, and the *S. shibatae* B3 paralog. The *S. solfataricus* MT4 B1 paralog was not included in phylogenetic analysis since it is 99.7% identical at the amino acid level to its ortholog from *S. solfataricus* P2. Two paralogs (called A and B in reference 49) from the crenarchaeote *P. occultum* were also included in analyses. *P. occultum* A and *S. solfataricus* P2 B1 are orthologs, but *P. occultum* B appears most related to *S. solfataricus* P2 B3 (see below). The two family B DNA polymerases from *P. occultum* have been renamed *P. occultum* B1 and *P. occultum* B3.

Regardless of the phylogenetic method used, the three *S. solfataricus* P2 DNA polymerases did not branch together, as would have been expected if they were related by recent gene duplications (Fig. 3). In both parsimony and distance analyses, the *S. solfataricus* P2 B3 and *S. shibatae* B3 sequences grouped with the *P. occultum* B3 DNA polymerase with moderate bootstrap support. The *S. solfataricus* P2 B1 and *S. acidocaldarius* B1 paralogs form a separate group with *P. occultum* B1. High bootstrap values for this group were obtained. Our results suggest that the *S. solfataricus* P2 B3, *S. shibatae* B3, and *P.*



Species excluded	Bootstrap at node A	B	C	D
None	90/69	68/80	66/37	76/96
Sulf P2 B1	92/71	64/83	67/47	61/100
Sulf P2 B2	87/94	85/89	×	67/100
Sulf P2 B3	86/76	47/57	46/42	×

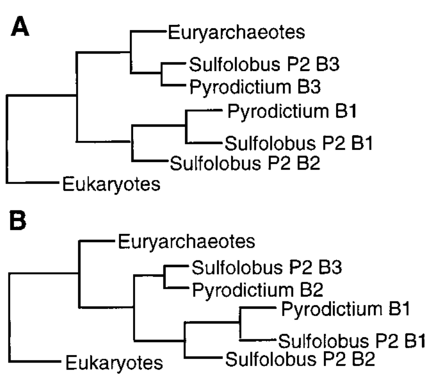
FIG. 4. Effect of removing rapidly evolving taxa from phylogenetic analysis. The optimal tree found by parsimony and distance analysis is drawn schematically. Important nodes are indicated by letters A through D. A, archaeal unity; B, support for group I; C, support for group II; D, affinity of *S. solfataricus* P2 B3 and *P. occultum* B3. Confidence for each node after removal of rapidly evolving taxa was measured by 100 bootstrap replicates by both parsimony and distance methods.

occultum B3 DNA polymerases are orthologs as are the *S. solfataricus* P2 B1, *S. acidocaldarius* B1, and *P. occultum* B1 polymerases. It is not clear to which DNA polymerase the *S. solfataricus* P2 B2 sequence is related. Low bootstrap support was found for this sequence grouping with the *S. solfataricus* P2 B1, *S. acidocaldarius* B1, and *P. occultum* B1 DNA polymerases. If this placement is correct, a crenarchaeote-specific gene duplication event must have occurred to give rise to the *S. solfataricus* P2 B2 paralog. Finding of an orthologous sequence in another crenarchaeote may help in resolving the phylogenetic position of this DNA polymerase.

Parsimony, distance, and maximum likelihood analyses split the archaeal DNA polymerases into two groups. We suggest calling the euryarchaeote and *S. solfataricus* P2 B3-*S. shibatae* B3-*P. occultum* B3 clade group I and the remaining crenarchaeote sequences group II (Fig. 3). This tree topology is consistent with the last common ancestor of archaea possessing at least two family B DNA polymerases. However, given the extremely rapid rates of evolution of the *S. solfataricus* B2 and B3 and the *S. shibatae* B3 sequences compared to that of other archaeal DNA polymerases, we were concerned that the tree topology found by all methods was artifactual. To test this possibility, we first eliminated the *S. shibatae* B3 sequence from phylogenetic analysis. Removal of this sequence did not result in tree topologies different from those obtained when it was included; we did not include the *S. shibatae* B3 sequence in any further phylogenetic analyses. The same rationale was applied to removing the *S. acidocaldarius* B1 sequence from further analyses. With this reduced data set, each of the three *S. solfataricus* P2 paralogs was separately removed from the analysis and the effect on tree topology was measured by both parsimony and distance bootstrap analyses. Figure 4 indicates that removal of the *S. solfataricus* P2 B3 sequence from the analysis had the greatest effect since bootstrap values at nodes supporting group I (node B) and group II (node C) were reduced.

The long branch lengths of the *S. solfataricus* P2 B2 and B3 DNA polymerases also concerned us, since rapidly evolving sequences such as these are known to be positively misleading in phylogenetic reconstruction (16). We were particularly interested in testing an alternative tree topology which would be consistent with the common ancestor of archaea possessing a single family B DNA polymerase. This tree topology, which unites all crenarchaeote sequences to the exclusion of euryarchaeote sequences, was not significantly worse than the shortest tree topology since both the Kishino-Hasegawa and Templeton tests did not reject the alternative topology at the 5% significance level (Fig. 5) (17).

We also removed the *S. solfataricus* P2 B2 and B3 sequences and performed parsimony, distance, and maximum likelihood analyses to find the best tree topology. If the branching order observed with all taxa is robust, removal of these two taxa should not result in a significantly different tree topology. Indeed, the best tree recovered by all methods was identical to the best tree recovered when all taxa were included, with the archaeal DNA polymerases split into two groups (Fig. 3B). However, an alternative topology uniting all crenarchaeote sequences as a group and consistent with the common ancestor of archaea possessing a single family B DNA polymerase was not significantly worse than the best tree (Fig. 5). The lack of phylogenetic resolution of the archaeal family B DNA polymerase data set cannot be attributed only to the rapidly evolving *S. solfataricus* P2 paralogs; other factors must also be contributing to the problem. Sequencing of additional paralogs from a diverse sampling of both crenarchaeotes and eury-



Tree	ln L	Δ ln l/S.E.	No. of steps	S.D.	Sig. worse?
A	-3166.1	0.0	943	0.0	BEST
A	-2447.2	0.0	725	0.0	BEST
B	-3169.6	-1.06	948	2.65	NO
B	-2451.6	-1.1	728	4.13	NO

FIG. 5. An alternative topology (tree B) consistent with the hypothesis that the common ancestor of archaea had only a single family B DNA polymerase is not significantly worse than the best tree (tree A). Abbreviations: ln L, log likelihood; ln L, difference in log likelihood; S.E., standard error. Alternative topologies are considered significantly worse if the Kishino-Hasegawa and Templeton tests reject these topologies at the 5% significance level (1, 17). The boxed values are those obtained when the *S. solfataricus* P2 B2 and B3 sequences were removed from the analysis.

archaeotes may help in resolving the phylogenetic relationship of archaeal DNA polymerases.

DISCUSSION

Previous studies on archaeal DNA replication have focused primarily on the biochemistry of purified DNA polymerases (20, 51, 52). These studies attempted to classify DNA polymerases as eukaryotic- or eubacterial-like on the basis of enzymatic properties and resistance or sensitivity to various inhibitors. The most common indicator of the presence of a eukaryotic-like DNA polymerase was sensitivity to aphidicolin, a fungal metabolite which inhibits eukaryotic DNA replication by allosterically binding to the replicative DNA polymerases (45). By using aphidicolin sensitivity as an indicator, *S. acidocaldarius* and *S. solfataricus* were found to possess a eukaryotic-like DNA polymerase activity as well as an unclassified aphidicolin-resistant DNA polymerase activity (14, 29, 35, 41, 49). Sequencing of the gene corresponding to the aphidicolin-sensitive activity from *S. solfataricus* MT4 confirmed that this DNA polymerase was a family B homolog more similar to eukaryotic than to eubacterial homologs (38). We call this paralog B1.

Prangishvili and Klenk (40) attempted to clone and sequence the gene for the aphidicolin-resistant DNA polymerase activity by designing a degenerate oligonucleotide against polymerase domain I of eukaryotic and archaeal family B homologs. The DNA polymerase they sequenced, and which we call B2, was significantly different on the amino acid level from the *S. solfataricus* MT4 B1 aphidicolin-sensitive DNA polymerase. Since the biochemical activities of the cloned *S. solfataricus* P2 B2 DNA polymerase were not studied, it is unclear if this DNA polymerase actually corresponds to the aphidicolin-resistant activity.

Since only two DNA polymerase activities were found in cell extracts of *S. acidocaldarius* and *S. solfataricus*, it is perhaps surprising that we have found a third DNA polymerase (B3) (44). The amino acid sequence of this DNA polymerase is extremely divergent, raising the question of whether this is actually the product of a functional gene. To address this question, we have cloned and sequenced the ortholog of this gene from *S. shibatae*, a closely related member of the *Sulfolobales* (23), reasoning that if *S. shibatae* also possesses this divergent DNA polymerase, it is likely to have some function. Database search scores (data not shown) and the alignment in Fig. 6 convincingly show that these proteins align over much of their length with other archaeal DNA polymerases. There is little doubt that the genes encoding these proteins evolved from an archaeal family B DNA polymerase. However, the number of amino acid differences between the orthologous *S. shibatae* B3 and *S. solfataricus* P2 B3 sequences is surprising given the close evolutionary relationship of these two organisms (23).

A possible explanation for the low amino acid identity between these two sequences is positive selection for a novel function(s). There are very few examples of positive selection based on molecular sequences and only a single possible example in *Archaea*, that of the superoxide dismutase genes of halophiles (15, 27). Evidence for positive selection can be assessed by taking the ratio of nonsynonymous (K_A) to synonymous (K_S) substitutions per site (15, 36). Ratios of >1 are considered strong evidence for positive selection, since the rate of nonsynonymous substitutions exceeds that which can be explained by neutral evolution. A ratio of <1 is taken as evidence for stabilizing or purifying selection, since deleterious nonsynonymous substitutions do not become fixed in the population. The K_A/K_S ratio for the B3 DNA polymerase genes is only 0.192 (Table 2), well below what is considered evidence for positive selection, but is higher than that of the β-galactosidase genes (0.042). It is clear from the alignment in Fig. 6 that the *S. solfataricus* and *S. shibatae* B3 DNA polymerases have undergone a high rate of nonsynonymous amino acid replacements after their divergence from a common ancestral sequence. Amino acid substitutions in catalytic domains suggest that we cannot be certain that the encoded proteins retain all or any of the ancestral exonuclease or polymerization functions, but the fact that both ORFs remain uninterrupted by nonsense mutations indicates that the genes encoding these proteins remain under some sort of selection.

The finding of multiple family B DNA polymerases in two crenarchaeotes but (as yet) only a single family B DNA polymerase in all euryarchaeotes studied, including the whole-genome sequence of *Methanococcus jannaschii* (7), raises a number of interesting questions concerning the evolution of the archaeal DNA replication apparatus. Our phylogenetic analysis, and previous analyses, cannot resolve what we view as two competing hypotheses of archaeal family B DNA polymerase evolution: (i) that the common ancestor of archaea possessed at least two family B DNA polymerases, one of which (the group II ortholog) was lost in the euryarchaeote lineage after the split of the two archaeal kingdoms; and (ii) that the common ancestor possessed a single family B DNA polymerase (orthologous to group I) and the multiple DNA polymerases of crenarchaeotes (group II) evolved by gene duplication after the split of the two archaeal kingdoms.

In the absence of data concerning the function and organismal distribution of family B DNA polymerases in archaea, both scenarios seem equally plausible. Scenario (i), which postulates a loss of a cellular encoded family B DNA polymerase, has been observed in two eubacteria. Neither *Haemophilus influenzae* Rd nor *Mycoplasma genitalium* possesses a family B homolog (18, 22), yet this DNA polymerase is present in *E. coli*, where it functions as a repair polymerase (4, 8, 26). This DNA polymerase must have been lost independently in the lineages leading to *H. influenzae* and *Mycoplasma genitalium.*

Scenerio (ii) would be favored by the finding of a single family B homolog in euryarchaeotes by genome sequencing projects. Indeed, only a single family B homolog is found in the recently completed genome sequence of *Methanococcus jannaschii* (7). However, two DNA polymerases activities have been described in another euryarchaeote, *Halobacterium halobium* (35, 46). One of these activities likely corresponds to a family B homolog, orthologous to the known family B DNA polymerases of euryarchaeotes. The second activity is aphidicolin resistant and unclassified. It is possible that the second polymerase activity could also correspond to a family B DNA polymerase since both aphidicolin-resistant and -sensitive family B homologs from *P. occultum* were recently characterized (49).

The finding of multiple family B DNA polymerases in crenarchaeotes is intriguing, given that eukaryotes use three family B DNA polymerases (α, δ, and ε) for nuclear DNA replication (2, 6, 31). The finding of three family B DNA polymerases in *S. solfataricus* P2 raises the obvious question of whether these homologs perform similar roles at the replication fork as the eukaryotic homologs. Conversely, if euryarchaeotes possess only a single family B homolog, as indicated by the genome sequence of *Methanococcus jannaschii*, this also raises questions concerning the function(s) of the single DNA polymerase at the replication fork. It is conceivable that a single DNA polymerase could replicate both the leading and lagging strands in euryarchaeotes, as does the family C DNA polymerase in *E. coli* (31).

		Exonuclease Domain I		A		Polymerase Domain IV	Exonuclease Domain II		B		C
S.solfataricus MT4 B1	[225]	IKRVAIDIEVY	[246]	QKAEFPIISI	[292]	EYELLGRFFDILLEY	.P.IVLTFNGDDFDLPYIYFR	[326]	.ALKLGY	[347]	KYLAGLHIDL
S.solfataricus P2 B1	[225]	IKRVAIDIEVY	[246]	QKAEFPIISI	[292]	EYELLGRFFDILLEY	.P.IVLTFNGDDFDLPYIYFR	[326]	.ALKLGY	[347]	KYLAGLHIDL
S.acidocaldarius B1	[223]	IKRVSLDIEVY	[244]	ERAEFPIISV	[290]	EKKLLARLFEIIREY	.P.MLLTFNGDDFDIPYIYFR	[324]	.ALRLNF	[344]	KFLAGIHIDL
S.solfataricus P2 B2	[16]	LEKLERIIERL	[51]	DELVDPYNLV	[151]	..............Y	FYYMRKRLNVVN.ETPTVLSQ	[172]	TLYRLGI	[208]	SLKGKV.FEV
S.solfataricus P2 B3	[166]	LRTIGVDFQIY	[183]	NPRKDPIVVM	[211]	DLKIIRRFVDYILNY	DPDIIFVYDSDLLPWKYITER	[247]	.ASSLGV	[273]	SISGRLNVDL
S.shibatae B3	[166]	LRAIGIDFQIY	[183]	NPRKDPIVVL	[211]	DLKIIRKFVDYILNY	DPDIIYVFDVDVFHWKYITER	[247]	.ANSLGV	[273]	SISGRLNVDL
P.occultum B1	[261]	PRRLAVDIEVF	[282]	STASYPVISV	[331]	ERALILEAFRLISNY	.P.VLLTFNGDNFDLPYLYNR	[365]	.AVKLGI	[385]	TLEYGFHIDL
P.occultum B3	[181]	MRLVAFDIEVY	[198]	NPARDPVIIV	[226]	DRRVLREFVEYVRAF	DPDIIVGYNSNHFDWPYLMER	[262]	.ARRLGI	[288]	SVQGRLNVDL
Pyrococcus fur.	[135]	LKILAFDIETL	[151]	EFGKGPIIMI	[187]	EREMIKRFLRIIREK	DPDIIVTYNGDSFDFPYLAKR	[223]	.AEKLGI	[251]	EVKGRIHFDL
Pyrococcus sp.	[135]	LKLLAFDIETL	[151]	EFAKGPIIMI	[187]	EREMIKRFLKVIREK	DPDVIITYNGDSFDLPYLVKR	[223]	.AEKLGI	[251]	EIKGRIHFDL
Pyrococcus KOD1	[135]	LKMLAFDIQTL	[151]	EFAEGPILMI	[187]	EREMIKRFLRVVKEK	DPDVLITYNGDNFDFAYLKKR	[223]	.CEKLGI	[251]	EVKGRIHFDL
Thermococcus lit.	[135]	LKLLAFDIETF	[151]	EFGKGEIIMI	[187]	EREMIKRFVQVVKEK	DPDVIITYNGDNFDLPYLIKR	[223]	.AEKLGV	[253]	EIKGRIHFDL
Thermococcus 9oN-7	[135]	LTMLAFDIETL	[151]	EFGTGPILMI	[187]	EKEMIKRFLRVVREK	DPDVLITYNGDNFDFAYLKKR	[223]	.CEELGI	[251]	EVKGRIHFDL
Methanococcus vol.	[175]	LNCIAFDMELY	[191]	NAKKDPIIMV	[231]	EKELIQKTIEIL..K	QYDVIYTYNGDNFDFPYLKKR	[265]	.ANIYEI	[298]	KIPGIIHIDL
Methanococcus jan.	[154]	LKSVAFDMEVY	[171]	NPERDPILMA	[207]	EKELIKKIIETL..K	EYDVIYTYNGDNFDFPYLKAR	[241]	.AKIYGI	[269]	YIPGRVHIDL

		Exonuclease Domain III		Polymerase Domain II		D		Polymerase Domain VI
S.solfataricus MT4 B1	[406]	LIEYNFRDAEITLQL	[496]	GAVVIDPPAGIFFNITVLDFASLYPSIIRT.WNLSYETV	[545]	KDETGEVLHIVCMDRP	[561]	GITAVITGLLRDFR..VKIYKKKA
S.solfataricus P2 B1	[406]	LIEYNFRDAEITLQL	[496]	GAVVIDPPAGIFFNITVLDFASLYPSIIRT.WNLSYETV	[545]	KDETGEVLHIVCMDRP	[561]	GITAVITGLLRDFR..VKIYKKKA
S.acidocaldarius B1	[403]	LIEYNLRDAEITLKL	[493]	GAVVIDPPAGVYFNVVVLDFASLYPSIIKN.WNISYETI	[542]	EDETGEKLHYVCMDKP	[558]	GITAVYQGLIRDFR..VKVYKKKA
S.solfataricus P2 B2	[226]	LIEWS..........	[326]	GGLILFPQPGCYDNVYQVDFSSMYPSLIVK.HNISAETV	[366]	CDDIKTELHSICLKEK	[382]	GIIPEALQWLIERKSELK......
S.solfataricus P2 B3	[331]	IREYSIENARSIYLL	[406]	KNIIIQPKVGIYTDVYVLDISSVY.SLVIRKFNIAPDTL	[451]	CYSSPISNYKFKREPS	[467]	GLYKTFLDELSNVR..........
S.shibatae B3	[331]	VKQYSLENAKSIYLL	[406]	KKTVIEPKIGIYSDVYVLDISSVY.LSVIRKFNISPDTL	[451]	CYVSTISNYKFKKEPS	[467]	GLYKTFLEELSNIQ..........
P.occultum B1	[444]	LVRYNVRDADLTLRL	[534]	GALVLDPPSGIYFNIVVLDFASLYPSIIKR.WNLSYETV	[583]	.EV.PDVGHKVCMSIP	[597]	GLTSQIVGLLRDYR..VKIYKKKA
P.occultum B3	[348]	LERYALDDVRATYGL	[421]	GAVVLKPLKGVHENVVVLDFSSMYPSIMIK.YNVGPDTI	[471]	CYVAPEVGHRFRRSPP	[487]	GFFKTVLENLLKLRRQVKEKMKEF
Pyrococcus fur.	[308]	VAKYSMEDAKATYEL	[387]	GGFVKEPEKGLWENIVYLDFRALYPSIIIT.HNVSPDTL	[432]	YDIAPQVGHKFCKDIP	[448]	GFIPSLLGHLLEERQKIKTKMKET
Pyrococcus sp.	[308]	VAKYSMEDAKVTYEL	[387]	GGYVKEPEKGLWEGLVSLDFRSLYPSIIIT.HNVSPDTL	[432]	YDVAPEVGHKFCKDFP	[448]	GFIPSLLKRLLDERQEIKRKMKAS
Pyrococcus KOD1	[308]	VARYSMEDAKVTYEL	[386]	GGYVKEPERGLWENIVYLDFRSLYPSIIIT.HNVSPDTL	[431]	YDVAPQVGHRFCKDFP	[447]	GFIPSLLGDLLEERQKIKKKMKAT
Thermococcus lit.	[310]	LAQYSMEDARATYEL	[389]	GGYVKEPEKGLWENIIYLDFRSLYPSIIVT.HNVSPDTL	[434]	YDVAPIVGYRFCKDFP	[450]	GFIPSILGDLIAMRQDIKKKMKST
Thermococcus 9oN-7	[308]	VARYSMEDAKVTYEL	[386]	GGYVKEPERGLWDNIVYLDFRSLYPSIIIT.HNVSPDTL	[431]	YDVAPEVGHKFCKDFP	[447]	GFIPSLLGDLLEERQKIKRKMKAT
Methanococcus vol.	[353]	LLRYAYEDALYTYKM	[432]	GGYVREPLKGIQEDIVSLDFMSLYPSILIS.HNISPETV	[477]	.ENM.EL.........	[482]	GIIPKTLNELLSRRKHIKMLLKDK
Methanococcus jan.	[324]	LIEYSLQDAKYTYKI	[403]	GGYVKEPEKGMFEDIISMDFRSLYPSIIIS.YNISPDTL	[442]	........CECCKDVS	[463]	GLIPKTLRNLIERRINIKRRMKKM

		Polymerase Domain III		Polymerase Domain I		Polymerase Domain VII		Polymerase Domain V
S.solfataricus MT4 B1	[599]	QRAMKVFINATYGVFGAETFPLYAPRVAESVTALGRY	[648]	GLTVLYGDTDSLFL	[692]	FVAFSGLKKNYFGVY	[707]	QDGKVDIKGMLVKKRNTPEFVK
S.solfataricus P2 B1	[599]	QRAMKVFINATYGVFGAETFPLYAPAVAESVTALGLR	[648]	GLTVLYGDTDSLFK	[692]	FVAFSGLKKNYFGVY	[707]	QDGKVDIKGMLVKKRNTPEFVK
S.acidocaldarius B1	[596]	QRAMKVFINATYGVFGAENFPLYAPAVAESVTAIGRY	[645]	NLKVIYGDTDSLFL	[689]	YVAYSGLKKNYFGVY	[704]	PDGKTEIKGMLAKKRNTPEFIK
S.solfataricus P2 B2	[406]	AEAIKWILVASFGYLGYRNSLFGKIEAYEMVTYLARK	[455]	GLRVLHGIIDSLVV	[482]	KET..GLRKRYNWII	[515]	MNGEMIAKGL..IRENMPNIVK
S.solfataricus P2 B3	[485]	IKVIEELISSFNDYVHWVNARWYSREIA.SAFDEFSN	[534]	GLDVILANDLLIFV	[577]	KSLLVLDNNRYAGLS	[592]	EGDKIDIARKGEEDMNLCELAR
S.shibatae B3	[485]	SKVIEELMSSFYDYIHWINSRWYSREIA.SAVDELSY	[534]	GFEVILANDFLVFV	[577]	RSLLILGNDRYAGLL	[592]	EGDKIDIARIGKEDRDLCELVR
P.occultum B1	[635]	QAAMKVYINASYGVFGAESFPFYAPPVAESVTAIGRY	[684]	GLRVLYGDTDSLFI	[728]	FVTFSGLKKRYIGAY	[743]	EDGSIDVKGMVAKKRNTPEFLK
P.occultum B3	[524]	QKALKVLANASYGYMGWSHARWYCKRCAEAVTAWGRN	[573]	GLKVIYGDTDSLFV	[616]	KVFFTEAKKRYVGLL	[631]	EDGRIDIVGFEAVRGDWCELAK
Pyrococcus fur.	[484]	QKAIKLLANSFYGYYGYAKARWYCKECAESVTAWGRK	[534]	GFKVLYIDTDGLYA	[586]	RGFFV.TKKRYAVID	[600]	EEGKVITRGLEIVRRDWSEIAK
Pyrococcus sp.	[484]	QRAIKILANSYYGYYGYAKARWYCKECAESVTAWGRE	[534]	GFKVLYIDTDGLYA	[586]	RGFFV.TKKKYALID	[600]	EEGKIITRGLEIVRRDWSEIAK
Pyrococcus KOD1	[483]	QRAIKILANSYYGYYGYARARWYCKECAESVTAWGRE	[533]	GFKVIYSDTDGFFA	[585]	RGFFV.TKKKYAVID	[599]	EEGKITTRGLEIVRRDWSEIAK
Thermococcus lit.	[486]	QRAIKLLANSFYGYMGYPKARWYSKECAESVTAWGRH	[536]	GFKVLYADTDGFYA	[588]	RGFFV.TKKRYAVID	[602]	EEGRITTRGLEVVRRDWSEIAK
Thermococcus 9oN-7	[483]	QRAIKILANSFYGYYGYAKARWYCKECAESVTAWGRE	[533]	GFKVLYADTDGLHA	[585]	RGFFV.TKKKYAVID	[599]	EEGKITTRGLEIVRRDWSEIAK
Methanococcus vol.	[522]	QKSIKVLANSHYGYLAFPMARWYSDKCAEMVTGLGRK	[571]	GFKVIYADTDGFYA	[645]	RGLFV.TKKKYALIE	[659]	DDGHIVVKGLEVVRRDWSNIAK
Methanococcus jan.	[503]	QKSLKILANSVYGYLAFPRARFYSRECAEIVTYLGRV	[522]	GFKVLYIDTDGFYA	[605]	RGIFV.TKKRYALVT	[627]	NGRVTVVKGLEFVRRDWSNIAK

FIG. 6. Amino acid alignment of archaeal family B DNA polymerases. Numbering of sequences is from the N to the C terminal and corresponds to the amino acid position at the start of each conserved domain. Exonuclease domain II and polymerase domain IV overlap, and a space has been inserted in the alignment to indicate the start of exonuclease domain II. Signature sequences which support a specific relationship of the *P. occultum* B3, *S. solfataricus* P2 B3, and *S. shibatae* B3 DNA polymerases with euryarchaeote DNA polymerases are boxed. Signature sequences which support a relationship of the *S. solfataricus* P2 B1 and *P. occultum* B1 DNA polymerases with euryarchaeotes are boxed and shaded. Amino acid residues which have been identified as functionally important by mutational studies are in bold type. Not all of the alignment was used for phylogenetic analysis (see Results), and the regions used are indicated by ~. Gaps introduced in the alignment are indicated by a period. Pyrococcus fur., *Pyrococcus furiosus*; Thermococcus lit., *Thermococcus litoralis*; Methanococcus vol., *Methanococcus voltae*; Methanococcus jan., *Methanococcus jannaschi*.

If the three *S. solfataricus* P2 family B DNA polymerases do perform function(s) analogous to those of the three eukaryotic family B DNA polymerases, it would not be unreasonable to expect them to be orthologs. For example, the *S. solfataricus* P2 B1 polymerase should branch with a specific group of eukaryotic polymerases in phylogenetic analyses (δ-type, for instance). However, this is not the relationship observed; the archaeal DNA polymerases form a monophyletic group with high bootstrap values (Fig. 3). The gene duplication events which gave rise to the multiple family B DNA polymerases of eukaryotes and crenarchaeotes must have occurred independently of one another.

ACKNOWLEDGMENTS

We thank members of the Doolittle lab for helpful discussion and comments on the manuscript and Melanie Dobson (Dalhousie University) for technical advice. This work was carried out in the context of the Sulfolobus Genome Sequencing project, and thanks are extended to all individuals involved in the project.

Work in the laboratory of W.F.D. is supported by the Medical Research Council of Canada (grant D4341). W.F.D. is a Fellow and Director of the Evolutionary Biology Program, Canadian Institute for Advanced Research.

REFERENCES

1. **Adachi, J., and M. Hasegawa.** 1992. MOLPHY: programs for molecular phylogenetics, version 2.2. Institute of Statistical Mathematics, Tokyo, Japan.
2. **Araki, H., P. A. Ropp, A. L. Johnson, L. H. Johnson, A. Morrison, and A. Sugino.** 1992. DNA polymerase II, the probable homolog of mammalian DNA polymerase ε replicates chromosomal DNA in the yeast *Saccharomyces cerevisiae*. EMBO J. **11:**733–740.
3. **Beese, L. S., and T. A. Steitz.** 1991. Structural basis for the 3′-5′ exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. EMBO J. **10:**25–33.
4. **Bonner, C. A., S. Hays, K. McEntee, and M. F. Goodman.** 1990. DNA polymerase II is encoded by the DNA damage-inducible *dinA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA **87:**7663–7667.
5. **Braithwaite, D. K., and J. Ito.** 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. **21:**787–802.
6. **Budd, M. E., and J. L. Campbell.** 1993. DNA polymerases delta and epsilon are required for chromosomal replication in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **13:**496–505.
7. **Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. Fitzgerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage,**

B. A. Dougherty, J.-F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. M. Geoghagen, J. F. Weidman, J. L. Fuhrmann, D. Nguyen, T. R. Utterback, J. M. Kelley, J. D. Peterson, P. W. Sadow, M. C. Hanna, M. D. Cotton, K. M. Roberts, M. A. Hurst, B. P. Kain, M. Borodovsky, H.-P. Klenk, C. M. Fraser, H. O. Smith, C. R. Woese, and J. C. Venter. 1996. Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. Science **273:**1058–1073.

8. **Chen, H., C. B. Lawrence, S. K. Bryan, and R. E. Moses.** 1990. Aphidicolin inhibits DNA polymerase II of *Escherichia coli*, an α-like DNA polymerase. Nucleic Acids Res. **18:**7185–7186.
9. **Copeland, W. C., and T. S.-F. Wang.** 1993. Mutational analysis of the human DNA polymerase α. J. Biol. Chem. **268:**11028–11040.
10. **Dalgard, J. Z., and R. A. Garret.** 1993. Archael hyperthermophilic genes, p. 535–564. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of archaea (archaebacteria). Elsevier Science Publishers, Amsterdam, The Netherlands.
11. **De Vega, M., J. Lazaro, M. Salas, and L. Blanco.** 1996. Primer-terminus stabilization at the 3′-5′ exonuclease active site of Φ29 DNA polymerase. Involvement of two amino acid residues highly conserved in proofreading DNA polymerases. EMBO J. **15:**1182–1192.
12. **Dong, Q., and T. S.-F. Wang.** 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.
13. **Doolittle, W. F., and J. R. Brown.** 1994. Tempo, mode, the progenote, and the universal root. Proc. Natl. Acad. Sci. USA **91:**6721–6728.

13a. **Edgell, D. R., and W. F. Doolittle.** Unpublished data.

14. **Elie, C., A. M. De Recondo, and P. Forterre.** 1989. Thermostable DNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. Eur. J. Biochem. **178:**619–626.
15. **Endo, T., K. Ikeo, and T. Gojobori.** 1996. Large-scale search for genes on which positive selection may operate. Mol. Biol. Evol. **13:**685–690.
16. **Felsenstein, J.** 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. **27:**401–410.
17. **Felsenstein, J.** 1996. PHYLIP (Phylogeny Inference Package), version 3.57c. Department of Genetics, University of Washington, Seattle.
18. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. Mckenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirely, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science **269:**496–512.
19. **Forterre, P.** 1992. The DNA polymerase from the archaebacterium *Pyrococcus furiosus* does not testify for a specific relationship between archaebacteria and eukaryotes. Nucleic Acids Res. **20:**1811.
20. **Forterre, P., and C. Elie.** 1993. Chromosome structure, DNA topoisomerases, and DNA polymerases in archaebacteria (archaea), p. 325–361. *In* M. Kates, D. J. Kushner, and A. T. Matheson (ed.), The biochemistry of archaea (archaebacteria). Elsevier Science Publishers, Amsterdam, The Netherlands.
21. **Forterre, P., C. Elie, and M. Kohiyama.** 1984. Aphidicolin inhibits growth and DNA synthesis in halophilic archaebacteria. J. Bacteriol. **159:**800–802.
22. **Fraser, C. M., J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, J. L. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J.-F. Tomb, B. A. Dougherty, K. F. Bott, P.-C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchinson III, and J. C. Venter.** 1995. The minimal gene comlement of *Mycoplasma genitalium*. Science **270:**397–403.
23. **Fuchs, T., H. Huber, S. Burggraf, and K. O. Stetter.** 1996. 16S rDNA-based phylogeny of the archael order *Sulfolobales* and reclassification of *Desulfurolobus ambivalens* as *Acidianus ambivalens* comb. nov. Syst. Appl. Microbiol. **19:**56–60.
24. **Hain, J., W.-D. Reiter, U. Hdepohl, and W. Zillig.** 1993. Elements of an archael promoter defined by mutational analysis. Nucleic Acids Res. **20:** 5423–5428.
25. **Hamal, A., P. Forterre, and C. Elie.** 1990. Purification and characterization of a DNA polymerase from the archaebacterium *Thermoplasma acidophilum*. Eur. J. Biochem. **190:**517–521.
26. **Iwasaki, H., A. Nakata, G. C. Walker, and H. Shinagawa.** 1990. The *Escherichia coli polB* gene, which encodes DNA polymerase II, is regulated by the SOS system. J. Bacteriol. **172:**6268–6273.
27. **Joshi, P., and P. P. Dennis.** 1993. Structure, function, and evolution of the family of superoxide dimutase proteins from halophilic archaebacteria. J. Bacteriol. **175:**1572–1579.
28. **Klenk, H.-P., B. Haas, V. Schwass, and W. Zillig.** 1986. Hybridization homology: a new parameter for the analysis of phylogenetic relations, demonstrated with the urkingdom of the archaebacteria. J. Mol. Evol. **24:** 167–173.
29. **Klimczak, L. J., F. Grummt, and K. J. Burger.** 1985. Purification and characterization of DNA polymerase from the archaebacterium *Sulfolobus acidocaldarius*. Nucleic Acids Res. **13:**5269–5281.
30. **Kong, H., R. B. Kucera, and W. E. Jack.** 1993. Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. J. Biol. Chem. **268:**1965–1975.
31. **Kornberg, A., and T. A. Baker.** 1992. DNA replication, 2nd ed. W. H. Freeman & Co., New York, N.Y.
32. **Kumar, S., K. Tamura, and M. Nei.** 1993. MEGA: molecular evolutionary genetics analysis, version 1.0. The Pennsylvania State University, University Park.
33. **Li, W.-H., and D. Graur.** 1991. Fundamentals of molecular evolution. Sinauer Associates, Inc., Sunderland, Mass.
34. **Mathur, E. J., M. W. Adams, W. N. Callen, and J. M. Cline.** 1991. The DNA polymerase gene from the hyperthermophilic marine archaebacterium, *Pyrococcus furiosus*, shows sequence homology with α-like DNA polymerases. Nucleic Acids Res. **19:**6952.
35. **Nakayama, N., and M. Kohiymama.** 1985. An α-like DNA polymerase from Halobacterium halobium. Eur. J. Biochem. **152:**293–297.
36. **Nei, M.** 1987. Molecular evolutionary genetics. Columbia University Press, New York, N.Y.
37. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. Mol. Biol. Evol. **3:**418–426.
38. **Pisani, F. M., C. De Martino, and M. Rossi.** 1992. A DNA polymerase from the archaeon *Sulfolobus solfataricus* shows sequence similarity to family B DNA polymerases. Nucleic Acids Res. **20:**2711–2716.
39. **Prangishvili, D. A.** 1986. DNA-dependent DNA polymerases of the thermoacidophilic archaebacterium *Sulfolobus acidocaldarius*. Mol. Biol. USSR **20:**477–488.
40. **Prangishvili, D. A., and H.-P. Klenk.** 1994. The gene for a 74 kDa DNA polymerase from the archaeon *Sulfolobus solfataricus*. Syst. Appl. Microbiol. **16:**665–671.
41. **Rossi, M., R. Rella, M. Pensa, S. Bartolucci, M. De Rosa, A. Gambacorta, C. A. Raia, and N. Dell'Aversano Orabona.** 1986. Structure and properties of a thermophilic and thermostable DNA polymerase isolated from *Sulfolobus solfataricus*. Syst. Appl. Microbiol. **7:**337–341.
42. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
43. **Schinzel, R., and K. J. Burger.** 1984. Sensitivity of halobacteria to aphidicolin, an inhibitor of eukaryotic α-type DNA polymerases. FEMS Microbiol. Lett. **25:**187–190.
44. **Sensen, C. W., H.-P. Klenk, R. K. Singh, G. Allard, C. C.-Y. Chan, Q. Y. Liu, S. L. Penny, F. Young, M. Schenk, T. Gaasterland, W. F. Doolittle, M. A. Ragan, and R. L. Charlebois.** 1996. Organizational characteristics and information content of an archaeal genome: 156 kbp of sequence from *Sulfolobus solfataricus* P2. Mol. Microbiol. **22:**175–191.
45. **Sheaff, R., D. Ilsely, and R. Kuchta.** 1991. Mechanism of DNA polymerase α inhibition by aphidicolin. Biochemistry **30:**8590–8597.
46. **Sorokine, I., K. Ben-Mahrez, M. Nakayama, and M. Kohiyama.** 1991. Exonuclease activity associated with DNA polymerases α and β of the archaebacterium *Halobacterium halobium*. Eur. J. Biochem. **197:**781–784.
47. **Swofford, D. L.** 1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Illinois Natural History Survey, Champaign.
48. **Uemori, T., Y. Ishino, K. Toh, I. Asada, and I. Kato.** 1993. Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. Nucleic Acids Res. **21:**259–265.
49. **Uemori, T., Y. Ishino, H. Doi, and I. Kat.** 1995. The hyperthermophilic archaeon *Pyrodictium occultum* has two α-like DNA polymerases. J. Bacteriol. **177:**2164–2177.
50. **Wong, S. W., A. F. Wahl, P.-M. Yuan, N. Arai, B. E. Pearson, K. Arai, D. Korn, M. W. Hunkapiller, and T. S.-F. Wang.** 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.
51. **Zabel, H.-P., H. Fischer, E. Holler, and J. Winter.** 1985. *In vivo* and *in vitro* evidence for eucaryotic α-type DNA polymerases in methanogens. Purification of the DNA polymerase of *Methanococcus vaneilii*. Syst. Appl. Microbiol. **6:**111–118.
52. **Zabel, H.-P., E. Holler, and J. Winter.** 1987. Mode of inhibition of the DNA polymerase of *Methanococcus vannielli* by aphidicolin. Eur. J. Biochem. **165:**171–175.