Evidence of Independent Gene Duplications During the Evolution of Archaeal and Eukaryotic Family B DNA Polymerases

David R. Edgell,¹ Shehre-Banoo Malik,² and W. Ford Doolittle

Canadian Institute for Advanced Research, Program in Evolutionary Biology, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia, Canada

Eukaryotes and archaea both possess multiple genes coding for family B DNA polymerases. In animals and fungi, three family B DNA polymerases, α , δ , and ϵ , are responsible for replication of nuclear DNA. We used a PCR-based approach to amplify and sequence phylogenetically conserved regions of these three DNA polymerases from *Giardia intestinalis* and *Trichomonas vaginalis*, representatives of early-diverging eukaryotic lineages. Phylogenetic analysis of eukaryotic and archaeal paralogs suggests that the gene duplications that gave rise to the three replicative paralogs occurred before the divergence of the earliest eukaryotic lineages, and that all eukaryotes are likely to possess these paralogs. One eukaryotic paralog, ϵ , consistently branches within archaeal sequences to the exclusion of other eukaryotic paralogs, suggesting that an ϵ -like family B DNA polymerase was ancestral to both archaea and eukaryotes. Because crenarchaeote and euryarchaeote paralogs do not form monophyletic groups in phylogenetic analysis, it is possible that archaeal family B paralogs themselves evolved by a series of gene duplications independent of the gene duplications that gave rise to eukaryotic paralogs.

Introduction

Sequencing of eubacterial, archaeal and eukaryotic genomes has revealed that many proteins encoded in these genomes are paralogs (Clayton et al. 1997), members of gene families that share a common evolutionary history because they evolved by gene duplication. Some of these paralogs, such as the elongation factors EF-Tu/ 1α and EF-G/2, are common to eubacteria, archaea, and eukaryotes, because the gene duplications which gave rise to these paralogs occurred in the cenancestor (Gogarten et al. 1989; Iwabe et al. 1989a). Other gene families, however, have more restricted phylogenetic distributions and are often characterized by the presence of multiple paralogs in eukaryotes, but only one or a few paralogs in archaea and eubacteria (for example, RNAdependent DNA polymerases; Iwabe et al. 1989b). Such gene families are of interest because some of these paralogs appear to function in cellular processes, or have biochemical activities, that are specific to eukaryotes.

One such gene family is the eukaryotic nuclear replicative DNA polymerases. Three DNA-dependent DNA polymerases, α , δ , and ϵ , have been identified through genetic and biochemical studies as essential for nuclear DNA replication in the budding yeast, *Saccharomyces cerevisiae* (Budd et al. 1989; Morrison et al. 1990; Budd and Campbell 1993). Sequencing of the genes corresponding to the catalytic subunits of these DNA polymerases revealed amino acid similarity of each of the proteins to one another, to other eukaryotic cellular-encoded DNA polymerases, to plasmid- and viral-encoded

¹ Present address: Center for Molecular Genetics, State University of New York at Albany.

² Present address: Department of Biology, University of Ottawa, Canada.

Key words: *Giardia, Trichomonas,* archaea, DNA polymerase, gene duplication.

Address for correspondence and reprints: David R. Edgell, Center for Molecular Genetics, Biology 338, State University of New York at Albany, 1400 Washington Avenue, Albany, New York 12222. E-mail: buddy@csc.albany.edu.

Mol. Biol. Evol. 15(9):1207-1217. 1998

© 1998 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

polymerases, to archaeal DNA polymerases, and to DNA polymerase II (*polB*) of *Escherichia coli* (Wong et al. 1988; Iwasaki et al. 1991). An additional cellularencoded DNA polymerase, Rev3, was identified in *S. cerevisiae* through a genetic screen for strains displaying reduced susceptibilities to UV mutagenesis (Morrison et al. 1989). Although this DNA polymerase shares sequence similarity with the three nuclear replicative polymerases, it is divergent in amino acid sequence and function. Collectively, these polymerases are classified as family B DNA polymerases (Braithwaite and Ito 1993).

Eukaryotic paralogs α , δ , ϵ , and Rev3 must be the result of gene duplication events that occurred prior to, at the time of, or after the divergence of eukaryotes from an archaea-like ancestor. However, the available collection of eukaryotic family B DNA polymerase sequences is limited to animals, fungi and a few protists. A search for family B DNA polymerase homologs in more deeply diverging eukaryotic lineages should provide information on the time of these duplications in two ways. First, the simple presence of multiple paralogs in a genome indicates that the relevant duplications had already occurred prior to the divergence of that lineage from the main eukaryotic trunk. Second, quantitative phylogenetic analyses should help establish the pattern and order of paralogous gene duplications of family B paralogs in the evolution of eukaryotes or in an archaea-like ancestor.

Curiously, some members of the archaea also possess multiple family B DNA polymerases (Uemori et al. 1995; Edgell, Klenk, and Doolittle 1997; Klenk et al. 1997). One possible interpretation of this observation would be that the multiple archaeal family B DNA polymerases are orthologs of the eukaryotic nuclear replicative DNA polymerases α , δ , and ϵ , with each ortholog performing a specific role at the replication fork. Although drug inhibition studies suggest that archaea use a family B DNA polymerase for DNA replication (Forterre, Elie, and Kohiyama 1984; Schinzel and Burger 1985; Zabel et al. 1985), there is no evidence pointing to a specific function(s) for each of the DNA polymerases at the replication fork, if indeed all archaeal paralogs function in replication. The distribution and numbers of family B DNA polymerases within archaea also confuse issues surrounding the evolutionary history of archaeal and eukaryotic family B paralogs. For instance, the completely sequenced genomes of two euryarchaeotes, Methanococcus jannaschii (Bult et al. 1996) and Methanobacterium thermoautotrophicum (Smith et al. 1997), possess only a single family B paralog, whereas another euryarchaeote, Archaeoglobus fulgidus (Klenk et al. 1997), possesses two family B paralogs. Moreover, the crenarchaeotes Pyrodictium occultum and Sulfolobus solfataricus P2 possess two and three family B DNA polymerases, respectively (Uemori et al. 1995; Edgell, Klenk, and Doolittle 1997). It is unclear if these archaeal family B paralogs arose independently of eukaryotic paralogs by gene duplications after the split of the eukaryotic and archaeal lineages, or if the present distribution of paralogs can be best explained by an ancestral set of gene duplications that occurred before the split of eukaryotes and archaea, followed by loss of certain paralogs in different lineages.

Here, we report the partial sequences of family B DNA polymerases from representatives of two earlybranching eukaryote lineages, the parabasalids and diplomonads, and results of phylogenetic analyses of eubacterial, archaeal, and eukaryotic family B DNA polymerases. Our results suggest that the gene duplications that gave rise to the three eukaryotic nuclear replicative family B DNA polymerases (α , δ , and ϵ) occurred before the divergence of the earliest eukaryotic lineages and that all eukarvotes are likely to possess these three paralogs. At least one of the eukaryotic paralogs, ϵ , appears to be ancestral to both archaea and eukaryotes, as it consistently branches within archaeal sequences to the exclusion of other eukaryotic paralogs. Both the organismal distribution and phylogeny of crenarchaeote and euryarchaote paralogs can be considered evidence supporting independent gene duplications during the evolution of archaeal and eukaryotic family B DNA polymerases.

Materials and Methods

Strains and DNA Extraction

Genomic DNA from *Trichomonas vaginalis* strain NIH-C1 (ATCC#30001) was a kind gift of Dr. Miklos Müller (Rockefeller University). Genomic DNA was isolated from *Giardia intestinalis* (commonly referred to as *G. lamblia* strain WB, ATCC#30957) grown in 15-ml glass culture tubes at 37°C in Keister's modified media supplemented with 250 μ g/ml streptomycin and 165 μ g/ml penicillin. When confluent growth was achieved, cells were pelleted into lysis buffer consisting of 0.5% SDS, 300 μ g/ml proteinase K, 0.1 M NaCl, and 1 mM EDTA and incubated at 50°C for 1 h. This mixture was then extracted with an equal volume of Tris-buffered phenol (pH 8.0) and with phenol/chloroform/isoamyl alcohol (25:24:1 ratio). DNA was precipitated by addition of 2 volumes of ethanol. To remove carbohydrates from

DNA preparations, the ethanol-precipitated DNA was resuspended in H₂O, and NaCl and cetyltrimethylammonium bromide (CTAB; Sigma) were added to final concentrations of 0.7 M and 1%, respectively. This mixture was incubated at 65°C for 30 min and extracted twice with an equal volume of chloroform. CTAB complexes with carbohydrates and forms an insoluble layer between the organic and aqueous layers. The aqueous layer was removed and DNA precipitated by the addition of 2 volumes ethanol and 0.1 volume sodium acetate (pH 5.0).

PCR Amplification, Cloning, and Sequencing

PCR conditions varied depending on the primer combinations and template but typically were carried out in 10 mM Tris-HCl, 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton X-100, 0.2 mg/ml BSA, 0.2 U/µl of Taq polymerase (Gibco-BRL), 5% acetamide (Reysenbach et al. 1992), with primers at 200 nM and genomic DNA at 10–100 ng. Primers are listed in table 1. Denaturation was done at 92°C for 1–2 min, annealing temperature was dependent on primer combinations (between 45°C and 55°C), and extension was done at 72°C for 1–5 min, depending on expected length of target sequence. Amplified bands of the correct molecular weight were gelpurified (Bio-Rad), ligated into a T-tailed vector (pCR2.1, Invitrogen), and either electroporated into DH5a or heat-shock transformed into INVaF (Invitrogen).

To check for insert-carrying plasmids, colonies were toothpicked directly into a $10-\mu l$ PCR reaction containing universal forward and reverse primers (Sandhu, Precup, and Kline 1989). Plasmids corresponding to PCR reactions with bands of the correct size (minus approximately 200 nt for the polylinker) were picked for manual sequencing analysis. Clones that carried fragments similar to family B DNA polymerases were identified by BLAST searches (Altschul et al. 1990). These clones were then used to screen genomic and cDNA libraries and used as probes in Southern hybridizations (Sambrook, Fritsch, and Maniatis 1989). Subclones, in pBluescript or M13, generated from screening of the libraries were sequenced manually and on Applied Biosystems and Licor automated sequencers.

Inverse PCR

Additional coding sequence outside of the original PCR product was obtained for DNA polymerase ϵ of *T. vaginalis* by inverse PCR (Ochman, Gerber, and Hartl 1988). The sequence of the initial PCR product indicated that the coding region contained a *Hin*dIII site; two sets of direct-match primers were designed, one that would amplify the 5' region of the open reading frame (ORF), and one that would amplify the 3' region of the ORF. Primer sequences are inv1: 5' TCT TCA GAG AAC CAT TCT CG 3'; inv2: 5' CGA ATA CTC TCC TTT ACC TG 3'; inv2: 5' AAG ATA CCA GAC TCG ATG GC 3'; and invD: 5' TTC ACC TGT AAT CAT GCA CC 3'. *Trichomonas vaginalis* genomic DNA was cut with *Hin*dIII and self-ligated. PCR reaction conditions were as above except that Tris pH 8.8 was includ-

Amino Acid Sequence	Nucleotide Sequence (5'–3')	Corresponding Amino Acid Positions in <i>Homo sapiens</i> Ortholog
δ-specific		
FDIEC YGFTGA DTDSVM DCPIFY	GGAATTCTTYGATATHGARTGC TATYGGNTTYTAYGGNC CGGGATCCATNACNGARTCNGTRTC GTARAADAGNGGRCARTC	315–319 701–706 755–760 1075–1080
α-specific		
DPDV (I) IV (I) GH DFNSLYPS KKKYAA	GAYCCNGAYRTNATTHRTNGGNC GAYTTYAATWSNCTNTAYCCNTG AARAARAARTAYGCNGC	628–635 855–862 1049–1054
€-specific		
QIMMISY NGDFFDWPF MYPNI ELDTDG	CAGATYATGATGATYTCNTAC AATGGNGAYTTYTTYGATTGGCCNTT GAATTCDATRTTNGGRTACAT CCRTCNGTRTCNAGG	265–271 336–344 601–605 829–834

Table 1 Eukaryotic Family B DNA Polymerase PCR Primers

ed. Annealing was done at 50°C for 1 min, and extension was done at 72°C for 4 min. Products were visualized on agarose gels, gel-purified, and cloned as described above.

Libraries and Southern Hybridizations

A G. intestinalis library in λ gt11 was a gift of Dr. T. Nash (NIH, Bethesda, Md.). Screening procedures were performed as described (Sambrook, Fritsch, and Maniatis 1988) except that filters were prewashed 2 \times 30 min with $0.1 \times SSC$, 1.0% SDS at 65°C. Stringency washes were 2 \times 20 min in 2 \times SSC, and 1 \times 20 min in $1 \times SSC$, 1.0% SDS, both at 65°C. Filters were then exposed to film at -70°C for 2-5 days. Trichomonas vaginalis genomic and cDNA libraries in λ ZAP (Stratagene) were gifts of Drs. Miklos Müller (Rockefeller University) and Patricia Johnson (UCLA). The screening procedure for these libraries was essentially the same as above, except that E. coli XL-1 blue was used as the host strain. In vivo excision of putative positive clones was performed as per manufacturers' instructions (Stratagene).

Southern hybridizations were also as described (Sambrook, Fritsch, and Maniatis 1988). Five micrograms of genomic DNAs from relevant early-branching eukaryotes were digested with various restriction enzymes (New England Biolabs), resolved on 0.7% agarose gels, and transferred to nylon membranes (DuPont). Hybridizations were carried out overnight at 65°C. Stringency washes were as above.

Gene Nomenclature

We propose to name the multiple family B DNA polymerases of crenarchaeotes and euryarchaeotes on the basis of their relationships to one of three family B paralogs that have been sequenced from *S. solfataricus* strain P2. The letter B after each species name refers to family B DNA polymerases, and the number after the letter B refers to the paralog of *S. solfataricus* P2 to which that polymerase appears most related. For instance, the complete genome sequence of the euryar-

chaeote *A. fulgidus* encodes two family B DNA polymerases, one of which appears to be an ortholog of the *S. solfataricus* P2 B2 polymerase. In our proposed nomenclature, this polymerase would be designated *A. fulgidus* B2. Some euryarchaeote genomes (e.g., *M. jannaschii*) encode a single family B DNA polymerase that does not appear to be an ortholog of any of the three *S. solfataricus* P2 paralogs. In these cases, we did not assign a qualifier (e.g., B2) after the species name. Eukaryotic family B DNA polymerases are classified as described (Braithwaite and Ito 1993). We do not refer to eukaryotic family B DNA polymerases as " α -type" (Wong et al. 1988) because, confusingly, one of the four eukaryotic paralogs is designated α .

Phylogenetic Analysis

Family B DNA polymerases are difficult to align at the amino acid level because short, highly conserved functional regions are separated by long stretches of low or no amino acid conservation. Initially, sequences thought to be orthologous were aligned with each other (e.g., all eukaryotic α paralogs) using the PILEUP option of GCG with default values, or with CLUSTAL W with modifed gap penalties (Thompson, Higgins, and Gibson 1994). Separate alignments of orthologs were then hand-edited and combined into a larger alignment consisting of the four eukaryotic paralogs (α , δ , ϵ , and Rev3), archaeal paralogs, and available eubacterial paralogs. The crystal structure of the RB69 phage family B DNA polymerase (Wang et al. 1997) was used to aid the alignment between paralogs. No phage, viral, or plasmid-encoded family B DNA polymerases were used in phylogenetic analysis. All sequences used were either generated in this study or downloaded from public databases, except for the A. fulgidus, Vibrio cholerae, and Pseudomonas aeruginosa sequences, which were obtained from prereleased genome sequences kindly provided by The Institute for Genomic Research (TIGR).

All phylogenetic analyses were performed using amino acids. For parsimony and distance analysis, the

final alignment consisted of 41 taxa and 161 positions. Parsimony analyses used PAUP 3.1 (Swofford 1993). One hundred random-addition replicates with TBR branch swapping were used to search for shortest trees. One hundred replicates of simple addition were used in bootstrap analysis. Distance analysis was performed with PHYLIP 3.57c (Felsenstein 1996) using the PROTDIST program with a Dayhoff weighting matrix. Trees were constructed from distance matrices using the NEIGHBOR program. SEQBOOT and CONSENSE were used in bootstrap analysis.

Two maximum-likelihood (ML) methods were employed. We first used the program PUZZLE (Strimmer and von Haeseler 1996) with 1,000 iterations, the Jones, Thorton, and Taylor substitution matrix, and eight rate categories. We used the resulting tree topology to aid in constraining taxa for the more computationally intensive ML program PROTML (Adachi and Hasegawa 1992). Exhaustive searches were carried out using the Jones, Thorton, and Taylor substitution matrix and user-defined trees. Bootstrap values were calculated using the RELL method on the 1,000 best trees.

Results

Orthologs of the Three Nuclear Replicative Family B DNA Polymerases of Animals and Fungi Are Found in Early-Diverging Eukaryotes

Based on multiple alignments of amino acid sequences of the catalytic subunits of eukaryotic family B DNA polymerases, we designed degenerate PCR primers to amplify paralogs from early-diverging eukaryotes (table 1). It is impossible to design a single primer set to amplify all three family B paralogs from one organism. It is possible, however, to design primer combinations which can do this. Using these primer combinations, we were able to amplify phylogenetically conserved regions of orthologs of DNA polymerases δ and ϵ from the parabasalid *T. vaginalis*, and an ortholog of DNA polymerase α from the diplomonad *G. intestinalis*. PCR products identified as putative family B DNA polymerases using BLAST (Altschul et al. 1990) were used as probes in Southern hybridizations against genomic DNAs from various protists. Each PCR product hybridized to genomic DNA of the organism that initial PCR reactions were performed with, and each DNA polymerase appeared single-copy (not shown). However, we were unable to amplify all three paralogs from a single organism. This result does not necessarily imply that early-diverging eukaryotes do not possess all three paralogs, as there could be a number of reasons why amplification was not successful (for instance, divergent target sequences or biased base composition of the genomic DNA).

It is also impossible to design PCR primers to amplify the entire coding regions of eukaryotic family B DNA polymerases; they are too divergent in sequence, and in *S. cerevisiae*, all paralogs are over 3 kb in coding sequence (the ϵ paralog is approximately 7 kb; Morrison et al. 1990). In addition, the number of phylogenetically informative sites shared between eubacterial, archaeal,

and eukaryotic homologs is less than 200 amino acids; between eukaryotic paralogs, the number of useful sites increases to around 300 amino acids. Additional coding sequence outside of the initial PCR products was obtained for DNA polymerase α of *G. intestinalis* by screening a genomic DNA library in λ gt11. Two *Bam*HI subclones, covering approximately 2.9 kb of coding sequence and all of the phylogenetically informative sites, were isolated from the library.

Screening of both cDNA and genomic DNA libraries of T. vaginalis failed to recover clones carrying additional coding sequence of DNA polymerase ϵ . Since the initial PCR product from T. vaginalis contained a single *Hin*dIII restriction site, two primer sets for use in inverse PCR, each flanking the HindIII site, were designed to amplify additional coding sequence 5' and 3' to that of the PCR product. Inverse PCR reactions resulted in the amplification of 2.1- and 0.9-kb fragments that were cloned. In all, 2.6 kb of coding sequence of DNA polymerase ϵ encompassing all of the phylogenetically informative sites from T. vaginalis was obtained. Attempts to obtain additional coding sequence for DNA polymerase δ of *T. vaginalis* by inverse PCR, screening of both genomic and cDNA libraries, and construction of a size-enriched subgenomic library were unsuccessful, even though the PCR product hybridized to T. vaginalis genomic DNA (not shown).

Eubacterial Family B DNA Polymerases Are Problematic in Phylogenetic Analysis

Phylogenetic analysis of archaeal and eukaryotic family B DNA polymerases is hindered by the lack of appropriate outgroup sequences. For instance, there are very few eubacterial sequences, and those available, E. coli, V. cholerae, and P. aeruginosa, are closely related members of the γ -subdivision of proteobacteria (Woese 1987). Both the V. cholerae and P. aeruginosa sequences were identified by TBLASTN (Altschul et al. 1997) searches of partially completed genome sequences from TIGR using the E. coli family B DNA polymerase as a query sequence. However, both of these sequences are incomplete and do not extend over all phylogenetically conserved regions; the V. cholerae sequence was not used in any analysis. BLASTP and TBLASTN searches of additional partial and complete eubacterial genome sequences available through GenBank and other WWWbased servers failed to identity any other potential eubacterial family B DNA polymerases.

Outgroup sequences with long branch lengths relative to ingroup sequences can be extremely problematic for parsimony analyses, but less so for distance analyses (Swofford et al. 1996). Preliminary PROTDIST analyses were thus performed with a single taxon, *E. coli*, as recommended (Swofford et al. 1996). This resulted in tree topologies that placed *E. coli* as a sister taxon to eukaryotic α , δ , and Rev3 paralogs to the exclusion of other eukaryotic and archaeal paralogs (not shown). For parsimony analysis, both eubacterial taxa (*E. coli* and *P. aeruginosa*) were included, and a topology similar to that in PROTDIST analysis was found. Support for eubacterial sequences grouping with eukaryotic α , δ , and Rev3 paralogs was low, 21% in parsimony analysis and 45% in distance analysis. Support for a clade consisting of α , δ , and Rev3 paralogs and excluding the eubacterial sequences was 28% in parsimony analysis and 30% in distance analysis. When eubacterial sequences were removed from the data set, bootstrap support for a α , δ , and Rev3 clade increased dramatically to 70% in parsimony and 56% in distance analysis. Tree stability as measured by confidence interval (CI) in parsimony analvsis also increased when eubacterial sequences were left out (CI = 0.646), as opposed to when they were included (CI = 0.632). By contrast, removal of other longbranch-length taxa, such as the rapidly evolving S. solfataricus and Sulfolobus shibatae B3 paralogs, did not affect confidence intervals as drastically (CI = 0.635). Because of the uncertainty of branching position of eubacterial sequences within eukaryotic paralogs, and because these sequences appear to decrease measures of tree quality (bootstrap values and CI), eubacterial sequences were not included in further detailed parsimony or distance analyses.

Euryarchaeote and Crenarchaeote Sequences Do Not Form Monophyletic Groups

We first performed phylogenetic analyses on a data set that included only archaeal sequences, and the eukaryotic α and δ paralogs as outgroup sequences. Parsimony and distance analyses always recover topologies that place the archaeal paralogs as a monophyletic assemblage, but they do not recover topologies that group the crenarchaeote and euryarchaeote sequences into two distinct monophyletic groups, as would be expected from other molecular data sets (Woese 1987: Woese et al. 1991). For instance, the recently released genome sequence of the euryarchaeote A. fulgidus contains two family B DNA polymerases; neither of these sequences group with other euryarchaeote sequences (fig. $\overline{2}$). One of the paralogs, which we call A. fulgidus B2, consistently groups with the divergent S. solfataricus P2 B2 paralog with 100% bootstrap support in both methods, while the other paralog groups with the S. solfataricus P2 B3/S. shibatae B3/P. occultum B3 paralogs with moderate bootstrap support (65% for parsimony and 50% for PROTDIST). We call this A. fulgidus paralog B3. This paralog groups with euryarchaeote sequences, excluding all other crenarchaeote sequences, in only 14% of 100 parsimony bootstrap replicates and in 16% of 100 distance bootstraps. The relationship among the remaining crenarchaeote paralogs, typified by the S. solfataricus P2 B1 and B2 sequences, is not well supported by bootstrap analysis. Only 44% of parsimony and 14% of 100 distance bootstrap replicates place the B1 and B2 paralogs as a sister group. However, there is less than 10% bootstrap support in both methods for placing the B2 paralogs alone as a sister group to euryarchaeotes sequences.

The Four Eukaryotic Paralogs Do Not Form a Monophyletic Group to the Exclusion of Archaeal Paralogs

We then performed phylogenetic analysis on a data set that included all archaeal paralogs and the eukaryotic α , δ , ϵ , and Rev3 paralogs. In both parsimony and distance methods, we found that the four eukaryotic paralogs do not form a monophyletic group to the exclusion of archaeal sequences, and one of the eukaryotic paralogs, ϵ , consistently branches within archaeal sequences as a sister group to eurvarchaeotes (fig. 2). This branching pattern is somewhat suspect because of extremely long branch lengths of the eukaryotic ϵ paralogs relative to other eukarvotic and archaeal paralogs, and is only found in 34% of parsimony and 41% of distance bootstrap replicates (fig. 3). The remainder of the bootstrap support was spread between various groupings of euryarchaeote, ϵ , and crenarchaeote B2 and B3 paralogs. However, only 8% of parsimony and 4% of distance bootstrap replicates grouped the four eukaryotic paralogs together to the exclusion of archaeal paralogs.

Maximum-likelihood analysis also suggests that the ϵ -type and euryarchaeote polymerases might be orthologs (figs. 2 and 3). PUZZLE analysis with all archaeal and eukaryotic taxa recovered an euryarchaeote/ ϵ grouping, as in parsimony and distance analyses, with 51% support (fig. 3). In none of the 1,000 quartets analyzed did the eukaryotic ϵ paralog group with α , δ , or Rev3 paralogs. PROTML analysis produced essentially the same results as other methods (figs. 2 and 3). Eukaryotic ϵ paralogs grouped with euryarchaeote sequences (excluding A. fulgidus paralogs), and in only 0.3% of the 1,000 most likely trees did the eukaryotic ϵ , α , δ , and Rev3 paralogs form a clade to the exclusion of archaeal sequences. In addition, visual inspection of the amino acid alignment of family B DNA polymerases indicates that in exonuclease domain II, ϵ and euryarchaeote sequences are remarkably similar (fig. 1) supporting the phylogenetic results.

Bootstrap support for a $\alpha/\delta/\text{Rev3}$ clade was surprisingly low, possibly due to the long branch lengths of the Rev3 paralogs relative to the α and δ paralogs. When Rev3 sequences were removed from the data set, bootstrap values for an α/δ clade increased for both parsimony and distance analyses (fig. 3). Low bootstrap values for an $\alpha/\delta/\text{Rev3}$ clade in distance analysis are also due to an attraction of the rapidly evolving *S. sol-fataricus* P2 B2/A. *fulgidus* B2 paralogs for the Rev3 paralogs. When the archaeal B2 paralogs are removed from the data set, bootstrap support increases to 90% in distance analysis (fig. 3).

As expected, the three DNA polymerases we amplified from representatives of early-diverging protist lineages, δ and ϵ paralogs from *T. vaginalis* and an α paralog from *G. intestinalis*, all grouped with orthologous sequences from other eukaryotes (fig. 2). However, in two of the three eukaryotic subgroups, the α and δ paralogs, sequences we obtained from early-diverging lineages were not basal to other eukaryotic sequences, as would be expected from other molecular phylogenies (Sogin et al. 1989; Cavalier-Smith 1993; Baldauf, Palmer, and Doolittle 1996; Cavalier-Smith and Chao 1996). In both α and δ subtrees, the *P. falciparum* sequences were basal to all other eukaryotes, possibly due to the divergent amino acid sequences and long branch length of the α paralog. When the *P. falciparum* α paralog was

	6		C	10	D	14	15	16	5	L			
	>	_		-==>		<u></u>	— <u> </u>		=>				
S.solfataricus P2 B1	DIKRVAIDIEVY	FPIISI	EYELLGRFFDILLEY	. P. IVLTFNGD	DFDLPYIYFR G	AVVIE	PPAGIFFI	VIY.V	VLDFASLY	PSIIRT.	WNLSYETV		
S.solfataricus P2 B2	ELEKLERIIERL	DPYNLV	Y	YY.MRKRLNVV	N.ETPTVLSQ G	GLILF	POPGCYDI	WY.	OVDFSSMY	PSLIVK.	HNISAETV		
S.solfataricus P2 B3	ELRTIGVDFQIY	DPIVVM	DLKIIRRFVDYILNY	DPDI IFVYDSD	LLPWKYITER K	NIIIQ)PKVGIYTI	vy.	VLDISSVY	SLVIRE	FNIAPDTL		
S.shibatae B3	ELRAIGIDFQIY	DPIVVL	DLKIIRKFVDYILNY	DPDIIYVFDVD	VFHWKYITER K	(TVI.E	PKIGIYSI	vy.	VLDISSVY	.LSVIRF	FNISPDTL	CREN	
S.acidocaldaríus Bl	DIKRVSLDIEVY	FPIISV	EKKLLARLFEIIREY	. P.MLLTFNGD	DFDIPYIYFR G	GAVVII	OPPAGVYFI	vvv.	VLDFASLY	PSIIKN.	WNISYETI		
P.occultum B1	KPRRLAVDIEVF	YPVISV	ERALILEAFRLISNY	. P. VLLTFNGD	NFDLPYLYNR G	GALVLE	PPSGIYFI	viv.	VLDFASLY	PSIIKR.	WNLSYETV		
P.occultum B3	PMRLVAFDIEVY	DPVIIV	DRRVLREFVEYVRAF	DPDIIVGYNSN	HEDWPYLMER G	AVVL	PLKGVHEI	WV.	VLDFSSM	PSIMIK.	YNVGPDTI		
A.fulgidus B2	ELKMLVFDCEML	.PIIVI	SY	DPDIIVGYNQD	AFDWPYLRKR G	GAFVLE	PERGLHEI	WAC	. LDFASMY	PSIMIA.	FNISPDT.		
A.fulgidus B3	LLEALETRYKVE		EKELLEELSGAMEDM	NPDIVVMEDAD	FWMQHLGLKA G	GMIFC	PEPGIYE	ww.	QLDFTSMY	PSIIVK.	HNLSPESI		
P.furiosus	ELKILAFDIETL	GPIIMI	EREMIKRFLRIIREK	DPDIIVTYNGD	SFDFPYLAKR G	GFVKE	EPEKGLWEI	NIV.	YLDFRALY	PSIIIT.	HNVSPDTL	EURY	
Pyrococcus sp.	ELKLLAFDIETL	GPIIMI	EREMIKRFLKVIREK	DPDVIITYNGD	SFDLPYLVKR G	GYVKE	PEKGLWE	GLV.	SLDFRSLY	PSIIIT.	HNVSPDTL		
M.voltae	ELNCIAFDMELY	DPIIMV	EKELIQKTIEILK	QYDVIYTYNGD	NFDFPYLKKR G	GYVRE	EPLKGIQEI	DIV.;	SLDFMSLY	PSILIS.	HNISPETV		
<u> </u>	KLKSVAFDMEVY	DPILMA	EKELIKKIIETLK	EYDVIYTYNGD	NFDFPYLKAR G	GYVKE	EPEKGMFEI	DII.	SMDFRSLY	PSIIIS.	YNISPDTL		
H.sapiens	FV.VLAFDIETT	DQIMMI	EAHLIQRWFEHVQET	KPTIMVTYNGD	FFDWPFVEAR G	GHVEA	ALESGVFR	SLIY	HLDVGAMY	PNIILT.	NRLQPAM	PDCTL	0.17
S.cerevisiae	FV.VMAFDIETT	DQIMMI	EVALLQRFFEHIRDV	RPTVISTFNGD	FFDWPFIHNR G	GHVES	SLEAGVFR	SLIY	HVDVASMY	PNIMTT.	.NRLQPDS	PLAT	
<u> </u>	FPRILAFDIECS	DOIMMI	EREMLKGWFNHIREV	KPHIFVTFNGD	FFDWEFIOER G	GRVEA	IESGLFR	/IIM	HLDVAAMY	PNIILT.	.NRLOPTA		
H.sapiens	PLRVLSFDIECA	DPVIQI	EEDLLQAWSTFIRIM	DPDVITGYNIQ	NFDLPYLIRG G	SATVIE	PLKGYYD	/PIA	FLDFSSLY	PSIMMA.	HNLCYTTL		
S.pombe	PLRIMSFDIECA	DPVIQI	EKTLLEAWNEFIIRI	DPDVLIGYNIC	NFDIPYLLRG G	SATVIE	SPIKGYYD:	PIA:	FLDFSSLY	PSIMMA.	HNLCYTTL	DFT.073	
S.cerevisiae	PLRIMSFDIECA	DPVIQI	EEEMLSNWRNFIIKV	DPDVIIGYNTT	NFDIPYLLRG G	SATVIE	SPIRGYYD'	/PIA	PLDFNSLY	PSIMMA.	HNLCYTTL	DEDIA	
P.falciparum	KLRILSFDIECI	DPIIQI	EKTLLEAWNEFIIRI	DPDFLTGYNII	NFDLPYILNR G	SATVLE	EPIKGYYII	SPIS	FLDFASLY	PSIMIA.	HNLCYSTL		
T.vaginalis	277777777777	??????	???????????????????????????????????????	277777777777	<u> ????????????????????????????????????</u>	??????	?????????	????	?????????	???MIG.	HNICYSTL		
S.cerevisiae Rev3	KIPDPAIDEVSM	IIW.	EFEMFEALTDLVLLL	DPDILSGFEIH	NFSWGYIIER V	/PLVME	EPESAFYKS	SPLIV	VLDFQSLY	PSIMIG.	YNYCYSTM	REV3	
M.musculus Rev3	???????????????????????????????????????	??????	EKALFOEITNIIKRY	DPDILLGYEIO	MHSWGYLLOR V	PLIME	PESRFYSI	ISVL	VLDFOSLY	PSIVIA.	YNYCFSTC		
H.sapiens	• • • • • • • • • • • • • •	• • • • • • •	ERTLLGFFLAKVHKI	DPDIIVGHNIY	GFELEVLLQR G	GLVLE	PKVGFYDI	FIL	LLDFNSLY	PSIIQE.	FNICFTTV		
0.nova	· · · · · · · · · · · · · · · ·		ERQLIEAFVAKIYQL	DPDLMVAHNLC	GGMFDLLLAR G	GLVIE	epkagfydi	4IIL)	LLDFNSLY	PSIIQE.	YNLCFTTV		
T.bruceii	• • • • • • • • • • • • •		ERALLTWFAETLAAL	DPDIIVGHNII	GYTVETLLNR G	GMVLE	PKSGLYSI	EVILI	LLDFNSLY	PSLIQE.	FNVCYTTI	ALPHA	
S.cerevisiae	• • • • • • • • • • • • • •	• • • • • •	EKAMLSCFCAMLKVE	DPDVIIGHRLQ	NVYLDVLAHR G	GLVFF	PEKGLHKI	4XATA	VMDFNSLY	PSIIQE.	FNICFTTV		
P.falciparum	· · · · · · · · · · · · · · ·		EKELLHTFLEKIKDL	DIDIYIGYNIL	NFDLEFLIHR G	GLVLI	PLCGYYD	CFVL	YLDFNSLY	PSIIIE.	YNVCFSTL		
G.intestinalis			ELOLYEELSKYLMHF	NPDIIMGHNIY	FHYNTMWHSR G	GYVML	DPVAGFHD	CIVIN	VLDFNSLY	PNIIRE.	YSLCFTTL		
E.coli	PLKWVS1DIETT	RIVYML	RPOLLEKLNAWFANY	DPDVIIGWNVV	QFDLRMLQKH G	GYVME	SRPGLYDS	S.VL	V.DYKSLY	PSIIRT.	FLIDPVGL	EUB	
P.aeruginosa	REREVSEDIETN	RQVYML	RAGLLERLNQWLAER	DPDAIIGWNLV	QFDLRVLHEH G	ASSLI	RGRGSYDI	RCWC	STN.KSSV	PSWSNR.	FGL		
	Exonuclease		Polymerase Dom	nain TV			Po	lyme	rase Don	main II			
	Domain I		Folymerase Dos	ain iv									
				EXODUCIEAS	e Domain II								
				Exonucleas	e Domain II								
				Exonucleas	e Domain II								
	N		P	Exonucleas	e Domain II Q		20 21		23	24	25		
	N		P	Exonucleas	e Domain II Q	38 -	20 21		23 	24	25 		
S.solfataricus P2 B1	N GITAVITGLLRDF	RVKIYKK	P KA QRAMKVFINAT	KRONUCIOAS YGVFGAE, TFPL	e Domain II Q ENTRODUCTION YAPAVAESVTALG	R VL	20 21	× ·		24	25 DIKGMLVKI	KRNTPEFVK	
S.solfataricus P2 B1 S.solfataricus P2 B2	N 	RVKIYKK KSELK	P KA QRAMKVFINAT AEAIKWILVASJ	EXONUCIONS YGVFGAE, TFPLY FGYLGYR, NSLF(Q Q CONTRACTOR YAPAVAESVTALG GKIEAYEMVTYLA	R VL R VL	20 21 YGDTDSLF HGIIDSLV	v F	23 VAYSGLK	24 KNYFGV KRYNWI	25 DIKGMLVKI IAKGLII	KRNTPEFVK RENMPNIVK	
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3	N — <u>EIIIZZZZ</u> GIIZZZ GIIPEALQWLIER GLYKTFLDELSNV	RVKIYKK KSELK R	P KA QRAMKVFINAT AEAIKWILVASI IKVIEELISSPI	EXONUCIERS YGVFGAE.TFPL FGYLGYR.NSLF NDYVHWV.NARW	Q Q ENTRACTOR YAPAVAESVTALG GKIEAYEMVTYLA YSREIA, SAFDEF	R VL R VL S VI	20 21 YGDTDSLF HGIIDSLV LANDLLIF	v F V F	23 VAYSGLK (ETGLR (SLLVLDN	24 KNYFGV KRYNWI NRYAGL	25 DIKGMLVKJ IAKGL.II DIARKGEEI	KRNTPEFVK RENMPNIVK OMNLCELAR	
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3	N GITAVITGLLRDF GIIPEALQWLIER GLYKTFLDELSNV GLYKTFLEELSNI	RVKIYKK KSELK Q	P CRAMKVFINAT AEAIKWILVASI IKVIEELISSFI . TFLEEELMSSF	KKONUCIERS YGVFGAE.TFPL FGYLGYR.NSLF NDYVHWV.NARW YDYIHWI.NSRW	Q Q CRATHERIC YAPAVAESVTALG GKIEAYEMVTYLA YSREIA.SAFDEF YSREIA.SAVDEF	R VL R VL S VI S VI	20 21 YGDTDSLF HGIIDSLV LANDLLIF LANDFLVF	V F	23 VAYSGLK ETGLR (SLLVLDN (SLLVLDN	24 KNYFGV KRYNWI NRYAGL DRYAGL	25 DIKGMLVKI IAKGL.II DIARKGEEI DIARIGKEI	KRNTPEFVK RENMPNIVK OMNLCELAR ORDLCELVR	CREN
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1	N GITAVITGLLRDF GIIPEALQWLIER GLYKTFLDELSNV GLYKTFLEELSNI GITAVYQGLIRDF	RVKIYKK KSELK Q RVKVYKK	P KA QRAMKVFINAT AEAIKWILVASJ IKVIEELISSFI KA QRAMKVFINAT	KXONUCLEAS YGVFGAE.TFPL FGYLGYR.NSLF NDYVHWV.NARW YGVFGAE.NFPL	Q YAPAVAESVTALG GKIEAYEMVTYLA YSREIA. SAFDEF YSREIA. SAVDEL YAPAVAESVTAIG	R VL R VL S VI S VI R VI	20 21 YGDTDSLF HGIIDSLV LANDLLIF LANDFLVF YGDTDSLF	V F V F V F	23 TVAYSGLK ETGLR SLLVLDN SLLILGN	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYFGT	25 DIKGMLVKI IAKGL.II DIARKGEEI DIARIGKEI EIKGMLAKI	KRNTPEFVK RENNPNIVK DMNLCELAR DRDLCELVR KRNTPEFIK	CREN
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1	N GITAVITGLLRDF GIIPEALQWLIER GLYKTFLDELSNU GITAVYQGLIRDF GLTSQIVGLLRDF	RVKIYKK KSELK R Q R.VKVYKK R.VKIYKK	P CRAMKVFINAT AEAIKWILVASI IKVIEELISSFI TFLEEELISSF KA QRAMKVFINAT KA QAAMKVFINAT	KKONUCIEAS YGVFGAE, TFPL YGVFGAE, NSLF(NDYVHWV, NARW YDYIHWI, NSRW YGVFGAE, NFPL YGVFGAE, SFPF	Q CHATTALIG SKIEAVEMUTYLA YSREIA.SAFDEF YSREIA.SAVDEL YAPAVAESVTAIG YAPAVAESVTAIG	R VL R VL S VI S VI S VI R VI	20 21 YGDTDSLF HGIIDSLV LANDLLIF LANDFLVF YGDTDSLF YGDTDSLF	K F V F V F V F	23 VVAYSGLK (SLLVLDN (SLLULGN (VAYSGLK VVTFSGLK	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYFGT KNYIGA	25 DIKGMLVKI IAKGL.II DIARKGEI DIARIGKEI EIKGMLAKI DVKGMVAKI	KRNTPEFVK RENMPNIVK OMNLCELAR ORDLCELVR KRNTPEFIK KRNTPEFIK	CREN
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B3 P.occultum B3	N GITAVITGLRDF GIIPEALQWLIER GLYKTPLDELSNV GLYKTPLEELSNI GITAVYQGLIRDF GLTSQIVGLLRDY GFFKTVLENLLKL	R. VKIYKK KSELK Q R. VKVYKK R. VKIYKK RROVKEKMK	P CRANKVFINAT: ARAIKWILVASI TFLEELMSSFI KA QRAMKVFINAT: KA QAAMKVFINAT: E OKALKVLANASI	EXONUCIESS YGVFGAE.TFPLY FGYLGYR.NSLFY DDYVHWV.NARW YDYIHWI.NSRW YGVFGAE.SFPFY YGVFGAE.SFPFY YGVFGAE.HARW	Q CREATER AND	R VL R VL S VI S VI R VI R VI R VI	20 21 YGDTDSLF HGIIDSLV LANDLLIF LANDFLVF YGDTDSLF YGDTDSLF YGDTDSLF	K F V F V F V F	23 TVAYSGLK (ETGLR (SLLVLDN (SLLILGN (VAYSGLK (VFFTEAK	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYFGT KNYIGA KRYVGL	25 DIKGMLVKJ IAKGL.II DIARKGEBJ EIKGMLAKJ DVKGMVAKJ DVKGFBAVJ	KRNTPEFVK RENMPNIVK OMNLCELAR DRDLCELVR KRNTPEFIK RGDWCELAK	CREN
S.solfataricus P2 B1 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1 <u>P.occultum B3</u> A.fulgidus B2	N GITAVITELERDF GIPBALGWIER GLYKTFLDELSNU GLYKTFLEELSNI GTAVYQCLIRDF GTSQIVGLERDY GFFFKILERLIEK	RVKIYKK KSELK RVKVYKK RVKVYKK R.OVKEKMK RRELKVELK	P APARKVFINAT APARKVFINAT IKVIEELISSFI TFLEELMSSFF A QRAKKVFINAT KA QAMKVYINAS EF OKALKVLANAST L QQTLKVLTNSF	KKONUCIEAB YGVFGAE.TFPLJ FGYLGYR.NSLF NDYVHWV.NARW YGVFGAE.NFPL YGVFGAE.SFPFJ YGYMGWS.HARW YGYMGWS.LARW	Q CALIFICATION VAPAVAESVTALG GKIEAYEMVTVLA YSREIA. SAVDEL YAPAVAESVTAIG YCKECARAVTAMG YCKECARAVTAMG	R VL R VL S VI S VI R VI R VI R VI	20 21 YGDTDSLF HGIIDSLV LANDLLIF LANDFLVF YGDTDSLF YGDTDSLF YGDTDSIF	K F V F V F V F	23 VAYSGLK ETGLR (SLLVLDN (SLLVLDN (SLLULGN (VAYSGLK (VAYSGLK (VFFTEAK AIFFVE.K	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYFGT KNYIGA KRYVGL KRYAGL	25 DIKGMLVKI IAKGL.II DIARIGKEI EIKGMLAKI DVKGMVAKI DIVGFEAVI VVKGLEVRI	KRNTPEFVK ESMMPNIVK OMLCELAR ORDLCELVR KRNTPEFIK KRNTPEFIK KRDTPEFLK GODWCELAK	CREN
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1 P.occultum B3 A.fulgidus B3 A.fulgidus B3	N 	RVKIYKK KSELK Q R.VKVYKK R.VKIYKK RROLKVELK RRELKVELK RIKTKSLKR	P A QRAMKVFINAT A RAIKWILVASI TFLEELMSSF KA QRAMKVFINAT E OKALKVLANAS NL QQTLKVLTNSF D SILKWALVTCI	EXONUCIESS YGVFGAE.TFPL PGYLGYR.NSLP(NSVHWV.NARW YDYHWI.NSRW YGVFGAE.SFPF YGYGGR.HARW YGYMGWN.LARW YGYMGWN.NAKF	Q YAPAVAESVTALG YKIEAVEMVTYLA YSREIA. SAVDEL YAPAVAESVTALG YAPAVAESVTALG YCKRCAEAVTANG YCKRCAEAVTANG YCKRCAEAVTANG	R VL R VL S VI S VI R VI R VI R VL R VL	20 21 YGDTDSLF HGIIDSLV YGDTDSLF YGDTDSLF YGDTDSLF HGIVDCLW	K FFV FV	23 VAYSGLK (SLLVLDN (SLLVLDN (VAYSGLK VVTFSGLK (VFFTEAK AIFFVE.K TGLLTEV	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYIGA KRYYGL KRYAGL EKFDWI	25 DIKGMLVKI IAKGL.II DIARKGEBI DIARIGKEBI EIKGMLAKI DVKGMVAKI DVVGLEVRI LRGVMARRI	KRNTPEFVK AENMPNIVK OMLCELAR ORDLCELVR KRNTPEFIK KRNTPEFIK KRNTPEFLK GOWCELAK RGDWCELAK	CREN
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1 A.fulgidus B2 A.fulgidus B2 P.furiosus	N GITAVITGLRDF GIIPEALQWLIER GLYKTPLDELSNV GLYKTPLEELSNI GITAVYQCLIRDF GLTSQIVGLIRDF GFFKTLERNLIEK GFFKRILRNLIEK GFFKRILRNLIEK GFFKGLGHLEE	R. VKIYKK KSELK Q R. VKYYKK R. VKIYKK RRELKVELK RIKTKSLKR RIKTKSLKR	P CRACKYFINAT A RAIKWILVASI TFIEEELMSSF KA QRAKKYFINAT KA QAAKKYFINAT FF OKALKVLANASI NL QQTLKVLINSF KD DSILKWALVTCI QKAIKLANSF	YGVFGAE. TFPL FGYLGYR. NSLF NDYVHWV. NARW YDY IHWI. NSRW YGVFGAE. SFPFJ YGVFGAE. SFPFJ YGYMGWS. HARW YGYMGWN. LARW YGYMGWN. LARW YGYYGYR. NARF	Q XAPAVAESVTALG SKIEAVEMVTVLA YSREIA.SAVDEF YSREIA.SAVDEF XAPVAESVTAIG YCRCARAVTANG YCRCARAVTANG YCRCCARAVTANG SRIEVHEEINRIG SRIEVHEEINRIG	R VL R VL S VI S VI R VI R VI R VI R VI R VI R VL R VL	20 21 YGDTDSLF HGIIDSLV YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF		23 VVAYSGLK (SLLVLDN (SLLVLDN) (VAYSGLK (VFFTEAK (VFFTEAK (SFFV.TK) (GFFV.TK	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYIGA KRYVGL KRYAGL EKFDWI KRYAVI	25 DIKGMLVKI IAKGL.II DIARKGEEI DIARIGKEI EIKGMLAKI DVKGMVAKI DVKGLEVKI LRGVMARRI ITRGLEIVI	KRNTPEFVK RENMPNIVK OMNLCELAR RODLCELVR KRNTPEFIK KRNTPEFIK KRNTPEFIK RODWCELAK RODWCELAK ROVELAK	CREN
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.scidocaldarius B1 P.occultum B1 P.occultum B3 A.fulgidus B3 P.furjosus Pyrococcus sp.	N GITAVITGLRDF GIJKATJELSNU GLYKTFLDELSNU GLYKTFLDELSNU GFFKTULSNLKG GFFKRILRMLIEK GFFFKILSRLLEE GFIFSLLGLLEE GFIFSLLGLLEE	RVKIYKK KSELK R RVKYYKK RVKYYKK RROVKEKMK RROVKEKMK RRELKVELK RIKTKSLK RQKIKTKMK QEIIRKMK	P A GRAKKVFINAT A A A KAKVFINAT TFLEELMSSF KA QRAKKVFINAT KA QAAKVVINAS EF OKALKVLANAST NL QQLKVLTNSF NL QQLKVLTNSF S QRAIKLANST	KONUCIESS IGVEGAE. TFPL FGYLCYR. NSLF GCYLCYR. NSLF IGVFGAE. NFPL IGVFGAE. SPPT IGVFGAE. SPPT IGVFGAE. LARW FGYTGYR. NAKF IGYTGYR. KARW	Q CENTRAL AND A CONTRAL AND	R VL R VL S VI S VI R VL R VL R VL R VL R VL R VL R VL	20 21 YGDTDSLF HGIIDSLV YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YIDTDGLY YIDTDGLY	KVVVLIV AFF	23 VAYSGLK (SLLVLDN SLLILGN VATSGLK (VFFTEAK AIFFVE.K CGFFV.TK CGFFV.TK	24 KNYFGV KRYMWI NRYAGL DRYAGL KNYFGT KNYIGA KRYAGL EKFDWI KRYAVI	25 DİKGMLVKİ IAKGL. II DIARKGEB EIKGMLAKİ DİVGFEAVİ UVVGLEVRİ LRGVARRI İTRGLEIVİ İTRGLEIVİ	RRNTPEFVK RENMPNIVK MNLCELAR RODLCELVR KRNTPEFIK KRNTPEFIK GOWCELAK ROVPEVVKR ROWSEIAK	CREN
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.acidocaldarius B1 P.occultum B1 P.occultum B3 A.fulgidus B2 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae	N GITAVITGLRDF GIIPEALQWLIER GLYKTFLEELSNI GITAVYQGLIRDF GFTSULERLIKL GFFKTVLENLLKL GFFKRTLENLIKL GFIPSLLKRLLEE GFIPSLLKRLLEE GFIPSLLKRLLEE	RVKIYKK KSELK Q RVKVYKK R.VKIYKK RRELKVELK RQKIKTKMK RQEIKRYKK RQEIKRYKK	P CALMANA CANANA A CALKWILVASI TFLEELISSFI KA QRAMKVFINAT KA QAAKKVFINAT FO KALKVLANASI D SILKWALVTCI ET QKAIKLANST AS QRAIKILANST M QKSIKVLANSH	KONUCIESS IGVEGAE. TFPLI OFVLGYR. NSLF NDYUHW. NARW IGVEGAE. SPPL IGVEGAE. SPPL IGVEGAE. SARW IGYMGWS. HARW IGYMGYA. KARW IGYYGYA. KARW IGYYGYA. KARW	Q CARAVARSVTALG SKIEAVENVTYLA SYREIA. SAPDEL YAPAVAESVTAIG (XRCARAVTANG VCHCCARATTANG SKIEVHEEINRIG C/CKECARSVTANG UCKECARSVTANG UCKECARSVTANG	R VL S VI S VI R VL R VL R VL R VL R VL R VL	20 21 YGDTDSLF HGIIDSLV LANDFLVF YGDTDSLF YGDTDSLF YGDTDSLF HGIVDCLW YIDTDGLY YIDTDGLY YIDTDGLY	K F V F V F V F V F V F V F V F V F V F V	23 VAYSGLK (ET GLR (SLLVLDN (SLLVLDN (VAYSGLK (VFFTBAK (IFFVE.K (GFFV.TK (GFFV.TK (GFFV.TK	24 KNYFGV KRYNWI DRYAGL DRYAGL KNYIGA KRYVGL KRYAGL EKFDWI KRYALI KRYALI	25 DIKGMLVKI IAKGL. TI DIARKGEBI DIARIGKEI DVKGWVAKI DVVKGLEVKI ITRGLEIVI ITRGLEIVI VVKGLEVKI	KRNT PEFVK RENMPNIVK OMLCELAR ORDLCELVR KRNT PEFIK KGDWCELAK KGDWCELAK ROWSELAK RRDWSEIAK RRDWSEIAK	CREN
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 P.occultum B1 P.occultum B1 A.fulgidus B3 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.jannaschii	K GITAVITELLRDF GITAVITELLRDF GLYKTPLEELSNI GLYKTPLEELSNI GLYKTPLEELSNI GLYSQUGLRDY GPFFRILRNLIK GPIPSLLGHLLEE GPIPSLLGHLLEE GPIPSLLGHLLEE GIIPKTLNELLSR	R. VKIYKK KSELK Q. VKVYKK R. VKIYKK RRELKVELK RIKTKSLKR RQKIKTKMK RQEIKRKMK RKHIKMLKK RKHIKMLKK	P 	KKONUC 1445 KKONGAE. TFPL YGVFGAE. TFPL YGVFGAE. NSRW YGVFGAE. SFPF YGVMGWE. SARW YGVFGAE. SFPF YGYMGWE. LARW FGYTGYR. NAKFY GYYGYR. KARW YGYLAFP. MARW YGYLAFP. RARP	Q APAVAESVTALG KIEAVEMVTYLA SKIEAVEMVTYLA SKIEAVEMVTYLA SKIEAVEMVTYLA SKIEAVEMVTYLA SKIEAVENTATIG VCHECAEVTANG SKIECAEVTANG VCHECAESVTANG VCHECAESVTANG VCHECAESVTANG VCHECAESVTANG	R VL S VI S VI R VL R VL R VL R VL R VL R VL	20 21 VGDTDSLF HGIIDSLV VGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSIF YGDTDSIF YGDTDSIF YGDTDGLY YIDTDGLY YIDTDGFY YIDTDGFY	K FFY FY FY A FFF	23 VAYSGLK CET. GLR CSLLVLDN SSLLILGN VVTFSGLK VVTFSGLK VTFSGLFV.TK GGFV.TK GGFV.TK	24 KNYFGV KRYNWI DRYAGL KNYFGT KNYIGA KRYAGL KRYAVI KRYAVI KKYALI KKYALI KKYALI	25 DIKGMLVKJ IAKGL. 11 DIARKGEEI DIKGMLAKJ DVKGMVAKI DIVGFEAVJ VVKGLEVRI ITRGLEIVI VVKGLEVVI VVKGLEVVI TVKGLEVVI	RRNT PEFVK RENMPAIVK MNLCELAR PRDLCELVR KRNT PEFLK RGDWCELAK RGDWCELAK RDV PEVVKR RRDWSEIAK RRDWSEIAK RRDWSEIAK RRDWSNIAK	CREN EURY
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1 P.occultum B3 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschi1 H.sapiens	N 	R. VKIYKK KSELK Q. VKIYKK R. VKIYKK RROLKVYLK RIKTKSLKR RQKIKTSLKR RQEIRKOM RKHIKMLLK RINIKRPMK RINIKRPMK	P A QRAMKVFINAT A ARAIKWILVASI TFUEBELMSSF KA QRAMKVFINAT E QRAMKVFINASI E QKAIKVLANASI NL QQTLKVLINASI S QRAIKILANSYI XK QLAHKULINSFI KM QKSLKILANSYI	EXONUCIESS VEGVEGAE. TFPL/ RGYLGYR.NSLF/ NDYUHW.NARW YGVFGAE.SPPL/ YGVFGAE.SPPL/ YGYMGME.HARW YGYMGME.HARW YGYYGYA.KARW/ YGYLAFP.NARF/ YGYLAFP.MARW/ YGYLAFP.MARW/ YGYLAFP.MARW/ YGYLAFP.MARW/ YGYLAFP.ARFF.	Q CARACTERISTIC Q VAPAVARSVTALG SKIRAYEMVTYLA SKRIA, SAPDEL VAPAVARSVTAIG VAPAVARSVTAIG VCRCARATTANG SKRCARSVTANG VCRCARSVTANG SSRCARIVTGLG SSRCARIVTGLG SSRCARIVTGLG	R VL R VL S VI R VL R VL R VL R VL R VL R VL A PL	20 21 YGDTDSLF HGIIDSLV YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSIF HGIVDCLM YIDTDGLY YIDTDGFY YIDTDGFY YIDTDGFY		23 VAYSGLK (ETGLR (SLLVLDN (VAYSGLK VYFFTBAK (VFFTBAK (GFFV.TK GGFV.TK GGFV.TK (GLFV.TK (GLFV.TK)	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYIGA KRYAGL EKFDWI KRYAUI KRYAUI KRYAUI KRYAUI KRYAU	25 DIARKGEL DIARKGEEI DIARIGKEE EIKGMLAKI DVKGMVAKI DVKGLEVRI LRGVMARRI ITRGLEIVI VVKGLEVVI TTKGLEIVI TVKGLEVVI ELKGFEVKI	RNTPEFVK RENMPNIVK DNLCELAR RDLCELVR KRNTPEFLK KGDWCELAK GDWCELAK RDWSEIAK RDWSEIAK RDWSEIAK RDWSEIAK RDWSIAK	CREN EURY
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B3 A.fulgidus B2 A.fulgidus B2 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschii K.sapiens S.cerevisiae	N GITAVITGLLRDF GITPALQWLIRE GIYATPLEELSNU GLYKTPLEELSNU GLTSQIVGLLRDY GFFKRILRNLIRK GFLAEVLEPLLRL GFIFRLGKLEE GFIFPLLKRLLDE GIIPFTLNELLSR GLIPKTLNLIEE GLIPKTLNLIEE GLAKTWKKKLSAXI	R. VKIYKK KSELK R. VKVYKK R. VKVYKK RROVKEKMK RRELKVELK RIKTKSLKR RQEITKYMK RQEITKYMK RQEITKYMK RUNIKRRMK VEVGDAAEV VEVGDAAEV DFSDKHARD	P CRAMERYFINAT: AEAIKWILVASI TFULEELMSSEF KA QRAMKVFINAT: KA QAAMKVFINAT: KA QAAMKVFINAT: KA QAAMKVFINAT: E QKAIKLANSFY KD DSILKWALVTCI E QKAIKLANSFY KK QLAHKCILMSEF; KR QLAHKCILMSEF; KR QLAHKCILMSEF; KR QLAHKCILMSEF; KR QLAHKCILMSEF;	EXOTUCIESS INTERNATIONAL INTERNATI	Q CALANCESTRAIG SKIRAYEMVTYLA SKIRAYEMVTYLA SKRIA, SAPUBE YSRIA, SAVDEL TAPAVAESUTAIG VCHCCAEATTAWG SRIEVHEEINRIG VCHCCAESUTAWG VCHCCAESUTAWG VCHCCAESUTAWG VCHCCAESUTAWG VCHCCAESUTAWG VCHCCAESUTAWG VCHCCAESUTAWG VCHCCAESUTAWG	R VL S VI S VI R VL R VL R VL R VL R VL R VL A PL	20 21 YGDTDSLF HGI IDSLV LANDLIF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDGLY YIDTDGLY YIDTDGLY YIDTDGLY YIDTDGFY ELDTDGIW		23 VAYSGLK (ETGLR (SLLVLDN (SLLULGN VVAYSGLK VTFSGLK VTFSGLK (GFFV.TK (GFFV.TK (GLFV.TK (GLFV.TK (GLFV.TK)) (LP.KKLK LP,KKLK	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYFGT KNYIGA KRYAGL KRYAU KRYAU KRYAU KRYAU KRYAU KRYAVF KRYAVF	25 DIKGMLVKI IAKGLII DIARKGEB DIARKGEB EIKGMLAKI DIVGFEAVI UVKGLEVVI ITRGLEIVI ITRGLEIVI TVKGLEEVI TVKGLEEVI ELKGFEVKI ELKGFELKI	RNTPEFVK RENMPNIVK OMNLCELAR RDLCELVR (KNTPEFIK RGDWCELAK RDWSELAK RDWSELAK RDWSELAK RDWSELAK RDWSELAK RDWSELAK RDWSNIAK RRDWSNIAK RRGELQLIK	CREN EURY EPSILON
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1 P.occultum B1 A.fulgidus B3 P.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschii K.sapiens S.cerevisiae T.vaginalis	N GITAVITGLRDF GIJYKTFLDELSNY GUYKTFLDELSNY GUYKTFLDELSNY GUYKTFLDELSNY GFFKRILRNLIK GFFKRILRNLIK GFIFSLIGLLEE GIJPKTLRNLIK GLIPKTLRNLIK GLIPKTLRNLIK GLAKWKKNLSAA GLAKWKKNLSAA	RVKIYKK KSELK Q Q RVKVYKK R.VKIYKK ROVKEKMK ROVKEKMK ROKIKYKM ROKIKYKM ROKIKYKM VEVGDAAEV DFSDKHARD	P A QRAKKVF INAT A A QRAKKVF INAT TFLEELMSSF KA QRAKVF INAT TFLEELMSSF KA QAAKVJINAS EF OKALKVJINSF D SILKWALVTCI ET QKAIKLIANSY DK OKSIKVLANSH KM OKSIKVLANSH KM QLAHKVILNSF EA QLAHKVILNSF EA QLAHKVILNSF	KRONUCIESS GOVERAL TPPLI GOVLGYR.NSLPY NDYUHWI.NSRW YGVFGALS.NPPLI YGVMGWS.HARW YGVMGWS.HARW YGYMGWS.HARW YGYLGYA.KARW YGYLGPP.MARW YGYLAPP.MARW YGYLAPP.RAPY YGYVMRX.GARW YGYVMRX.GARW	Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q Q	R VL S VI S VI R VL R VL R VL R VL R VL A PL	20 21 YGDTDSLE HG I IDSLV LANDFLUF YGDTDSLE YGDTDSLE YGDTDSLE YGDTDSLE HGIVDCLW YIDTDGLY YIDTDGLY YIDTDGLY YIDTDGFY YLDTDGFY ELDTDGIW ELDTDGIW	V FUV FUV FUV FUV FUV FUV FUV FUV FUV FU	23 VVAYSGLK (SLLVLDN (SLLVLDN VVFYESK (VFFTESK (SFFV.TK (GFV).TK (GF	24 KNYFGV KRYNWI NRYAGL DRYAGL KNYFGT KNYIGA KRYAGL EKFPWI KRYAJI KRYAJI KRYAJI KRYAJV KRYAVF KRYAVF KRYAVF	25 DIRKGELUKI IARGL.II DIARKGEEL DIARKGEEL DIARKGEEL DIARKGEEL DIARGEELVI ITRGLEIVI ITRGLEIVI TIRGLEIVI ELKGFELKI ELKGFELKI ELKGFELKI	KRNTPEFVK RENMPNIVK MNLCELAR NDLCELVR KRNTPEFLK KGDWCELAK KGDWCELAK BUVPEVVKR RDWSEIAK RDWSEIAK RDWSEIAK RDWSNIAK RGBLQLIK RGGELQLIK	CREN EURY EPSILON
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.acidocaldarius B1 P.occultum B1 P.occultum B3 A.fulgidus B2 A.fulgidus B2 A.fulgidus B2 P.furiosus Pyrococcus sp. M.voltae M.iannaschii K.sapiens S.cerevisiae T.xaqinalis H.sapiens	N GITAVITGLERF GIYAVTLDELSNU GUYKTPLEELSNI GITAVVQCLIRDF GLTSQIVGLIRDF GFFKTVLENLLKL GFFKTVLENLLKL GFIPSLLGHLLEE GFIPSLLGHLLEE GFIPSLLGHLLES GLIPFTLENLLSR GLAKTWKKLSAA GLAKTWKGNLSXI AKYVKORLNNCK	R. VKIYKK KSELK R. VKIYKK R. VKIYKK RRELKVELK ROXIEXMK ROXIKTKMK RQEIKRVMK RINIKMLK RINIKRMK KINIKRMK DPSDXHARD KHYORPKVE RINALA	P CALAMANIA MANA KA QRAMKVFINATI TFLEELMSSF KA QRAMKVFINAT KA QRAMKVFINAT F QKAIKUAINASI KA QCILKVLAINASI KA QCILKVLAINASI KA QLAIKCLINASI KA QLAIKCLINASI KE QLAIKVILINSI KE QLAIKVILINSI	KONUCIESS GOVERS . TFPL' GOVERS . NSLF DYYHW. NARW YGVFGAB . NFPL' YGVFGAB . NFPL' YGVFGAB . SFPF' YGYMGWS . HARW YGYYGYA . KARW YGYYGYA . KARW YGYYAFP. RAFF' YGYYAFK . GARW YGYVAFK . GSRW YGYVARK . GSRW	Q CENTRAL AND AND AND AND AND AND AND AND AND AND	R VL R VL S VI R VL R VL R VL R VL R VL R VL R VL R VL	20 21 YGDTDSLF HGIIDSLV LANDLIF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YIDTDGLY YIDTDGLY YIDTDGLY YIDTDGLY ELDTDGTM ELDTDGIM ELDTDGIM	V FUV FUV FUV FUV FUV FUV FUV FUV FUV FU	23 VAYSGLK VETGLR SSLLULON VAYSGLK VVTFSGLK VVTFSGLK KOFFV.TK GGFV.TK GGFV.TK GGFV.TK GGFV.TK GGFV.TK LP.KKLK LP.KKLK	24 KINYFGV KRYMMI NRYAGL DRYAGL KNYFGT KNYIGA KRYYGL KRYAGL KRYAGL KKYALI KRYAVI KRYAVI KRYAVF KRYAVF KRYAVF	25 DIKGMLVKI IAKGL. I DIARIGKE DIARIGKE DIKOEPAV UVKGLEVI ITRGLEIVI VVKGLEVI ELKOFPEKI ELKOFPEKI ELKOFPEKI	RNTPEFVK ERNMPNIVK MNLCELAR RRTPEFIK KRNTPEFIK GOWCELAK GOWCELAK ROWSEIAK RRDWSEIAK RRDWSEIAK RRDWSEIAK RRDWSNIAK RRGELQLIK RGELQLIK XGELLIK	CREN EURY EPSILON
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 P.occultum B1 P.occultum B1 A.fulgidus B3 A.fulgidus B3 P.furiosus Pyrococus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae T.vaginalis H.sapiens S.pombe	N GITAVITGLRDF GITAVITGLRDF GLYKTPLBELSNI GLYKTPLBELSNI GLYKTPLBELSNI GLTSQIVGLRDY GPFFRILRNLIEK GPIAEVLEPLLRL GFIPSLLKRLLDE GIIPKTLNELSR GLIPKTLNELSR GLIPKTLNELSR GLAKTWKGNNTCK GLLPQILENLISXI AKYYKGRLNNTCK GLLPQILENLISXI	R. VKIYKK KSELK Q. VKIYKK R. VKIYKK RCVYEEVK RELVYEKK RIKTKSLKR RKIKTKSLKR RKIKTKSLK VEVGDAEV DPSDKHARD KHYOENKY RKRAKAELA	P A GRAMKVFINAT A ARAKVFINAT A ARAKVFINAT TFLEELMSSF KA QAAKVFINAT FF OKALKVLANAS FF OKALKVLANAS KA QAAKVLANAS KA QLAHKVLLANSF KA QLAHKVLLANSF KE QLAHKVLLNSF KE QLALKVSANSV KE QLALKVSANSV	KKONUCIESS (GVFGAE.TPFL) PGYLGYR.NSLP(NDY1HWI.NSRW (GVFGAE.SPF) (GYGGAE.SPF) (GYGGAE.SPF) (GYGGYA.KARW (GYGYAYA.KARW (GYLAPP.ARAP) (GYVMEN.GARW (GYVMEN.VDRW (GYVMEN.VDRW) (GYMGA.VGRL) (GYMGA.VGRL) (GYGGA.VGRL)	Q APADVAESUTALG SKIEAVEMUTYLA SKREIA. SAPDEL SKIEAVEMUTYLA SKREIA. SAPDEL SKIEAVEMUTYLA SKREIA. SAVDEL KAPVAESUTALG SKIEVHEEINRIG VCKECAESUTANG SKIEVHEEINRIG VCKECAESUTANG SKIEVAESUTANG S	R VL R VL S VI R VL R VL R VL R VL R VL R VL R VL R VL	20 21 VGDTDSLE HGIIDSLV LANDLLIF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VIDTDGFV VIDTDGFV VIDTDGFV VIDTDGFV VIDTDGFV VIDTDGFV VGDTDSVW VGDTDSVW	K FRVVV FYVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVV FYVVVVV FYVVVVV FYVVVVV FYVVVVV FYVVVVVVVV	23 VAYSGLK CETGLR (SLLVLDN) (SLLVLDN) (SLLVLDN) (SLLVLDN) (SLLVLDN) (SLLVLN) (SLLVLN) (SLVLN	24 KNYFGV KRYMVI NRYAGL DRYAGL KNYIGT KNYIGT KRYAGL KRYAGL KRYAVI KRYAVI KRYAVF KRYAGL KRYAYF KRYAGL KRYAGL	25 DIKGMLVROI LAKGLII DIARICKEE EIKCMLAKI DIVOKGMVAKI DIVOKGMVAKI DIVOKGMVAKI DIVOKGLEVRI ITROLEIVI ITROLEIVI TVKGLEFVI ELKOFPSKKI ELKOFPSKKI CKGLEAVI DSKGIETVI	RRNTPEFVK RENMPNIVK OMNLCELAR ORDLCELVR KRNTPEFIK RGDWCELAK RODWCELAK RDWSEIAK RRDWSEIAK RRDWSEIAK RRDWSEIAK RRDWSEIAK RGDELQLIK RGELQLIK RGELQLIK RRGELQLIK	CREN EURY EPSILON
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.acidocaldarius B1 P.occultum B1 A.fulgidus B2 A.fulgidus B2 A.fulgidus B2 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae T.vaginalis S.pombe S.cerevisiae	X 	R. VKIYKK KSELK R. VKIYKK ROVKEKMK RROVKEKMK RROVKEKMK ROVKEKMK ROVKEKMK ROVKEKMK ROVIKTON ROVIK RINIKANK RINIKANK RINIKANK RINIKANK RINIKANKA RINIKANKA RINIKANKA RINIKANKAN RINIKANKAN RINIKANKAN RINIKANKAN RINIKANKAN RINIKANKAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKANAN RINIKAN RINIKANAN RINIKAN RINI	P 	EXONUCIESS VECTOR REVEAL REVIEW, NARW REVEAL REVE	Q CARAVARSVTALG VARAVARSVTALG VARAVARSVTALG VSREJA. SAPDEL VARAVARSVTALG VSREJA. SAVDEL VARAVARSVTALG VCHCCARSVTAWG VCHCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCKCCARSVTAWG VCLSISSVTFGG VCLSISSVTFGG VCLAISSSVTAYG	R VL S VI S VI S VI R VL S VI R VL R VL A PL A PL R VI	20 21 yGDTDSLF HGIIDSLV LANDLIF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLY YIDTDGLY YIDTDGLY YIDTDGFY YIDTDGFY YGDTDSVM YGDTDSVM YGDTDSVM		23 VAYSGLK (ETGLR (SILVLIDN (VIAYSGLK VTFSGLK VTFSGLK (VIFYE.K (GFFV.TK (GIFV.TK (F) (GIFV.TK (F) (F) (F) (F) (F) (F) (F) (F)	24 KNYFGV KRYMWI NRYAGL DRYAGL KRYAGL KRYAGL KRYAGL KRYAU KRYAU KRYAU KRYAVF KRYAVF KRYAYF KRYAGL KRYAGL KRYAGL KRYAGL	25 DIKGMLVKI IAKGL.T DIARIGKE EIKGMLAKI DVKGMVAKI DIVGEDAVI VVKGLEVVI ITRGLEIVI VVKGLEVI ELKGFPEKKI ELKGFPEKKI ELKGFELKI ELKGFELKI DCKGLEAVI DSKGLETVI DSKGLETVI DSKGLETVI	RNTPEFVK RENMPNIVK MNLCELAR PRDLCELVR RRNTPEFIK GOWCELAK GOWCELAK ROWSEIAK RDWSEIAK RDWSEIAK RCDWSEIAK RGELQLIK RGELQLIK RGELQLIK RGELQLIK RGDNCPLVS	CREN EURY EPSILON DELTA
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 P.occultum B1 P.occultum B1 P.occultum B1 A.fulgidus B3 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae T.vaginalis S.cerevisiae P.falciparum	N GITAVITCLED GITAVITCLESN GITAVITCLESN GLYKTPLEELSN GLYKTPLEELSN GLYKTPLEELSN GLYKTPLEELSN GUTSQIVGLRDY GFFKRILRALIE GFIPSLLGRLLEE GFIPSLLKRLLDE GIIPKTLNELSR GLAKTWKKNLSAN GLARTWKORLNNFCK GLLPQILENLESA GLIPILDELISA	R. VKIYKK KSELK QQ. R. VKIYKK RROVKEKMK RROVKEKMK RIKTKSLKR QKIKTKMK ROSIKKMK KHIMLK KHYOKPKVE RRNAKALLA RRNAKKLJLR RRNAKKLJLR	P 	KKONUCIESS KKONGAE.TFPL PGYLGYR.NSLF PGYLGYR.NSLF PGYLGYR.NSLF YGVFGAE.SFPF YGVMGWB.LSF YGVTGYR.SARF YGYTGYR.NAKFY GYYGYR.KARW YGYLAFP.RAFY YGYVMRK.GARW YGYLAFP.RAFY YGYVMRK.GGRW GYYMGL.VDRW YGFTGAT.NGRLI YGFTGAT.NGRLI YGFTGAT.VGSGOLI	Q CALANCE AND A CONTRACT OF A	RR VL SS VI SS VI RR VL RR VL RR VL VI RR VL VI RR VL VI AA PL AR VV VI	20 21 VGDTDSLE HGIIDSLU LANDLLIF VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLF VGDTDSLF VGDTDSLF VIDTDGLV VIDTDGLV VIDTDGLV VIDTDGLV VIDTDGLV VGDTDSVM VGDTDSVM VGDTDSVM		23 VVAYSGLK (SLLVLDN SLLIUGN VVTSGLK VVFTBAK GFFV.TK GGFFV.TK GGFFV.TK GGFFV.TK GGFFV.TK GGFFV.TK HP.KLK LP.KGLK PYLLISK PYLLISK PYLLINK	24 KNYFGV KRYMJ NRYAGL KNYFGT KNYIGL KRYAGL KRYAGL KRYAU KRYAU KRYAVF KRYAVF KRYAGL KRYAGL KRYAGL KRYAGL	25 DIKOMLVKI LAKGL. TI DIARKGEBI DIARIGKEJ EIKOMLAKI DVKGLEVVI UVKGLEVVI TVKGLEVVI TVKGLEFVI ELKOFPLKI ELKOFPLKI ELKOFPLKI DCKGLEVU DQKGLASVI DCKGLEVI DCKGLEVI	KRNTPEFVK SENMPNIVK OMNLCELAR ORDLCELVR KRNTPEFIK KRNTPEFIK KRDVELAK ROWCELAK ROWCELAK ROWCELAK ROWCELAK ROWCELAK ROBUCLIK ROBUCLIK RODUCPLVA REDUCPLVA REDUCPLVA REDUCPLVA	CREN EURY EPSILON DELTA
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1 P.occultum B1 A.fulgidus B3 P.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschii K.sapiens S.cerevisiae P.falciparum T.vaginalis T.vaginalis	N GITAVITGLRDF GITAVITGLRDF GIYKTFLDELSNI GIYKTFLDELSNI GIYKTURNLKL GFFKRILRNLBK GFFKRILRNLBK GFIPSLLGRLLEE GIPKTLRLLDE GLIPKTLRLLE GLIPKTLRLLE GLIPKTLRLLE GLIPKTLRLLE GLIPKTLRLLE GLIPKTLRLLE GLIPKTLRLLE GLIPKTLRLE GLIPKTLRLE GLIPKTLRLE GLIPKTLRLE GLIPUTURLE GLIPILDELSA GLLPILDELSA GILPILDELSA	R. VKIYKK KSELK R. VKIYKK RECVKEKKK RRELKVELK RRELKVELK ROSI RROWK RKHIKMLK RINIKREMK VSVGDAREV BYSORAREV BYSORAREV RKRAKLL RKKAKADLK RKKAKLLIK RKKVKLLIK	P A QRAMKVFINAT A QRAMKVFINAT TFLEELMSSF KA QRAMKVFINAT TFLEELMSSF KA QAAKKVINAS EF OKALKVLANAS NL QQTLKVLANAS NL QQTLKVLANAS KA QAAKKULANSF KK QLAHKCILMSF KK QLAHKCILMSF KE QLALKVSANSV KE QLALKVSANSV DE QLALKISANSV NE QLALKISANSV NE QLALKISANSV	EXOTUCI 1445 (GYUGRA . TFPL) RGYLGYR . NSLF NDYYHWI . NSLF NDYYHWI . NSLF (GYUGYR . SFPF (GYMGWS . HARW (GYUGYS . NARF (GYUGYR . NARF (GYUGYA . KARW (GYLAP . MARW (GYLAP . MARW (GYLAP . MARW (GYLAP . MARW (GYLAP . NGRL (GYTGA . VGKL) (GFTGA . VGKL) (GFTGA . VGKL) (GFTGA . VGKL)	Q CAN A Constant of the second secon	RR VIL SS VII SS VI VI RR VI VI RR VI VI RR VI VI RR VI VI RR VI VI RR VI VI RR VI VI RR VI VI RR VI VI RR VI VI VI VI VI VI VI VI VI VI VI VI VI V	20 21 YGDTDSLE HGTIDSL/ LANDFLUF YGDTDSLE YGDTDSLE YGDTDSLE YGDTDSLE YGDTDSLE YGDTDSLE YGDTDSLE YGDTDSLE YLDTDGLW YLDTDGLW YGDTDSW YGDTDSW YGDTDSW YGDTDSW YGDTDSW YGDTDSW		23 VVAYSGLK (SLLVLDN SLLILGN VTFSGLK VTFSGL	24 KNYFGV KRYNWI KRYAGL KNYFGT KNYIGA KRYAGL KRYAU KRYAVI KRYAVI KRYAVF KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL	25 DIKGMLVKI IAKGL. II DIARIGKE EIKGMLAKI DVKGWLAKI DVKGWLAKI ITRGLEIVI ITRGLEIVI ITRGLEIVI VVKGLEVI UKGLEVI CKGLEAVI DSKGIETVI DSKGIETVI DSKGIETVI 2222222	CRNTPEFVK SENMPNIVK MNLCELAR SPDLCELVR KRNTPEFIK KGDWCELAK GOWCELAK SODWCELAK SODWCELAK RDWSEIAK RCDWSEIAK RCDWSEIAK RCDWSIAK RGELQLIK RGELQLIK RGELQLIK RCDCPLVS RCDSCPLVS RCDSCPLVS	CREN EURY EPSILON DELTA
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B3 A.fulgidus B2 A.fulgidus B2 A.fulgidus B2 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vaginalis S.cerevisiae	N GITAVITGLLRDF GITPALQWLIRE GITAVITGLLRDF GIYRTPLEELSNU GLYKTPLEELSNU GJTAVYQGLIRDF GLTSQIVGLLRDY GFFRRILRNLIRE GFIFRLRNLIRE GFIFRLIGGLLEE GFIFRLINGE GLIPRTLNLIEE GLIPRTLNLIEE GLARTWKCNLSXI AKYWCRLNNICK GLLPILLDELISA GILPIILDELISA GILPIILDELISA SILPRILEELLAA	R. VKIYKK KSELK QQ. R. VKIYKK RCVYEKK RCVYEKKK RCVEKEKK RCSIKKOMK ROEIKROMK ROEIKROMK ROEIKROMK VEVGDAEV DFSDRHAB RKRAKALLA RKRAKALLA RKRAKALLA RKRAKALLA RKRAKKLME RKAKKLME RKAKKLME	P 	KONUCIESS IGVEGAE. TFPLI IGVEGAE. TFPLI PGYLGYR. NSLF IGVFGAE. NSPH IGVFGAE. SPPLI IGVFGAE. SPPLI IGVFGAE. HARW IGYMGWS. HARW IGYMGWS. HARW IGYMGWS. HARW IGYMGWS. GARW IGYMGWS. GSRW IGYMGK. GSRW IGYMGK. GSRW IGYMGK. GSRW IGYTGAT. VGRLI IGFTGAT. VGRLI IGFTGAT. VGRLI IGFTGAT. VGRLI IGFTGAT. SGGQLI IGFTGAT. SGGQLI IGFTGAT. SGGQLI IGFTGAT. SGGQLI IGFTGAT. SGGQLI IGFTGAT. SGGQLI IGFTGAT. SGGQLI IGFTGAT. SGGGQLI IGFTGAT. SGGGGLI IGFTGAT. SGGGGGLI IGFTGAT. SGGGGGLI IGFTGAT. SGGGGGLI IGFTGAT. SGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	Q CENTRAL ANALOS VTALG SKIEAYEMVTYLA SYREIA. SAPDEL YAPAVAESVTAIG YAPAVAESVTAIG YAPAVAESVTAIG YAPAVAESVTAIG GYCKECAESVTAWG GYCKECAESVTAWG GYCKECAESVTAWG SYBECAENUTYLG SYBECAENUTYLG SYBECAENUTYLG SYBECAENUTYLG SYBECAENUTYLG SYBECAENUTYLG CIESSVTAYG PCLEISSVTAYG PCLEISSVTAYG PCLEISSVTAYG PCLEISSVTAYG PCLESSVTAYG PCLESSVTAYG	R VL S VI S VI S VI VL R VL R VL S VI VL R VL VL R VL VL R VV VI R VV VI R VV	20 21 VGDTDSLE HGIIDSLU LANDELLIE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE ELDTDGLV VGDTDSVE VGDTDSVE VGDTDSVE VGDTDSVE VGDTDSVE VGDTDSVE		23 VVAYSGLK (SLLVLDN VSLLILGN VVTSGLK (VFFTBAK (UF	24 KNYFGV KRYJWI NRYAGL DRYAGL KNYFGT KRYYGL KRYAGL KRYAGL KRYAF KRYAF KRYAGL KRYA	25 DIKGMLVKI IAKGL.I DIARIGKE DIARIGKE DIARIGKE DIKGEDAVI VVKGLEVVI ITRGLEIVI VVKGLEVVI ITRGLEIVI VVKGLEVVI ELKGPEVKI ELKGFESKI ECKGFESKI DCKGLEAVI DSKGIETVI DCKGIESV	RRNTPEFVK RENMPNIVK ORLCELAR ROLCELVR (RNTPEFIK ROLVELAK RODVELAK RODVELAK RODVELAK RODVELAK RRDWSEIAK RRDWSEIAK RRDWSEIAK RRDUCELUS RRDELLIK RRDLCELVS RRDSCPLVS RRDSCPLVS RRDSCPLVS	CREN EURY EPSILON DELTA
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.shibatae B3 P.occultum B1 P.occultum B1 P.occultum B3 A.fulgidus B3 P.furiosus Byrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vaginalis S.cerevisiae Rev3 M.musculus Rev3	N GITAVITGLRDF GITAVITGLRDF GUYKTPLDELSNU GUYKTPLDELSNU GUYKTPLDELSNU GFFKRILKNLIEK GFFKRILKNLIEK GFIPSLLGRLEF GIPFRLERNLIEF GLHKTVMKKKLSAN GLAKTWKONLSKI GLHKTLMLIEK GLHKTLMLIELSN GLIPFILDLLNA GILPILDELLSN GILPILDELSN GULPFLERLLSN GULPFLERLSN	R. VKIYKK KSELK R. VKIYKK SELK R. VKIYKK RCVYEKKK RKVYKK RKTKSLKR RKTKSLK RKTKSLK RKTKSLK RKTKALK RKRAKALL RKRAKALL RKRAKALL RKRAKALL RKRAKALL	P 	KRONUCIESS GUIDEN GYUFGAE. TFPLI PGYLGYR.NSLFY NDYUHWI.NSRW YGVFGAE.SFPF YGYMGWS.HARW YGVFGAE.SFPF YGYMGWS.HARW YGYLAFP.NARW YGYLAFP.NARW YGYLAFP.NARW YGYLAFP.NARW YGYLAFP.NARW YGYTGAT.NGRLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGYTGASSGGQLI YGYTGASFSGRM	Q Q Q Q Q Q Q Q Q Q Q Q Q Q	R VLI S VI S VI S VI S VI VLI R VLI VLI R VLI VLI R VI VI R VI VV V V V V V V V V V V V V V V V V V	20 21 VGDTDSLF HGTIDSLV LANDLLTE LANDFLVF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VIDTDGLV VIDTDGLV VIDTDGLV VIDTDGLV VGDTDSVF VGDTDSVF VGDTDSVF VGDTDSVF VGDTDSVF VGDTDSVF VGDTDSVF VGDTDSVF VGDTDSVF		23 VAYSGLK (SLLVLDN (SLLVLDN (VAYSGLK VTFSGLF (VFFTBAK (GFFV.TK (GFF	24 KNYFGV KRYMGL KRYMGL KRYAGL KRYAGL KRYAGL KRYALI KRYALI KRYALI KRYAUF KRYAGL KR	25 DIKGMLVKI IARGLII DIARIGKE EIKGMLAKI DIVGOWAKI UVKGLEVKI ITRGLEIVI ITRGLEIVI ITRGLEIVI TVKGLEVVI VVKGLEVVI UVKGELEVI ELKGFESKI ELKGFESKI ECKGFESKI DCKGLEAVI DCKGLEAVI DCKGLEAVI DCKGLETVI DCKGLETVI DAKGIETVI DAKGIETVI	RRNTPEFVK RENMPNIVK MNLCELAR RDLCELVR KRNTPEFIK KGDWCELAK KRNTPEFIK KGDWCELAK RDWSSIAK RDWSSIAK RDWSSIAK RDWSNIAK RGELQLIK RGELQLIK RGELQLIK RGELQLIK RGELQLIK RDSCPLVS RDSCPLVS RDSCPLVS RDSCPLVS RDSCPLVS RDSCPLVS RDSCPLVS	CREN EURY EPSILON DELTA REV3
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1 P.occultum B3 A.fulgidus B2 A.fulgidus B2 A.fulgidus B2 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vacinalis S.cerevisiae P.falciparum T.vacinalis S.cerevisiae P.falciparum T.vacinalis S.cerevisiae Rev3 M.musculus Rev3 H.sapiens	N GITAVITCLLRDF GITPALQWLIER GITAVITCLLRDF GIYKTPLEELSNU GTAVYQGLIRDF GLTSQIVGLLRDY GFFKRILRALIEK GFIRSLIKRLIDE GIIPKTLNELLSR GLIPKTLENLLSR GLIPKTLENLLSR GLLPTILADLLMA GILPIILDELLSA GILPIILVEELLSA GILPIILVEELLSA GILPIILVEELLSA STLSKMITVEELSA	R. VKIYKK KSELK Q R. VKIYKK RCVYEKK RCVYEKKK RCVEKEKK RCVEKEKK RCBIKRCK RCBIKRCK RCBIKRCK RCBIKRCK RCBIKRCK RCBIKRCK RCBIKRCK RCBIK RCK RCBIK RCK RCK RCK RCK RCBIK RCK RCK RCK RCK RCK RCK RCK RCK RCK RC	P CALANCELEVICE A QRAMCVFINAT: A REALKWILVASI TFULEELISSFI KA QRAMKVFINAT: B QAAMKVFINAT: C QRAKVLANASI D SILKWAINTCI ET QKAIKLLANST QCAIKILANST KM QLAHKCILNSFI EA QLAHKVILNSFI EA QLAHKVILNSSI EE QLALKVSANSV KE QLAHKVSANSV KE QLAHKVSANSV EQ LALKUSANSV EQ CLALKUSANSV EA	EXONUCIESS CONTRACTOR REVIEW, NARW REVEAL NSUP CONTRACTOR CONTRACTON CONTRACTON CON	Q CENTROLOGY VAPAVAESVTALG VAPAVAESVTALG VSREIA, SAPDEL VAPAVAESVTALG VAPAVAESVTALG VAPAVAESVTALG VAPAVAESVTALG VARCABAUTANG VCHECAESVTAWG VCHECAESVTAWG VCHECAESVTAWG VCHECAESVTAWG VCHECAESVTAWG VCHECAESVTAWG VCHECAESVTAWG CLEISSSVTAVG PCLEISSVTAVG PCL	R VLISS VI S VI S VI R VLISS VI R VLISS VI R VLISS VI VI R VI VI R VI VI R VI VI R VI VI R VI VI R VI VI R VI VI R VLISS VI VI VI R VLISS VI VI VI VI VI VI VI VI VI VI VI VI VI V	20 21 VGDTDSLE HGTIDSL/ LANDE/LVF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF YGDTDSLW VIDTDGLV YIDTDGLV YDDTDGTW YGDTDSVW YGDTDSVW YGDTDSVW YGDTDSVW YGDTDSVW YGDTDSVW YGDTDSVW		23 VAYSGLK (SLLVLDN (SLLVLDN (SLLVLDN (SLLVLDN (SLFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (GIFV.TK (GIFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK (SLFV.TK) (SLFV.TK (SLFV.TK (SLFV.TK) (SLFV.TK (SLFV.TK) (SLFV.TK (SLFV.TK) (S	24 KNYFGV KRYJWI NRYAGL DRYAGL KRYYGT KRYYGT KRYYGT KRYYGT KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYYGT KRYVTF	25 DİKGMLVKI IAKGL. Tİ DIARICKE EIKGMLAKI DVKOMVAKI DIVGERAVY VVKGLEVVI ITRGLEIVI VVKGLEVI ELKGPEKI ELKGPEKI ELKGPEKI DCKGLEAVI DCKGLEVI DCKGLETVI DAKGIETVI DAKGIETVI DAKGIETVI	RNTPEFVK RENMPNIVK OMNLCELAR RRDTEFLK RRNTPEFLK RRDTEFLK RODVEELAK RODVEELAK RODVEEVKR RRDWSEIAK RRDWSEIAK RRDWSEIAK RRDELLIK RGELQLIK RGELQLIK RCDNCPLVS RRDCPLVS RR	CREN EURY EPSILON DELTA REV3
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 P.occultum B1 P.occultum B1 P.occultum B1 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vaginalis S.cerevisiae Rev3 M.musculus Rev3 H.sapiens O pova	N GITAVITGLRDF GITAVITGLRDF GLYKTPLBELSNI GLYKTPLBELSNI GLYKTPLBELSNI GLYKTPLBELSNI GFFKRILRNLIEK GFIRSLIKRLLDE GIIPKTLNELLSR GLIPKTLNELLSR GLIPKTLNELSR GLAKVWKKLSAN GLAFUKGLNNFCK GLIPKILRLIDE GLIPKUKRLIADLINA GLIPILDELISA GLIPILDELISA GLIPILDELISA GLIPILDELISA GULPILKELLAN STLSKMLTDILDV	RVKIYKK KSELK Q Q RVKIYKK RCVYEKYK RIKTKSLKR RKTKSLKR RKTKSLKR RKTKKLK RKTKKLK RKATADLK RKATADLK RKATADLK RKATALK RKATKLM RKATKLM RKATKLM RKATKLM RKATKLM	P 	KKONUCIESS KKONGLESSE PGYLGYR.NSLPG NDYJHWI.NSLPG NDYJHWI.NSLPG YGVFGAE.SPFP YGYMGWN.LARWF YGYTGYR.NARFF YGYYGYA.KARWF YGYYGYA.KARWF YGYYGYA.KARWF YGYYGYA.KARWF YGYYMER.GRWF YGYYMER.GRWF YGYYMER.VER YGYYMEN.VDRWF YGYTGAT.NGRLI YGFTGAT.NGRLI YGFTGAT.VGKLI YGYTGASFSGRWF YGYYASFSGRWF YGYLGPC SSFF YGYLGPC SSFF YGYLGPC SSFF YGYLGPC SSFF	Q A DOMAIN II Q CANADACASUMATICA SCRIA, SAPDEL SCRIA, SAPDEL VAPAVAESUTAIG VAPAVAESUTAIG VAPAVAESUTAIG VAPAVAESUTAIG SCRIALATIANG SCRIALATIANG SCRIALATIANG SCRIALATIANG SCRIALASUTAIG SCR	R VLI S VI VI R VVI R VVI R VVI VVI R VVI R VVI R VVI R VVI R VVI R VVI R VVI	20 21 VGDTDSLF HGTIDSLV LANDLLTF UGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSV VGDTDSVW VGDTDSVW VGDTDSVW VGDTDSVW VGDTDSVW VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF VGDTDSLF		23 VVAYSGLK SSLLIGN VVAYSGLK VVTFSGLK KUFFTEAK KUFFTEAK KUFFTEAK KUFFTEAK KUFYLISK PYLLISK PYLLISK PYLLISK PYLLISK TYLLISK TYLLISK TYLLISK	24 KNYFGV KRYJMU NRYAGL DRYAGL KNYIGA KRYAGL KRYAGL KRYAVI KRYAVI KRYAVF KRYAGL	25 DIKGMUVRI LAKGLII DIARIGKE EIKGMLAKI DIVKGVAKI DIVKGUSVRI LRGVMARI DIVKGLEVRI ITRGLEIVI ITRGLEIVI ITRGLEIVI TYKGLESVI DKGIETVI DKGIETVI DKGIETVI DAKGIETVI DAKGIETVI ELKGPSKI	KRNTPEFVK RENMPNIVK MNLCELAR NDLCELVR KRNTPEFIK KGDWCELAK KGDWCELAK KGDWCELAK KRDWSEIAK RRDWSEIAK RRDWSEIAK RRDWSIAK KRGELQLIK KRGELQLIK KRGELQLIK KRGELQLIK KRDCPLVS RRDNCPLVS RRDS	CREN EURY EPSILON DELTA REV3
S.solfataricus P2 B1 S.solfataricus P2 B2 S.solfataricus P2 B3 S.solfataricus P2 B3 S.acidocaldarius B1 P.occultum B1 A.fulgidus B2 A.fulgidus B2 A.fulgidus B2 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vaginalis S.cerevisiae P.falciparum T.vaginalis S.cerevisiae Rev3 M.musculus Rev1 M.sapiens O.nova T.bruceii	X GITAVITGLEDF GITAVITGLEDF GIYKTFLDELSNI GIYKTFLDELSNI GIYKTLBELSNI GFYKTULRULKL GFFKRULRULE GFIPSLLGRLLEE GIPKTLRULE GIPKTLRULE GLIPKTLRULE GLIPKTLRULE GLIPKTLRULE GLIPKTLRULE GLIPKTLRULE GLIPKTLRULE GLIPKTLRULE GLIPKTLRULE GLIPKTLRULE GLIPILENLSA GLEPILDELSA GILPILDELSA GILPILDELSA GILPILDELSA GILPRIEKLKE STLSKMITDILDZ GUPKELELAL	R. VKIYKK KSELK R. VKIYKK RECVECKENK RRELKVELK RRELKVELK RRELKVELK RINIKRENK VSVGDAAEV GOEIRROW RINIKRENK RINIKRENK RINIKRENK RINIKRENK RINIKRALLIK RINIKRALLIK RINIKRENK RINIKCOMK RINIKCOMK RINIKCOMK RINIKCOMK	P 	EXONUCIESS STORED ST	Q 	R VLI S VI VI S VI VLI R VLI VLI VLI VLI VLI VLI VLI VLI VLI VLI	20 21 YGDTDSLF HGTIDSLV LANDFLUF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YIDTDGLY YIDTDGLY YIDTDGLY YGDTDSW YGDTDSW YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF		23 VVAYSGLK (SLLVUDN (SLLVUDN VAYSGLK VVFFTBAK (VFFTBAK (GFFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (GLFV.TK (FYLLISK (FY	24 KNYFGV KRYJMU NRYAGL DRYAGL KRYJGA KRYJGA KRYJGA KRYAGL	25 DIKGMLVKI IAKGL. II DIARIGKE EIKGMLAKI DVKGMVAKI DVKGMVAKI DIVGEAU ITRGLEIVI ITRGLEIVI ITRGLEIVI VVKGLEVI UKGLEVI DKGLESVI DCKGLEVI DCK	CRNTPEFVK RENMPNIVK DNLCELAR RODCELVR KRNTPEFLK GODWCELAK GODWCELAK ROWSEIAK ROWSEIAK ROWSEIAK ROWSEIAK REDWSEIAK REDWSEIAK REDUCPLVS REDSCPLVS	CREN EURY EPSILON DELTA REV3
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.solfataricus P2 B3 S.solfataricus P2 B3 S.acidocaldarius B1 P.occultum B1 P.occultum B1 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.furiosus S.cerevisiae P.falciparum T.vaginalis S.cerevisiae P.falciparum T.vaginalis S.cerevisiae P.falciparum T.vaginalis S.cerevisiae M.musculus Rev3 H.sapiens O.nova T.bruceii S.cerevisiae	X GITAVITGLLRDF GITPALQWLIRE GITAVITGLLRDF GITPALQWLIRE GUYKTPLEELSNU GUYKTPLEELSNU GUTSQUVGLLRDY GUTSQUVGLLRDY GPFFRILRAULTRA GFFRFILRAULTRA GIPPTLAGULEPLLRAL GIPPTLAGULEPLLRAL GULPQILENLLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELLSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA GULPTILDELSA	R. VKIYKK KSELK QQ. R. VKIYKK RCVYEKMK RCVYEKMK RCVYEKMK RCVFEKMK RCVFEKMK RCVFEKMK RCVFEKMK RCVFEKMK RCVFE RCVFEKMK RCVFELMK RCVFE	P 	KKONUCIESS KKONGAL TFPL PGYLGYR NSLF PGYLGYR NSLF YGYUGYR AS NFPL YGYUGYR AS SPF YGYMGWR AS NFF YGYUGYR AS NFF YGYUGYR AS NFF YGYUGYR AS NFF YGYUGYR AS SG YGYUGYR ASS YGYUGYR ASS YGYUGYR ASS YGYUGYS ASS YGYTASS YGYTASS YGUSS YGYTASS YGUSS YG	Q CALANCE AND A CONTRACT OF A CONTRACT ON A CONTRACT OF A	R VL R VL S VI S VI R VL S VI R VL	20 21 VGDTDSLE HGIIDSLU LANDELLIE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE ELDTDGLW VGDTDSVW VGDTDSVW VGDTDSVW VGDTDSVM VGDTDSVM VGDTDSSVM VGDTDSSVM VGDTDSSVM		23 VVAYSGLK (SLLVLDN VSLLILGN VVTFSGLK (VFFTBAK (IFFTBAK GFFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (IP, KIL	24 KNYFGV KRYJWI NRYAGL DRYAGL KNYFGT KRYYGL KRYAGL KRYAGL KRYAK KRYAK KRYAK KRYAGL KRYAG	25 DIKGMLVKI IAKGL.I DIARIGKE DIARIGKE DIARIGKE DIKGEDAVI VVKGLEVVI ITRGLEIVI VVKGLEVVI ITRGLEIVI VVKGLEVVI ELKGPEVKI ELKGFESKI DCKGLEAVI DSKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI ELKGLDVI ELKGLDVI	RNTPEFVK RENMPNIVK MNLCELAR RDLCELVR (KNTPEFIK KRNTPEFIK RODVELAK RODVELAK RODVELAK RDWSEIAK RRDWSEIAK RRDWSEIAK RRDWSEIAK RRDUCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLVS RRDSCPLSK RRDWCPLSK RRDWCPLSK	CREN EURY EPSILON DELTA REV3 ALPHA
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.solfataricus P2 B3 S.scidocaldarius B1 P.occultum B1 P.occultum B1 A.fulgidus B3 P.furiosus Byrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vaginalis S.cerevisiae Rev3 M.musculus Rev3 H.sapiens C.nova T.bruceii S.cerevisiae P.falciparum	X 	R. VKIYKK KSELK R. VKIYKK RCVKEKK RRUKVEKK RRUKVEKK RRUKVEKK RRUKTKKK RKIKKK RKIKK RKIKK RKIKK RKRAKKLL RKRAKKLL RKRAKKLL RKRAKKLL RKRAKKLL RKRAKKLL RKRAKKLL RKRAKKLL RKRAKKLL RKRAKKLL RKRAKKLL RKATKLM RKAYKLL RKATKKLM RKAYKLL RKATKKLM RKAYKLL RKAYKLL RKAYKLL RKAYKLL	P 	EXOTUCI 1886 (GYGFGAE TFPL) GYLGYR NSLFY GYUGYR NSLFY (GYUGYGAE SFPF) (GYMGWS HARW (GYUGYS NARW (GYLGYS NARW (GYLAPP MARW (GYLAPP MARW (GYLAPP MARW (GYLAPP MARW (GYLAPP NARW (GYLAPP NARW (GYLAPP NARW (GYLAPP NARW (GYLAPT NGRL) (GYTGAT NGRL) (GYTGAT NGRL) (GYTGAT NGRL) (GYTGAT NGRL) (GYTGAT SFSGRM (GCLGFE SFSFF (GCLGFE SFSFF (GCLGFE SFSFF (GCLGFE SFSFF) (GCLGFE	Q CARACTERISTICS Q VAPAVAESVTALG VSREAVENVTYLA VSREAVENVTYLA VSREAVENVTYLA VSREAVENT VAPAVAESVTATG VCRECAESVTAWG VCRECAESVTAWG VCRECAESVTAWG VCRECAESVTAWG VCRECAESVTAWG VCRECAESVTAWG CCLEVAESVTAWG CCLEVAEVTYNEG CCL	R VILVER	20 21 YGDTDSLF HGTIDSL/ LANDFL/F YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YDDTDGLY YIDTDGLY YDTDGLY YGDTDSVM YGDTDSVM YGDTDSVM YGDTDSVM YGDTDSVM YGDTDSLF YGDTSLF YGDTSL		23 VVAYSGLK (SLLVUDN (SLLVUDN (VAYSGLK VVFFSEAK (UFFFEAK (GFFV.TK (GFFV.TK (GFFV.TK (GFV.TK	24 KNYFGY KRYJWI NRYAGL DRYAGL KRYJGA KRYJGA KRYJGA KRYJGA KRYAGL	25 DIKEMLVKI IARGL.II DIARIGKE BIKOMLAKI DVKOMLAKI DVKOMLAKI ITROLEIVI ITROLEIVI ITROLEIVI TVKGLEVI VVKGLEVI VVKGLEVI DVKGLEVI DXKGIETVI DXKGIETVI DXKGIETVI DXKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI DAKGIETVI	CRNTPEFVK RENMPNIVK MNLCELAR RODLCELVR KRNTPEFIK GODWCELAK GODWCELAK GODWCELAK GODWCELAK ROVSEIAK RCDWSEIAK RCDWSEIAK RCDWSNIAK RCDWSNIAK RCDWSNIAK RCDWSNIAK RCDUCLIX RCDCLVS RCDSCLUS RCD	CREN EURY EPSILON DELTA REV3 ALPHA
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 S.acidocaldarius B1 P.occultum B1 P.occultum B3 A.fulgidus B2 A.fulgidus B2 A.fulgidus B2 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vaginalis S.cerevisiae P.falciparum T.vaginalis S.cerevisiae P.falciparum T.vaginalis S.cerevisiae P.falciparum T.suginalis S.cerevisiae P.falciparum T.bruceii S.cerevisiae	X GITAVITGLLRDF GITPALQWLIRE GITAVITGLLRDF GITPALQWLIRE GUYKTPLEELSNU GUYKTPLEELSNU GUTSQIVGLLRDY GFFRRILRALIEK GFIRSLKRLLDE GFIFRLARVLEPLLRL GIIPFTLRELLSR GLIPFTLRELLSR GLLPFILDELISLA GLLPFILDELISLA GILPFILDELISLA GILPFILDELISLA GILPFILDELISLA GILPFILDELISLA GILPFILDELISLA GILPFILDELISLA GILPFILDELISLA GILPFILDELISLA GUPFILRELLA	R. VKLYKK KSELK QQ. R. VKLYKK RCVYEKK RCVYEKKK RCVEKEKK RCVEKEKK RCELKEKK RCELKEKK RCELKEKK RCELKEKK RCELKEKK RCELKEKK RCATALE RKRAKA RKRAKALE RKRAKALE RKRAKALE RKRAKALE RKRAKA RKRAKALE RKRAK	P 	EXONUCIESS TOTAL REVIEW, NALF REVIEW, REVIE	Q Q CENTROLOGY VAPAVAESVTALG VSREALSAVDEL VAPAVAESVTALG VSREALSAVDEL VAPAVAESVTALG VCRECAESVTAWG VCREC	RRS VII SRRVI VI VI VI VI VI VI VI VI VI VI VI VI V	20 21 VGDTDSLE HGTIDSL/ LANDPLUF VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE VGDTDSLE ELDTDGLW VGDTDSVE VGDTSVE VG		23 VVAYSGLK (SLLVLDN (SLLVLDN (SLLVLDN (SLLVLDN (SLFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (GIFV.TK (G	24 KNYFGV KRYJWI NRYAGL DRYAGL KRYAGL EKFDWI KRYAGL EKFDWI KRYAGL	25 DIRGMLVKI IAKGL.I DIARIGRE EIKGMLAKI DVKOMVAKI DVKGLEVI ITRGLEIVI VVKGLEVI LRGVMARKI ELKGPEVKI ELKGPEKI ELKGPEKI DCKGLEVI DCKG	RRNTPEFVK RENNTPNIVK OWNLCELAR RCDLCELVR (RNTPEFIK GOWCELAK GOWCELAK RCDWSEIAK RCDWSEIAK RCDWSEIAK RCDWSEIAK RCDWSEIAK RCDWSEIAK RCDWSEIAK RCDUCPLVA RCDUCPLVA RCDUCPLVA RCDCCPLVS RCDCCLLIQ 2722727222 RCDCCPLVS RCDCCLVS RCDCCLVS RCDCCLVS RCDCCLVS RCDCCLVS RCDCCLVS RCDCCLSK RCDWCCLSK RCD	CREN EURY EPSILON DELTA REV3 ALPHA
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.shibatae B3 P.occultum B1 P.occultum B1 P.occultum B1 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vaginalis S.cerevisiae Rev3 M.musculus Rev3 H.sapiens O.nova T.bruceii S.cerevisiae P.falciparum G.intestinalis	N GITAVITGLRDF GITAVITGLRDF GLYKTPLBELSNI GLYKTPLBELSNI GLYKTPLBELSNI GLYKTPLBELSNI GFFKRILKNLIEK GFIFSLICHLEE GIPSLICHLEE GIPSLICHLEE GLIPKTLNELSR GLIPKTLNELSR GLIPKTLNELSR GLIPKTLNELSR GLIPKTLNELSR GLIPKTLNELSN GLIPILDELISA GILPIILDELISA GILPIILEELIAA STLSMIITDILDV GULPREIKKUER GULPREIKKUES GULPREIKKUES GULPREIKKUES GULPREIKKUES GULPREIKKUES GULPREIKKUES GULPREIKKUES GULPREIKSUES GULPREIKSUES	RVKIYKK KSELK RVKIYKK RVKIYKK R.VKIYKENK RIKTKSLKR RKTKSLKR RKTKSLKR RKTKSLKR RKTKKLK RKTKKLLK RKTKKLLK RKRAKKDLR RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKKATADLK RKATKKLM RKATKKIM RKATKKIM RKATKKIM RKATKKIM RKATKKIM RKATKKIM RKATKKIM	P 	KKORUCIESE KKORUCIESE PGYLGYR NSLFG NDYJHWI NSLFG NDYJHWI NSLFG YGVFGAE.SFPF YGYMGYB.S.FARW YGVFGAE.SFPF YGYMGYB.NAKF YGYYGYA.KARW YGYLGPP.ARPP MGYVMRL.GRW YGYLGPP.ARPP YGYVMRL.GRW YGYTGAT.NGRLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGFTGAT.VGKLI YGCLGFS.YSFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.ANFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.APF YGCLGYT.NSFF YGCLGYT.NSFF YGCLGYT.APF	Q CONSTRUCTION	R VLL R VLL S VI VI VI VL VL VL VL VL VL VL VL VL VL	20 21 YGDTDSLF HGTIDSL/ LANDFLUF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YDDTDGLY YIDTDGLY YIDTDGY YGDTDSW YGDTDSW YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSLF YGDTDSVF YGDTDSLF YGDTDSVF YGDTDSLF YGDTDSVF YGDTDSLF YGDTDSVF YGDTDSLF YGDTDSVF YGDTDSLF YGDTDSVF YGDTDSLF YGDTDSVF YGDTSVF YGDTDSVF YGDTSVF YGD YGDTSVF YGDTSVF YGDTSVF YGDTSVF YGD YGDTSVF		23 VAYSGLK (SLLVLDN (SLLVLDN (SLLVLDN VAYSGLK VTFSGLF (VFFTPBAK (GFFV.TK (GFFV).TK (GFFV.TK (GFFV.TK (GFFV.TK (GFFV).TK (GFFV.TK (GFFV).TK (GFV).TK	24 KNYFGV KRYJMU NRYAGL DRYAGL KNYIGA KRYAGL KRYAGL KRYAGL KRYANI KRYANI KRYANI KRYANI KRYAGL	25 DIKGMLVKI IAKGLII DIARICKE EIKGMLAKI DIVGPEAVI VVKGLEVKI ITROLEIVI ITROLEIVI ITROLEIVI TROLEIVI VVKGLEVVI VVKGLEVVI DIKGIEVI DIKGIEVI DIKGIEVI DIKGIEVI DIKGIEVI DIKGIEVI DIKGLDVI ELKGFESKI EKGEDI EVKGLDVI VKGLEVI	RRNTPEFVK RENMPNIVK MNLCELAR NDLCELVR KRNTPEFLK KGDWCELAK KRNTPEFLK KGDWCELAK RDWSEIAK RDWSEIAK RDWSEIAK RRDWSEIAK RRDWSNIAK RRGELQLIK RGELQLIK RGELQLIK RCDCCPLVS RCDSCPLVS RCDSCPLVS RCDSCPLVS RCDSCPLVS RCDSCPLVS RCDSCPLVS RCDSCPLSK RCDWCPLSK RCDWCPLSK RCDWCPLSK RCDWCPLSK RCDWCPLSQ REFSTLSK	CREN EURY EPSILON DELTA REV3 ALPHA
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.solfataricus P2 B3 S.acidocaldarius B1 P.occultum B1 P.occultum B3 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae P.falciparum T.vacinalis S.cerevisiae P.falciparum T.vacinalis S.cerevisiae P.falciparum T.vacinalis S.cerevisiae P.falciparum T.bruceii S.cerevisiae P.falciparum G.intestinalis E.coli	X GITAVITGLLRDF GITPALQWLIER GITAVITGLLRDF GITPALQWLIEL GTAVYQGLIRDF GLTSQIVGLLRDY GFFKRILRALIEK GFIRSLIKRLIDE GFIFSLLKRLIDE GFIFSLLKRLIDE GIIPKTLNNLER GLIRTTLNNLER GLIRTTLNNLER GLIRTTLNNLER GLIRTTLNNLER GLIPTLLNLER GLIPTLLNLER GLIPTLLNLER GLIPTLLNLER GLIPTLLNLER GLIPTLENLSA GLIPILVEELISA GILPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILVEELISA GULPILADUVA GULPRILRUNG GULP	R. VKIYKK KSELK Q R. VKIYKK RROVVEKK RROVVEKK RROVEKK RKIKSLK ROEIKROWK ROEIKROWK ROEIKROWK RKIIKROK RKIKKLL RKRAKEL	P 	EXONUCIESS STORES STORESS S	Q Q CENTRAL AND CONTRAL AND	R VILING	20 21 YGDTDSLE HGTIDSLU LANDFLUF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF ELDTDGLW YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSSTF YGDTSSTF Y	KVVVLIIVVVAAAACCCCCVVVV2VLIIIIIIIIIIIIIIIIIIIIVV	23 VVAYSGLK (SLLVLDN (SLLVLDN VVAYSGLK VVFFTBAK (UFFTBAK (GFFV.TK (GFFV.TK (GFFV.TK (GFFV.TK (GIFV.TK (GLFV.TK (LP.KKLK PYLLISK PYLLISK (CUCTK TINLLKK TINLLKK (GAPTGSK (GAPTGSK (GAPTGSK (GAPTGSK (GAPTGSK)	24 KNYFGV KRYJWI NRYAGL DRYAGL KRYYGT KRYYGT KRYYGT KRYYGT KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAAL KKYAAL KKYAAL	25 DIKGMLVKI IAKGL.T DIAKIG.ET DIAKIGKE EIKGMLAKI DVKGMLVKI UVKGLEVVI ITRGLEIVI VVKGLEVI ITRGLEIVI VVKGLEVI ELKGFVEKI ELKGFEKI ELKGFEKI ELKGFEKI DXGLEVI DXGLETVI DXGLETVI UCKGLEVI VKGLEVI VKGLEVI VKGLEVI VFKGLETVI VFKGLETVI	RRNTPEFVK RRNTPEFVK RRNTPEFIK RRNTPEFIK RRNTPEFIK RRNTPEFIK RRDWSELAK RDWSELAK RDWSELAK RDWSELAK RDWSELAK RGELQLIK RGELQLIK RGELQLIK RGELQLIK RGELQLIK RGECLUS RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV3 RCDCFLV4 RCDCFLV4 RCDCFLV4 RCDCFLV4 RCDCFLV4 RCDCFLV5 RCDCFLV7 RCDCFL	CREN EURY EPSILON DELTA REV3 ALPHA EUB
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.solfataricus P2 B3 S.solfataricus P2 B3 S.acidocaldarius B1 P.occultum B1 P.occultum B1 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae T.vaginalis S.cerevisiae P.falciparum T.vaginalis S.cerevisiae P.falciparum T.vaginalis S.cerevisiae P.falciparum C.nova T.bruceii S.cerevisiae P.falciparum C.intestinalis S.cerevisiae D.falciparum C.intestinalis E.coli P.aeruginosa	N GITAVITCLESS GITAVITCLESS GITAVITCLESS GLYKTPLEELSNI GLYKTPLEELSNI GLYKTPLEELSNI GLTSQIVGLRDY GFFKRILRNLIEK GFIPSLLGRLLDE GFIPSLLGRLLDE GFIPSLLGRLLDE GIIPKTLRELSE GLIPKTLRELSE GLIPKTLRELSE GLIPKUERLISE GLIPKUERLSELSS GLLPTILDELSS STLSKMLTDILDVI GULPFILKELLSA STLSKMLTDILDVI GULPFLKELSLAS STLSKMLTDILDVI GULPFLKELSLAS STLSKMLTDILDVI GULPFLKELSLAS GULPFLKELSLAS GULPFLKELSLAS GULPFLKELSLAS	R. VKIYKK KSELK QQ. R. VKIYKK RROVKENKK RROVKENKK RROKEKNK RKIKSLRR QKIKTKMK RQKIKTKMK RKIKKMK KHIKMLK RKAKADLK RKAKADLK RKAKADLK RKAKADLK RKAKADLK RKAKADLK RKAKKDLR RKAKALK RKAKKLM RKAKKLM RKAKKLM RKAKKLM RKALK RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKALK RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKAKLM RKALM RKAKLM RKAKLM RKAKLM RKALM RKAKLM RKAKLM RKAKLM RKALM RKAKLM R	P 	KKONUCIAS KKONUCIAS KKONGAL TFPL PGYLGYR NSLP(KUYUGAL SPP) KUYUGAL SPP KUYUGAL SPP KUYUGAL SPP KUYUGYL ARAW KUYUAPP. ARAW KUYUAPP. ARAW KUYUAPP. ARAW KUYUAPP. ARAW KUYUAPP. ARAW KUYUAPP. ARAW KUYUAPP. ARAW KUYUASP. SGRU KUYUGADA. VORU KUPTGAT. VGLU KUPTGAT. VGKLI KUTGAG. SSRFI KULGPC S. SSRFI KULGLGYL SSRFI KULGLGYL SSRFI KULGLGYL SSRFI KULGLGYL SSRFI KULGLGYL SSRFI KULGLGYL ARAW	Q POMAIN II Q VAPAVAESVTALG SKIEAVEMVTVLA SGRIA. SAPDEL XAPAVAESVTALG VAPAVAESVTALG VAPAVAESVTALG VAPAVAESVTALG VCHCCAEASTTANG SCHCAEATTANG VCHCCAEASTTANG SGRCAEIVTYLG VSHCAEATTANG SGRCAEIVTYLG VSHCAEATTANG SGRCAEIVTYLG VSHCAEATTANG SGRCAEIVTYLG VSHCAEATTANG SGRCAEIVTYLG VSHCAEATTANG SGRCAEIVTYLG VSHCAEATTANG SGRCAEIVTYLG VSHCAEATTANG SGRCAEIVTYLG VSHCAEATTANG CLEISSVTAYG CLEISSVTAYG CLEISSVTAYG CLEISSVTAYG CLEISSVTAYG CLEISSVTAYG CLEISSVTAYG CLEISSVTAYG CLEISSVTAYG VARLALVTYKG VARLALVTYKG VARLALVTYKG VARLALVTYKG VARLALVTYKG VARLASITHSK CCPHVAARLASITHSK C	RR VLL RR VL VI VI VL VL VL VL VL VL VL VL VL VL VL VL VL	20 21 YGDTDSLF HGTIDSLV LANDLLTF LANDLLTF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDST ELDTDGIW ELDTDGIW YGDTDSW YGDTDSW YGDTDSVM YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSSW YGDTDSST YGDTDSST YGDTDSST YGDTDSST YGDTDSST	KVVVLIILLVVVAAAACCCCCVVV2VLIILLLVV	23 VVAYSGLK VSLILGN VSLILGN VVAYSGLR VVFFGLAK VFFTEAK VFFTEAK VFFTEAK VLP.KKLK KGFFV.TK GGFFV.TK GGFV.TK GGFV.TK KGLFV.TK VLP.KKLK VLP.KLISK YFYLLISK YFYLLISK YFYLLISK YFYLLISK YTGLLLKK TTGLLLKK TTGLLLKK TTGLLLKK KTGLLLKK KTGLLLKK KTGLLLKK KTGLLLKK	24 KNYFGV KRYAGL DRYAGL KNYFGT KNYIGA KRYAGL KRYAGL KRYAGL KRYAK KRYAK KRYAK KRYAK KRYAK KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL KRYAGL	25 DIKGMLVEO LARGL.II DIARICKE DIARICKE DIVEGNAK DVKGUEVEI URGUARI UVKGLEVEI TRGLEIU ITRGLEIU ITRGLEIU UKGLEVEI VVKGLEVEI DKGEFEKE ELKGFESKE DCKGLEVEI DCKGIETU DAKGIETU DAKGIETU DAKGIETU DKGLEVEI VKGLDUK EKGLDUK	RRNTPEFVK RRNTPEFVK PRDLCELVR (RNTPEFIK RRNTPEFIK ROWCELAK RODVELAK ROWSEIAK RRDWSEIAK RRDWSEIAK RRDWSEIAK RRDUCPLVA RRDNCPLVS RRDSCPLSK RDSCPLSK	CREN EURY EPSILON DELTA REV3 ALPHA EUB
S.solfataricus P2 B1 S.solfataricus P2 B3 S.solfataricus P2 B3 S.solfataricus P2 B3 S.scidocaldarius B1 P.occultum B1 P.occultum B1 A.fulgidus B3 P.furiosus Pyrococcus sp. M.voltae M.iannaschii H.sapiens S.cerevisiae T.vacinalis S.cerevisiae P.falciparum T.vacinalis S.cerevisiae Rev3 M.musculus Rev3 M.musculus Rev3 S.cerevisiae P.falciparum T.bruceii S.cerevisiae P.falciparum G.intestinalis E.coli	N GITAVITGLERDF GITAVITGLERDF GUYKTPLBELSNI GUYKTPLBELSNI GUTSQIVGULDDY GPFFRILRMLIEK GPIPSLLGHLEBE GIPFTLNELSR GLERTLRMLIEK GLERTLRMLIEK GLERTLRMLIEK GLERTLRMLIEK GLERTLRMLIEK GLERTLRMLIEK GLERTLRMLIELSA GLERTLRMLELSA GLEPTLRELLSA GLEPTLRELLSA GULPILBELSA GULPILBELSA GULPILBELSA GULPILBELSA GULPILBELSA GULPILSELSA GULPILSELSA GULPILSELSA GULPILSELSA GULPILSELSA GULPILSELSA GULPILSELSA	R. VKIYKK KSELK R. VKIYKK SELK R. VKIYKK REVVEENK RIKTKSLKR RKTKSLKR RKTKSLKR RKTKKSLKR RKTKKL VEVGDAAEV DFSDKHARD KHYOKFKC RKRAKALL RKRAKALL RKRAKALL RKRAKALL RKRAKALL RKKAKALL RKRAKALL SKATKKLME RKAKKOLK RKAKALL SKATKKLME RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKKOLK RKAKALL RKAKKOLK RKAKAL RKAKKOLK RKAKAL RKAKKOLK RKAKAL RKAKA RKAKA RKAKAL RKAKA RKAKAL RKAKAL RKAKA RKAKA RKAKAKA RKAKA	P CALIFORMA CALL	KONNELISS KONFALL TFPLI PGYLGYR NSLFY NDYUHWI NARW YGVFGAE SFPFI YGYMGWS HARW YGVFGAE SFPFI YGYMGWS HARW YGYLGYL NARFY YGYLGYL NARFY YGYLGYL ARAW YGYLGYL ARAW YGYLGYL ARAW YGYLGYL ARAW YGYLGYL ARAW YGYLGYL ARAW YGYLGYL ARAW YGYLGYL SFSGRM YGCLGFS SFSF YGCLGFS SFSFF YGCLGYL SFSF YGCLGYL SFF YGCLGYL NRFF YGCLGYL NRFF YGCLGYL NRFF YGCLGYL NRFF YGCLGYL NRFF YGCLGYL ARAW YGCLGYL ARAW YGCLGYL SFSF YGCLGYL SF YGCLGYL	Q CANADAL STANDARD CANADAL STANDARD CANADA	RR VIII RR VIIII VIIIII VIIII VIIII VIIII VIIII VIIIIII VIIII VIIII VIIIIIIII	20 21 YGDTDSLF HGTIDSL/ LANDFLVF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSLF YGDTDSVF YGDTDSVF YGDTDSVF YGDTDSLF YGDTSLF Y	KVVVLIIILIIIVVVAAAACCCCCVVV2VVIIIILIIIVVV	23 VAYSGLK (SLLVLDN (SLLVLDN (VAYSGLK VTFSGLFVLTK (GFFV.TK	24 KNYFGV KRYJMU NRYAGL DRYAGL KRYMGT KRYAGL EKFPWI KRYAGL	25 DIKGMLVKI IARGL.II DIARIGKE BIKGMLAKI DVKGMLAKI DVKGMLAKI DVKGMLAKI ITRGLEIVI ITRGLEIVI TVKGLEVVI VVKGLEVVI VVKGLEVVI DKGIEVI DKGIEVI DKGIEVI DKGIEVI DAKGIEVI DAKGIEVI PAKGLEVI VFKGLEVI VFKGLEVI PPOJme	CRNTPEFVK RENMPNIVK MNLCELAR RDLCELVR KRNTPEFIK SGDWCELAK SGDWCELAK SGDWCELAK RDVSEINK RCDVSEINK RCDVSEINK RGELQLIK RGELQLIK RGELQLIK RGELQLIK RGELQLIK RGELQLIK RGDCELVS RCDVCLVS RCDVCLVS RCDVCLVS RCDVCLX RDDCLX RCDVCLX RCDVCLX RCDVCLX RCDVCLS RCDVCS RC	CREN EURY EPSILON DELTA REV3 ALPHA EUB

FIG. 1.—Amino acid alignment of eubacterial, archaeal, and selected eukaryotic family B DNA polymerase sequences. Numbering of conserved functional domains is as previously published (Wong et al. 1988). Secondary-structure elements corresponding to the family B DNA polymerase of bacteriophage RB69 are indicated by an arrow (for β -sheets) or a hatched rectangle (for α -helices). Lines represent unstructured regions of the protein. Each structural element is assigned a letter or number corresponding to its position in the RB69 DNA polymerase amino acid sequence (6): Sheet 6, 109–117; 10, 211–216; 14, 395–399; 15, 403–405; 16, 407–412; 20, 626–621; 21, 622–626; 23, 700–703; 24, 707–710; 25, 726–728. Helix C, 194–208; D, 222–230; L, 417–424; N, 471–491; P, 547–571; Q, 581–597. The single-amino-acid deletions in exonuclease domain II of crenarchaeote polymerases that support a grouping of B1 orthologs are indicated by boxes. Signature sequences that support a grouping of archaeal and eukaryotic ϵ polymerases are indicated by shaded boxes. Gaps introduced in the alignment are indicated by periods. Missing data are indicated by question marks.

removed from parsimony analyses, the *G. intestinalis* sequence was basal to other eukaryotic α paralogs (not shown).

Discussion

Phylogenetic Analysis of Family B DNA Polymerases Is Confounded by Rapid Rates of Sequence Evolution

Phylogenetic analysis of family B DNA polymerases is problematic for two reasons. First, the number of amino acids that can be aligned with confidence between eubacterial, archaeal, and eukaryotic paralogs is small. Second, a number of archaeal and eukaryotic paralogs exhibit extremely long branch lengths relative to other paralogs. Because these sequences have experienced high rates of amino acid replacements relative to other family B paralogs, they may artifactually attract other rapidly evolving taxa during phylogenetic reconstruction (Felsenstein 1978; Huelsenbeck 1997). One method for dealing with such long-branch effects is to obtain sequence from taxa thought to be intermediate in branching position to the rapidly and slowly evolving



FIG. 2.—Phylogenetic analysis of eukaryotic and archaeal family B DNA polymerases. The tree shown is 1,250 steps (consistency index = 0.644, homoplasy index [HI] = 0.356) and is one of 66 shortest trees found by parsimony analysis with all eukaryotic and archaeal sequences. Bootstrap values are indicated above nodes in the order parsimony/PROTDIST/PROTML. Nodes constrained during PROTML analysis are indicated by an oversized line; no PROTML bootstrap values are indicated at these nodes. The dashed lines connecting the *A. fulgidus* B3 and *P. occultum* B3 sequences to a common branch with *S. solfataricus* P2 B3 and *S. shibatae* B3 sequences indicate that this topology was found preferentially in PROTDIST analysis. Bootstrap values in the order parsimony/PROTDIST/PROTML to the right of the circled and shaded letter "A" indicate support for this topology. The circled and shaded letter "B" refers to bootstrap values (parsimony/distance) supporting a grouping of the α , δ , and Rev3 eukaryotic paralogs when all taxa were included in the analyses. The circled and shaded letter "C" refers to bootstrap values for the same grouping obtained when either Rev3 paralogs (above line) or B2 paralogs (below line) were removed from parsimony and distance analyses. Bootstrap values with asterisks indicate support for nodes found in PROTML analysis when eubacterial outgroup sequences were used.

taxa (Swofford et al. 1996). We attempted to do this by sequencing the DNA polymerase ϵ paralog from *T. vaginalis* and by including eubacterial and archaeal paralogs from recently released genome sequences.

Of all phylogenetic reconstruction methods commonly used, likelihood-based methods perform with higher accuracy and consistency than either parsimonyor distance-based methods in simulations that approximate conditions similar to those seen with the family B DNA polymerase data set, a small number of informative sites, and rapid rates of sequence evolution (Hasegawa and Fujiwara 1993; Kuhner and Felsenstein 1994; Huelsenbeck 1995; Swofford et al. 1996). It is noteworthy, then, that in both likelihood methods employed, PUZZLE and PROTML, eukaryotic ϵ paralogs were never observed to branch with other eukaryotic paralogs, but instead always branched with archaeal paralogs (figs. 2 and 3). Given that this result was also recovered by parsimony and distance methods, it is possible that this topology accurately reflects the evolutionary history of family B DNA polymerases. Low bootstrap support for this grouping is probably due to long branch attractions between various archaeal (e.g., crenarchaeote B2) and eukaryotic (e.g., Rev3) paralogs.

Eubacterial Family B DNA Polymerases: Multiple Independent Losses or Lateral Transfer?

The replicative DNA polymerases of eubacteria (DNA polymerase III, a family C polymerase, and DNA polymerase I, a family A polymerase) and the replicative DNA polymerases of eukaryotes (all family B polymerases) do not share significant primary sequence similar-

Taxa deleted	Tree topology	Bootstrap support at node (pars/dis/ML)
none	eubacteria $\delta/rev3$ α α β β β β β β β β	 重 21/45 重量 100 至 28/30/53 (15) 36/30/40 (41) 至 24/42/78 (40) 章 22/40/77 (51) 40/19/23 (17)
eubacteria	$ \begin{array}{c} \delta/rev3\\ \alpha\\ \hline a\\ B3\\ \hline a\\ B3\\ \hline a\\ B1/B2 \end{array} $	重 70/40/58 茎 26/12/27 茎 30/57/43 醒 34/41/53
eubacteria, cren B2	$ \begin{array}{c} \delta / rev3 \\ \alpha \\ \overline{\alpha} \\ $	董 53/90/97 2 68/70/82 3 52/46/58
eubacteria, rev3	$\begin{bmatrix} \delta \\ \alpha \\ \vdots \\ \vdots \\ B1/B2 \end{bmatrix} = \begin{bmatrix} \delta \\ \alpha \\ \vdots \\ B1/B2 \end{bmatrix}$	1 80/58/56 2 45/14/24 3 33/55/45 3 34/41/55

FIG. 3.—Summary and comparison of bootstrap values found by parsimony, distance, and maximum-likelihood methods for various tree topologies. Each tree topology is a consensus of optimal trees found by parsimony and distance and does not include branch lengths. Boot-strap values are in the order parsimony (pars)/distance (dis)/maximum likelihood (ML). ML values are those found by PUZZLE analysis, except for the tree topology found when all taxa were included, for which both PROTML and PUZZLE (in brackets) values are stated. In addition, node 1a and the dashed line leading to eubacteria indicates that PUZZLE analysis placed the eubacterial sequences as an outgroup to eukaryotic and archaeal paralogs. Euryarchaeote is abbreviated as eury and refers to all euryarchaeote sequences except the *A. fulgidus* paralogs. Crenarchaeote paralogs are abbreviated to B1, B2, and B3 include the paralogs from the euryarchaeote *A. fulgidus*.

ity (Braithwaite and Ito 1993; Edgell and Doolittle 1997). However, one homologous DNA-dependent DNA polymerase, a family B polymerase, is present in representatives of all three domains (Braithwaite and Ito 1993; Iwasaki et al. 1991) suggesting that it must have been present in the genome of the cenancestor. Yet, the cenancestral function of this DNA polymerase is unclear; a change of function(s) must have occurred in either the eubacterial lineage (where the family B homolog *pol*B now functions primarily as a repair polymerase; Bonner et al. 1990; Iwasaki et al. 1990) or the archaeal/eukaryotic lineage (where family B homologs now function as replicative polymerases; Budd et al. 1989: Morrison et al. 1990: Budd and Campbell 1993). Interestingly, the E. coli polB protein can interact with the eubacterial processivity factor, $Pol\beta$ (encoded by the dnaN gene), and with the clamp loading γ complex (Hughes et al. 1991; Bonner et al. 1992). These proteins primarily associate with the *E. coli* replicative polymerase, *pol*C, implying that association of these proteins with *pol*B could confer processive replication on *pol*B. Recent experimental evidence has demonstrated that *pol*B does replicate chromosomal and episomal (F') DNA in dividing cells, but only in the presence of an antimutator allele of *pol*C (Rangarajan et al. 1997). However, it is not clear if *pol*B is used in a replicative function in logarithmically growing wild-type cells.

If a family B DNA polymerase was encoded in the genome of the cenancestor, the present distribution of eubacterial homologs is most confusing. Family B homologs have only been found in three eubacteria, E. coli, V. cholerae, and P. aeruginosa, all closely related members of the y-subdivision of proteobacteria (Woese 1987). A family B DNA polymerase is missing, or has diverged so much in primary sequence as to be unrecognizable by common database search algorithms, from the completely sequenced eubacterial genomes of Haemophilus influenzae (Fleischmann et al. 1995), Mycoplasma genitalium (Fraser et al. 1995), Mycoplasma pneumoniae (Himmelreich et al. 1996), Synechocystis sp. strain PCC6803 (Kaneko et al. 1996), Helicobacter pylori (Tomb et al. 1997), and Borrelia burgdorferi (Fraser et al. 1997) and from many partially sequenced eubacterial genomes (see Results). The observed distribution of eubacterial family B homologs suggests that multiple independent losses have occurred along eubacterial lineages, except in the lineage leading to the γ subdivision of proteobacteria, after eubacteria diverged from a common ancestor with archaea and eukaryotes. It is also possible that a family B homolog was present in the cenancestor, was lost only once in the common ancestor of eubacteria after the divergence of the eubacterial and archaeal/eukaryotic lineages, and has since been reacquired by lateral transfer in the γ -proteobacterial lineage from a noneubacterial source. Alternatively, a family B DNA polymerase(s) might not have been present in the cenancestor at all, but might have evolved by gene duplication along the archaea/eukaryotic lineage after the divergence of eubacterial and archaeal/ eukaryotic lineages from the cenancestor. A family B DNA polymerase would subsequently have been acquired by γ -proteobacteria through lateral transfer from a noneubacterial source.

Phylogenetic Analysis is Suggestive of Multiple Independent Gene Duplications During the Evolution of Archaeal and Eukaryotic Family B DNA Polymerases

The finding of multiple family B DNA polymerases in eukaryotes and some (Uemori et al. 1995; Edgell, Klenk, and Doolittle 1997; Klenk et al. 1997), but not all, archaea (Bult et al. 1996; Smith et al. 1997) raises a number of interesting questions concerning the evolution of archaeal and eukaryotic DNA polymerases. The foremost question is whether the gene duplications that gave rise to the multiple archaeal and eukaryotic paralogs occurred independently of one another after the split of the archaeal and eukaryotic lineages, or whether the gene duplications occurred in a common ancestor of archaea and eukaryotes. Our phylogenetic analyses are suggestive of multiple independent gene duplication events in the evolution of archaeal and eukaryotic family B paralogs, all of which occurred after the divergence of archaea and eukaryotes from the cenancestor. Although phylogenetic analysis of archaeal and eukaryotic family B DNA polymerases is problematic because some paralogs have extremely long branch lengths and because there are a limited number of phylogenetically informative sites, we can, with some certainty, make the following conclusions concerning the evolutionary history of family B DNA polymerases.

First, the gene duplications that gave rise to the three eukaryotic replicative paralogs (α , δ , and ϵ) occurred prior to the divergence of the earliest eukaryotic lineages. We feel that this conclusion is well supported, because we have sequenced phylogenetically conserved regions of these paralogs from representatives of earlydiverging lineages (fig. 1) and because these eukaryotic paralogs form monophyletic groups in phylogenetic analysis (fig. 2). Another gene duplication, which involved a δ -type paralog and give rise to the Rev3 paralogs found in M. musculus and S. cerevisiae, occurred at some point during eukaryotic evolution. As this paralog has only been sequenced from representatives of late-diverging groups (animals and fungi), it is impossible to determine if this duplication event also occurred early in eukaryotic evolution.

Second, the common ancestor of archaea and eukaryotes likely possessed two family B DNA polymerases. The observations that archaeal sequences are split into two phylogenetic groups, that representatives of both eurvarchaeotes and crenarchaeotes possess multiple paralogs that fall into each of these two groups, and that one eukaryotic paralog, ϵ , consistently branches within archaeal sequences all support this conclusion. One of the ancient polymerases was ancestral to present-day euryarchaeote, archaeal B3, and eukaryotic ϵ paralogs, because these sequences group together in phylogenetic analysis. The second ancient family B polymerase was probably ancestral to present-day archaeal B2 paralogs, because B2 orthologs are found in both crenarchaeotes (e.g., S. solfataricus P2 B2) and in euryarchaeotes (e.g., A. fulgidus B2).

Third, one of the ancestral family B DNA polymerases was lost from some, but not all, euryarchaeotes after the split of the two archaeal kingdoms. The finding of only a single family B paralog in the completely sequenced genomes of *M. jannaschii* (Bult et al. 1996) and *M. thermoautotrophicum* (Smith et al. 1997), but two family B paralogs in the completely sequenced *A. fulgidus* genome (Klenk et al. 1997), suggests that the ancestral B2 paralog was lost from these archaeal lineages. Crenarchaeote B1 paralogs evolved as the result of a gene duplication event that occurred after the divergence of the two archaeal kingdoms.

However, because the branching pattern between some archaeal paralogs is not supported by high bootstrap values, we can envision an alternative explanation that as well explains the distribution and phylogenetic relationships of archaeal and eukaryotic paralogs as

those presented above. In this alternate scenario, only a single family B DNA polymerase predates the divergence of archaea and eukaryotes and was ancestral to present-day euryarchaeote, crenarchaeote B3, and eukaryotic ϵ paralogs, as well as one of the two A. fulgidus paralogs. All remaining crenarchaeote, euryarchaeote, and eukaryotic family B paralogs evolved by duplication(s) from this ancestral paralog before the divergence of archaeal and eukaryotic lineages. This scenario would imply that present-day archaeal and eukaryotic paralogs are actually orthologs (for example, eukaryotic α and crenarchaeote B1 sequences), but for reasons stated above, we cannot accurately reconstruct the phylogenetic relationship of these family B DNA polymerases. The absence of B2 paralogs from euryarchaeote genomes (excluding A. fulgidus) is best explained by loss of this paralog from the genome of the common ancestor of euryarchaeotes after the divergence of the lineage that gave rise to A. fulgidus.

Specialization of DNA Polymerase Function in Archaea and Eukaryotes

Another question of interest concerns the function(s) of eukaryotic and archaeal family B DNA polymerases that are likely orthologs, the euryarchaeote polymerases (excluding A. fulgidus paralogs) and ϵ -type paralogs of eukaryotes. Since their divergence, euryarchaeote and eukaryotic ϵ polymerases have independently undergone numerous changes in structure and function. For instance, the catalytic subunits of S. cer*evisiae* and *H. sapiens* DNA polymerase ϵ are both over 2,200 amino acids in length (Morrison et al. 1990; Kesti, Frantti, and Syvaoja 1993), yet euryarchaeote and crenarchaeote B3 polymerases are under 900 amino acids in length (see, e.g., Uemori et al. 1993). These additional amino acids of DNA polymerase ϵ , present as a long carboxy-terminal extension relative to other family B paralogs, are implicated in cell cycle regulation, as deletion of this region interferes with a DNA replication checkpoint in S phase (Navas, Zhou, and Elledge 1995). Since archaea are not likely to possess a eukaryote-like cell cycle (or the elaborate checkpoint controls associated with one), it is probable that this region of DNA polymerase ϵ was acquired early in the evolution of eukaryotes.

Regardless of the exact cellular function(s) and biochemical activities of archaeal and eukaryotic family B paralogs, it is clear that archaea and eukaryotes share many more similarities in DNA replication machinery than either do with eubacteria (Edgell and Doolittle 1997). Most of the protein components shared between archaea and eukaryotes must have been present, and functioning in a replicative function, in a common ancestor. At least one of these components, family B DNA polymerases, has been subject to selection for expanded function(s), whether by duplication or addition of function-specific domains, since archaea and eukaryotes diverged.

Acknowledgments

We would like to thank members of the Doolittle lab for helpful discussions and comments on the manuscript and Sandie Baldauf for help with phylogenetic analyses. W.F.D. is a Fellow of the Evolutionary Biology Program, Canadian Institute for Advanced Research. This work was supported by a grant (MT4467) from the Medical Research Council of Canada to W.F.D.

Sequence Availability

The *G. intestinalis* DNA polymerase α and *T. vaginalis* DNA polymerase δ and ϵ sequences have been deposited in GenBank with the accession numbers AF067402, AF067403, and AF067404.

LITERATURE CITED

- ADACHI, J., and M. HASEGAWA. 1992. MOLPHY, programs for molecular phylogenetics I-PROTML, maximum likelihood inference of protein phylogeny. Comput. Sci. Monogr. No. 28.
- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403–410.
- ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG, W. MILLER, and D. J. LIPMAN. 1997. Gapped BLAST and PSI-BLAST—a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.
- BALDAUF, S., J. D. PALMER, and W. F. DOOLITTLE. 1996. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proc. Natl. Acad. Sci. USA 93:7749–7754.
- BONNER, C. A., S. HAYS, K. MCENTEE, and M. F. GOODMAN. 1990. DNA polymerase II is encoded by the DNA damageinducible *dinA* gene of *Escherichia coli*. Proc. Natl. Acad. Sci. USA 87:7663–7667.
- BONNER, C. A., P. T. STUKENBERG, M. RAJAGOPALAN, R. ER-ITJA, M. O'DONNELL, K. MCENTEE, H. ECHOLS, and M. F. GOODMAN. 1992. Processive DNA synthesis by DNA polymerase II mediated by DNA polymerase III accessory proteins. J. Biol. Chem. 267:11431–11438.
- BRAITHWAITE, D. K., and J. ITO. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. Nucleic Acids Res. **21**:787–802.
- 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.
- BUDD, M. E., K. D. WITRUP, J. E. BAILEY, and J. L. CAMPBELL. 1989. DNA polymerase I is required for premeiotic DNA replication and sporulation but not X-ray repair in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **9**:365–376.
- BULT, C. J., O. WHITE, G. J. OLSEN et al. (38 co-authors). 1996. The complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. Science 273:1058– 1073.
- CAVALIER-SMITH, T. 1993. Kindgom protozoa and its 18 phyla. Microbiol. Rev. **57**:953–994.
- CAVALIER-SMITH, T., and E. E. CHAO. 1996. Molecular phylogeny of the free-living archezoan *Trepomonas agilis* and the nature of the first eukaryote. J. Mol. Evol. **43**:551–562.
- CLAYTON, R. A., O. WHITE, K. A. KETCHUM, and J. C. VEN-TER. 1997. The first genome from the third domain of life. Nature **387**:459–462.
- EDGELL, D. R., and W. F. DOOLITTLE. 1997. Archaea and the origin(s) of DNA replication proteins. Cell **89**:995–998.
- EDGELL, D. R., H.-P. KLENK, and W. F. DOOLITTLE. 1997. Gene duplications in evolution of archaeal family B DNA polymerases. J. Bacteriol. **179**:2632–2640.

- FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. 27:401–410.
- FELSENSTEIN, J. 1996. PHYLIP (phylogeny inference package). Version 3.57c. Department of Genetics, University of Washington, Seattle.
- FLEISCHMANN, R. D., M. D. ADAMS, O. WHITE et al. (37 coauthors). 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. Science 269:496– 512.
- FORTERRE, P., C. ELIE, and M. KOHIYAMA. 1984. Aphidicolin inhibits growth and DNA synthesis in halophilic archaea. J. Bacteriol. 159:800–802.
- FRASER, C. M., S. CASJENS, W. M. HUANG et al. (35 co-authors). 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature **390**:580–586.
- FRASER, C. M., J. D. GOCAYNE, O. WHITE et al. (25 co-authors). 1995. The minimal gene complement of *Mycoplasma genitalium*. Science **270**:397–403.
- GOGARTEN, J. P., H. KIBAK, P. DITTRICH et al. (13 co-authors). 1989. Evolution of the vacuolar H⁺-ATPase: implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA 86:6661–6665.
- HASEGAWA, M., and M. FUJIWARA. 1993. Relative efficiencies of the maximum-likelihood, maximum parsimony, and neighbor-joining methods for estimating protein phylogenies. Mol. Phylogenet. Evol. 2:1–5.
- HIMMELREICH, R., H. HILBER, H. PLAGENS, E. PIRKL, B. C. LI, and R. HERRMANN. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. Nucleic Acids Res. 24:4420–4449.
- HUELSENBECK, J. P. 1995. Performance of phylogenetic methods in simulation. Syst. Biol. 44:17–48.
- HUELSENBECK, J. P. 1997. Is the Felsenstein zone a fly trap? Syst. Biol. 46:69–74.
- HUGHES, A. J. JR., S. K. BRYAN, H. CHEN, R. E. MOSES, and C. S. MCHENRY. 1991. *Escherichia coli* DNA polymerase II is stimulated by DNA polymerase II holoenzyme auxiliary subunits. J. Biol. Chem. **266**:4568–4573.
- IWABE, N., K. I. KUMA, M. HASEGAWA, S. OSAWA, and T. MIYATA. 1989a. Evolutionary relationships of archaea, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA 86:9355– 9399.
- IWABE, N., K. I. KUMA, H. KISHINO, M. HASEGAWA, and T. MIYATA. 1989b. Evolution of RNA polymerases and branching patterns of the three major groups of archaea. J. Mol. Evol. 32:70–78.
- IWASAKI, H., Y. ISHINO, H. TOH, A. NAKATA, and H. SHINA-GAWA. 1991. *Escherichia coli* DNA polymerase II is homologous to α-like DNA polymerases. Mol. Gen. Genet. 226:24–33.
- IWASAKI, H., A. NAKATA, G. C. WALKER, and H. SHINAGAWA. 1990. The *Escherichia coli pol*B gene, which encodes DNA polymerase II, is regulated by the SOS system. J. Bacteriol. 172:6268–6273.
- KANEKO, T., S. SATO, A. L. KOTANI et al. (21 co-authors). 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 3:109–136.
- KESTI, T., H. FRANTTI, and J. E. SYVAOJA. 1993. Molecular cloning of the cDNA for the catalytic subunit of human DNA polymerase ϵ . J. Biol. Chem. **268**:10238–10245.
- KLENK, H.-P., R. A. CLAYTON, J.-F. TOMB et al. (48 co-authors). 1997. The complete genome sequence of the hy-

perthermophilic, sulphate-reducing archaeon Archaeoglobus fulgidus. Nature **390**:364–370.

- KUHNER, M. K., and J. FELSENSTEIN. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rate. Mol. Biol. Evol. **11**:459–468.
- MORRISON, A., H. ARAKI, A. B. CLARK, R. K. HAMATAKE, and A. SUGINO. 1990. A third essential DNA polymerase in *S. cerevisiae*. Cell **62**:1143–1151.
- MORRISON, A., R. B. CHRISTENSEN, J. ALLEY, A. K. BECK, E. G. BERNSTINE, J. F. LEMONTT, and C. W. LAWRENCE. 1989. *REV3*, a Saccharomyces cerevisiae gene whose function is required for mutagenesis, is predicted to encode a nonessential DNA polymerase. J. Bacteriol. **171**:5659–5667.
- NAVAS, T. A., Z. ZHOU, and S. J. ELLEDGE. 1995. DNA polymerase epsilon links the DNA replication machinery to the S-phase checkpoint. Cell **80**:29–39.
- OCHMAN, H., A. S. GERBER, and D. L. HARTL. 1988. Genetic applications of an inverse polymerase chain reaction. Genetics 120:621–623.
- RANGARAJAN, S., G. GUDMUNDSSON, Z. QIU, P. L. FOSTER, and M. F. GOODMAN. 1997. *Escherichia coli* DNA polymerase II catalyzes chromosomal and episomal DNA synthesis *in vivo*. Proc. Natl. Acad. Sci. USA **94**:946–951.
- REYSENBACH, A. L., L. J. GIVER, G. S. WICKHAM, and N. R. PACE. 1992. Differential amplification of rRNA genes by polymerase chain reaction. Appl. Envrion. Microbiol. 58: 3417–3418.
- SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
- SANDHU, G. S., J. W. PRECUP, and B. C. KLINE. 1989. Rapid one-step characterization of recombinant vectors by direct analysis of transformed *Escherichia coli* colonies. Bio-Techniques 7:689–690.
- SCHINZEL, R., and K. J. BURGER. 1985. Sensitivity of halobacteria to aphidicolin, an inhibitor of eukaryotic α-type DNA polymerases. FEMS Microbiol. Lett. 25:187–190.
- SMITH, D. R., L. A. DOUCETTE-SMITH, C. DELOUGHERY et al. (34 co-authors). 1997. Complete genome sequence of *Meth-anobacterium thermoautotrophicum* ΔH: Functional analysis and comparative genomics. J. Bacteriol. **179**:7135–7155.
- SOGIN, M. L., J. H. GUNDERSON, H. J. ELWOOD, R. A. ALONSO, and D. A. PEATTIE. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia*. Science 243:75–77.
- STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964–969.

- SWOFFORD, D. L. 1993. PAUP, phylogenetic analysis using parsimony, Version 3.1. Illinois Natural History Survey, Champaign.
- SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–515 in D. M. HILLIS, G. MORITZ, and B. K. MABLE, eds. Molecular systematics. 2nd edition. Sinauer, Sunderland, Mass.
- THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTALW, improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
- TOMB, J.-F., O. WHITE, A. R. KERLAVAGE et al. (40 co-authors). 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. Nature **388**:539–547.
- 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.
- UEMORI, T., Y. ISHINO, K. TOH, I. ADADA, and I. KATO. 1993. Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. Nucleic Acids Res. 21:259–265.
- WANG, J., A. K. SATTAR, C. C. WANG, J. D. KARAM, W. H. KONIGSBERG, and T. A. STEITZ. 1997. Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. Cell 89:1087–1099.
- WOESE, C. R. 1987. Bacterial evolution. Microbiol. Rev. 51: 221–271.
- WOESE, C. R., L. ACHENBACH, P. ROUVIERE, and L. MANDEL-CO. 1991. Archaeal phylogeny: reexamination of the phylogenetic position of *Archaeoglobus fulgidus* in light of certain composition-induced artifacts. Syst. Appl. Microbiol. 14:364–371.
- WONG, S. W., A. F. WAHL, P.-M. YUAN, N. ARAI, B. E. PEAR-SON, 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.
- 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.

THOMAS H. EICKBUSH, reviewing editor

Accepted June 5, 1998