Extremophiles (2009) 13:817–826
DOI 10.1007/s00792-009-0269-8

ORIGINAL PAPER

Molecular diversity and catalytic activity of *Thermus* DNA polymerases

Moreland D. Gibbs · Rosalind A. Reeves · David Mandelman · Qingli Mi · Jun Lee · Peter L. Bergquist

Received: 31 March 2009 / Accepted: 25 June 2009 / Published online: 12 July 2009

Abstract *Thermus aquaticus* DNA polymerase (*Taq* polymerase) made the polymerase chain reaction feasible and led to a paradigm shift in genomic analysis. Other *Thermus* polymerases were reported to have comparable performance in PCR and there was an analysis of their properties in the 1990s. We re-evaluated our earlier phylogeny of *Thermus* species on the basis of 16S rDNA sequences and concluded that the genus could be divided into eight clades. We examined 22 representative isolates and isolated their DNA polymerase I genes. The eight most diverse polymerase genes were selected to represent the eight clades and cloned into an expression vector coding for a His-tag. Six of the eight polymerases were expressed so that there was sufficient protein for purification. The proteins were purified to homogeneity and examination of the biochemical characteristics showed that although they were competent to perform PCR, none was as thermostable as commercially available *Taq* polymerase; all had similar error-frequencies to *Taq* polymerase and all showed the expected 5′–3′ exonuclease activity. We conclude that the initial selection of *T. aquaticus* for DNA polymerase purification was a far-reaching and fortuitous choice but simple mutagenesis procedures on other *Thermus*-derived polymerases should provide comparable thermostability for the PCR reaction.

Keywords Thermus · Enzymes · DNA polymerase · Thermophile

Communicated by A. Driessen.

M. D. Gibbs · R. A. Reeves · P. L. Bergquist (✉)
Department of Chemistry and Biomolecular Sciences, Macquarie University, North Ryde, Sydney, NSW 2109, Australia
e-mail: peter.bergquist@mq.edu.au

M. D. Gibbs
e-mail: moreland.gibbs@mq.edu.au

D. Mandelman · Q. Mi · J. Lee
Life Technologies Corporation, 1610 Faraday Avenue, Carlsbad, CA 92130, USA

P. L. Bergquist
Department of Molecular Medicine & Pathology, University of Auckland Medical School, Auckland, New Zealand

Introduction

Thermophilic DNA polymerases have revolutionized molecular biology. They find uses in gene isolation, gene cloning, sequencing and mutagenesis. Thermostable DNA polymerase (DNA pol) from *Thermus aquaticus* (*Taq* pol) made the polymerase chain reaction (PCR) feasible, and introduced a powerful technology that complemented recombinant DNA studies and aided in the diagnosis of inherited and infectious diseases (reviewed in Innis et al. 1990). Thermostable DNA polymerases have played an important role in the isolation of genes from eukaryotic messenger RNA following reverse transcription to give a cDNA copy of the target gene. Basic PCR has become commonplace in research laboratories where it has been used to amplify DNA fragments and detect DNA or RNA sequences but it has evolved far beyond simple amplification and detection. Many extensions to the basic PCR approach have been described and a number of enzymes showing similar properties to *Taq* polymerase have been isolated from other strains of *Thermus* (for example, Carballeira et al. 1990; Myers and Gelfand 1991; Harrell and Hart 1994; Perler et al. 1996; Kwon et al. 1997; Kim et al. 1998).

T. aquaticus DNA pol has reverse transcriptase activity (Jones and Foulkes 1989) but the reverse transcriptase activity of a recombinant DNA polymerase from *T. thermophilus* (*Tth* pol, Myers and Gelfand 1991) has been reported to be 100-fold more efficient, although these two strains are closely related phylogenetically on the basis of 16S rDNA analysis (Saul et al. 1993). This result, along with others such as the Phe667Tyr mutation in *Taq*, which overcame the discrimination against dideoxynucleotides (Tabor and Richardson 1995), indicated that one or several residue differences between polymerases could have a dramatic effect on their biochemical characteristics. We have reported previously the isolation of a number of thermostable DNA polymerases from thermophiles and extreme thermophiles (Shandilya et al. 2004). In the course of this work, we isolated a novel polymerase from *Thermus* Rt4A.1 (*Tfil* polymerase) that was phylogenetically distinct from *Taq* and *Tth* polymerases, matching the taxonomic distinctions between the isolates. The majority of the new polymerases were insufficiently thermostable to support PCR but *Tfil* polymerase retained significant replicative activity at 90°C but had negligible RT activity. We re-examined our original phylogenetic information on the genus *Thermus* and re-evaluated it with the addition of a number of new 16S rDNA sequences that had become available since the original publication by Saul et al. 1993. In this communication, we report on the isolation and expression of DNA polymerase genes from *Thermus* strains selected on the basis of their phylogenetic relationship to *T. aquaticus* as judged by 16S rDNA analysis and have formulated a phylogeny on the basis of DNA polymerase sequences. We report also on some of the enzymatic properties of the gene products and their performance in the polymerase chain reaction. We conclude that the isolation of *T. aquaticus* and the selection of its' polymerase for purification was a fortuitous event when the thermostability of other native *Thermus* polymerases are evaluated.

Materials and methods

Thermus isolates

Thermus strains were obtained from the collection of Prof. Hugh Morgan, Waikato University, Hamilton, New Zealand and genomic DNA was prepared as previously described (Morris et al. 1998).

Amplification of polymerase genes

PCR products from genomic amplification and genomic walking PCR (GWPCR) were prepared and sequenced as described in Shandilya et al. (2004) where a detailed description is found.

16S Gene Isolation, Sequencing and Phylogenetic analysis

The isolation and analyses of 16S SSU gene sequences were performed following the basic methodology of Saul et al. (1993).

Degenerate Primer design and PCR

Degenerate primers were designed using the CODEHOP strategy (Rose et al. 1998). We previously used this technique to isolate thermostable DNA polymerases from other bacterial species; it is advantageous as it does not rely on expression of the polymerase gene for gene selection (Shandilya et al. 2004).

GWPCR

Genomic walking PCR was performed as described by Shandilya et al. (2004).

Gene cloning

Full-length polymerase genes were expressed in a modified His-tag version of the expression vector, pJLA602 (Schauder et al. 1987). The His-tag was incorporated into pJLA602 by amplifying a 95 bp PCR fragment (*Xho*I to *Nco*I) from pJLA602 using primers PJLAXHOI and HISTAG (Table 2). The HISTAG primer includes a 6xhis tag and repositioned the *Nco*I site downstream of the start methionine and his-tag.

Purification of DNA polymerases

E. coli cells (DH5α, Invitrogen) bearing the plasmid pJLA602NHIS with the DNA polymerase gene from either OHA.2, ATCC22737, Wai28A.1, Tok23A.1, NMX2A.1, or Fiji3A.1, were grown in 2.8 L Fernbach flasks in LB broth containing 100 μg/mL Ampicillin at 28°C. After the culture reached an A_{600} of 1.0, expression of DNA polymerase was induced by shifting the growth temperature to 42°C for 3 h. Cells were harvested by centrifugation and stored at −70°C.

The frozen cell paste was thawed on ice and resuspended in 3 mL of lysis buffer (50 mM Tris–HCl (pH 7.5), 1 mM EDTA, 5 mM β-mercaptoethanol, 8% (v/v) glycerol, 50 μg/mL phenylmethylsulfonyl fluoride) per gram of wet cell paste and lysed by sonication (70–80% lysis based on OD_{600}). The lysate was heat-treated for 30 min at 75°C, then immediately placed on ice and NaCl was added to a

final concentration of 250 mM. Polyethylenimine (PEI; 2% v/v) was added drop-wise to the lysate at 4°C to final concentration of 0.15% (v/v) and allowed to mix for 30 min at 4°C. The lysate was centrifuged for 1 h in a Sorvall SS-34 rotor at 17 500 rpm (36,000×*g*). The pellet was resuspended in low imidazole buffer (Buffer A: 10 mM K_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM imidazole) and loaded onto a 5 mL HiTrap chelating column charged with Ni^{2+} ions and equilibrated in buffer A. After washing the column with four column volumes (CV) of buffer A, the protein was eluted in a 10 CV gradient from 10 mM to 500 mM imidazole (Buffer B: 10 mM K_2PO_4 (pH 8.0), 300 mM NaCl, 500 mM imidazole). Fractions containing the protein of interest were confirmed by SDS-PAGE and unit activity assays. These fractions were then dialyzed overnight into low salt EMDSO_4 buffer (30 mM Tris–HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 50 mM NaCl).

The protein was then loaded onto a 6 mL EMD-SO_4 column equilibrated in low salt buffer (Fractogel; Merck, Germany). After washing with 5 CV of low salt buffer, the protein was eluted in a 10 CV salt gradient ranging from 50 to 500 mM NaCl (high salt buffer: 30 mM Tris–HCl (pH 7.5), 1 mM EDTA, 1 mM DTT, 10% (v/v) glycerol, 1000 mM NaCl).

Fractions containing the particular polymerase were confirmed by SDS-PAGE and unit activity assays. They were pooled and dialyzed overnight into polymerase storage buffer (20 mM Tris–HCl (pH 8), 40 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5% Nonidet P40, 0.5% Tween-20, 50% (v/v) glycerol) and stored at −20°C. The storage buffer for the positive control, *rTaq*, was identical except it lacked KCl. SDS-PAGE analysis showed that the polymerases had been purified to homogeneity by this procedure (data not shown).

The yields of protein from YSP2A.1 and Tp10A.2 were very low and these two polymerases were not investigated further.

DNA-directed DNA polymerase unit activity

Enzyme assays for DNA polymerase activity were performed as described in Shandilya et al. (2004). One unit of polymerase activity is defined as the amount of enzyme required to incorporate 10 nmol of dNTP into an acid insoluble form at 74°C in 30 min.

Thermostability assay

Twenty units of enzyme were incubated in PCR buffer (20 mM Tris–HCl (pH 8.4), 50 mM KCl) at 95°C for 5, 10, 20, and 45 min. At the times specified, 4 µL (2%) was removed from the reaction and diluted 10-fold in assay dilution buffer on ice to stop the reaction (25 mM Tris–HCl (pH 8.0), 50 mM KCl, 100 µg/mL gelatin, 0.1 mM EDTA, 5 mM β-mercaptoethanol). 5 µL of the diluted reaction was assayed for activity as previously described (Shandilya et al. 2004).

Fidelity assay

The assay is based on streptomycin resistance of *rpsL* mutants (Fujii et al. 1999), modified from Mo et al. (1991), as described in Shandilya et al. (2004). The mutation frequency was determined by dividing the total number of mutations by the total number of transformed cells. The error frequency was determined by dividing the mutation frequency by 130 (the number of amino acids that cause phenotypic changes in *rps*L) and the number of template doublings (Fujii et al. 1999). All fidelity assays were performed with 25 PCR cycles.

5′–3′ exonuclease activity

This assay was performed according to Kaisert et al. (1999).

Accession numbers

The DNA sequences for each PolI polymerase have accession numbers as follows: ATCC27737, FJ358541; YSP2A.1, FJ358542; NMX2A.1, FJ358543; OHA.2, FJ358544: Tok23A.1, FJ358545; Tp10A.2, FJ358546; Fiji3A.1, FJ358547; Wai28A.1, FJ358548.

Results

Isolation of DNA polymerase genes

An alignment of known *Thermus* PolI-type polymerase sequences (Table 1 and Fig. 1) indicated that a simple PCR strategy could be devised for rapidly screening polymerase genes from most *Thermus* species. Preliminary experiments were performed to see if almost full-length polymerase genes could be amplified for sequencing. Degenerate primers for highly conserved motifs (Fig. 1) were designed to amplify the entire genes from short regions of identity near the N- and C-termini of each gene (Table 2). Initial PCR experiments using the primers THERMUSF1CH and THERMUSCR1 (Table 2) were not successful in amplifying the approximately 2.5 kb product from each of the *Thermus* strain template genomic DNA's. We suspected that this was because of imperfect conservation of the motifs upon which the primers were based, resulting in the inability of one or both primers to

successfully bind and prime at their predicted optimal annealing temperatures. Therefore, the primers THERMUSNR1CH and THERMUSCF1 (Table 2) were designed to amplify short N- and C- terminal regions of each gene. These primers worked successfully in combination with the primers THERMUSF1CH and THERMUSCR1 respectively, and each product was sequenced. The two sequence sets were aligned with published *Thermus* PolA sequences and used for phylogenetic analysis. The sequences were used also for the design of the genomic walking primers, THERMUSGWR and THERMUSGWF, for completing the 5′- and 3′-terminal sequences of the genes (see Fig. 1).

The final protocol for the isolation of full-length polymerase genes was a four-step PCR procedure following the preparation of genomic DNA from the selected strains after preliminary experiments showed that it was difficult to isolate close to full-length genes directly from genomic DNA. *Thermus* DNA has a high percentage G + C content of around 70% that can affect the specificity and efficiency of PCR using degenerate primers. Therefore, first, we amplified key small 5′- and 3′-regions with *Thermus*-specific degenerate primers to establish the sequence diversity of the polymerase gene sequences. The primers were designed on the basis of an alignment of seven *Thermus* genes and two *Deinococcus* genes. The *Deinococcus* group is the most closely related group to *Thermus* and their PolI genes were included in the alignments to ensure the motifs we had selected were also conserved outside the *Thermus* group. This addition gives increased confidence that the conserved motifs would be found in all or nearly all *Thermus* species (Fig. 1). At this point, phylogenetic comparisons of the N- and C-terminal sequences with known *Thermus* polymerase sequences were undertaken to ensure that the genes were not identical to known sequences. Secondly, we then used the DNA sequences obtained from the N- and C-terminal PCR products to design gene-specific PCR primer sets that allowed amplification of an almost full-length copy of each polymerase gene. Thirdly, we designed primers to genomic-walk 45–100 bp upstream and 40 bp downstream in order to obtain the complete gene sequences. Finally, specific primers then were designed to amplify each complete gene for transfer into a plasmid expression vector.

We expressed each full-length polymerase gene as an N-terminal his-tag fusion protein for easy isolation and purification. Each purified enzyme was assayed for thermostability to ensure that they meet the high thermostability requirements of the PCR (see later). We used the phylogenetic information derived from the DNA polymerase sequences to eliminate closely related genes from further examination.

Table 1 Known *Thermus* polymerase gene sequences

GenBank entry	Source organism
TAU62584	*Thermus caldophilus*
TTHPOLA	*Thermus thermophilus*
AE017303	*Thermus thermophilus*.
TTHDNAP	*Thermus aquaticus*
TTHTAQP1A	*Thermus aquaticus*.
TFPOLDNA	T*hermus flavus*.
AF030320	*Thermus filiformis*
AY247645	*Thermus filiformis* strain Rt4A.1

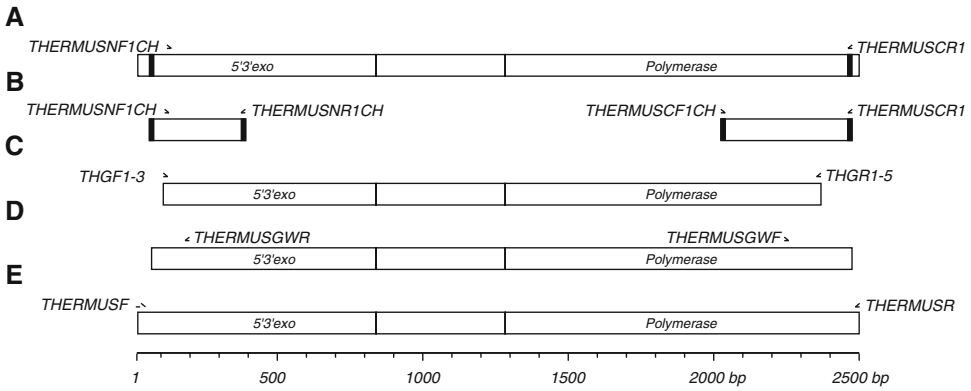

Fig. 1 Primer sets for the amplification of *polI* genes from novel *Thermus* isolates. **a** A diagrammatic representation of the basic three-domain structure of bacterial PolI DNA polymerases and the relative binding positions of the degenerate primers initially used to amplify almost full-length *Thermus* polymerase genes. **b** Degenerate primer sets designed to amplify small portions of the N- and C-terminal regions of *Thermus* polymerase genes. **c** Specific primers designed based upon the DNA sequence of the N- and C-terminal portions that allow amplification of almost full-length *polI* genes. **d** Gene-specific primers designed based upon the DNA sequence of the N- and C-terminal portions that enabled genomic walking PCR (GWPCR) to obtain the N- and C-terminal portions of each gene

Table 2 Oligonucleotide primers used in this study

Primer name	Sequence	Conserved motif
THERMUSNF1CH	5′-GGCCCGGTCCTCCTGRTNGAYGGNCA	GRVLL[I/V]DGH
THERMUSNR1CH	5′-GGGTGGCCAGGACGTCRTCNGCYTC	EADDVLATL
THERMUSCF1	5′-GGCGTCCTCTACGGCATGTCNGCNCA	GVLYGMSAH
THERMUSCR1	5′-ATCCCCACCTCCACYTCNAGNGGNAC	VPLEVEVG
THERMUSGWF	5′-GCATGGCCTTCAACATGCC	MAFNM
THERMUSGWR	5′-GGGGGCCTTGGCGTCAAA	FDAKA
THGF1	5′-CCCTGAGCCTCACCACCT	LSLTT
THGF2	5′-CCTGAAGGGCCTCACCAC	LKGLTT
THGF3	5′-CCTCAAGGGCCTCACCAC	LKGLTT
THGR1	5′-CAGGACCAGCTCGTCGTG	HDELVL
THGR2	5′-GAGGACCAGTTCGTCGTG	HDELVL
THGR3	5′-GAGGACCAGCTCGTCGTG	HDELVL
THGR4	5′-CCAAAAGGAGCTCGTCGTG	HDELLL
THGR5	5′-CAAGACCAGCTCGTCGTG	HDELVL

Thermus polymerase N- and C-terminal sequence determination

In total, DNA sequences from 19 *Thermus polI* genes were obtained. A further three sequences were mixed and not used in the following analysis. We used standard phylogenetic analysis of the translated sequences to show that a number of the sequences were very closely related and could be eliminated from further consideration (Swofford 2002). Two members of the closely related group (*T. filiformis* and *Thermus* sp. T351) were used to represent the group in phylogenetic analysis using translated protein sequences. The informative subset of amino acids was used to generate a consensus phylogram derived from 16 most parsimonious trees. A phylogenetic analysis was performed on an alignment of the related DNA sequences with the inclusion of *T. thermophilus* HB8 sequence as the outgroup to more clearly resolve the relationships of the closely related group. The minimal informative data set was analyzed by parsimony and maximum likelihood (data not shown). A similar analysis was performed on the small N- and C-terminal portions of the genes and the results were consistent with the tree pattern from the entire gene sequences. This information was used to generate the tree shown in Fig. 3 (discussed below).

Analysis of *Thermus* 16S rDNA sequences and their corresponding DNA polymerase sequences

The genus *Thermus* represents a deep bacterial branch and clades A-H were previously defined by Saul et al. (1993) from 16S rDNA data and are included on the phylogram in Fig. 2. Several new isolates from New Zealand have been included that we had not examined in our earlier publication. The sequences can be separated into a number of clades related to established species on the basis of 16S rDNA analysis. The genes from some strains were isolated and sequenced and the phylogenetic relationships of all available *Thermus* strains have been re-examined and a new phylogenetic tree has been established (Fig. 2). We have divided the culturable *Thermus* isolates into eight clades. An initial alignment of 83 *Thermus* 16S SSU sequences was used to define those that were closely related to typed *Thermus* strains. The tree was rooted using *Meiothermus ruber* as the outgroup. A smaller subset of 43 sequences was analyzed by parsimony and further close relatives culled from the alignment. The final consensus phylogram was derived using parsimony and from examination of the relationships of the sequences we have now defined a further set of clades, termed I-L. The phylogram presented in Fig. 2 shows the phylogenetic relationships of the various isolates. It was generated using maximum parsimony analysis (PAUP) and all branches had greater than 50% bootstrap support (100 replicates).

The phylogenetic distribution broadly matches the geographical locations from which the various strains were isolated, as detailed in Table 1. However, members of most clades have been isolated from widely separated regions. It is unclear whether the geographical association reflects to some degree the culturing techniques used to isolate the various strains. Two isolates, *Thermus* sp. strain C4 (GenBank entry AY028380) and *Thermus* sp. strain SRI-248 (GenBank entry AF255591) do not fall into any of the defined clades and the phylogram may warrant further definition once they have been fully described.

We conclude from the phylogenetic analysis of four new 16S rDNA sequences obtained in this study, taken in combination with a selected subset of existing *Thermus* species, that *Thermus* isolate ZVIEA.3 is closely related to *T. oshimai* (Group K); *Thermus* ZHGICA.1 is very closely

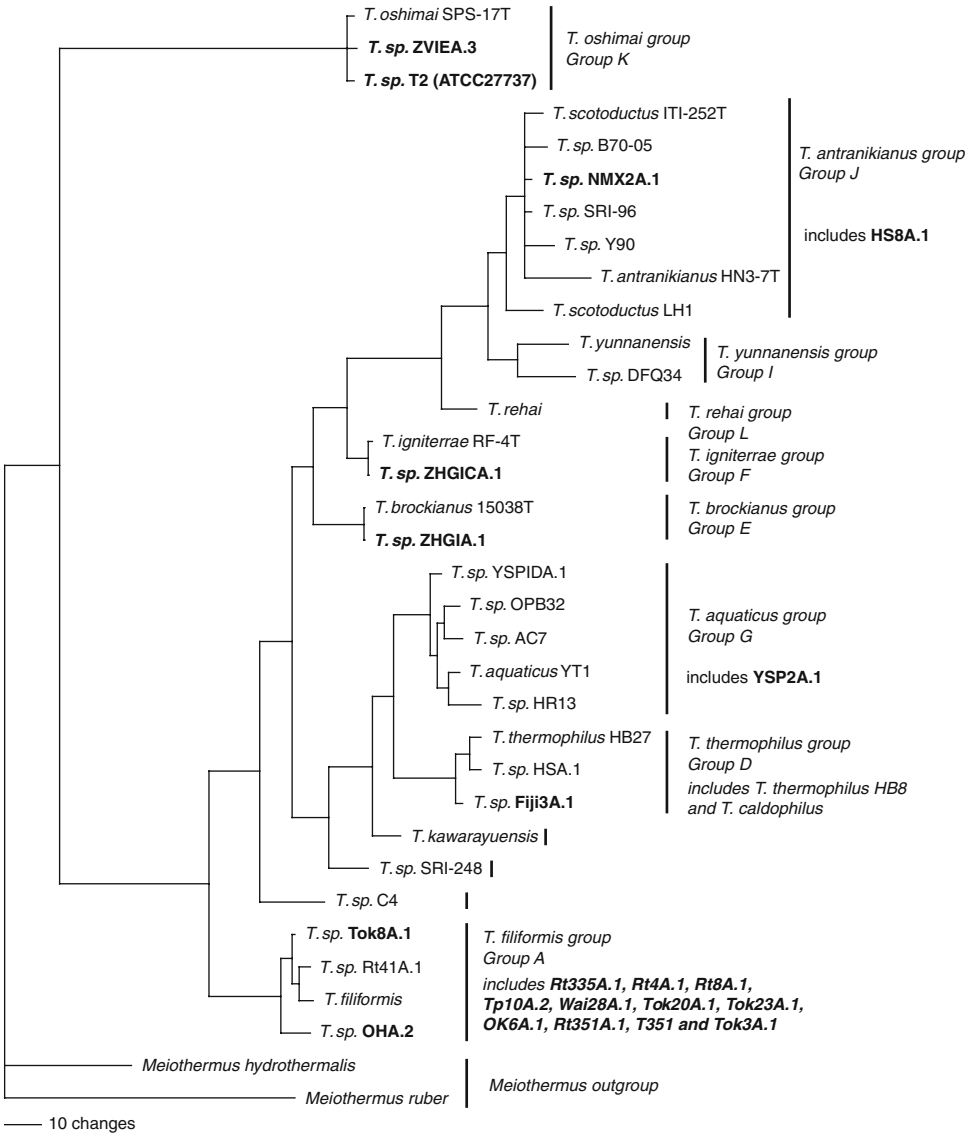

Fig. 2 Phylogenetic tree obtained by parsimony analysis of 16S SSU gene sequence (approximately 600 bp of the 5′ end of the SSU gene). The organisms whose polymerases have been examined in this study are shown in bold. The groups defined by Saul et al. (1993) are marked with vertical bars (groups A–H). We have defined groups I–L in this paper. The probable phylogenetic location for organisms where 16S sequence data is not available is given to the right of the vertical bar. Placement is based on the preceding phylogenetic analysis of partial DNA polymerase sequences. An initial alignment of 83 *Thermus* 16S SSU sequences was used to define sequences that were closely related to typed *Thermus* strains. A smaller subset of 43 sequences was analysed by parsimony and further close relatives culled from the alignment. The final consensus phylogram shown below was derived using parsimony

related to *T. igniterrae* (Group F); *Thermus* isolate ZHGIA.1 is very closely related to *T. brockianus* (Group E) and *Thermus* OHA.2 is very closely related to *T. filiformis* (Group A). All New Zealand isolates, with the exception of strain HSA.1 and HS8A.1, cluster closely with the group A filamentous strain *Thermus filiformis*, although the majority of this group do not have a filamentous morphology. The strains HSA.1 and HS8A.1 were both isolated from Hamner Springs, New Zealand, and are associated with clades D and J respectively.

The phylogeny of the *Thermus* DNA polymerases sequenced shows a reasonable correlation with the 16S rDNA tree (Fig. 3). All polymerases from the New Zealand isolates cluster together in the *T. filiformis* clade, with the exception of HS8A.1. The New Zealand *T. filiformis*-like isolates probably can be further broken down into at least two distinct clades. However, further polymerase gene sequence information is required to delineate the divisions more clearly within this group of isolates.

Several of our *Thermus* polymerase sequences appeared to be mixed in one or both of the DNA polymerase fragments examined. It was unclear whether this was due to mixed cultures, template contamination, or the presence of trace amounts of the *Taq* polymerase gene in commercial preparations of the enzyme mixture used for PCR amplification (a phenomenon we have previously observed, see Shandilya et al. 2004). 16S SSU genes were amplified and analyzed from four genomic DNAs, *Thermus* ZVIEA.3, *Thermus* ZHGICA.1, *Thermus* ZHGIA.1 and *Thermus* OHA.2 in an attempt to determine if the mixed sequence data obtained was due to mixed cultures. In all four cases, unambiguous 16S rDNA sequences were obtained, implying that the genomic DNAs were not derived from mixed cultures.

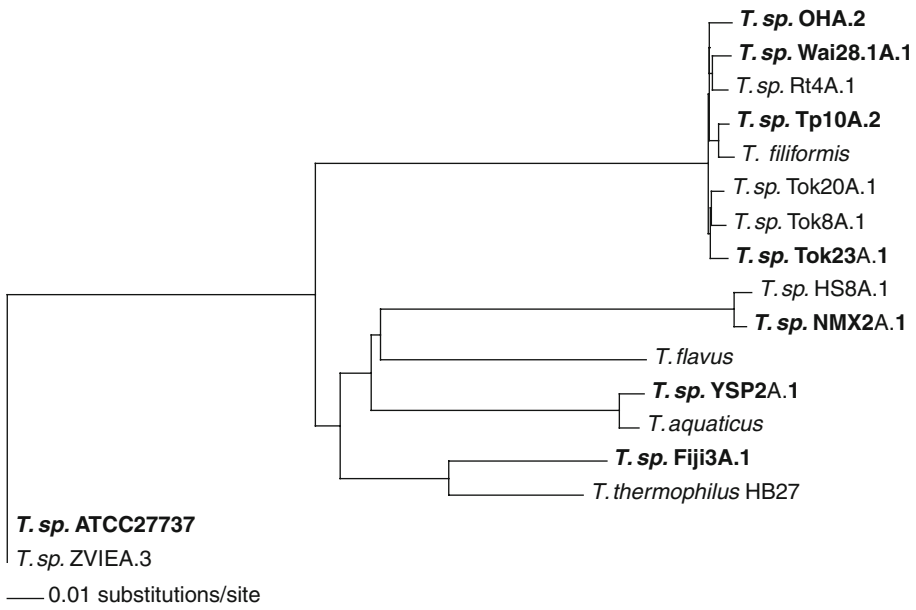

Fig. 3 Phylogeny of Thermus DNA polymerases based on Maximum Likelihood analysis. The tree is based on protein sequences and the eight newly sequenced genes reported here are shown in bold type

Negative control PCR reactions using the N- and C-terminal primer sets without template were performed. The N-terminal primers gave a correctly sized PCR product that was sequenced and were found to be mixed, indicating that the Eppendorf TripleMaster enzyme blend used may contain bacterial DNA polymerase DNA.

The ZVIEA.3 16S rDNA result suggested that the polymerase sequence amplified was because of contamination by *Taq* polymerase DNA. On the basis of the 16S SSU analysis, ZVIEA.3 is very distantly related to most other *Thermus* strains. Subsequent examination of sequences and primers revealed that our degenerate polymerase primers were not appropriate for amplification of Pol I from this strain.

In conclusion, based on 16S phylogeny in combination with the polymerase sequence phylogeny we have obtained, all the *Thermus* examples we have examined appear to fit within the clades defined by the 16S sequences, as outlined in Table 3.

Table 3 Phylogenetic placement of *Thermus* isolates studied in this work according to 16S DNA sequence data

Group	Saul clade	Isolates studied in this work
T. oshimai	New K	ZVIEA.3, *T. sp.* T2 (ATCC27737)
T. antranikianus	New J	*T. sp.* NMX2A.1 and *T. sp.* HS8A.1
T. yunnanensis	New I	–
T. rehai	New L	–
T. igniterrae	F	*T. sp.* ZHGICA.1
T. brockianus	E	*T. sp.* ZHGIA.1
T. aquaticus	G	*T. sp.* YSP2A.1
T. thermophilus	D	*T. sp.* Fiji3A.1 and NMX2A.3
T. kawarayuensis	New	-
T. filiformis	A/B	*T. sp.* Rt335A.1, Rt4A.1, Rt8A.1, Tp10A.1, Wai28A.1, Tok20A.1, Tok8A.1, Tok23A.1, OK6A.1 and T351

Biochemical characterization of the polymerases

The *rTaq* used as a control for these experiments was a commercial preparation expressed as a recombinant protein but the gene sequence was otherwise wild-type. The six DNA polymerases selected for detailed study were purified to homogeneity as judged by SDS gel electrophoresis but showed a four-fold range in specific activities (Table 4). Their thermostability at 95°C was compared to *rTaq* and none showed the thermostability of the control (Table 4). All polymerases were able to carry out PCR with a variety of target sequences ranging from 0.4 to 4 kb but yields were generally lower than the control *rTaq* polymerase even with twice as many units present in the reactions (Fig. 4). These results presumably reflect the lower thermostability of the new polymerases. The six polymerases showed 5′–3′ exonuclease activity similar to a control of commercial polymerase (*Tfil*, Invitrogen, Fig. 5), and all the new DNA polymerases are more or less similar to that of *Taq* DNA polymerase with respect to fidelity of replication (within twofold range, Table 5).

Discussion

Since the first description of *Thermus aquaticus* by Brock and Freeze (1969), members of the genus *Thermus* have been readily isolated from neutral-pH geothermal areas around the world. A number of *Thermus* polymerases were isolated and their genes cloned in the early 1990s, including the enzymes from *T. flavus*, *T. thermophilus* and *T. caldophilus* (reviewed in Perler et al. 1996). These enzymes showed similar capabilities to *Taq* in thermostability and ability to perform the polymerase chain reaction and some were produced commercially but for several reasons, including patent issues, *Taq* in its several formulations became the most widely used polymerase in molecular biological applications. To our knowledge, there has been no systematic survey of the DNA polymerase I enzymes from the diverse group of isolates classified as *Thermus* strains or species.

Table 4 Specific activity and thermostability

Polymerase	Specific activity (μ mg^{-1} protein)	Half-life at 95°C (min)
rTaq	ND	20
OHA.2	87.6	2.5
ATCC27737	110.4	5
Wai28A.1	67.9	4
Tok23A.1	30.4	4
NMX2A.1	219.5	3.5
Fiji3A.1	133.3	6

ND not done

We previously described the taxonomy of the three validly named species known at time in relation to a further 18 *Thermus* isolates (Saul et al. 1993). Since then, at least a further 10 species have been proposed and numerous undescribed strains have been isolated and identified. We have presented here an update of our previous phylogeny to include more recent *Thermus* 16S rRNA SSU gene sequence data and show here that the isolates can be divided into eight clades on this basis. The majority of the isolates we investigated fell into Clade A, the *T. filiformis* group, which reflects the origin of the isolates available to us. A phylogeny based on the DNA polymerase sequences of representative members of the clades did not show any major discrepancies compared to the 16S rDNA sequences.

We have reported here on 22 *Thermus* strain DNA polymerases that were screened by degenerate PCR to determine their relatedness to published *Thermus* polymerase gene sequences. The full-length sequences of all candidate genes were obtained and eight of the most diverse genes were selected as candidates for full-length

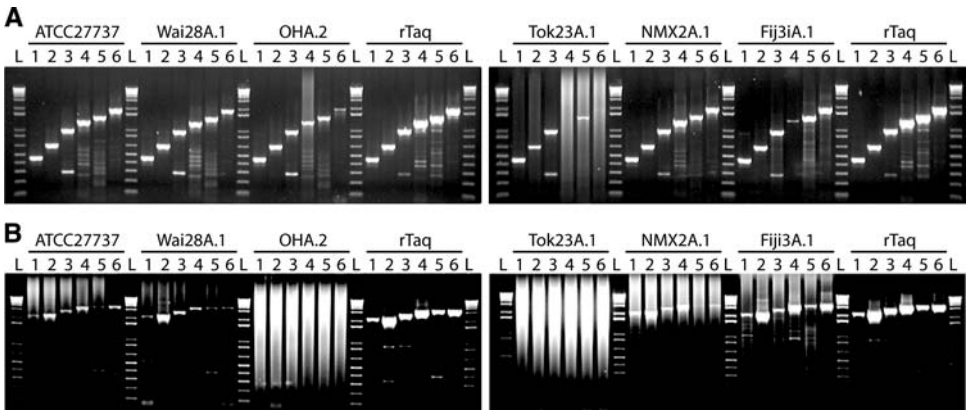

Fig. 4 PCR performance on genomic targets. Twelve amplicons ranging from 455 bp to 3560 bp were amplified from 50 ng of cell-line K562 genomic DNA. All reactions were assembled on ice. *Thermus aquaticus* (rTaq) DNA polymerase PCR reactions were assembled with 0.2 µM each primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 1× PCR buffer (20 mM Tris (pH 8.4), 50 mM KCl) and 1 unit rTaq per reaction. Taq-"like" DNA polymerases PCR reactions were assembled with 0.2 µM each primer, 0.2 mM dNTPs, 1.5 mM $MgCl_2$, 1× PCR buffer, 2 µl enzyme (1 U/µL) per reaction. PCR cycling conditions for all DNA polymerases were 1 cycle at 94°C, 2 min; 35 cycles: 94°C, 15 s; 60°C, 30 s; 72°C, 1 min per kb. Agarose gel conditions were 1% agarose in 0.5× TBE. The gels were loaded using 5 µL of each 50 µL PCR reaction. Molecular weight markers are Invitrogen 1 kb Plus ladder. **a** The gene target names and their accession numbers are as follows: *Lane 1*; Hbg2, AY662983, *Homo sapiens* hemoglobin gamma-G chain variant Hb.M-Circleville gene (455 bp); *Lane 2* CCR5, U66285, Human CC chemokine receptor gene (625 bp); *Lane 3* CDK4, U37022 Human cyclin-dependent kinase gene (997 bp); *Lane 4* Hbg, NG_000007, Human beta globin region on chromosome 11 (1314 bp); *Lane 5* Rhod, HSU49742, Human rhodopsin (1502 bp); *Lane 6* Rhod, HSU49742, Human rhodopsin (1945 bp). **b** *Lane 1* Fabp HSU17081, Human fatty acid binding protein (FABP3) gene (2176 bp); *Lane 2* HC1; NC_000001, *Homo sapiens* chromosome 1 (2260 bp); *Lane 3* Fabp HSU17081, Human fatty acid binding protein (FABP3) gene (2691 bp); *Lane 4* Hmhc, NM_002471, Human myosin heavy chain (3352 bp); *Lane 5* AF039704, *Homo sapiens* lysosomal pepstatin insensitive protease (CLN2) gene (3432 bp); *Lane 6* Fabp HSU17081, Human fatty acid binding protein (FABP3) gene (3560 bp)

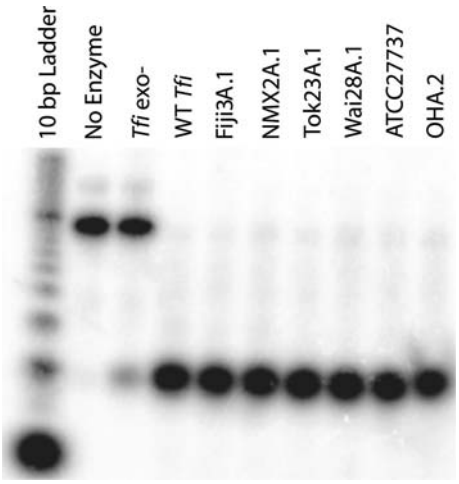

Fig. 5 Detection of 5′–3′ exonuclease activity. Five units of each enzyme were incubated with 0.44 nmols of 5′-labeled (^{32}P) oligonucleotide, 1× PCR buffer, and 5 mM $MgCl_2$ for 30 min at 72°C. The reactions were stopped with 3× sequencing stop buffer (97% Formamide, 10 mM NaOH, 5 mM EDTA, 2% glycerol, .05% xylene cyanol/bromophenol blue) and placed on ice. One-third of the reaction was run on a 15% TBE-Urea gel (Invitrogen Corpn) for 1 h at 200 V. The autoradiograph was visualized with a phosphor imager

Table 5 Comparison of fidelity of different *Thermus* DNA polymerases

DNA polymerase	Mutation frequency (%)[a]	Template doubling[b]	Error rate ($\times 10^{-5}$)[c]
rTaq	2.3	6	3
OHA.2	1.3	5.3	1.9
ATCC27737	1.2	5.7	1.6
Wai28A.1	0.85	5.3	1.3
Tok23A.1	1.8	2.7	5.1
NMX2A.1	3	5.7	4
Fiji3A.1	2.3	5.7	3.1

[a] Mutation frequencies were calculated using the equation: mf = total mutants/total colonies × 100

[b] Template doublings were determined using the equation: TD = Log (amount of PCR product)/Log 2

[c] Error rates were calculated using the equation: ER = mutation frequency/(template doublings × 130 bp)

isolation, cloning and characterization. Gene sequences from eight isolates were selected for full-length isolation and transfer into the expression vector pJLA602NHIS: Wai28A.1; Tp10A.2; Tok23A.1; OHA.2; NMX2A.1; YSP2A.1; Fiji3A.1; and ATCC27737. These genes were selected based upon their being the most divergent sequences observed in a phylogenetic analysis of their sequences.

Six of the eight *Thermus* polymerase genes selected were expressed in sufficient amounts for purification and their biochemical characteristics were measured in comparison with *Taq* polymerase as a control. All showed a comparable performance to *Taq* in their possession of 5′–3′ exonuclease activity, their ability to perform PCR and their fidelity of replication. However, none was as thermostable as *Taq* and had shorter half-lives at 95°C, which may limit their utility in PCR. However, this characteristic could be enhanced relatively easily by any one of several directed evolution techniques.

Other reports have suggested that some polymerases from *Thermus* strains may be superior in some aspects to *Taq*. For example, Park et al. (1993) reported that the half-life of *T. caldophilus* DNA polymerase (*Tcal*) was 70 min at 95°C but no side-by-side comparisons have been published. In another paper from the same group, Jung et al. (1997) cloned and sequenced the DNA polymerase from *Thermus filiformis* but no biochemical data were reported. This polymerase, presumably from the type strain of *Thermus filiformis*, has 99.3% identity (six differences over 833 amino acids) with the polymerase from Rt4A.1 sequenced by us that is included in the *T. filiformis* clade (group A, Fig. 2) indicating that the sequence diversity between clade members may account for observed differences in thermostability.

Recent developments in the search for improved and enhanced DNA polymerases have been reviewed by Pavlov et al. (2004) who concluded that there is considerable potential in the construction of genetically modified polymerases that may have advantages over their naturally occurring counterparts. For example, the mutation of Arg660 of *Taq* polymerase based on crystal structure considerations eliminated the bias in favour of ddGTP in dideoxynucleotide incorporation (Li et al. 1999). Other changes have been achieved through directed evolution techniques as demonstrated by the use of compartmentalized self-replication to increase both the thermostability of *Taq* polymerase and its resistance to heparin without compromising catalytic efficiency (Ghadessey et al. 2001, 2004). In another example using compartmentalized replication, DNA polymerases have been evolved that can extend multiple mismatches and by-pass lesions found in ancient DNA (d' Abbadie et al. 2007). These results plus our survey of sequences from bioprospecting suggest that discovery of further natural variation of *Taq*-like DNA polymerases may be limited and that a directed evolution approach using the enzyme scaffolds already available for site-specific and/or directed evolution may be a more profitable pursuit. For example, it should be relatively simple to increase the thermostability of our cloned DNA polymerases by random mutation and directed evolution.

Acknowledgments Research reported here was funded by a Macquarie University Collaborative Research Grant.

References

Brock TD, Freeze H (1969) *Thermus aquaticus* gen. n. and sp. n., a non-sporulating extreme thermophile. J Bacteriol 98:289–297

Carballeira N, Nazabal M, Brito J, Garcia O (1990) Purification of a thermostable DNA polymerase from *Thermus thermophilus* HB8, useful in the polymerase chain reaction. Biotechniques 9:276–281

d' Abbadie M, Hofreiter M, Vaisman A, Loakes D, Gasparutto D, Cadet J, Woodgate R, Pääbo S, Holliger P (2007) Molecular breeding of polymerases for amplification of ancient DNA. Nat Biotechnol 25:939–943

Fujii S, Akiyama M, Aoki K, Sugaya Y, Higuchi K, Hiraoka M, Miki Y, Saitoh N, Yoshiyama K, Ihara K, Seki M, Ohtsubo E, Maki H (1999) DNA replication errors produced by the replicative apparatus of *Escherichia coli*. J Mol Biol 289:835–850

Ghadessey FJ, Ramsay N, Boudsocq F, Loakes D, Brown A, Iwai S, Vaisman A, Woodgate R, Holliger P (2004) Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. Nat Biotechnol 22:755–759

Ghadessy FJ, Ong JL, Holliger P (2001) Directed evolution of polymerase function by compartmentalized self-replication. Proc Natl Acad Sci USA 98:4552–4557

Harrell RA, Hart RP (1994) Rapid preparation of *Thermus flavus* DNA polymerase. PCR Meth Appl 3:372–375

Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds) (1990) PCR protocols: a guide to methods and applications. Academic press, San Diego, 482 pp

Jung SE, Choi JJ, Kim HK, Kwon S-T (1997) Cloning and analysis of the DNA polymerase-encoding gene from *Thermus filiformis*. Mol Cells 7:769–776

Kaisert MW, Lyamicheva N, Ma W, Miller C, Neri B, Fors L, Lyamichev VI (1999) A comparison of Eubacterial and Archaeal structure-specific 5′-exonucleases. J Biol Chem 274:21387–21394

Kim JS, Koh S, Kim JJ, Kwon ST, Lee DS (1998) Top DNA polymerase from *Thermus thermophilus* HB27: gene cloning, sequence determination, and physicochemical properties. Mol Cells 8:157–162

Kwon ST, Kim JS, Park JH, Kim HK, Lee DS (1997) Cloning and analysis of the DNA polymerase-encoding gene from *Thermus caldophilus* GK24. Mol Cells 7:264–271

Li Y, Mitaxov V, Waksman G (1999) Structure-based design of Taq DNA polymerases with improved properties of dideoxynucleotide incorporation. Proc Natl Acad Sc USA 96:9491–9496

Mo J-Y, Maki H, Sekiguchi M (1991) Mutational specificity of the *dna*E1733 mutator associated with a defect in the catalytic subunit of DNA polymerase III of *Escherichia coli*. J Mol Biol 222:925

Morris DD, Gibbs MD, Chin CW, Koh MH, Wong KKY, Allison RW, Nelson PJ, Bergquist PL (1998) Cloning of the *xynB* gene from *Dictyoglomus thermophilum* strain Rt46B.1 and action of the gene-product on kraft pulp. Appl Environ Microbiol 64:1759–1765

Myers TW, Gelfand DH (1991) Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. Biochemistry 30:7661–7666

Park JH, Kim JS, Kwon ST, Lee DS (1993) Purification and characterization of *Thermus caldophilus* GK24 DNA polymerase. Eur J Biochem 214:135–140

Pavlov AR, Pavlova NV, Kozyavkin SA, Slesarev AI (2004) Recent developments in the optimization of thermostable DNA polymerases for efficient applications. Trends Biotechnol 22:253–260

Perler FB, Kumar SJ, Kong H (1996) Thermostable DNA polymerases. Adv Protein Chem 48:377–435

Rose TM, Schultz ER, Henikoff JG, Pietrokovski S, McCallum CM, Henikoff S (1998) Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res 26:1637–1644

Saul DJ, Rodrigo AG, Reeves RA, Williams LC, Borges KM, Morgan HW, Bergquist PL (1993) Phylogeny of twenty *Thermus* isolates constructed from 16S rRNA gene sequence data. Int J Syst Bacteriol 43:754–760

Schauder B, Blöcker H, Frank R, McCarthy JEG (1987) Inducible expression vectors incorporating the *Escherichia coli atpE* transcriptional initiation region. Gene 52:279–283

Shandilya H, Griffiths K, Flynn EK, Astatke M, Shih P-J, Lee JE, Gerard GF, Gibbs MD, Bergquist PL (2004) Thermophilic bacterial DNA polymerases with reverse transcriptase activity. Extremophiles 8:243–251

Swofford D (2002) PAUP*: Phylogenetic analysis using parsimony (*and other methods), Sinauer Associates

Tabor S, Richardson CC (1995) A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. Proc Natl Acad Sci USA 96:6339–6343