(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 December 2007 (13.12.2007)

PCT

(10) International Publication Number
WO 2007/143436 A2

(51) International Patent Classification:
C12N 9/12 (2006.01)

(21) International Application Number:
PCT/US2007/069837

(22) International Filing Date: 29 May 2007 (29.05.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/803,424 30 May 2006 (30.05.2006) US

(71) Applicant *(for all designated States except US)*: **GE HEALTHCARE BIO-SCIENCES CORP.** [US/US]; 800 Centennial Avenue, Piscataway, New Jersey 08855 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **HU, Guolu** [US/US]; 3511 Sutton Road, Pepper Pike, Ohio 44122 (US). **GARNOVA, Elena** [RU/US]; 42 Tyler Avenue, Apt. 1101, Iselin, NJ 08830 (US). **XIAO, Haiguang** [US/US]; 120 York Drive, Princeton, New Jersey 08540 (US).

(74) Agents: **JI, Yonggang** et al.; GE Healthcare Bio-Sciences Corp., 800 Centennial Avenue, Piscataway, NJ 08855 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

[Continued on next page]

(54) Title: NOVEL DNA POLYMERASE FROM SPIROCHAETA THERMOPHILA

(57) Abstract: Disclosed is a DNA polymerase from Spirochaeta thermophila. This enzyme is useful for procedures requiring strand-displacing DNA synthesis such as SDA, for DNA sequencing, and polymerase chain reaction. Included within the scope of the present invention are various mutants (deletion and substitution) that retain the ability to replicate DNA as the native Spirochaeta thermophila polymerase.

Nucleic acid sequence of the wild-type *Spirochaeta thermophila* DNA polymerase gene.

(SEQ ID NO:1)

ATGAAACCGCTCTTCCTCATAGATGCCTACGGCCTCATCTATCGCGCCTACTTCGCCCTCATCAGGGCCCCCATGAGGAC
CAAGGACGGGCGCAACACGTCTGCGATCTACGGCTTCTTTCGTATGCTTTTCAGATTCGTGAAGGAGTACGATCCTGTCT
ATTGTGGCGTGGTGCTCGATTCCCTCACGCCCACCTTCAGGGAGAAGACCTTCGAGGCCTACAAGGCCACCAGGCAGAAG
ACACCGGAGGATCTCCATCCGCAGATCCCGGTGATCGAAGAGATCCTCGAGGCCTTGGGGTTCGCTACCGTCCGGATGAA
CGGGTTCGAGGCGGACGACGTGATGGGTGCGCTCGCCGCCATGGCCCAGGCGGAGGGACGCCCGTGTTTCGTGGTATCGG
GTGACAAGGACCTCGCCCAGCTGGTCACGAAGGGAGTGAGGATCGTCCGGATGGACAAGGACGATTTCGTGATTCTCGAC
GAGGAGGGGGTGAAGGAGAAGTGGGGGGTTCGGGCGGATCAGATCGTGGACTTCCTGGCCCTGGTGGGGGATGCCTCCGA
CAACATCCCGGGTGTGGAGGGGATCGGTGAGAAGACGGCCCGGCGCCTCCTCGCCGAGTATGGGTCGCTCGAGGGAGTCT
ATGCCCACCTCGACGAGCTCTCTCCCTCGTTGAGGAAGAAGCTCGAGGGGGGACGAGAGCGGGCCTTCCTGAGCAGGGAT
CTCGCCACGATACGGACGGACCTCGAACTCTCCCTCACGATCGAGGATCTCAGGCGGAGGGAACCCGAGGTACAGCGGGC
GAGGGAGCTCTTTCTCGCCTACGACATACGGAGCTTGGCCCAGGAGGTGGGAGGGAGTCCTGCCGTGGGGGAGAAGCCGA
CCCTCGAGGTGGAGGGTCCCTCCAGGGCGGGTGTCTCCTACGAGGTGGTCACCGACCGGGCCTCCCTCGCCCGATGGGTG
GACAAGGCCCTGGAGAGGGGGGTGGTGGCCGTGGATACCGAGACCGACGGGCTCGACCCTCTCACGTGCCGTCTCGTGGG
TGTCTCGCTCGCGGTGGACGAGGGGCGGGCGTGCTACGTGCCGCTCGCGGCTCCGGATGTGCGGCCGCTTCCCGCCGACG
TGGTGCGCGAGGTCCTCACGCCGCTTCTTGCCTCACAGAAGGTGTCGAAGGTGGGGCAGAACCTGAAGTTCGACTATCAC
GTGCTCAGGCGATGGGGCGTCCGTCCGGAGGGGCCGTTCTTCGATACCATGGTGGCGGCGTGGCTGCTGGAGTCGGATGC
CGGGCGCTACAACCTCGACAGGCTCGCGGAGAAGTATCTGGGGTGGCGCACCATCCATTACAAGGACGTAGTGGAAAAGG
GGGCTTCGTTCGAGACGGTGCCGGTGGCCGAGGCAGGGGTGTACGCGGCCGAGGATGCGGACATCGCCTTGCGGCTCTCC
CGGGTCTTGAAGGAACGTCTCGAGGCCGGGGATCTCGGGCGGGTCTTCTACGAGATGGAGATGCCGCTCCTCTCCATCCT
CGCTCGGATGGAGGAGTGGGGGATCGGGCTCGACGGGGAGGCTCTGTCGGCCTTCGGCAGGGAACTCGATGAGCGCATCT
CTAATATAGAGAATGAGATATTCGATCTCGTCGGGCACGAGTTCAACGTGGGGTCCACCAAGCAGCTCCAGGAGGTGCTG
TTCGTGGAGCGGGGGCTGCCGCCGGTGAAGAAGACCAAGACCGGTTTTTCCACGGACACGAGCGTGTTGGAGGAGCTGGC
TGCTCAGGATCCGGTGGCGGCGAAGGTCCTGGAGTACCGAAGTCTCACCAAGTTGAAGAACACCTATGTGGACGTGCTTC
CCGGGCTCGTGAATCCGGAGACGGGCCGGCTTCATACCTGGTTCAGCCAGGTGGGGACGGCCACGGGACGGCTCTCGAGC
AAGGATCCCAACCTCCAGAACATCCCGATCAAGACCGAGGAGGGTCGCCGGATCAGGGAGGCGTTCGTGCCGCAGCGGGG
CTGGTGGTTCCTGAGCGCGGACTACTCTCAGATCGAGCTCGTGATCCTGGCCCATCTCTCGGGGGATCCGGCTCTCTGCA
GGGCGTTCAGGGAGGGCGAGGACGTACACCGGGCTACGGGGAGCCTCCTGTTCGGGGTGCCGCCCGAGACAGTGACGCCC
GAGCAGCGGAGGATCGCCAAGACGATCAACTTCGGCGTGATCTACGGGATGTCGGCCTTCCGGCTCTCGAGGGAGCTCGG
TATCCCGCGCAAGGAGGCCGAACGGTTCATCAATGCGTACTTCACCACGTATGCGGGGGTGAGGGCGTTCATCGAACGGA
CGATCGCCGAGGCGGAGGAGAAGGGGTACGTGACCACGCTCTTCGGGAGGAAGCGGCCGCTTCCCTACATCACGAGCCGG
AACAAGACGCAGAAGACAGGAGCGGAGCGAATCGCGGTGAACACGCCGATCCAGGGCTCGGGCGCGGATATCATCAAGCT
GGCGATGATCGACCTGGACGCCCAGATGCGGAGGATGGGGCTCGCCTCCCGGATGATCCTCCAGGTGCACGACGAGCTCA
TCTTCGAGGTGCCTCCCGAGGAGCTCGAGGTGATGAAGCGGCTGGTGAGGGAGCGGATGGAGGGGGTGGTGCGGCTCTCG
GTGCCGCTTCGGGTGGAGATAGGGGTGGGGAGGAACTGGGGAGAGGCCCACTAG

WO 2007/143436 A2

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL DNA POLYMERASE FROM *SPIROCHAETA THERMOPHILA*

Cross-Reference to Related Applications

This application claims priority to United States provisional patent application number 60/803,424 filed May 30, 2006; the entire disclosure of which is incorporated herein by reference in its entirety.

Field of the Invention

The present invention relates to novel DNA polymerases obtainable from the thermophilic organism *Spirochaeta thermophila*, to certain deletions and mutants of this enzyme, to genes and vectors encoding the wild type and mutant polymerases and their use in strand displacement activity, polymerase chain reaction, DNA sequencing and as reverse transcriptases.

Background of the Invention

DNA polymerases are a family of enzymes involved in DNA repair and replication. DNA polymerases have been isolated from *E.coli* (e.g. *E.coli* DNA polymerase I and the Klenow fragment thereof) and T4 DNA polymerase and more recently thermostable DNA polymerases have been isolated (e.g. from *T. aquaticus,* U.S. Patent 4,889,818, and from *T. litoralis).* Thermostable DNA polymerases have been suggested (U.S. Patent 4,683,195) for use in amplifying existing nucleic acid sequences in amounts that are large compared to that originally present. The polymerase chain reaction (PCR) and strand displacement amplification (SDA) are two methods of amplifying nucleic acid sequences.

PCR is based on the hybridization of oligonucleotide primers to specific sequences on opposite strands of the target DNA molecule, and subsequent extension of these primers

with a DNA polymerase to generate two new strands of DNA which themselves can serve as a template for a further round of hybridization and extension. In PCR reactions the product of one cycle serves as the template for the next cycle such that at each repeat of the cycle the amount of the specific sequence present in the reaction doubles leading to an exponential amplification process.

PCR relies on a process of temperature cycling to promote the amplification reaction. This temperature cycling is necessary to separate the strands of DNA formed in one cycle of the reaction to allow hybridization of the oligonucleotides required to initiate the next cycle. DNA strand separation is usually achieved by melting the DNA at temperatures in the range 90-100°C followed by cooling to a lower temperature to allow oligonucleotide hybridization followed by extension or ligation, depending on the reaction process. This cycling process may be repeated 20-50 times, again depending on the process and the degree of amplification required.

Temperature cycling is normally achieved by the use of specialized equipment, commonly termed thermocyclers, which can operate by a wide variety of mechanical, electrical or hydraulic means, but serve a common purpose in heating and cooling a small container or a number of such containers in which the amplification reaction is performed.

In order for the amplification reactions to proceed with the desired efficiency and specificity it is necessary to perform the temperature cycling process within strictly defined and reproducible limits of temperature and time. Failure of the temperature cycling apparatus to achieve the required conditions will result in the partial or total failure of the amplification reaction. These strict requirements on time and temperature of cycling can impose severe restrictions if the handling of large numbers of reactions is required. If it is desired to perform several hundred or more reactions simultaneously use of conventional thermocyclers would be extremely expensive in terms of the capital investment required in

equipment, and would in any case be prone to variations between individual thermocyclers. Alternative solutions for large scale use have been constructed based on fan assisted oven or multiple water baths, but such solutions are inevitably cumbersome in use and in any case still require the user to take the normal stringent steps to avoid evaporation which become extremely difficult to apply to a large number of cycles.

In reverse transcription/polymerase chain reaction (RT/PCR), a DNA primer is hybridized to a strand of the target RNA molecule, and subsequent extension of this primer with a reverse transcriptase generates a new strand of DNA, which can serve as a template for PCR. Preparation of the DNA template is preferably carried out at an elevated temperature to avoid early termination of the reverse transcriptase reaction caused by RNA secondary structure. There is a lack of efficient reverse transcriptases that act at elevated temperatures, e.g. above 50°C.

SDA differs from PCR in being an isothermal amplification process, i.e. all reactions occur at the same temperature without the need for elevated temperature to melt DNA strands. This is made possible by adoption of a reaction scheme which uses the ability of certain DNA polymerases when extending along a DNA template strand to displace any DNA molecules already hybridized to the template. In SDA this strand displacement is used to separate the double stranded DNA produced during the reaction process and hence to maintain continuous amplification of the target DNA sequence (Walker, G.T., Little, M.C., Nadeau, J.G. and Shank D.D. (1992) Proc. *Natl. Acad. Sci. USA* 89:392-396). SDA is therefore in principle more suited to use with large numbers of samples than PCR as the isothermal process, which is performed at temperatures of 37°C to 60 °C, does not require stringent precautions to be taken to avoid evaporation and can be performed with simple temperature control equipment, for example in a standard laboratory incubator.

DNA polymerases, e.g. Sequenase, Klenow, Taq, etc, have also been extensively

used in DNA sequencing, see for example "Molecular Cloning: A Laboratory Manual" (Sambrook, Fritsch, and Maniatis, 2nd edition, Cold Spring Harbor laboratory Press, 1989).

Brief Description of the Invention

The present invention provides a DNA polymerase from *Spirochaeta thermophila*. This enzyme is useful for procedures requiring strand-displacing DNA synthesis such as SDA, for DNA sequencing, reverse transcription and polymerase chain reaction. Included within the scope of the present invention are various mutants (deletion and substitution) that retain the ability to replicate DNA with substantially the same efficiency as the native *Spirochaeta thermophila* polymerase.

Brief Description of the Drawings

Fig. 1 is the DNA sequence from *Spirochaeta thermophila* encoding a full length novel DNA polymerase (SEQ ID NO:1).

Fig. 2 is the amino acid sequence encoding the full length DNA polymerase from *Spirochaeta thermophila* (SEQ ID NO:2). Translation is of the open reading frame spanning SEQ ID NO:1 as shown in Figure 1, encoding a native polymerase.

Figure 3 is a DNA sequence encoding the DNA polymerase from *Spirochaeta thermophila*, containing Y74C, F731Y and E745R mutations (SEQ ID NO:3).

Figure 4 is the amino acid sequence of the DNA polymerase from *Spirochaeta thermophila*, containing Y74C, F731Y and E745R mutations (SEQ ID NO:4).

Figure 5 is DNA sequence encoding a truncated version of a DNA polymerase from *Spirochaeta thermophila*, containing F731Y and E745R mutations (SEQ ID NO:5).

Figure 6 is the amino acid sequence of the truncated version of DNA polymerase

from *Spirochaeta thermophila*, containing F731Y and E745R mutations (SEQ ID NO:6).

Figure 7 is DNA sequence encoding a preferred, truncated version of a DNA polymerase from *Spirochaeta thermophila*, containing F731Y and E745R mutations (SEQ ID NO:7).

Figure 8 is an alignment of several DNA polymerase protein sequences. From the top, *Spirochaeta thermophila* (SEQ ID NO:2), NCBI AAR11879 (SEQ ID NO:8), *Pseudomonas aeruginosa* UCBPP-PA14 (SEQ ID NO:9), *Pseudomonas aeruginosa* PAO1 (SEQ ID NO:10), *Methylococcus capsulatus* str. Bath (SEQ ID NO:11), *Nitrococcus mobilis* Nb-231 (SEQ ID NO:12), *Erwinia carotovora* subsp. *Atroseptica* (SEQ ID NO:13), *Photorhabdus luminescens* subsp. *laumondii* TTO1 (SEQ ID NO:14) and *Yersinia bercovieri* ATCC 43970 (SEQ ID NO:15).

Detailed Description of the Invention

In a first aspect, the present invention provides a DNA polymerase or fragment thereof having the DNA polymerase activity of *Spirochaeta thermophila* and having at least 80% amino acid homology, preferably at least 90% homology, most preferably at least 96% homology to at least a contiguous 40 amino acid sequence shown in Fig. 2 (SEQ ID NO:2). Fig. 2 represents the translation of the open reading frame of DNA sequence encoding a DNA polymerase from *Spirochaeta thermophila* (Fig. 1) (SEQ ID NO:1) potentially encoding the native polymerase.

When used herein, the term amino acid homology means the amino acid identity of the parent enzyme or conservative amino acid changes thereto. The DNA polymerase can be encoded by a full-length nucleic acid sequence or any portion of the full-length nucleic acid sequence, so long as a functional activity of the polypeptide is retained. The amino acid sequence will be substantially similar to the sequence shown in Fig. 2, or fragments

thereof. A sequence that is substantially similar will preferably have at least 80% identity (more preferably at least 90% and most preferably 98-100%) to the sequence of Fig. 2.

By "identity" is meant a property of sequences that measures their similarity or relationship. Identity is measured by dividing the number of identical residues by the total number of residues and multiplying the product by 100. Thus, two copies of exactly the same sequence have 100% identity, but sequences that are less highly conserved, and have deletions, additions, or replacements, may have a lower degree of identity. Those skilled in the art will recognize that several computer programs are available for determining sequence identity.

The enzyme of the present invention has a molecular weight of approximately 100,000 daltons. It possesses a 5'-3' exonuclease activity.

When used herein, the term a DNA polymerase or fragment thereof having the DNA polymerase activity of *Spirochaeta thermophila* means a DNA polymerase or fragment thereof (as hereinafter defined) which has the ability to replicate DNA with substantially the same efficiency as the enzyme encoded by the SEQ ID NO:1. By substantially the same efficiency is meant at least 80% and preferably at least 90% of the efficiency of the enzyme encoded by SEQ ID NO:1 to incorporate deoxynucleotides.

The invention also encompasses a stable enzyme composition which comprises a purified DNA polymerase from *Spirochaeta thermophila* in a buffer.

The DNA polymerases of the present invention are preferably in a purified form. By purified is meant that the DNA polymerase is isolated from a majority of host cell proteins normally associated with it; preferably the polymerase is at least 10% (w/w), e.g. at least 50% (w/w) , of the protein of a preparation, even more preferably it is provided as a homogeneous preparation, e.g. homogeneous solution. Preferably the DNA polymerase is a single polypeptide on an SDS polyacrylamide gel.

Buffers around neutral pH (5-9) such as 5-100 mM TrisHC1, phosphate or MES are suitable for use in the current invention.

The present invention also provides a gene encoding a polymerase of the present invention. Fig. 1 represents the DNA sequence of a full length native polymerase from *Spirochaeta thermophila* (SEQ ID NO:1).

It has been found that the entire amino acid sequence of the polymerase is not required for enzymatic activity. Thus, for example, the exonuclease domain of the enzyme has been deleted to give an enzyme of molecular weight of approximately 70,000 daltons which retains enzyme activity. This exonuclease-free enzyme is analogous to the Klenow fragment of *E. coli* DNA polymerase I. Thus, the present invention also provides fragments of the polymerase which retain the DNA polymerase activity of *Spirochaeta thermophila* but have one or more amino acids deleted, preferably from the amino-terminus, while still having at least 80% amino acid homology to at least a 40 contiguous amino acid sequence shown in Fig. 2 (SEQ ID NO: 2) .

In a further aspect, the present invention provides a DNA polymerase which corresponds to the DNA polymerase from *Spirochaeta thermophila* in which up to one third of the amino acid sequence at the amino-terminus has been deleted. In particular, fragments of *Spirochaeta thermophila* having N-terminal deletions to give 630 amino acids (See Fig. 6) (SEQ ID NO:6) have been found to retain enzyme activity.

It is preferred that the 5'-3' exonuclease activity of the DNA polymerase is removed or reduced. This may be achieved by deleting the amino acid region of the enzyme responsible for this activity, e.g. by deleting up to one third of the amino acid sequence at the amino terminus (Fig. 6) (SEQ ID NO:6), or by appropriate amino acid changes (Y74C, See Fig. 4) (SEQ ID NO:4).

In addition to the N-terminal deletions and amino acid changes to remove the

exonuclease activity, the enzyme may have conservative amino acid changes compared with the native enzyme which do not significantly influence enzyme activity. Such changes include substitution of like charged amino acids for one another or amino acids with small side chains for other small side chains, e. g. ala for val. More drastic changes may be introduced at non-critical regions where little or no effect on polymerase activity is observed by such a change.

Joyce and Steitz, *Annu. Rev. Biochem*, 63:777-822, 1994, discuss various functions of DNA polymerases including the catalytic center, the binding site for the 3' terminus of the primer, and the dNTP binding site. In particular, it mentions mutations that affect the binding of dNTP in the ternary complex. European Patent Application 0655506A discloses that the presence of a polar, hydroxyl containing amino acid residue at a position near the binding site for the dNTP substrate is important for the polymerase being able to efficiently incorporate a dideoxynucleotide. Applicant has discovered that the modification of the dNTP binding site for the dNTP substrate in DNA polymerase obtainable from *Spirochaeta thermophila* by the inclusion of a polar, hydroxyl containing amino acid residue at a position near the binding site increases the efficiency of the polymerase to incorporate a dideoxynucleotide. Preferably the polar, hydroxyl containing amino acid is tyrosine. It has also been found that replacing the phenylalanine at the position corresponding to 731 of the native enzyme with tyrosine improves the incorporation of dideoxynucleotides when the enzyme is used for sequencing. In particular, a polymerase from *Spirochaeta thermophila* in which the exonuclease activity has been deleted e.g. by point mutation or deletion and which has the phenylalanine at the position corresponding to 731 of the native enzyme replaced by an amino acid which increases the efficiency of the enzyme to incorporate dideoxynucleotides at least 20 fold compared to the wild type enzyme, e.g. tyrosine, is a particularly preferred enzyme for use in sequencing. Preferably, this modified enzyme has

from between 592 and 634 amino acids, for example 612 to 634 amino acids and preferably 630 amino acids.

Several modified DNA polymerase sequences are provided, containing the FY mutation and other, preferable point mutations and truncations (See Fig. 3-6) (SEQ ID NO:3-6).

The DNA polymerases of the present invention can be constructed using standard techniques familiar to those who practice the art. By way of example, in order to prepare a polymerase with the phenylalanine to tyrosine mutation, mutagenic PCR primers can be designed to incorporate the desired Phe to Tyr amino acid change (FY mutation in one primer). Deletion of the exonuclease function is carried out by PCR to remove the amino terminus, or standard techniques of site directed mutagenesis to generate point mutations.

Improved expression of the DNA polymerases of the present invention can be achieved by introducing silent codon changes (i.e., the amino acid encoded is not changed). Such changes can be introduced by the use of mutagenic PCR primers. Silent codon changes such as the following increase protein production in E. *coli:*

substitution of the codon GAG for GAA;

substitution of the codon AGG, AGA, CGG or CGA for CGT or CGC;

substitution of the codon CTT, CTC, CTA, TTG or TTA for CTG; substitution of the codon ATA for ATT or ATC;

substitution of the codon GGG or GGA for GGT or GGC.

Figure 7 (SEQ ID NO:7) provides a nucleic acid sequence encoding the N-terminal truncated *Spirochaeta thermophila* DNA polymerase enzyme, with F731Y, E745R mutations, that incorporates such silent codon substitutions. The sequence of this clone was generated based on analysis performed using a software program: *E.coli* codon usage analysis 2.0 (by Morris Maduro).

Genes encoding the DNA polymerase from *Spirochaeta thermophila* polymerase in which up to one third of the amino acid sequence at the amino terminus has been deleted and which have the exonuclease activity removed by point mutation and such polymerases which incorporate the phenylalanine to tyrosine modification are also provided by the present invention.

The DNA polymerases of the present invention are suitably used in SDA, preferably in combination with a thermostable restriction enzyme. Accordingly, the present invention provides a composition which comprises a DNA polymerase of the present invention in combination with a thermostable restriction enzyme, for example *BsoBI* from *Bacillus stearothermophilus.* The invention also features a kit or solution for SDA comprising a DNA polymerase of the present invention in combination and a thermostable restriction enzyme.

The polymerases of the present invention are also useful in methods for generating and amplifying a nucleic acid fragment via a strand displacement amplification (SDA) mechanism. The method generally comprises:

a) specifically hybridizing a first primer 5' to a target nucleic acid sequence, the first primer containing a restriction enzyme recognition sequence 5' to a target binding region;

b) extending the 3' ends of the hybridized material using a DNA polymerase of the present invention, preferably one in which the exonuclease activity has been removed, in the presence of three dNTPs and one dNTPα S;

c) nicking at the hemiphosphorothioate recognition site with a restriction enzyme, preferably;

d) extending the 3' end at the nick using a DNA polymerase of the present invention, displacing the downstream complement of the target strand; and

e) repeating steps (c) and (d).

This SDA method proceeds at a linear amplification rate if one primer is used as above. However, if two primers are used which hybridize to each strand of a double-stranded DNA fragment, then the method proceeds exponentially (Walker, G.T., Little, M.C., Nadeau, J.G. and Shank D.D. (1992) Proc. *Natl. Acad. Sci. USA* 89:392-396).

The present invention also provides a method for determining the nucleotide base sequence of a DNA molecule. The method includes providing a DNA molecule, annealing with a primer molecule able to hybridize to the DNA molecule; and incubating the annealed molecules in a vessel containing at least one, and preferably four deoxynucleotide triphosphate, and a DNA polymerase of the present invention preferably one containing the phenylalanine to tyrosine mutation. Also provided is at least one DNA synthesis terminating agent which terminates DNA synthesis at a specific nucleotide base. The method further includes separating the DNA products of the incubating reaction according to size, whereby at least a part of the nucleotide base sequence of the DNA molecule can be determined.

In preferred embodiments, the sequencing is performed at a temperature between 40°C and 75°C.

In other preferred embodiments, the DNA polymerase has less than 1000, 250, 100, 50, 10 or even 2 units of exonuclease activity per mg of polymerase (measured by standard procedure, see below) and is able to utilize primers having only 4, 6 or 10 bases; and the concentration of all four deoxynucleoside triphosphates at the start of the incubating step is sufficient to allow DNA synthesis to continue until terminated by the agent, e.g. a ddNTP.

Preferably, more than 2, 5, 10 or even 100 fold excess of a dNTP is provided to the corresponding ddNTP.

In a related aspect, the invention features a kit or solution for DNA sequencing including a DNA polymerase of the present invention and a reagent necessary for the

sequencing such as dITP, deaza dGTP, a chain terminating agent such as a ddNTP, and optionally a pyrophosphatase.

The DNA polymerases of the present invention containing the phenylalanine to tyrosine mutation are suitably used in sequencing, preferably in combination with a pyrophosphatase. Accordingly, the present invention provides a composition which comprises a DNA polymerase of the present invention containing the phyenylalanine to tyrosine mutation in combination with a pyrophosphatase, preferably a thermostable pyrophosphatase from *Thermoplasma acidophilum*.

In another related aspect, the invention features a method for sequencing a strand of DNA essentially as described above with one or more (preferably 2, 3 or 4) deoxyribonucleoside triphosphates, a DNA polymerase of the present invention, and a first chain terminating agent. The DNA polymerase causes the primer to be elongated to form a first series of first DNA products differing in the length of the elongated primer, each first DNA product having a chain terminating agent at its elongated end, and the number of molecules of each first DNA products being approximately the same for substantially all DNA products differing in length by no more than 20 bases. The method also features providing a second chain terminating agent in the hybridized mixture at a concentration different from the first chain terminating agent, wherein the DNA polymerase causes production of a second series of second DNA products differing in length of the elongated primer, with each second DNA product having the second chain terminating agent at its elongated end. The number of molecules of each second DNA product is approximately the same for substantially all second DNA products differing in length from each other by from 1 to 20 bases, and is distinctly different from the number of molecules of all the first DNA products having a length differing by no more than 20 bases from that of said second DNA products.

In preferred embodiments, three or four such chain terminating agents can be used to make different products and the sequence reaction is provided with a magnesium ion, or even a manganese or iron ion (e. g. at a concentration between 0.05 and 100 mM, preferably between 1 and 10mM)i and the DNA products are separated according to molecular weight in four or less lanes of a gel.

In another related aspect, the invention features a method for sequencing a nucleic acid by combining an oligonucleotide primer, a nucleic acid to be sequenced, between one and four deoxyribonucleoside triphosphates, a DNA polymerase of the present invention, and at least two chain terminating agents in different amounts, under conditions favoring extension of the oligonucleotide primer to form nucleic acid fragments complementary to the nucleic acid to be sequenced. The method further includes separating the nucleic acid fragments by size and determining the nucleic acid sequence. The agents are differentiated from each other by intensity of a label in the primer extension products.

In a further aspect, the present invention provides a method for preparing complementary DNA by combining an oligonucleotide primer, a sample of RNA, a DNA polymerase of the present invention, and between one and four deoxyribonucleoside phosphates, under conditions favoring preparation of the complementary DNA.

The DNA polymerases of the present invention which act as reverse transcriptases lack appreciable, and preferably have no RNaseH activity and, as such, are useful in RT/PCR, the generation of hybridization probes and RNA sequencing. In a yet further aspect, the present invention provides a purified reverse transcriptase having a reverse transcriptase activity of greater than 1000 units per milligram. Preferably, the reverse transcriptase lacks RNaseH activity; the reverse transcriptase is from *Spirochaeta thermophila*; the reverse transcriptase has an N-terminal deletion or amino acid changes that remove the exonuclease function. In a further aspect, the invention features a method

for reverse transcription/polymerase chain reaction (RT/PCR) which utilizes a DNA polymerase of the present invention and a DNA polymerase suitable for PCR in the same reaction vessel. Preferably, the DNA polymerase of the present invention is from *Spirochaeta thermophila*, the polymerase has one or more amino acids deleted from the amino terminus or amino acid changes to remove the exonuclease activity, the DNA polymerase suitable for PCR is Taq DNA polymerase. In another aspect, the present invention features a kit or solution for RT/PCR comprising a DNA polymerase of the present invention and a DNA polymerase suitable for PCR. Preferably, the DNA polymerase of the present invention is from *Spirochaeta thermophila*, the polymerase has one or more amino acids deleted from the amino terminus or amino acid changes to remove the exonuclease function, and the DNA polymerase suitable for PCR is Taq DNA polymerase.

In a further preferred embodiment the invention features polymerases that have exonuclease activity removed by replacing an aspartic acid residue at position 8 with an alanine residue. This is a preferred mutation for it not only removes the exonuclease activity but removes an RNaseH activity, which may effect reverse transcription.

In another aspect, the invention features a method for polymerase chain reaction in the presence of a polymerase stabilizing agent, utilizing an enzymatically active DNA polymerase having at least 80% identity in its amino acid sequence to the DNA polymerase of *Spirochaeta thermophila* and an exonuclease activity removed. The polymerase stabilizing agents include, but are not limited to, glycerol (10 to 50% final concentration), trimethylamine-N-oxide *(TMANO;* up to 4 M final concentration), and N-methylmorpholine-N-oxide *(MMO;* up to 3 M final concentration). Preferably, glycerol is used at a final concentration of 30%. By polymerase stabilizing agent is meant an agent which allows the use of the polymerase in PCR and RT/PCR. These agents reduce the

denaturing temperature of the template and stabilize the polymerase. By stabilize is meant make temperature stable. By final concentration is meant the final concentration of the agent in the PCR or RT/PCR solution.

The DNA polymerases of the present invention can be used to carry out polymerase chain reaction (PCR) when the reaction is carried out in the presence of a polymerase stabilizing agent, such as glycerol, TMANO, or MMO, and the like. Preferably, the glycerol concentration is in the range of 10 to 50%, and most preferably it is at a final concentration of 30%. TMANO and MMO can be used at concentrations up to 4 M and 3 M final concentration, respectively. These agents have been shown to stabilize the polymerase reaction. Thus, the polymerases of the present invention are also suitable for reverse transcription/polymerase chain reaction (RT/PCR) under these conditions.

Thus, in another aspect the invention features a method for reverse transcription/polymerase chain reaction, in the presence of a polymerase stabilizing agent, utilizing an enzymatically active DNA polymerase having at least 80% identity in its amino acid sequence to the DNA polymerase of *Spirochaeta thermophila* and an exonuclease activity removed. In preferred embodiments, the polymerase stabilizing agents include, but are not limited to, glycerol (10 to 50% final concentration), TMANO (up to 4 M final concentration), and MMO (up to 3 M final concentration). Preferably, glycerol is used at a final concentration of 30%.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments, and from the claims.

Examples

The following examples serve to illustrate the DNA polymerases of the present invention and are not intended to be limiting.

Example 1: Cloning of native polymerase and generation of mutant enzymes

The *Spirochaeta thermophila* bacterial strain was obtained from ATCC culture collection (ATCC 700085) as a freeze-dried culture, and was dissolve in 50 ul of sterile water. Genomic DNA amplification was performed as described (GenomiPhi™, GE Healthcare). The polymerase gene was then amplified by PCR and cloned into PKK vector at the EcoRI and HindIII Site. The primers were designed using the known genomic DNA sequence, with the introduction of restriction enzyme sites. The primer at N-terminus (with EcoRI site) is 5'- GCG CGA ATT CAT GAA ACC GCT CTT CCT CAT AGA TGC CTA CGG - 3' (SEQ ID NO:16). The primer at the C-terminus (with HindIII site) is 5' – GCG CAA GCT TGT CGA CCC CCT AGT GGG CCT CTC CCC AG – 3' (SEQ ID NO:17). The PCR product was generated using Taq DNA polymerase and cloned into PKK vector.

The full length clone was sequenced to identify the DNA sequence (SEQ ID NO:1) that encodes the full length DNA polymerase (SEQ ID NO:2). We note that the sequence we obtained is substantially different from the published sequence from another group (GenBank Accession numbers AAR11879 and AY247648 for protein and nucleic acid sequences respectively; see Shandilya et al, Extremophiles, 8:243-251, 2004). The amino acid sequences of the two are only 96 % homologous and 98 % identical. Our repeated sequence analyses of the clones prove that we obtained the novel, correct sequence for the DNA polymerase.

Point mutants were generated by using the Quickchange site-directed mutagenesis kit (Stratagene). The following primers were used: (1) for FY, forward: 5'- GCC AAG ACG ATC AAC TAC GGC GTG ATC TAC G -3' (SEQ ID NO:18), and reverse: 5'- CGT AGA TCA CGC CGT AGT TGA TCG TCT TGG C -3' (SEQ ID NO:19); (2) ER, forward: 5'- GGC TCT CGA GGC GGC TCG GTA TCC CGC GCA AG -3' (SEQ ID NO:20), and reverse: 5'- CTT GCG CGG GAT ACC GAG CCG CCT CGA GAG CC -3' (SEQ ID

NO:21); (3) YC, forward: 5'- GAA GAC CTT CGA GGC CTG CAA GGC CAC CAG -3' (SEQ ID NO:22), and reverse: 5'- CTG GTG GCC TTG CAG GCC TCG AAG GTC TTC -3' (SEQ ID NO:23). Pfu High-Fidelity polymerase and standard PCR conditions were used.

The N-terminal deletion versions of the FY/DR mutant were also generated. The truncated clones start at amino acid sequence 269 of SEQ ID NO:2. These clones were generated from the double FY/DR mutant by PCR using EcoRI and HindIII restriction sites and the following primers. To generate the mutants as shown in Figure 5 (SEQ ID NO:5), the following primers are used: Forward: 5'- GGA ATT CAT GGA GCT CTT TCT CGC CTA CGA CAT ACG GAG CTT GG -3' (SEQ ID NO:24); Reverse: SEQ ID NO:17. To generate the mutants as shown in Figure 7 (SEQ ID NO:7), with preferred codon for E.coli expression, the following primers are used: Forward: 5'- GGA ATT CAT GGA GCT GTT TCT GGC CTA CGA CAT ACG GAG CTT GG -3' (SEQ ID NO:25); Reverse: SEQ ID NO:17. Amplified products were cloned into PKK vector.

Example 2: Purification of Cloned Enzymes

The full length enzyme clones were fermented using 2X LB medium. Overnight cultures of the plasmid-containing strains were inoculated into 2X LB medium. The polymerase expression was induced at 1.0 OD by adding 1mM IPTG. The cells were harvested after growth for 4 hours at 37°C.

Cell broth was centrifugation to collect cell pellets. The cell paste containing full length enzymes was resuspended in 4 volume lysis buffer (50 mM Tris-Cl, pH 7.5; 50 mM NaCl; 0.5 mM EDTA; 0.2% NP-40; 0.2% Tween 20, 1mM PMSF) and normalized to OD. Lysis was carried out by heat at 75 °C for 15 min, then 1mg/ml lysozyme was added and the cell lysate was centrifuged at 6,000 rpm for 20 min. The clear supernatant was collected

and analyzed on SDS-PAGE. A protein of expected size (about 100,000 daltons) can be detected.

The supernatant was loaded onto 5ml HiTrap Heparin Sepharose HP. The column was washed with 10 column volumes with Buffer A (50mM Tris pH 8.5; 1mM EDTA; 50mM NaCl). The column was further washed with gradient: 0 – 60% Buffer B (50mM Tris pH 8.3; 1mM EDTA; 1M NaCl) in 10 column volumes. Elution was continued by stepping up to 100% buffer B for 5 column volumes. 5ml fractions were collected.

Fractions showing polymerase activity were dialyzed overnight against Buffer A so that the conductivity is less than 10mS/cm. The resulting sample was loaded onto a 5ml HiTrap HiTrap Q column, washed with 10 column volumes of Buffer A, then washed with gradient: 0 – 60% Buffer B in 10 column volumes. Elution was continued by stepping up to 100% buffer B for 5 column volumes. 5ml fractions were collected.

Peak fractions from HiTrapQ were dialyzed overnight against Buffer A until conductivity is less than 10mS/cm. Sample was loaded onto 5ml HiTrap Q Sepharose HP, washed with 10 column volumes of Buffer A, then with gradient: 0 – 60% Buffer B in 10 column volumes. Elution was continued by stepping up to 100% buffer B for 5 column volumes. 5ml fractions were collected.

Fractions containing polymerase activity was pooled and dialyzed against final storage buffer (20 mM Tris-Cl, pH 8.5, 25 mM KCL, 0.1 mM EDTA, 50% glycerol, 0.5% NP-40, 0.5% Tween 20 and 1mM DTT).

Example 3: Characterization of Enzymes

Thermostability

Thermostability of the enzymes is assayed at 95°C, as 50% activity at time determined by DAPI assay. The enzyme is diluted to 20ng/μl in enzyme dilution buffer

(25mM Tris pH 8.0, 50mM KCl, 0.5% (v/v) Tween 20, and 0.5% (v/v) NP-40). A 20µl reaction is set up with the following final composition (50mM Tris pH 8.0, 5mM $MgCl_2$, 50ng M13 DNA, 50µM dNTPs and 20ng enzyme). The reaction is incubated at 95˚C for 0, 1, 2, 4, 6, 8, 10 and 12 minutes.

DAPI assay is performed by taking 5µl of the 20µl reaction, adding it to 195µl DAPI reaction buffer (80mM Tris pH 9.0, 2.4mM $MgCl_2$, 5µM FAS42 template, 7.5µM DAPI and 250µM dNTPs). DAPI plate assay is run at 37°C, and the maximum slopes are plotted.

DNA sequencing

The sequencing premix composition is formulated with DYEnamic ET terminators and dITP/dA, T, C (2500 µM dITP, 500 µM dCTP, 500 µM dATP, and 500 µM dTTP). The ratio of dNTP:ddNTP is 156. The composition also contains 20 units of TS II enzyme mix or *Spirochaeta thermophila* polymerase per reaction.

5 pmoles of -40 M13 primer, 200 ng M13 DNA, 8 µl of terminator premix composition and water are added to a total volume of 20 µl. Reaction mixtures are cycled through 95°C, 20 seconds; 50°C, 30 seconds; and 60°C, 60 seconds, repeated 30 times. Reactions are then held at 4°C until purification and analysis.

The samples are purified and analyzed according to manufacturer's instructions, and run on ABI 3100 capillary sequencing instrument (Applied Biosystems). The resulting electropherograms using TS II enzyme mix and *Spirochaeta thermophila* polymerase are compared.

One skilled in the art would readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as

those inherent therein. The methods, kits, solutions, and molecules, described herein are presently representative of preferred embodiments are exemplary and are not intended as limitations on the scope of the invention. Changes therein and other uses will occur to those skilled in the art which are encompassed within the spirit of the invention are defined by the scope of the claims.

It will be readily apparent to one skilled in the art that varying substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

All patents and publications mentioned in the specification are indicative of the levels of those skilled in the art to which the invention pertains. All patents and publications are herein incorporated by reference to the same extent as if each individual publication was specifically and individually indicated to be incorporated by reference.

Other embodiments are within the following claims.

What is claimed is:

1. A novel DNA polymerase protein comprising:

(a) a DNA polymerase from *Spirochaeta thermophila* having amino acid sequence of SEQ ID NO:2, SEQ ID NO:4 or SEQ ID NO:6; or

(b) a polypeptide having a sequence with at least 96% similarity to that of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8.

2. The DNA polymerase of claim 1 with strand displacement activity.

3. The DNA polymerase of claim 1 with reverse transcriptase activity.

4. A nucleic acid sequence encoding a DNA polymerase, comprising:

(a) a nucleotide having the sequence of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:7; or

(b) a polynucleotide having a sequence with at least 90% similarity to that of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5 or SEQ ID NO:7.

5. A vector containing a nucleic acid sequence encoding a protein of claim 1.

6. A transformed cell containing a vector of claim 5.

7. A vector containing a nucleic acid sequence of claim 4.

8. A transformed cell containing a vector of claim 7.

9. In a method for reverse transcription of a RNA strand to a DNA strand, the improvement comprises using the novel DNA polymerase of claim 1 as the reverse transcriptase.

10. In a strand displacement amplification method, the improvement comprises using the novel DNA polymerase of claim 1 as the amplification enzyme.

11. A novel DNA polymerase protein comprising:

a polypeptide having a sequence with at least 98% similarity to that of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6 or SEQ ID NO:8.

Figure 1

Nucleic acid sequence of the wild-type *Spirochaeta thermophila* DNA polymerase gene.
(SEQ ID NO:1)

ATGAAACCGCTCTTCCTCATAGATGCCTACGGCCTCATCTATCGCGCCTACTTCGCCCTCATCAGGGCCCCCATGAGGAC
CAAGGACGGGCGCAACACGTCTGCGATCTACGGCTTCTTTCGTATGCTTTTCAGATTCGTGAAGGAGTACGATCCTGTCT
ATTGTGGCGTGGTGCTCGATTCCCTCACGCCCACCTTCAGGGAGAAGACCTTCGAGGCCTACAAGGCCACCAGGCAGAAG
ACACCGGAGGATCTCCATCCGCAGATCCCGGTGATCGAAGAGATCCTCGAGGCCTTGGGGTTCGCTACCGTCCGGATGAA
CGGGTTCGAGGCGGACGACGTGATGGGTGCGCTCGCCGCCATGGCCCAGGCGGAGGGACGCCCGTGTTTCGTGGTATCGG
GTGACAAGGACCTCGCCCAGCTGGTCACGAAGGGAGTGAGGATCGTCCGGATGGACAAGGACGATTTCGTGATTCTCGAC
GAGGAGGGGGTGAAGGAGAAGTGGGGGGGTTCGGGCGGATCAGATCGTGGACTTCCTGGCCCTGGTGGGGGATGCCTCCGA
CAACATCCCGGGTGTGGAGGGGATCGGTGAGAAGACGGCCCGGCGCCTCCTCGCCGAGTATGGGTCGCTCGAGGGAGTCT
ATGCCCACCTCGACGAGCTCTCTCCCTCGTTGAGGAAGAAGCTCGAGGGGGGACGAGAGCGGGCCTTCCTGAGCAGGGAT
CTCGCCACGATACGGACGGACCTCGAACTCTCCCTCACGATCGAGGATCTCAGGCGGAGGGAACCCGAGGTACAGCGGGC
GAGGGAGCTCTTTCTCGCCTACGACATACGGAGCTTGGCCCAGGAGGTGGGAGGGAGTCCTGCCGTGGGGGAGAAGCCGA
CCCTCGAGGTGGAGGGTCCCTCCAGGGCGGGTGTCTCCTACGAGGTGGTCACCGACCGGGCCTCCCTCGCCCGATGGGTG
GACAAGGCCCTGGAGAGGGGGGTGGTGGCCGTGGATACCGAGACCGACGGGCTCGACCCTCTCACGTGCCGTCTCGTGGG
TGTCTCGCTCGCGGTGGACGAGGGGCGGGCGTGCTACGTGCCGCTCGCGGCTCCGGATGTGCGGCCGCTTCCCGCCGACG
TGGTGCGCGAGGTCCTCACGCCGCTTCTTGCCTCACAGAAGGTGTCGAAGGTGGGGCAGAACCTGAAGTTCGACTATCAC
GTGCTCAGGCGATGGGGCGTCCGTCCGGAGGGGCCGTTCTTCGATACCATGGTGGCGGCGTGGCTGCTGGAGTCGGATGC
CGGGCGCTACAACCTCGACAGGCTCGCGGAGAAGTATCTGGGGTGGCGCACCATCCATTACAAGGACGTAGTGGAAAAGG
GGGCTTCGTTCGAGACGGTGCCGGTGGCCGAGGCAGGGGTGTACGCGGCCGAGGATGCGGACATCGCCTTGCGGCTCTCC
CGGGTCTTGAAGGAACGTCTCGAGGCCGGGGATCTCGGGCGGGTCTTCTACGAGATGGAGATGCCGCTCCTCTCCATCCT
CGCTCGGATGGAGGAGTGGGGGATCGGGCTCGACGGGGAGGCTCTGTCGGCCTTCGGCAGGGAACTCGATGAGCGCATCT
CTAATATAGAGAATGAGATATTCGATCTCGTCGGGCACGAGTTCAACGTGGGGTCCACCAAGCAGCTCCAGGAGGTGCTG
TTCGTGGAGCGGGGGCTGCCGCCGGTGAAGAAGACCAAGACCGGTTTTTCCACGGACACGAGCGTGTTGGAGGAGCTGGC
TGCTCAGGATCCGGTGGCGGCGAAGGTCCTGGAGTACCGAAGTCTCACCAAGTTGAAGAACACCTATGTGGACGTGCTTC
CCGGGCTCGTGAATCCGGAGACGGGCCGGCTTCATACCTGGTTCAGCCAGGTGGGGACGGCCACGGGACGGCTCTCGAGC
AAGGATCCCAACCTCCAGAACATCCCGATCAAGACCGAGGAGGGTCGCCGGATCAGGGAGGCGTTCGTGCCGCAGCGGGG
CTGGTGGTTCCTGAGCGCGGACTACTCTCAGATCGAGCTCGTGATCCTGGCCCATCTCTCGGGGGATCCGGCTCTCTGCA
GGGCGTTCAGGGAGGGCGAGGACGTACACCGGGCTACGGGGAGCCTCCTGTTCGGGGTGCCGCCCGAGACAGTGACGCCC
GAGCAGCGGAGGATCGCCAAGACGATCAACTTCGGCGTGATCTACGGGATGTCGGCCTTCCGGCTCTCGAGGGAGCTCGG
TATCCCGCGCAAGGAGGCCGAACGGTTCATCAATGCGTACTTCACCACGTATGCGGGGGTGAGGGCGTTCATCGAACGGA
CGATCGCCGAGGCGGAGGAGAAGGGGTACGTGACCACGCTCTTCGGGAGGAAGCGGCCGCTTCCCTACATCACGAGCCGG
AACAAGACGCAGAAGACAGGAGCGGAGCGAATCGCGGTGAACACGCCGATCCAGGGCTCGGGCGCGGATATCATCAAGCT
GGCGATGATCGACCTGGACGCCCAGATGCGGAGGATGGGGCTCGCCTCCCGGATGATCCTCCAGGTGCACGACGAGCTCA
TCTTCGAGGTGCCTCCCGAGGAGCTCGAGGTGATGAAGCGGCTGGTGAGGGAGCGGATGGAGGGGGTGGTGCGGCTCTCG
GTGCCGCTTCGGGTGGAGATAGGGGTGGGGAGGAACTGGGGAGAGGCCCACTAG

Figure 2

Amino acid sequence of the wild-type *Spirochaeta thermophila* DNA polymerase enzyme. **(SEQ ID NO:2)**

MKPLFLIDAYGLIYRAYFALIRAPMRTKDGRNTSAIYGFFRMLFRFVKEYDPVYCGVVLDSLTPTFREKTFEAYKATRQK
TPEDLHPQIPVIEEILEALGFATVRMNGFEADDVMGALAAMAQAEGRPCFVVSGDKDLAQLVTKGVRIVRMDKDDFVILD
EEGVKEKWGVRADQIVDFLALVGDASDNIPGVEGIGEKTARRLLAEYGSLEGVYAHLDELSPSLRKKLEGGRERAFLSRD
LATIRTDLELSLTIEDLRRREPEVQRARELFLAYDIRSLAQEVGGSPAVGEKPTLEVEGPSRAGVSYEVVTDRASLARWV
DKALERGVVAVDTETDGLDPLTCRLVGVSLAVDEGRACYVPLAAPDVRPLPADVVREVLTPLLASQKVSKVGQNLKFDYH
VLRRWGVRPEGPFFDTMVAAWLLESDAGRYNLDRLAEKYLGWRTIHYKDVVEKGASFETVPVAEAGVYAAEDADIALRLS
RVLKERLEAGDLGRVFYEMEMPLLSILARMEEWGIGLDGEALSAFGRELDERISNIENEIFDLVGHEFNVGSTKQLQEVL
FVERGLPPVKKTKTGFSTDTSVLEELAAQDPVAAKVLEYRSLTKLKNTYVDVLPGLVNPETGRLHTWFSQVGTATGRLSS
KDPNLQNIPIKTEEGRRIREAFVPQRGWWFLSADYSQIELVILAHLSGDPALCRAFREGEDVHRATGSLLFGVPPETVTP
EQRRIAKTINFGVIYGMSAFRLSRELGIPRKEAERFINAYFTTYAGVRAFIERTIAEAEEKGYVTTLFGRKRPLPYITSR
NKTQKTGAERIAVNTPIQGSGADIIKLAMIDLDAQMRRMGLASRMILQVHDELIFEVPPEELEVMKRLVRERMEGVVRLS
VPLRVEIGVGRNWGEAH

Figure 3

Nucleic acid sequence of the *Spirochaeta thermophila* DNA polymerase gene, with Y74C, F731Y, E745R mutations. **(SEQ ID NO:3)**

ATGAAACCGCTCTTCCTCATAGATGCCTACGGCCTCATCTATCGCGCCTACTTCGCCCTCATCAGGGCCCCCATGAGGAC
CAAGGACGGGCGCAACACGTCTGCGATCTACGGCTTCTTTCGTATGCTTTTCAGATTCGTGAAGGAGTACGATCCTGTCT
ATTGTGGCGTGGTGCTCGATTCCCTCACGCCCACCTTCAGGGAGAAGACCTTCGAGGCCTGCAAGGCCACCAGGCAGAAG
ACACCGGAGGATCTCCATCCGCAGATCCCGGTGATCGAAGAGATCCTCGAGGCCTTGGGGTTCGCTACCGTCCGGATGAA
CGGGTTCGAGGCGGACGACGTGATGGGTGCGCTCGCCGCCATGGCCCAGGCGGAGGGACGCCCGTGTTTCGTGGTATCGG
GTGACAAGGACCTCGCCCAGCTGGTCACGAAGGGAGTGAGGATCGTCCGGATGGACAAGGACGATTTCGTGATTCTCGAC
GAGGAGGGGGTGAAGGAGAAGTGGGGGGTTCGGGCGGATCAGATCGTGGACTTCCTGGCCCTGGTGGGGGATGCCTCCGA
CAACATCCCGGGTGTGGAGGGGATCGGTGAGAAGACGGCCCGGCGCCTCCTCGCCGAGTATGGGTCGCTCGAGGGAGTCT
ATGCCCACCTCGACGAGCTCTCTCCCTCGTTGAGGAAGAAGCTCGAGGGGGGACGAGAGCGGGCCTTCCTGAGCAGGGAT
CTCGCCACGATACGGACGGACCTCGAACTCTCCCTCACGATCGAGGATCTCAGGCGGAGGGAACCCGAGGTACAGCGGGC
GAGGGAGCTCTTTCTCGCCTACGACATACGGAGCTTGGCCCAGGAGGTGGGAGGGAGTCCTGCCGTGGGGGAGAAGCCGA
CCCTCGAGGTGGAGGGTCCCTCCAGGGCGGGTGTCTCCTACGAGGTGGTCACCGACCGGGCCTCCCTCGCCCGATGGGTG
GACAAGGCCCTGGAGAGGGGGGTGGTGGCCGTGGATACCGAGACCGACGGGCTCGACCCTCTCACGTGCCGTCTCGTGGG
TGTCTCGCTCGCGGTGGACGAGGGGCGGGCGTGCTACGTGCCGCTCGCGGCTCCGGATGTGCGGCCGCTTCCCGCCGACG
TGGTGCGCGAGGTCCTCACGCCGCTTCTTGCCTCACAGAAGGTGTCGAAGGTGGGGCAGAACCTGAAGTTCGACTATCAC
GTGCTCAGGCGATGGGGCGTCCGTCCGGAGGGGCCGTTCTTCGATACCATGGTGGCGGCGTGGCTGCTGGAGTCGGATGC
CGGGCGCTACAACCTCGACAGGCTCGCGGAGAAGTATCTGGGGTGGCGCACCATCCATTACAAGGACGTAGTGGAAAAGG
GGGCTTCGTTCGAGACGGTGCCGGTGGCCGAGGCAGGGGTGTACGCGGCCGAGGATGCGGACATCGCCTTGCGGCTCTCC
CGGGTCTTGAAGGAACGTCTCGAGGCCGGGGATCTCGGGCGGGTCTTCTACGAGATGGAGATGCCGCTCCTCTCCATCCT
CGCTCGGATGGAGGAGTGGGGGATCGGGCTCGACGGGGAGGCTCTGTCGGCCTTCGGCAGGGAACTCGATGAGCGCATCT
CTAATATAGAGAATGAGATATTCGATCTCGTCGGGCACGAGTTCAACGTGGGGTCCACCAAGCAGCTCCAGGAGGTGCTG
TTCGTGGAGCGGGGGCTGCCGCCGGTGAAGAAGACCAAGACCGGTTTTTCCACGGACACGAGCGTGTTGGAGGAGCTGGC
TGCTCAGGATCCGGTGGCGGCGAAGGTCCTGGAGTACCGAAGTCTCACCAAGTTGAAGAACACCTATGTGGACGTGCTTC
CCGGGCTCGTGAATCCGGAGACGGGCCGGCTTCATACCTGGTTCAGCCAGGTGGGGACGGCCACGGGACGGCTCTCGAGC
AAGGATCCCAACCTCCAGAACATCCCGATCAAGACCGAGGAGGGTCGCCGGATCAGGGAGGCGTTCGTGCCGCAGCGGGG
CTGGTGGTTCCTGAGCGCGGACTACTCTCAGATCGAGCTCGTGATCCTGGCCCATCTCTCGGGGGATCCGGCTCTCTGCA
GGGCGTTCAGGGAGGGCGAGGACGTACACCGGGCTACGGGGAGCCTCCTGTTCGGGGTGCCGCCCGAGACAGTGACGCCC
GAGCAGCGGAGGATCGCCAAGACGATCAACTACGGCGTGATCTACGGGATGTCGGCCTTCCGGCTCTCGAGGCGGCTCGG
TATCCCGCGCAAGGAGGCCGAACGGTTCATCAATGCGTACTTCACCACGTATGCGGGGGTGAGGGCGTTCATCGAACGGA
CGATCGCCGAGGCGGAGGAGAAGGGGTACGTGACCACGCTCTTCGGGAGGAAGCGGCCGCTTCCCTACATCACGAGCCGG
AACAAGACGCAGAAGACAGGAGCGGAGCGAATCGCGGTGAACACGCCGATCCAGGGCTCGGGCGCGGATATCATCAAGCT
GGCGATGATCGACCTGGACGCCCAGATGCGGAGGATGGGGCTCGCCTCCCGGATGATCCTCCAGGTGCACGACGAGCTCA
TCTTCGAGGTGCCTCCCGAGGAGCTCGAGGTGATGAAGCGGCTGGTGAGGGAGCGGATGGAGGGGGTGGTGCGGCTCTCG
GTGCCGCTTCGGGTGGAGATAGGGGTGGGGAGGAACTGGGGAGAGGCCCACTAG

Figure 4

Amino acid sequence of the *Spirochaeta thermophila* DNA polymerase enzyme, with Y74C, F731Y, E745R mutations. **(SEQ ID NO:4)**

MKPLFLIDAYGLIYRAYFALIRAPMRTKDGRNTSAIYGFFRMLFRFVKEYDPVYCGVVLDSLTPTFREKTFEACKATRQK
TPEDLHPQIPVIEEILEALGFATVRMNGFEADDVMGALAAMAQAEGRPCFVVSGDKDLAQLVTKGVRIVRMDKDDFVILD
EEGVKEKWGVRADQIVDFLALVGDASDNIPGVEGIGEKTARRLLAEYGSLEGVYAHLDELSPSLRKKLEGGRERAFLSRD
LATIRTDLELSLTIEDLRRREPEVQRARELFLAYDIRSLAQEVGGSPAVGEKPTLEVEGPSRAGVSYEVVTDRASLARWV
DKALERGVVAVDTETDGLDPLTCRLVGVSLAVDEGRACYVPLAAPDVRPLPADVVREVLTPLLASQKVSKVGQNLKFDYH
VLRRWGVRPEGPFFDTMVAAWLLESDAGRYNLDRLAEKYLGWRTIHYKDVVEKGASFETVPVAEAGVYAAEDADIALRLS
RVLKERLEAGDLGRVFYEMEMPLLSILARMEEWGIGLDGEALSAFGRELDERISNIENEIFDLVGHEFNVGSTKQLQEVL
FVERGLPPVKKTKTGFSTDTSVLEELAAQDPVAAKVLEYRSLTKLKNTYVDVLPGLVNPETGRLHTWFSQVGTATGRLSS
KDPNLQNIPIKTEEGRRIREAFVPQRGWWFLSADYSQIELVILAHLSGDPALCRAFREGEDVHRATGSLLFGVPPETVTP
EQRRIAKTINYGVIYGMSAFRLSRRLGIPRKEAERFINAYFTTYAGVRAFIERTIAEAEEKGYVTTLFGRKRPLPYITSR
NKTQKTGAERIAVNTPIQGSGADIIKLAMIDLDAQMRRMGLASRMILQVHDELIFEVPPEELEVMKRLVRERMEGVVRLS
VPLRVEIGVGRNWGEAH

Figure 5

Nucleic acid sequence of one N-terminal truncated *Spirochaeta thermophila* DNA polymerase gene, with F731Y, E745R mutations. **(SEQ ID NO:5)**

ATGGAGCTCTTTCTCGCCTACGACATACGGAGCTTGGCCCAGGAGGTGGGAGGGAGTCCTGCCGTGGGGGAGAAGCCGAC
CCTCGAGGTGGAGGGTCCCTCCAGGGCGGGTGTCTCCTACGAGGTGGTCACCGACCGGGCCTCCCTCGCCCGATGGGTGG
ACAAGGCCCTGGAGAGGGGGGTGGTGGCCGTGGATACCGAGACCGACGGGCTCGACCCTCTCACGTGCCGTCTCGTGGGT
GTCTCGCTCGCGGTGGACGAGGGGCGGGCGTGCTACGTGCCGCTCGCGGCTCCGGATGTGCGGCCGCTTCCCGCCGACGT
GGTGCGCGAGGTCCTCACGCCGCTTCTTGCCTCACAGAAGGTGTCGAAGGTGGGGCAGAACCTGAAGTTCGACTATCACG
TGCTCAGGCGATGGGGCGTCCGTCCGGAGGGGCCGTTCTTCGATACCATGGTGGCGGCGTGGCTGCTGGAGTCGGATGCC
GGGCGCTACAACCTCGACAGGCTCGCGGAGAAGTATCTGGGGTGGCGCACCATCCATTACAAGGACGTAGTGGAAAAGGG
GGCTTCGTTCGAGACGGTGCCGGTGGCCGAGGCAGGGGTGTACGCGGCCGAGGATGCGGACATCGCCTTGCGGCTCTCCC
GGGTCTTGAAGGAACGTCTCGAGGCCGGGGATCTCGGGCGGGTCTTCTACGAGATGGAGATGCCGCTCCTCTCCATCCTC
GCTCGGATGGAGGAGTGGGGGATCGGGCTCGACGGGGAGGCTCTGTCGGCCTTCGGCAGGGAACTCGATGAGCGCATCTC
TAATATAGAGAATGAGATATTCGATCTCGTCGGGCACGAGTTCAACGTGGGGTCCACCAAGCAGCTCCAGGAGGTGCTGT
TCGTGGAGCGGGGGCTGCCGCCGGTGAAGAAGACCAAGACCGGTTTTTCCACGGACACGAGCGTGTTGGAGGAGCTGGCT
GCTCAGGATCCGGTGGCGGCGAAGGTCCTGGAGTACCGAAGTCTCACCAAGTTGAAGAACACCTATGTGGACGTGCTTCC
CGGGCTCGTGAATCCGGAGACGGGCCGGCTTCATACCTGGTTCAGCCAGGTGGGGACGGCCACGGGACGGCTCTCGAGCA
AGGATCCCAACCTCCAGAACATCCCGATCAAGACCGAGGAGGGTCGCCGGATCAGGGAGGCGTTCGTGCCGCAGCGGGGC
TGGTGGTTCCTGAGCGCGGACTACTCTCAGATCGAGCTCGTGATCCTGGCCCATCTCTCGGGGGATCCGGCTCTCTGCAG
GGCGTTCAGGGAGGGCGAGGACGTACACCGGGCTACGGGGAGCCTCCTGTTCGGGGTGCCGCCCGAGACAGTGACGCCCG
AGCAGCGGAGGATCGCCAAGACGATCAACTACGGCGTGATCTACGGGATGTCGGCCTTCCGGCTCTCGAGGCGGCTCGGT
ATCCCGCGCAAGGAGGCCGAACGGTTCATCAATGCGTACTTCACCACGTATGCGGGGGTGAGGGCGTTCATCGAACGGAC
GATCGCCGAGGCGGAGGAGAAGGGGTACGTGACCACGCTCTTCGGGAGGAAGCGGCCGCTTCCCTACATCACGAGCCGGA
ACAAGACGCAGAAGACAGGAGCGGAGCGAATCGCGGTGAACACGCCGATCCAGGGCTCGGGCGCGGATATCATCAAGCTG
GCGATGATCGACCTGGACGCCCAGATGCGGAGGATGGGGCTCGCCTCCCGGATGATCCTCCAGGTGCACGACGAGCTCAT
CTTCGAGGTGCCTCCCGAGGAGCTCGAGGTGATGAAGCGGCTGGTGAGGGAGCGGATGGAGGGGGTGGTGCGGCTCTCGG
TGCCGCTTCGGGTGGAGATAGGGGTGGGGAGGAACTGGGGAGAGGCCCACTAG

Figure 6

Amino acid sequence of one N-terminal truncated *Spirochaeta thermophila* DNA polymerase enzyme, with F731Y, E745R mutations. **(SEQ ID NO:6)**

MELFLAYDIRSLAQEVGGSPAVGEKPTLEVEGPSRAGVSYEVVTDRASLARWVDKALERGVVAVDTETDGLDPLTCRLVG
VSLAVDEGRACYVPLAAPDVRPLPADVVREVLTPLLASQKVSKVGQNLKFDYHVLRRWGVRPEGPFFDTMVAAWLLESDA
GRYNLDRLAEKYLGWRTIHYKDVVEKGASFETVPVAEAGVYAAEDADIALRLSRVLKERLEAGDLGRVFYEMEMPLLSIL
ARMEEWGIGLDGEALSAFGRELDERISNIENEIFDLVGHEFNVGSTKQLQEVLFVERGLPPVKKTKTGFSTDTSVLEELA
AQDPVAAKVLEYRSLTKLKNTYVDVLPGLVNPETGRLHTWFSQVGTATGRLSSKDPNLQNIPIKTEEGRRIREAFVPQRG
WWFLSADYSQIELVILAHLSGDPALCRAFREGEDVHRATGSLLFGVPPETVTPEQRRIAKTINYGVIYGMSAFRLSRRLG
IPRKEAERFINAYFTTYAGVRAFIERTIAEAEEKGYVTTLFGRKRPLPYITSRNKTQKTGAERIAVNTPIQGSGADIIKL
AMIDLDAQMRRMGLASRMILQVHDELIFEVPPEELEVMKRLVRERMEGVVRLSVPLRVEIGVGRNWGEAH

Figure 7

Nucleic acid sequence of the N-terminal truncated *Spirochaeta thermophila* DNA polymerase gene, with F731Y, E745R mutations, further incorporating preferred codon usage for protein expression in *E. coli*. **(SEQ ID NO:7)**

ATGGAGCTGTTTCTGGCCTACGACATACGGAGCTTGGCCCAGGAGGTGGGAGGGAGTCCTGCCGTGGGGGAGAAGCCGAC
CCTCGAGGTGGAGGGTCCCTCCAGGGCGGGTGTCTCCTACGAGGTGGTCACCGACCGGGCCTCCCTCGCCCGATGGGTGG
ACAAGGCCCTGGAGAGGGGGGTGGTGGCCGTGGATACCGAGACCGACGGGCTCGACCCTCTCACGTGCCGTCTCGTGGGT
GTCTCGCTCGCGGTGGACGAGGGGCGGGCGTGCTACGTGCCGCTCGCGGCTCCGGATGTGCGGCCGCTTCCCGCCGACGT
GGTGCGCGAGGTCCTCACGCCGCTTCTTGCCTCACAGAAGGTGTCGAAGGTGGGGCAGAACCTGAAGTTCGACTATCACG
TGCTCAGGCGATGGGGCGTCCGTCCGGAGGGGCCGTTCTTCGATACCATGGTGGCGGCGTGGCTGCTGGAGTCGGATGCC
GGGCGCTACAACCTCGACAGGCTCGCGGAGAAGTATCTGGGGTGGCGCACCATCCATTACAAGGACGTAGTGGAAAAGGG
GGCTTCGTTCGAGACGGTGCCGGTGGCCGAGGCAGGGGTGTACGCGGCCGAGGATGCGGACATCGCCTTGCGGCTCTCCC
GGGTCTTGAAGGAACGTCTCGAGGCCGGGGATCTCGGGCGGGTCTTCTACGAGATGGAGATGCCGCTCCTCTCCATCCTC
GCTCGGATGGAGGAGTGGGGGATCGGGCTCGACGGGGAGGCTCTGTCGGCCTTCGGCAGGGAACTCGATGAGCGCATCTC
TAATATAGAGAATGAGATATTCGATCTCGTCGGGCACGAGTTCAACGTGGGGTCCACCAAGCAGCTCCAGGAGGTGCTGT
TCGTGGAGCGGGGGCTGCCGCCGGTGAAGAAGACCAAGACCGGTTTTTCCACGGACACGAGCGTGTTGGAGGAGCTGGCT
GCTCAGGATCCGGTGGCGGCGAAGGTCCTGGAGTACCGAAGTCTCACCAAGTTGAAGAACACCTATGTGGACGTGCTTCC
CGGGCTCGTGAATCCGGAGACGGGCCGGCTTCATACCTGGTTCAGCCAGGTGGGGACGGCCACGGGACGGCTCTCGAGCA
AGGATCCCAACCTCCAGAACATCCCGATCAAGACCGAGGAGGGTCGCCGGATCAGGGAGGCGTTCGTGCCGCAGCGGGGC
TGGTGGTTCCTGAGCGCGGACTACTCTCAGATCGAGCTCGTGATCCTGGCCCATCTCTCGGGGGATCCGGCTCTCTGCAG
GGCGTTCAGGGAGGGCGAGGACGTACACCGGGCTACGGGGAGCCTCCTGTTCGGGGTGCCGCCCGAGACAGTGACGCCCG
AGCAGCGGAGGATCGCCAAGACGATCAACTACGGCGTGATCTACGGGATGTCGGCCTTCCGGCTCTCGAGGCGGCTCGGT
ATCCCGCGCAAGGAGGCCGAACGGTTCATCAATGCGTACTTCACCACGTATGCGGGGGTGAGGGCGTTCATCGAACGGAC
GATCGCCGAGGCGGAGGAGAAGGGGTACGTGACCACGCTCTTCGGGAGGAAGCGGCCGCTTCCCTACATCACGAGCCGGA
ACAAGACGCAGAAGACAGGAGCGGAGCGAATCGCGGTGAACACGCCGATCCAGGGCTCGGGCGCGGATATCATCAAGCTG
GCGATGATCGACCTGGACGCCCAGATGCGGAGGATGGGGCTCGCCTCCCGGATGATCCTCCAGGTGCACGACGAGCTCAT
CTTCGAGGTGCCTCCCGAGGAGCTCGAGGTGATGAAGCGGCTGGTGAGGGAGCGGATGGAGGGGGTGGTGCGGCTCTCGG
TGCCGCTTCGGGTGGAGATAGGGGTGGGGAGGAACTGGGGAGAGGCCCACTAG

Figure 8: Alignment of *Spirochaeta thermophila* DNA polymerase with several other DNA-polymerases

```
001 ------MKPLFLIDAYGLIYRAYFALIRAPMRTKDGRNTSAIYGFFRMLFRFVKEYDPVYCGVVLDSLTPTFREKTFEAY Spirochaeta thermophila
001 ------MKPLFLIDAYGLIYRAYFALIRAPMRTKDGRNTSAIYGFFRMLFRFVKEYDPVYCGVVLDSLTPTFREKTFEAY NCBI AAR11879
001 ----MSQAPLVLVDGSSYLYRAFHALP--PLTTSTGKPTGAVKGVLNMLLSLRKQYPDSPFAVVFDAKGPTFRDELFAEY Pseudomonas aeruginosa UCBPP-PA14
001 ----MSQAPLVLVDGSSYLYRAFHALP--PLTTSTGKPTGAVKGVLNMLLSLRKQYPDSPFAVVFDAKGPTFRDELFAEY Pseudomonas aeruginosa PAO1
001 MPVSDTSPLLVLVDGSSFLYRAFFALP--PLTNSHGEPTGAVYGVINMLRKLLQTYDGAHIAVVFDAPGRNFRDELFEHY Methylococcus capsulatus str. Bath
001 --MSEQYPLLILVDGSSYLYRAYYALP--PLTTPAGEPTGAVYGVVSMLRKLLAEYDTAYFGVVFDAKGKTFRDELFEAY Nitrococcus mobilis Nb-231
001 -MAQIAENPLILVDGSSYLYRAYHAFP--PLTNSAGEATGAMYGVLNMLRSLLLQYHPSHVAVVFDAKGKTFRDELFENY Erwinia carotovora subsp. atroseptica
001 -MAQIADNPLILVDGSSYLYRAYHAFP--PLTNSAGEPTGAMYGVLNMLRSLIMQYKPSHVAVVFDAKGKTFRDELFAEY Photorhabdus luminescens subsp. laumondii TTO1
001 -MAQIAENPLILVDGSSYLYRAYHAFP--PLTNGSGEPTGAMYGVLNMLRSLLLQYRPSHVAVVFDAKGKTFRDELFAEY Yersinia bercovieri ATCC 43970

075 KATRQKTPEDLHPQIPVIEEILEALGFATVRMNGFEADDVMGALAAMAQAEGRPCFVVSGDKDLAQLVTKGVRIVRMDKD Spirochaeta thermophila
075 KATRQKTPEDLHPQIPVIEEILEALGFATVRMNGFEADDVMGTLAAMAQAEGRPCFVVSGDKDLAQLVTKGVKIVRMDKD NCBI AAR11879
075 KANRPSMPDDLRVQVEPLHASVRALGLPLLCVEGVEADDVIGTLARSSAAAGRPVVISTGDKDMAQLVDGHITLVNTMTG Pseudomonas aeruginosa UCBPP-PA14
075 KANRPSMPDDLRVQVEPLHASVRALGLPLLCVEGVEADDVIGTLARSSAAADRPVVISTGDKDMAQLVDGHITLVNTMTG Pseudomonas aeruginosa PAO1
079 KAHRPPMPDDLRSQIEPLHQVVRAMGLPMLIEPGVEADDVIGTLARQAVERGFRVVISTGDKDMAQLVCDRVTLENTMFD Methylococcus capsulatus str. Bath
077 KANRAPMPDELRVQLEPLKRIVCALGLSRIEIEGVEADDVIGTLAVRASVAGLRTIIATGDKDMAQIVNDNVVLLNTMSG Nitrococcus mobilis Nb-231
078 KAHRPPMPEDLREQIEPLHSMVKAMGLPLLAVSGVEADDVIGTLAVQAEKAGKSVLISTGDKDMAQLVTPSVTLINTMNN Erwinia carotovora subsp. atroseptica
078 KSHRPPMPDDLRTQIEPLHQMVEAMGLPLLVVPGVEADDVIGTLALQAEKEGRAVLISTGDKDMAQLVTPNITLINTMNN Photorhabdus luminescens subsp. laumondii TTO1
078 KSHRPPMPDDLRSQIEPLHQMVKAMGLPLLVVSGVEADDVIGTLAQEAEKAGHSVLISTGDKDMAQLVTPNITLINTMNN Yersinia bercovieri ATCC 43970

155 DFVILDEEGVKEKWGVRADQIVDFLALVGDASDNIPGVEGIGEKTARRLLAEYGS-LEGVYAHLDELSP-------SLRK Spirochaeta thermophila
155 DFVVLDEEGVKEKWGVRADQIVDFLALVGDASDNIPGVEGIGEKTAQRLLAEYGS-LEGVYAHLDELSP-------SLRR NCBI AAR11879
155 --SRLDVDGVKEKFGVGPELIIDFLALMGDKVDNIPGVPGVGEKTALGLLTGVGGGLEVLYASLDKVPELPIRGAKGLPA Pseudomonas aeruginosa UCBPP-PA14
155 --SRLDVDGVKEKFGVGPELIIDFLALMGDKVDNIPGVPGVGEKTALGLLTGVGGGLEVLYASLDKVPELPIRGAKGLPA Pseudomonas aeruginosa PAO1
159 --SRLDVNGVIVKFGVPPERIVDYLALVGDTSDNIPGVPKVGPKTAAKWLAEYGS-LDALIARAGEIGG-------KTGE Methylococcus capsulatus str. Bath
157 --TWLDAAAVQDKFGVPPERMVDYLALVGDAVDNIPGVPRVGPKTAAKWLGQYGS-LDALLECVDEIKG-------KVGE Nitrococcus mobilis Nb-231
158 --TILGPQEVCDKYGIPPELIIDFLALMGDASDNIPGVPGVGEKTAQALLQGLGG-LDSLYANLDKIAELSFRGAKTMAP Erwinia carotovora subsp. atroseptica
158 --TILGPEEVVEKYGVPPELIIDFLALMGDSSDNIPGVPGVGEKTALGLLQGIGG-MDDIFANLDKIAGLSFRGAKTLAD Photorhabdus luminescens subsp. laumondii TTO1
158 --AILGPQEVCEKYGVPPELIIDFLALMGDSSDNIPGVPGVGEKTAQALLQGLGG-LDSLFSNLDKIPTLTFRGAKTMSA Yersinia bercovieri ATCC 43970

227 KLEGGRERAFLSRDLATIRTDLELSLTIEDLRRREPEVQRARELFLAYDIRSLAQEVGG----------SPAVGEKPT-- Spirochaeta thermophila
227 KLEEGRERAFLSRDLATIRTDLDLSLTIGDLRRREPEVQRARELFLAYDIRSLAQEVGG----------SPSVGEKPT-- NCBI AAR11879
233 KLEENREQAFLSYQLATIKTDVELDVEIDKLYPGEPQREALIALYRELEFKNWLDDLLR----------EAKEAGENG-- Pseudomonas aeruginosa UCBPP-PA14
233 KLEENREQAFLSYQLATIKTDVELDVEIDKLYPGEPQREALIALYRELEFKNWLDDLLR----------EAKEAGENG-- Pseudomonas aeruginosa PAO1
229 NLRASLELIPLSRELATIRCDLPLPLTPESLERRPPDKAALRELYTRLEFKTLLRQLDAEAD-------DPKAAPPV--- Methylococcus capsulatus str. Bath
227 SLREHLEQLALSRELATIRCDLDLDLDPKAMRPRAPDVDALARLYRQHAFSSWLTQLGDEH---------PGAGDPE--- Nitrococcus mobilis Nb-231
235 KLEQHKEVAYLSYQLATIKTDVELELSCEQLTVNELDVDELHRLFSRYEFKRWLSDIESGTWMQGKKSSQPVQTVSNAVV Erwinia carotovora subsp. atroseptica
235 KMVQHKDVAYLSYQLATIKTDVELDKTCVALSVKEPDTEKLHQLFSRYEFKRWLSDVANGDWLAEK-GKKPVSTGNDESS Photorhabdus luminescens subsp. laumondii TTO1
235 KLEQNKDVAYLSYKLATIKTDVELDVTCDELTVSPPDDEQLHQLFSRYEFKRWLADVEAGKWLDGKKERPTGQTSNKAFV Yersinia bercovieri ATCC 43970
```

Figure 8 (continued): Alignment of *Spirochaeta thermophila* DNA polymerase with several other DNA-polymerases

```
295 ------LEVEGPSRAGVSYEVVTDRASLARWVDKALERGVVAVDTETDGLDPLTCRLVGVSLAVDEGRACYVPLAAPDVR Spirochaeta thermophila
295 ------LEVEGPSRAGVSYEVVTDRASLSRWVEEALERGAVAVDTETDGLDPLTCRLVGVSFAVDEGRACYVPLAAPDVR NCBI AAR11879
301 -------EAETPIQAEVDYDVVLDQAGFDAWLKKLEEAELIAFDTETTSIDAQQAQVVGVSFAVKEGEAAYVPLAHSYMG Pseudomonas aeruginosa UCBPP-PA14
301 -------EAETPIQAEVDYDVVLDQAGFDAWLKKLEEAELIAFDTETTSIDAQQAQVVGVSFAVKEGEAAYVPLAHSYMG Pseudomonas aeruginosa PAO1
299 ----------AAAPEVQTRYQAVMDPAAFDRWLEKLEAAELFAFDTETSSLDYMRAEVIGLSFAVEPGAAAYVPLAHDYPG Methylococcus capsulatus str. Bath
295 ----------QAA----VHYETVLEEAGLQQWLQRLAGAEVFAFDLETTSLNYMDAEIVGLSFAVEPARAAYVPVAHIGQG Nitrococcus mobilis Nb-231
315 --EQAVEEDNAPTLSADGYVTILDEKTLLDWIERLKQAEVFAFDTETDGLDTLTANLIGLSFAIKPGEAAYLPLAHDYLD Erwinia carotovora subsp. atroseptica
314 --PTEDKSAAATLSSDNYQTILDQQSFDEWIEKLKKASVFAFDTETDSLDTLTANLVGVSFAISAGDAAYLPLGHDYLG Photorhabdus luminescens subsp. laumondii TTO1
315 AAEPAPAAAEVTAVLSQENYQTILDEKSLADWIERLKKSEVFAFDTETDGLDTLSSNLIGLSFAVAPGEAAYLPLAHDYLD Yersinia bercovieri ATCC 43970

369 P---LPADVVREVLTPLLASQKVSKVGQNLKFDYHVLRRWG--VREPEGPFFDTMVAAWLLESDA--GRYNLDRLAEKYLG Spirochaeta thermophila
369 P---LPADVVREVLSPLLVSTEVVKVGQNLKFDYHVLRRWG--VREPEGPFFDTMVAAWLLESDA--GRYNLDRLAEKYLG NCBI AAR11879
374 VPEQLDRDAVLRALKPLLEDPKKLKVGQHAKYDINVLANASTPIALRGVAYDTMLESYVLDSTA--TRHDMDSLALKYLG Pseudomonas aeruginosa UCBPP-PA14
374 VPEQLDRDAVLRALKPLLEDPKKLKVGQHAKYDINVLANASTPIALRGVAYDTMLESYVLDSTA--TRHDMDSLALKYLG Pseudomonas aeruginosa PAO1
370 APPQLDRAMVLERLRPLLEDPGKAKLGQHLKYDANVLLNHG--IMLRGIRHDTMLESYVLNSTA--TRHDMDSLAERYLD Methylococcus capsulatus str. Bath
362 VSPQLERGHVLEQLRPLLEDPRRAKIGQNLKYDMSVLAHYD--IALRGVRFDTMLESYILDSTA--TRHDMDSLALKYLG Nitrococcus mobilis Nb-231
393 APEQLDRAQVLALFKPLLEDEKLLKIGQNLKFDKGVMQRYD--IDLRGIAFDTMLESYVLDSVA--GRHDMDSLAERYLK Erwinia carotovora subsp. atroseptica
391 APEQLDREVVLAALKPLLEDSNLLKIGQNIKFDRGVLARYG--IELQGIAYDTMLESYVLNSVAGMGRHDMDSLAERHLN Photorhabdus luminescens subsp. laumondii TTO1
395 APAQLDRDWVLATLKPLLEDEKALKVGQNLKFDQSMLARYG--IELRGIAFDTMLESYVLNSVA- GRHDMDSLAERYLN Yersinia bercovieri ATCC 43970

442 WRTIHYKDVVEKGAS---FETVPVAEAGVYAAEDADIALRLSRVLKERLEAG-DLGRVFYEMEMPLLSILARMEEWGIGL Spirochaeta thermophila
442 WRTIHYKDVVEKGAS---FETVPVAEAGVYAAEDADIALRLSRVLKERLEAE-GLGRVFYEMEMPLLSILARMEEWGIGL NCBI AAR11879
452 HSTIRFEDIAGKGAKQLTFDQIAIEQAGPYAAEDADVTLRLHQALWQKLEAIPSLARVLTDIEMPLVPVLARIERNGALV Pseudomonas aeruginosa UCBPP-PA14
452 HSTIRFEDIAGKGAKQLTFDQIAIEQAGPYAAEDADVTLRLHQALWQKLEAIPSLARVLTDIEMPLVPVLARIERNGALV Pseudomonas aeruginosa PAO1
446 RKTLHYEDVAGKGAKQIPFARVSVEDACRYAAEDADVTLCLHRALWPRLESIPSLRAVYETIEIPLVPVLSRIERAGVLV Methylococcus capsulatus str. Bath
438 HRTIKFEDVAGKGAKQLTFDQVAIERATEYAAEDADIALRLHRVLYPKLRQLPGPERVFHAIEMPLLPVLSRMERTGVLV Nitrococcus mobilis Nb-231
469 HKTITFEEIAGKGKNQLTFNQIALEQAGPYAAEDADVTLHLHQTLWGKLQPHADLCQVFQAIDMPLVPVLSRIERTGVLI Erwinia carotovora subsp. atroseptica
469 HKTITFEEIAGKGKNQLTFNQVAVEQAAMYAAEDADVTLLLHQAMWPQLEKAPTLEKVFLNTEMPLVMVLSRMERTGVLI Photorhabdus luminescens subsp. laumondii TTO1
471 HKTITFEEIAGKGKNQLTFNQIALEEAGPYASEDADVTLQLHLVLWPKLQESEGLKRVFQEIEMPLLPILSRIERTGVLI Yersinia bercovieri ATCC 43970

518 DGEALSAFGRELDERISNIENEIFDLVGHEFNVGSTKQLQEVLFVERGLPPVKKTKTG-FSTDTSVLEELAA-QDPVAAK Spirochaeta thermophila
518 DGEALSAFGRELEERISTIEKEIFELVGHEFNVGSTKQLQEVLFVERGLPPVKKTKTG-FSTDTSVLEELAA-RDPVAAK NCBI AAR11879
532 DANLLGIQSRELGEKMVALERQAYDLAGQEFNLGSPKQLGAILYDKLGLPVLSKTATGQPSTAEAVLAELAEQDFELPKV Pseudomonas aeruginosa UCBPP-PA14
532 DANLLGIQSRELGEKMVALERQAYDLAGQEFNLGSPKQLGAILYDKLGLPVLSKTATGQPSTAEAVLAELAEQDFELPKV Pseudomonas aeruginosa PAO1
526 DVYKLAEQSRELERRMAEVEAEARAVAGETFNLGSPKQIQTILYDKLGLPVIKKTPTGQPSTDESVLQDLAE-TFELPRL Methylococcus capsulatus str. Bath
518 SRELLAAQSAELAQRLAEIEASVYRAAGCEFNLGSPKQIQAVLYDRQGLPVLAKTPKGAPSTAESVLQELAA-TYPLPRM Nitrococcus mobilis Nb-231
549 DPAILAEHSKELTTRLAGLETQAYELAGEEFNLASTKQLQGILYEKQKLPILKKTPKGAPSTNEEVLAELAL-DYPLPKL Erwinia carotovora subsp. atroseptica
549 DANILAEHSKEITARLDELEKEAHELAGEVFNLASPKQLQTILFEKLQLPVVKKTPNGAPSTNEEVLEELAD-THALPKV Photorhabdus luminescens subsp. laumondii TTO1
551 DQNILAAHSKELTIRLDELEKQAHELAEEPFNLASPKQLQVILYEKQKLPILKKTPGGAASTNEEVLAELAL-DYPLPKV Yersinia bercovieri ATCC 43970
```

Figure 8 (continued): Alignment of *Spirochaeta thermophila* DNA polymerase with several other DNA-polymerases

```
596 VLEYRSLTKLKNTYVDVLPGLVNPETGRLHTWFSQVGTATGRLSSKDPNLQNIPIKTEEGRRIREAFVPQRGWWFLSADY Spirochaeta thermophila
596 VLEYRSLTKLKNTYVDVLPGLVNPGTGRLHTWFSQVGTATGRLSSKDPNLQNIPIKTEEGRRIREAFVPQRGWWFLSADY NCBI AAR11879
532 IMQYRSMSKLKSTYTDRLPEQINPRTGRIHTSYHQAVAATGRLSSSDPNLQNIPIRTAEGRRIRQAFVAPQGYKLLAADY Pseudomonas aeruginosa UCBPP-PA14
532 IMQYRSMSKLKSTYTDRLPEQINPRTGRIHTSYHQAVAATGRLSSSDPNLQNIPIRTAEGRRIRQAFVAPQGYKLLAADY Pseudomonas aeruginosa PAO1
526 ILEYRSLSKLKSTYTDKLPHQVNPVSGRVHTSYHQAVAATGRLSSSDPNLQNIPVRTEEGRRIRQAFVAPPGHKLLAADY Methylococcus capsulatus str. Bath
518 ILEHRALAKLKSTYTDKLPLLIHPRTGRVHTSYRQAVAATGRLSSSDPNLQNIPVRSPEGRRIRRGFIARQGYRLLAVDY Nitrococcus mobilis Nb-231
549 ILEYRGLAKLKSTYTDKLPLMINPATKRVHTSYHQAVTATGRLSSSDPNLQNIPVRNDEGRRIRQAFIASKGYSIVAADY Erwinia carotovora subsp. atroseptica
549 ILEHRGLAKLKSTYTDKLPQMVHPLTKRVHTSYHQAVTATGRLSSRDPNLQNIPVRNKEGRRIRQAFVAPEGYRIMAADY Photorhabdus luminescens subsp. laumondii TTO1
551 ILEYRGLAKLKSTYTDKLPLMINPVSGRVHTSYHQAVTATGRLSSRDPNLQNIPVRNEEGRRIRQAFIAPKGYRIMAADY Yersinia bercovieri ATCC 43970

676 SQIELVILAHLSGDPALCRAFREGEDVHRATGSLLFGVPPETVTPEQRRIAKTINFGVIYGMSAFRLSRELGIPRKEAER Spirochaeta thermophila
676 SQIELVILAHLSEDPALCRAFREGEDVHRATGSLLFGVPPEAVTPEQRRIAKTINFGVIYGMSAFRLSRELGIPRKEAER NCBI AAR11879
692 SQIELRIMAHLAKDDGLLDAFRHDLDVHRATAAEVFGVPLEDVSGDQRRSAKAINFGLIYGMSAFGLAKQIGVERKEAQA Pseudomonas aeruginosa UCBPP-PA14
692 SQIELRIMAHLAKDDGLLDAFRHDLDVHRATAAEVFGVPLEDVSGDQRRSAKAINFGLIYGMSAFGLAKQIGVERKEAQA Pseudomonas aeruginosa PAO1
685 SQIELRIMAHLSGDANLLAAFAEDADVHRATAAEVFGVALEEVTSSQRRSAKAINFGLIYGMSAFGLAKQLGIQQKLAQG Methylococcus capsulatus str. Bath
677 SQVELRIMAHPSGDERLHRAFAAGEDIHRATAAEVFATALEQVTAEQRRAAKAINFGLIYGMSAYGLARQLGIERDAAQH Nitrococcus mobilis Nb-231
708 SQIELRIMAHLSGDKGLLNAFANGLDIHRATASEVFGIALDKVTSEQRRSAKAINFGLIYGMSAFGLSRQLNIPRSESQK Erwinia carotovora subsp. atroseptica
708 SQIELRIMAHLSQDQGLLKAFAEGKDIHRATAAEVFGLPLEQVTSEQRRSAKAINFGLIYGMSAFGLSRQLGIPRGEAQR Photorhabdus luminescens subsp. laumondii TTO1
710 SQIELRIMAHLSQDEGLLAAFAAGKDIHRATAAEVFGLPLEKVTTEQRRSAKAINFGLIYGMSAFGLARQLNIPRGEAQR Yersinia bercovieri ATCC 43970

756 FINAYFTTYAGVRAFIERTIAEAEEKGYVTTLFGRKRPLPYITSRNKTQKTGAERIAVNTPIQGSGADIIKLAMIDLDAQ Spirochaeta thermophila
756 FIDAYFTTYAGVRAFIERTIAEAEEKGYVTTLFGRKRPLPYITSRNKTQKTGAERIAVNTPIQGSGADIIKLAMIDLDAQ NCBI AAR11879
772 YIDRYFARYPGVLAYMERTRAQAAEQGFVETLFGRRLYLPEIHSKNGAMRKAAERTAINAPMQGTAADIMKRAMVAVDNW Pseudomonas aeruginosa UCBPP-PA14
772 YIDRYFARYPGVLAYMERTRAQAAEQGFVETLFGRRLYLPEIHSKNGAMRKAAERTAINAPMQGTAADIMKRAMVAVDNW Pseudomonas aeruginosa PAO1
765 YIDLYFARYPGVRAYMDRTRESARELGYVETLFGRRLHVPDIQSRNGQRRQYAERTAINAPMQGTAADIIKRAMIALDAW Methylococcus capsulatus str. Bath
757 YVERYFERYPGVKAFMERIRREARERQYVETVFGRRLYLPEINAGNAQRRQAAERTAINAPMQGSAADIIKRAMVRVDQW Nitrococcus mobilis Nb-231
788 YMNLYFERYPGVQDYMERTRQQAAEHGYVSTLDGRRLYLPDIHSRNAMARKGAERAAINAPMQGTAADIIKKAMISIDDW Erwinia carotovora subsp. atroseptica
788 YMDLYFERYPGVLAYMERTRKQAEEQGYVETLDGRRLYLSDIKSRNAMRRKAAEREAINAPMQGTAADIIKKAMIAVDDW Photorhabdus luminescens subsp. laumondii TTO1
790 YMDLYFERYPGVLEYMERTRKQAADQGYVTTLDGRRLYLPDIHSRNANRRKAAEREAINAPMQGTAADIIKRAMIAVDAW Yersinia bercovieri ATCC 43970

836 MRRMG-LASRMILQVHDELIFEVPPEELEVMKRLVRERMEGVVRLSVPLRVEIGVGRNWGEAH Spirochaeta thermophila (SEQ ID NO:2)
836 MRRMG-LASRMILQVHDELIFEVPPEELEVMKRLVRERMEGVVRLSVPLRVEMGVGRNWGEAH NCBI AAR11879 (SEQ ID NO:8)
852 LQESG-LDARVILQVHDELVLEVREDLVEQVREGIRPLMSGAATLDVPLVVEAGVGSNWDEAH Pseudomonas aeruginosa UCBPP-PA14 (SEQ ID NO:9)
852 LQESG-LDARVILQVHDELVLEVREDLVEQVCEGIRPLMSGAATLDVPLVVEAGVGSNWDEAH Pseudomonas aeruginosa PAO1 (SEQ ID NO:10)
845 IESSG-APLRMIMQVHDELVFEVAEDFVSEATATVREHMSRAAELAVPLMVDIGTGDNWDEAH Methylococcus capsulatus str. Bath (SEQ ID NO:11)
837 LQSER-EDVPMVLQVHDELVFEVPTPDVQAVAERIRAFMVGAAELSVPLTVEAGWGDNWEEAH Nitrococcus mobilis Nb-231 (SEQ ID NO:12)
868 LQKDT-PLVRMIMQVHDELVFEIHDSVIEESISKIKILMEGCMQLNVPLQVDIGTGMNWDQAH Erwinia carotovora subsp. Atroseptica (SEQ ID NO:13)
868 IVSEE-PKVRMIMQVHDELVFEVHESELEYAEQKIRELMEKSMVLDVPLKVDVGIGDNWDQAH Photorhabdus luminescens subsp. laumondii TTO1 (SEQ ID NO:14)
870 LQQESEPLVRVIMQVHDELVFEVHESVLESAEQKIRELMEQSMQLAVPLKVDVGVGDNWDQAH Yersinia bercovieri ATCC 43970 (SEQ ID NO:15)
```