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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 September 2009 (03.09.2009)

PCT

(10) International Publication Number
WO 2009/106795 A1

(51) **International Patent Classification**:
C07K 14/195 (2006.01) ***C12N 9/12*** (2006.01)

(21) **International Application Number**:
PCT/GB2009/000411

(22) **International Filing Date**:
13 February 2009 (13.02.2009)

(25) **Filing Language**: English

(26) **Publication Language**: English

(30) **Priority Data**:
0803628.7 28 February 2008 (28.02.2008) GB

(71) **Applicant** *(for all designated States except US)*: **GENESYS LTD** [GB/GB]; Innovation House, Albany Park, Camberley, Surrey GU16 7PL (GB).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: **CLARK, Duncan, Roy** [GB/GB]; GeneSys Limited, Innovation House, Albany Park, Camberley, Surrey GU16 7PL (GB). **MORANT, Nicholas** [GB/GB]; GeneSys Limited, Innovation House, Albany Park, Camberley, Surrey GU16 7PL (GB).

(74) **Agent**: **TURNER, Rhiannon, R.**; GREAVES BREWSTER LLP, Indigo House, Cheddar Business Park, Wedmore Road, Cheddar, Somerset BS27 3EB (GB).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, 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, ME, 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, ST, 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, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2009/106795 A1

(54) **Title**: ENZYME

(57) **Abstract**: There is provided a polypeptide having thermostable DNA polymerase activity and comprising or consisting of an amino acid sequence with at least 55% identity to *Thermodesulfatator indicus* DNA polymerase I Large fragment shown in SEQ ID NO: 1 or in SEQ ID NO:32.

Enzyme

Field of Invention

The present invention relates to novel polypeptides having DNA polymerase activity, and their uses.

Background

DNA polymerases are enzymes involved *in vivo* in DNA repair and replication, but have become an important *in vitro* diagnostic and analytical tool for the molecular biologist. *E. coli* DNA polymerase I encoded by the gene "DNA polA" was discovered in 1956, and cloned and characterised in the early 1970s. The enzyme has a variety of uses including DNA labelling by nick translation, second-strand cDNA synthesis in cDNA cloning, and DNA sequencing. The so-called "Klenow" or "Large" fragment of *E. coli* DNA polymerase I is a large protein fragment originally produced upon cleavage of the native enzyme by the protease enzyme subtilisin. This Large fragment exhibits 5'→3' polymerase activity and 3'→5' exonuclease proofreading activity, but loses 5'→3' exonuclease activity which mediates nick translation during DNA repair in the native enzyme.

Since being discovered in *E. coli*, DNA polymerase I-like enzymes have been characterised in many prokaryotes, although the non-*E. coli* counterparts do not always have a 3'→5' exonuclease proofreading function. Certain DNA polymerase I -like enzymes obtained from various thermophilic eubacteria, for example *Thermus flavus, Thermus aquaticus, Thermus brockianus, Thermus ruber, Thermus thermophilus, Thermus filiformis, Thermus lacteus, Thermus rubens, Bacillus stearothermophilus, Bacillus caldotenax* and *Thermotoga maritima*, have been found to be thermostable, retaining polymerase activity at around 45°C to 100°C.

In general, thermostable DNA polymerases have found wide use in methods for amplifying nucleic acid sequences by thermocycling amplification reactions such as the polymerase chain reaction (PCR) or by isothermal amplification reactions such as strand displacement amplification (SDA), nucleic acid sequence-based amplification

(NASBA), self-sustained sequence replication (3SR), and loop-mediated isothermal amplification (LAMP; see Notomi et al., 2000, Nucleic Acids Res. 28: e63). Thermostable DNA polymerases have different properties such as thermostability, strand displacement activity, fidelity (error rate) and binding affinity to template DNA and/or free nucleotides, and are therefore typically suited to different types of amplification reaction.

Isothermal amplification reactions require a DNA polymerase with strong strand displacement activity, and DNA polymerase I enzymes such as *Bst* DNA polymerase I Large fragment and *Bca* DNA polymerase I Large fragment are preferred in reactions such as LAMP (see Notomi et al., 2000, *supra*).

On the other hand, thermocycling amplification reactions such as PCR require a DNA polymerase with reasonable processivity and thermostability at the cycling temperatures used (typically up to 94°C). Many of the commercially used DNA polymerases for PCR are DNA polymerase II-like enzymes (for example, Vent, Deep Vent, *Pwo*, *Pfu*, KOD, 9N7, *Tfu* DNA polymerases) which lack 5'→3' exonuclease activity but have proofreading 3'→5' exonuclease activity. Some DNA polymerase I enzymes (typically those from *Thermotoga* and *Thermus* species, for example *Taq* DNA polymerase) are used in PCR, but *Taq* DNA polymerase, for example, has insufficient strand displacement activity to function adequately in isothermal amplification reactions.

WO2007/127893 discloses thermostable DNA polymerases from *Thermotoga naphthophila* and *Thermotoga petrophellia.*

Moussard *et al.* (Int. J. Systemic & Evolutionary Microbiol. (2004) 54: 227-233) discloses the discovery of the genus *Thermodesulfatator*, with *Thermodesulfatator indicus* as the type species.

The present invention provides a novel thermostable DNA polymerase I and Large fragment thereof for use in reactions requiring DNA polymerase activity such as

nucleic acid amplification reactions. The polymerase, particularly its Large fragment, has surprisingly and advantageously been found to be useful in both thermocycling and isothermal amplification reactions. Included within the scope of the present invention are various mutants (deletion and substitution) that retain thermostability and the ability to replicate DNA.

Summary of Invention

According to one aspect of the present invention there is provided a polypeptide having thermostable DNA polymerase activity and comprising or consisting essentially of an amino acid sequence with at least 51% identity, for example at least 55%, 56%, 57%, 58%, 59%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or even 99% identity, to *Thermodesulfatator indicus* DNA polymerase I Large (or "Klenow") fragment shown in SEQ ID NO:1. Preferably, the polypeptide is isolated.

The Large fragment of *T. indicus* DNA polymerase I has the following amino acid sequence:

MGLLKELPATKTLSMTRYELVLDPDKVKEIVEKAKGAEVVAIDLESDTKDPM
RGKIVGVSLCFNPPKAYYFPFRHEGLEAQKQLPWEAFTHLASLIEDPSVKKIG
HNIKYDLIILARYGVTLKGLEGDTMLASYLLDPTRRTHGLDELAEEVLGHTMI
FYKEVTKELAKGESFARVPLEKAKVYACEDAHVTYLLYQYFWPKLKEESLW
KVFTEIDRPLIEVLAHMEMVGIKIDTAYLRGLSREMAEKLKELEEKIYTLAGE
KFNINSSKQLGQILFEKLKLPTVKKTPKKTAYSTDNEVLEELSAVHELPRLILE
YRTLAKLKSTYVDALPKMVNPETGRLHTSFNQTVTATGRLSSSDPNLQNIPVR
GEEGLKIRQAFVPEEIFAADYTQIDLRVLAHYSGDETLIKAFWQGEDIHRRTA
AEIFGIPPEEVTPEMRRMAKTINFGIVYGMSPYGLAKELKIGRREAKAFIERYF
ERYPGVKRYMEQIVAEAREKGYVETLFGRKRPLPDINSPNRTAREFAERTAIN
TPIQGTAADIIKLAMIKIHRIFKEKGFGTRMLLQVHDELIFEAPKEIEEIQPIVRQI
MEGVVELKVPLKVNLAIGKNWAEAKA (SEQ ID NO:1).

An alternative amino acid sequence, identified by further and improved sequencing analysis, for the Large fragment of *T. indicus* DNA polymerase I is SEQ ID NO:32 as follows:

MGLLKELPATKTLSYDQYELVLDPDKVKEIVEKAKGAEVVAIDLESDTKDPM
RGKIVGVSLCFNPPKAYYFPFRHEGLEAQKQLPWEAFTHLASLIEDPSVKKIG
HNIKYDLIILARYGVTLKGLEGDTMLASYLLDPTRRTHGLDELAEEVLGHTMI
FYKEVTKELAKGESFARVPLEKAKVYACEDAHVTYLLYQYFWPKLKEESLW

KVFTEIDRPLIEVLAHMEMVGIKIDTAYLRGLSREMAEKLKELEEKIYTLAGE
KFNINSSKQLGQILFEKLKLPTVKKTPKKTAYSTDNEVLEELSAVHELPRLILE
YRTLAKLKSTYVDALPKMVNPETGRLHTSFNQTVTATGRLSSSDPNLQNIPVR
GEEGLKIRQAFVPEEIFAADYTQIDLRVLAHYSGDETLIKAFWQGEDIHRRTA
AEIFGIPPEEVTPEMRRMAKTINFGIVYGMSPYGLAKELKIGRREAKAFIERYF
ERYPGVKRYMEQIVAEAREKGYVETLFGRKRPLPDINSPNRTAREFAERTAIN
TPIQGTAADIIKLAMIKIHRIFKEKGFGTRMLLQVHDELLFEVPEKEIEEIQPIVR
QIMEGVVELKVPLKVNLAIGKNWAEAKA (SEQ ID NO:32)

This sequence is 99% identical to SEQ ID NO:1.

The predicted molecular weight of the 613 amino acid residue *T. indicus* DNA polymerase I Large fragment shown in SEQ ID NO:32 is about 69,990 Daltons. The predicted molecular weight of the 612 amino acid residue sequence shown in SEQ ID NO:1 is about 69,820 Daltons.

The amino acid sequence for inclusion in the polypeptide according to the invention may be an amino acid sequence with at least 51% identity, for example at least 55%, 56%, 57%, 58%, 59%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or even 99% identity, to the sequence shown in SEQ ID NO:32.

The percentage sequence identity may be determined using the BLASTP computer program with SEQ ID NO:1 or 32 as the base sequence. This means that SEQ ID NO:1 or 32, as appropriate, is the sequence against which the percentage identity is determined. The BLAST software is publicly available at http://blast.ncbi.nlm.nih.gov/Blast.cgi (accessible on 11 February 2009).

T. indicus is a thermophilic chemolithoautotrophic sulphate-reducing bacterium isolated from a deep-sea hydrothermal vent site, and has a reported temperature range for growth of 55-80°C and an optimum growth temperature of 70°C (see Moussard et al., 2004, Int. J. Syst. Evol. Microbiol. 54: 227-233). The inventors have isolated genomic DNA (gDNA) from *T. indicus* and used a sophisticated gene walking technique to clone a DNA polymerase A (polA) gene encoding a DNA polymerase I and corresponding Large fragment thereof. The Large fragment having the amino acid

sequence as shown in SEQ ID NO: 1 has been shown to be surprisingly efficient in both PCR and LAMP amplification reactions when compared with the different preferred DNA polymerases for these reactions. The ability of the *T. indicus* DNA polymerase I Large fragment to be sufficiently thermostable to function in PCR, with temperatures rising to around 94°C, could not have been predicted based on the optimum growth temperature of 70°C for this bacterium.

The polypeptide of the invention may exhibit strand displacement activity. The polypeptide may accordingly be suitable for carrying out isothermal amplification reactions such as LAMP.

The polypeptide may additionally or alternatively be suitable for carrying out thermocycling amplification reactions such as PCR.

The polypeptide as described herein may be about 613 amino acid residues in length, for example from about 610 to about 620, about 600 to about 630, about 550 to about 650, or about 500 to about 750 amino acids in length.

The polypeptide may comprise or consist essentially of the amino acid sequence SEQ ID NO:1 or 32, or of the amino acid sequence of SEQ ID NO:1 or 32 with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, about 20, about 30, about 40, about 50, about 100, about 200, about 250, about 260, about 270, 280, 281, 282, 283, 284, 285, 286, 287 or 288 contiguous amino acids added to or removed from any part of the polypeptide and/or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, about 20, about 30, about 40, about 50, about 100, about 150, about 200, about 250, about 260, about 270, 280, 281, 282, 283, 284, 285, 286, 287 or 288 amino acids added to or removed from the N-terminus region and/or the C-terminus region.

In one embodiment where the polypeptide of the invention includes an N-terminal His tag, the full length may be 619 amino acid residues.

According to a further aspect of the invention, there is provided an isolated polypeptide having thermostable DNA polymerase activity and comprising or consisting essentially of an amino acid sequence with at least 55% identity, for example at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or even 99% identity, to *T. indicus* DNA polymerase I as shown in SEQ ID NO:2. Preferably, the polypeptide according to this aspect of the invention is a polypeptide according to the first aspect of the invention and, therefore, has at least 51% identity to *T. indicus* DNA polymerase I Large fragment shown in SEQ ID NO:1

T. indicus DNA polymerase I has a full length amino acid sequence as follows:

MAQKSLFPKKLPFKDDKDPIFVIDGSSFVYRAYYAIRGHLSNRKGLPTKAVFG
FTQMLLKLLREMNPEYVVVCFDAKGPTFRHEMYKEYKANRPPMPDDLSVQI
PYIKEVTRAFGVPILEIEGFEADDLIAAIATRMERPIVIVGGDKDLFPLISEKVV
MWDPMKDELIDESWIKKRFGIEPKKLLDVRALAGDSIDNVPGVPGIGEKTAL
RLIKEYGSLEEVLNHAEEIKQKRLRENLIKHAGDALISKKLVELEAKAPIPLEP
DFYRKRPLNALKLRELFLELEFKKLLKELPATKTLSMTRYELVLDPDKVKEIV
EKAKGAEVVAIDLESDTKDPMRGKIVGVSLCFNPPKAYYFPFRHEGLEAQKQ
LPWEAFTHLASLIEDPSVKKIGHNIKYDLIILARYGVTLKGLEGDTMLASYLLD
PTRRTHGLDELAEEVLGHTMIFYKEVTKELAKGESFARVPLEKAKVYACEDA
HVTYLLYQYFWPKLKEESLWKVFTEIDRPLIEVLAHMEMVGIKIDTAYLRGLS
REMAEKLKELEEKIYTLAGEKFNINSSKQLGQILFEKLKLPTVKKTPKKTAYS
TDNEVLEELSAVHELPRLILEYRTLAKLKSTYVDALPKMVNPETGRLHTSFNQ
TVTATGRLSSSDPNLQNIPVRGEEGLKIRQAFVPEEIFAADYTQIDLRVLAHYS
GDETLIKAFWQGEDIHRRTAAEIFGIPPEEVTPEMRRMAKTINFGIVYGMSPYG
LAKELKIGRREAKAFIERYFERYPGVKRYMEQIVAEAREKGYVETLFGRKRPL
PDINSPNRTAREFAERTAINTPIQGTAADIIKLAMIKIHRIFKEKGFGTRMLLQV
HDELIFEAPEKEIEEIQPIVRQIMEGVVELKVPLKVNLAIGKNWAEAKA (SEQ ID NO: 2).

An alternative amino acid sequence, identified by further and improved sequencing analysis, for full length *T. indicus* DNA polymerase I is SEQ ID NO:34 as follows:

MAQKSLFPKKLPFKDDKDPIFVIDGSSFVYRAYYAIRGHLSNRKGLPTKAVFG
FTQMLLKLLREMNPEYVVVCFDAKGPTFRHEMYKEYKANRPPMPDDLSVQI
PYIKEVTRAFGVPILEIEGFEADDLIAAIATRMERPIVIVGGDKDLFPLISEKVV
MWDPMKDELIDESWIKKRFGIEPKKLLDVRALAGDSIDNVPGVPGIGEKTAL
RLIKEYGSLEEVLNHAEEIKQKRLRENLIKHAGDALISKKLVELEAKAPIPLEP
DFYRKRPLNALKLRELFLELEFKKLLKELPATKTLSYDQYELVLDPDKVKEIV
EKAKGAEVVAIDLESDTKDPMRGKIVGVSLCFNPPKAYYFPFRHEGLEAQKQ
LPWEAFTHLASLIEDPSVKKIGHNIKYDLIILARYGVTLKGLEGDTMLASYLLD
PTRRTHGLDELAEEVLGHTMIFYKEVTKELAKGESFARVPLEKAKVYACEDA
HVTYLLYQYFWPKLKEESLWKVFTEIDRPLIEVLAHMEMVGIKIDTAYLRGLS

REMAEKLKELEEKIYTLAGEKFNINSSKQLGQILFEKLKLPTVKKTPKKTAYS
TDNEVLEELSAVHELPRLILEYRTLAKLKSTYVDALPKMVNPETGRLHTSFNQ
TVTATGRLSSSDPNLQNIPVRGEEGLKIRQAFVPEEIFAADYTQIDLRVLAHYS
GDETLIKAFWQGEDIHRRTAAEIFGIPPEEVTPEMRRMAKTINFGIVYGMSPYG
LAKELKIGRREAKAFIERYFERYPGVKRYMEQIVAEAREKGYVETLFGRKRPL
PDINSPNRTAREFAERTAINTPIQGTAADIIKLAMIKIHRIFKEKGFGTRMLLQV
HDELLFEVPEKEIEEIQPIVRQIMEGVVELKVPLKVNLAIGKNWAEAKA (SEQ ID NO: 34)

This sequence is 99.44% identical to SEQ ID NO:2.

The predicted molecular weight of this 900 amino acid residue *T. indicus* DNA polymerase I shown in SEQ ID NO:34 is about 102,900 Daltons. The predicted molecular weight of the 900 amino acid residue sequence shown in SEQ ID NO:2 is about 102,850 Daltons.

The amino acid sequence for inclusion in the polypeptide according to the invention may be an amino acid sequence with at least 51% identity, for example at least 52%, 53%, 54%, 55%, 56%, 57%, 58%, 59%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or even 99% identity, to the sequence shown in SEQ ID NO:34.

The percentage sequence identity may be determined using the BLASTP computer program with SEQ ID NO:2 or 34 as the base sequence. This means that SEQ ID NO:2 or 34, as appropriate, is the sequence against which the percentage identity is determined.

The polypeptide may comprise or consist essentially of the amino acid sequence SEQ ID NO:2 or 34, or of the amino acid sequence of SEQ ID NO:2 or 34 with 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, about 20, about 30, about 40, about 50, about 100, about 150, about 200, about 250, about 260, about 270, 280, 281, 282, 283, 284, 285, 286, 287 or 288 contiguous amino acids added to or removed from any part of the polypeptide and/or 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, about 20, about 30, about 40, about 50, about 100, about 150, about 200, about 250, about 260, about 270,

280, 281, 282, 283, 284, 285, 286, 287 or 288 amino acids added to or removed from the N-terminus region and/or the C-terminus region.

The polypeptide according to this aspect of the invention may be an isolated thermostable DNA polymerase I obtainable from *T. indicus* and having a molecular weight of about 102,500 to 103,500 Daltons (preferably about 102,900 or about 103,000 Daltons), or an enzymatically active fragment thereof. The term "enzymatically active fragment" means a fragment of such a polymerase obtainable from *T. indicus* and having enzyme activity which is at least 60%, preferably at least 70%, more preferably at least 80%, yet more preferably 90%, 95%, 96%, 97%, 98%, 99% or 100% that of the full length polymerase being compared to. The given activity may be determined by any standard measure, for example, the number of bases of nucleotides of the template sequence which can be replicated in a given time period. The skilled person is routinely able to determine such properties and activities.

Residues 3-612 of the *T. indicus* DNA polymerase I Large fragment shown in SEQ ID NO: 1 correspond with residues 290-900 of the full length DNA polymerase I shown in SEQ ID NO:2. Residues 1-2 of SEQ ID NO:1 are artificially introduced compared to the sequence of SEQ ID NO:2 to allow *in vitro* expression of the Large fragment in a host cell (see Examples below). Similarly, residues 3-613 of the *T. indicus* DNA polymerase I Large fragment shown in SEQ ID NO: 32 correspond with residues 290-900 of the full length DNA polymerase I shown in SEQ ID NO:34.

The polypeptide according to the invention may be greater in size where, according to a further aspect of the invention, it comprises additional functional or structural domains, for example an affinity purification tag (such as an His purification tag), or DNA polymerase activity-enhancing domains such as the proliferating cell nuclear antigen homologue from *Archaeoglobus fulgidus*, T3 DNA polymerase thioredoxin binding domain, DNA binding protein Sso7d from *Sulfolobus solfataricus*, Sso7d-like proteins, or mutants thereof, or helix-hairpin-helix motifs derived from DNA topoisomerase V. The DNA polymerase activity-enhancing domain may also be a Cren7 enhancer domain or variant thereof, as defined and exemplified in co-pending

International patent application no. PCT/GB2009/000063, which discloses that this highly conserved protein domain from Crenarchaeal organisms is useful to enhance the properties of a DNA polymerase. International patent application no. PCT/GB2009/000063 is incorporated herein by reference in its entirety.

The polypeptides of the invention may be suitable for use in one or more reactions requiring DNA polymerase activity, for example one or more of the group consisting of: nick translation, second-strand cDNA synthesis in cDNA cloning, DNA sequencing, thermocycling amplification reactions such as PCR, and isothermal amplification reactions for example strand displacement amplification (SDA), nucleic acid sequence-based amplification (NASBA), self-sustained sequence replication (3SR) and LAMP.

Also provided according to the present invention is a polypeptide with thermostable 5'→3' exonuclease activity and having at least 55% identity, for example at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or even 99% identity, to residues 1-289 of *T. indicus* DNA polymerase I as shown in SEQ ID NO: 2 or 34.

Based on sequence comparisons with known DNA polymerases, in one aspect the polypeptides of the invention have 3'→5' exonuclease proofreading activity.

In a further aspect of the invention the polypeptides exhibit high fidelity polymerase activity during a thermocycling amplification reaction (such as PCR). High fidelity may be defined as a PCR error rate of less than 1 nucleotide per 300 x 10^6 amplified nucleotides, for example less than 1 nucleotide per 250 x 10^6, 200 x 10^6, 150 x 10^6, 100 x 10^6 or 50 x 10^6 amplified nucleotides. Alternatively, the error rate of the polypeptides may be in the range 1-300 nucleotides per 10^6 amplified nucleotides, for example 1-200, 1-100, 100-300, 200-300, 100-200 or 75-200 nucleotides per 10^6 amplified nucleotides. Error rate may be determined using the opal reversion assay as described by Kunkel et al. (1987, Proc. Natl Acad. Sci. USA 84: 4865-4869).

In another aspect of the invention there is provided a composition comprising the polypeptide as described herein. The composition may for example include a buffer, most or all ingredients for performing a reaction (such as a DNA amplification reaction for example PCR or LAMP), a stabiliser (such as *E. coli* GroEL protein, to enhance thermostability), and/or other compounds.

The invention further provides an isolated nucleic acid encoding the polypeptide with identity to the *T. indicus* DNA polymerase I Large fragment. The nucleic acid may, for example, have a sequence as shown below (5'-3'):

atgggcctcttaaaggaacttccagctactaaaaccctttcgatgaccagatacgagctggttcttgacccggataaagtaaa
agaaattgtagaaaaggccaaaggggccgaagtggtggctattgaccttgaaagtgatacgaaagaccccatgcgtggg
aaaatagtaggggtctcgctttgttttaacccgcccaaagcctattatttccctttagacatgaaggccttgaggcccaaaag
cagcttccctgggaggcctttactcatctggccagcctcattgaagacccctcagttaaaaagataggccacaatatcaagta
tgacttgattattcttgctcgctacggcgtaactttaaagggccttgaaggggataccatgctggcttcgtatctccttgatcca
acacgtcgtacccacggccttgatgagctggccgaagaggtcctggggcataccatgatttttacaaggaagtgactaaa
gaactggccaaaggagagagctttgccagggtccctcttgaaaaggcaaaagtttacgcctgtgaagacgcccacgttac
ctatctgctttatcaatatttctggcccaaactcaaagaggaaagcctctggaaggtctttacggaaattgatcgacctttaata
gaagttttggcccacatggaaatggtaggtattaagattgacaccgcctatcttagaggactttcgcgagaaatggctgaaaa
gttaaaggagcttgaagaaaaaatttacaccctggctggtgaaaaatttaatatcaattccagcaaacaactgggccagatttt
atttgaaaagctaaaactccctacggttaaaaagaccccaaaaaaaacggcctattcaacggataacgaagtattagagga
actttctgcggtccacgaacttccgcgtctgatacttgagtatagaactctggctaaactcaaatctacttatgttgatgccctcc
cgaagatggttaatcctgaaactggtcgtcttcatacttcctttaaccagacggttacggccactggaagactttcaagcagtg
accctaatcttcaaaatattcctgtgcgtggtgaagaggggcttaagattcgccaggcctttgtgccggaggagAtttttgct
gccgattacactcagatcgatctgcgagttttagcccattactcgggagatgaaaccttgattaaggccttctggcaggggg
aagacattcaccggcgcacggctgcagaaatttttggtatcccgccagaagaagtaactcctgagatgcggcgtatggcca
agactataaactttggcattgtttacggcatgagtccttacggtctggcgaaagaactcaaaattggccgccgtgaggccaa
ggcctttattgagcgctattttgaacgctacccaggtgtgaaacgctatatggaacaaatcgtggctgaagcccgagaaaag
ggctacgtggagacccttttcggacgcaaaaggcctcttcctgacatcaatagccctaatcgtacggcgcgcgagtttgcc
gagcgcacggctataaacactcctattcaggggacagccgctgatattatcaagctcgccatgataaaaattcaccggatttt
taaagaaaaaggctttgggacaaggatgcttcttcaggtgcatgacgagcttatttttgaagcgccaaaagagattgaagaa
atccagccaattgtccgacaaatcatggaaggagtggttgaattgaaggttcctctaaaagtaaacctggcaatagggaaaa
attgggcagaggcaaaggcataa (SEQ ID NO:3).

The nucleotide of SEQ ID NO:3 encodes the *T. indicus* DNA polymerase I Large fragment of SEQ ID NO:1 as follows:

```
   1 atgggcctcttaaaggaacttccagctactaaaaccctttcgatgaccagatacgagctg
   1  M  G  L  L  K  E  L  P  A  T  K  T  L  S  M  T  R  Y  E  L

  61 gttcttgacccggataaagtaaaagaaattgtagaaaaggccaaaggggccgaagtggtg
  21  V  L  D  P  D  K  V  K  E  I  V  E  K  A  K  G  A  E  V  V

 121 gctattgaccttgaaagtgatacgaaagaccccatgcgtgggaaaatagtaggggtctcg
  41  A  I  D  L  E  S  D  T  K  D  P  M  R  G  K  I  V  G  V  S
```

```
181 ctttgttttaacccgcccaaagcctattatttccctttagacatgaaggccttgaggcc
 61  L  C  F  N  P  P  K  A  Y  Y  F  P  F  R  H  E  G  L  E  A

241 caaaagcagcttccctgggaggcctttactcatctggccagcctcattgaagacccctca
 81  Q  K  Q  L  P  W  E  A  F  T  H  L  A  S  L  I  E  D  P  S

301 gttaaaaagataggccacaatatcaagtatgacttgattattcttgctcgctacggcgta
101  V  K  K  I  G  H  N  I  K  Y  D  L  I  I  L  A  R  Y  G  V

361 actttaaagggccttgaaggggataccatgctggcttcgtatctccttgatccaacacgt
121  T  L  K  G  L  E  G  D  T  M  L  A  S  Y  L  L  D  P  T  R

421 cgtacccacggccttgatgagctggccgaagaggtcctggggcataccatgatttttac
141  R  T  H  G  L  D  E  L  A  E  E  V  L  G  H  T  M  I  F  Y

481 aaggaagtgactaaagaactggccaaaggagagagctttgccagggtccctcttgaaaag
161  K  E  V  T  K  E  L  A  K  G  E  S  F  A  R  V  P  L  E  K

541 gcaaaagtttacgcctgtgaagacgcccacgttacctatctgctttatcaatatttctgg
181  A  K  V  Y  A  C  E  D  A  H  V  T  Y  L  L  Y  Q  Y  F  W

601 cccaaactcaaagaggaaagcctctggaaggtctttacggaaattgatcgacctttaata
201  P  K  L  K  E  E  S  L  W  K  V  F  T  E  I  D  R  P  L  I

661 gaagttttggcccacatggaaatggtaggtattaagattgacaccgcctatcttagagga
221  E  V  L  A  H  M  E  M  V  G  I  K  I  D  T  A  Y  L  R  G

721 ctttcgcgagaaatggctgaaaagttaaaggagcttgaagaaaaaatttacaccctggct
241  L  S  R  E  M  A  E  K  L  K  E  L  E  E  K  I  Y  T  L  A

781 ggtgaaaaatttaatatcaattccagcaaacaactgggccagattttatttgaaaagcta
261  G  E  K  F  N  I  N  S  S  K  Q  L  G  Q  I  L  F  E  K  L

841 aaactccctacggttaaaaagaccccaaaaaaaacggcctattcaacggataacgaagta
281  K  L  P  T  V  K  K  T  P  K  K  T  A  Y  S  T  D  N  E  V

901 ttagaggaactttctgcggtccacgaacttccgcgtctgatacttgagtatagaactctg
301  L  E  E  L  S  A  V  H  E  L  P  R  L  I  L  E  Y  R  T  L

961 gctaaactcaaatctacttatgttgatgccctcccgaagatggttaatcctgaaactggt
321  A  K  L  K  S  T  Y  V  D  A  L  P  K  M  V  N  P  E  T  G

1021 cgtcttcatacttcctttaaccagacggttacggccactggaagactttcaagcagtgac
 341  R  L  H  T  S  F  N  Q  T  V  T  A  T  G  R  L  S  S  S  D

1081 cctaatcttcaaaatattcctgtgcgtggtgaagaggggcttaagattcgccaggccttt
 361  P  N  L  Q  N  I  P  V  R  G  E  E  G  L  K  I  R  Q  A  F

1141 gtgccggaggagatttttgctgccgattacactcagatcgatctgcgagttttagcccat
 381  V  P  E  E  I  F  A  A  D  Y  T  Q  I  D  L  R  V  L  A  H

1201 tactcgggagatgaaaccttgattaaggccttctggcaggggggaagacattcaccggcgc
 401  Y  S  G  D  E  T  L  I  K  A  F  W  Q  G  E  D  I  H  R  R

1261 acggctgcagaaatttttggtatcccgccagaagaagtaactcctgagatgcggcgtatg
 421  T  A  A  E  I  F  G  I  P  P  E  E  V  T  P  E  M  R  R  M

1321 gccaagactataaactttggcattgtttacggcatgagtccttacggtctggcgaaagaa
 441  A  K  T  I  N  F  G  I  V  Y  G  M  S  P  Y  G  L  A  K  E
```

```
1381 ctcaaaattggccgccgtgaggccaaggcctttattgagcgctattttgaacgctaccca
 461  L  K  I  G  R  R  E  A  K  A  F  I  E  R  Y  F  E  R  Y  P

1441 ggtgtgaaacgctatatggaacaaatcgtggctgaagcccgagaaaagggctacgtggag
 481  G  V  K  R  Y  M  E  Q  I  V  A  E  A  R  E  K  G  Y  V  E

1501 acccttttcggacgcaaaaggcctcttcctgacatcaatagccctaatcgtacggcgcgc
 501  T  L  F  G  R  K  R  P  L  P  D  I  N  S  P  N  R  T  A  R

1561 gagtttgccgagcgcacggctataaacactcctattcaggggacagccgctgatattatc
 521  E  F  A  E  R  T  A  I  N  T  P  I  Q  G  T  A  A  D  I  I

1621 aagctcgccatgataaaaattcaccggatttttaaagaaaaaggctttgggacaaggatg
 541  K  L  A  M  I  K  I  H  R  I  F  K  E  K  G  F  G  T  R  M

1681 cttcttcaggtgcatgacgagcttatttttgaagcgccaaaagagattgaagaaatccag
 561  L  L  Q  V  H  D  E  L  I  F  E  A  P  K  E  I  E  E  I  Q

1741 ccaattgtccgacaaatcatggaaggagtggttgaattgaaggttcctctaaaagtaaac
 581  P  I  V  R  Q  I  M  E  G  V  V  E  L  K  V  P  L  K  V  N

1801 ctggcaatagggaaaaattgggcagaggcaaaggcatga     (SEQ ID NO:3)
 601  L  A  I  G  K  N  W  A  E  A  K  A  *     (SEQ ID NO:1).
```

Alternatively, the nucleic acid has the sequence shown below (5'-3'):

atgggcctcttaaaggaacttccagctactaaaacccttt cgtatgaccagtacgagctggttcttgacccggataaagtaaa
agaaattgtagaaaaggccaaaggggccgaagtggtggctattgaccttgaaagtgatacgaaagaccccatgcgtggg
aaaatagtaggggtctcgctttgttttaacccgcccaaagcctattatttccctttagacatgaaggccttgaggcccaaaag
cagcttccctgggaggcctttactcatctggccagcctcattgaagacccctcagttaaaaagataggccacaatatcaagta
tgacttgattattcttgctcgctacggcgtaactttaaagggccttgaaggggataccatgctggcttcgtatctccttgatcca
acacgtcgtacccacggccttgatgagctggccgaagaggtcctggggcataccatgatttttacaaggaagtgactaaa
gaactggccaaaggagagagctttgccagggtccctcttgaaaaggcaaaagtttacgcctgtgaagacgcccacgttac
ctatctgctttatcaatatttctggcccaaactcaaagaggaaagcctctggaaggtctttacggaaattgatcgacctttaata
gaagttttggcccacatggaaatggtaggtattaagattgacaccgcctatcttagaggactttcgcgagaaatggctgaaaa
gttaaaggagcttgaagaaaaaatttacaccctggctggtgaaaaatttaatatcaattccagcaaacaactgggccagatttt
atttgaaaagctaaaactccctacggttaaaaagaccccaaaaaaaacggcctattcaacggataacgaagtattagagga
actttctgcggtccacgaacttccgcgtctgatacttgagtatagaactctggctaaactcaaatctacttatgttgatgccctcc
cgaagatggttaatcctgaaactggtcgtcttcatacttcctttaaccagacggttacggccactggaagactttcaagcagtg
accctaatcttcaaaatattcctgtgcgtggtgaagaggggcttaagattcgccaggcctttgtgccggaggagatttttgctg
ccgattacactcagatcgatctgcgagttttagcccattactcgggagatgaaaccttgattaaggccttctggcaggggaa
gacattcaccggcgcacggctgcagaaatttttggtatcccgccagaagaagtaactcctgagatgcggcgtatggccaag
actataaactttggcattgtttacggcatgagtccttacggtctggcgaaagaactcaaaattggccgccgtgaggccaagg
cctttattgagcgctattttgaacgctacccaggtgtgaaacgctatatggaacaaatcgtggctgaagcccgagaaaaggg
ctacgtggagacccttttcggacgcaaaaggcctcttcctgacatcaatagccctaatcgtacggcgcgcgagtttgccgag
cgcacggctataaacactcctattcaggggacagccgctgatattatcaagctcgccatgataaaaattcaccggatttttaa
agaaaaaggctttgggacaaggatgcttcttcaggtgcacgacgaacttctttttgaagtgcctgaaaaagagattgaagaa
atccagccaattgtccgacaaatcatggaaggagtggttgaattgaaggttcctctaaaagtaaacctggcaatagggaaaa
attgggcagaggcaaaggcataa (SEQ ID NO:33).

The nucleotide of SEQ ID NO:33 encodes the *T. indicus* polymerase I Large fragment of SEQ ID NO:32 as follows:

```
   1 atgggcctcttaaaggaacttccagctactaaaaccctttcgtatgaccagtacgagctg
   1  M  G  L  L  K  E  L  P  A  T  K  T  L  S  Y  D  Q  Y  E  L

  61 gttcttgacccggataaagtaaaagaaattgtagaaaaggccaaaggggccgaagtggtg
  21  V  L  D  P  D  K  V  K  E  I  V  E  K  A  K  G  A  E  V  V

 121 gctattgaccttgaaagtgatacgaaagaccccatgcgtgggaaaatagtaggggtctcg
  41  A  I  D  L  E  S  D  T  K  D  P  M  R  G  K  I  V  G  V  S

 181 ctttgttttaacccgcccaaagcctattatttccctttagacatgaaggccttgaggcc
  61  L  C  F  N  P  P  K  A  Y  Y  F  P  F  R  H  E  G  L  E  A

 241 caaaagcagcttccctgggaggcctttactcatctggccagcctcattgaagacccctca
  81  Q  K  Q  L  P  W  E  A  F  T  H  L  A  S  L  I  E  D  P  S

 301 gttaaaaagataggccacaatatcaagtatgacttgattattcttgctcgctacggcgta
 101  V  K  K  I  G  H  N  I  K  Y  D  L  I  I  L  A  R  Y  G  V

 361 actttaaagggccttgaaggggataccatgctggcttcgtatctccttgatccaacacgt
 121  T  L  K  G  L  E  G  D  T  M  L  A  S  Y  L  L  D  P  T  R

 421 cgtacccacggccttgatgagctggccgaagaggtcctggggcataccatgatttttac
 141  R  T  H  G  L  D  E  L  A  E  E  V  L  G  H  T  M  I  F  Y

 481 aaggaagtgactaaagaactggccaaaggagagagctttgccagggtccctcttgaaaag
 161  K  E  V  T  K  E  L  A  K  G  E  S  F  A  R  V  P  L  E  K

 541 gcaaaagtttacgcctgtgaagacgcccacgttacctatctgctttatcaatatttctgg
 181  A  K  V  Y  A  C  E  D  A  H  V  T  Y  L  L  Y  Q  Y  F  W

 601 cccaaactcaaagaggaaagcctctggaaggtctttacggaaattgatcgacctttaata
 201  P  K  L  K  E  E  S  L  W  K  V  F  T  E  I  D  R  P  L  I

 661 gaagttttggcccacatggaaatggtaggtattaagattgacaccgcctatcttagagga
 221  E  V  L  A  H  M  E  M  V  G  I  K  I  D  T  A  Y  L  R  G

 721 ctttcgcgagaaatggctgaaaagttaaaggagcttgaagaaaaaatttacaccctggct
 241  L  S  R  E  M  A  E  K  L  K  E  L  E  E  K  I  Y  T  L  A

 781 ggtgaaaaatttaatatcaattccagcaaacaactgggccagattttatttgaaaagcta
 261  G  E  K  F  N  I  N  S  S  K  Q  L  G  Q  I  L  F  E  K  L

 841 aaactccctacggttaaaaagaccccaaaaaaaacggcctattcaacggataacgaagta
 281  K  L  P  T  V  K  K  T  P  K  K  T  A  Y  S  T  D  N  E  V

 901 ttagaggaactttctgcggtccacgaacttccgcgtctgatacttgagtatagaactctg
 301  L  E  E  L  S  A  V  H  E  L  P  R  L  I  L  E  Y  R  T  L

 961 gctaaactcaaatctacttatgttgatgccctcccgaagatggttaatcctgaaactggt
 321  A  K  L  K  S  T  Y  V  D  A  L  P  K  M  V  N  P  E  T  G

1021 cgtcttcatacttcctttaaccagacggttacggccactggaagactttcaagcagtgac
 341  R  L  H  T  S  F  N  Q  T  V  T  A  T  G  R  L  S  S  S  D
```

```
1081 cctaatcttcaaaatattcctgtgcgtggtgaagaggggcttaagattcgccaggccttt
 361  P  N  L  Q  N  I  P  V  R  G  E  E  G  L  K  I  R  Q  A  F

1141 gtgccggaggagatttttgctgccgattacactcagatcgatctgcgagttttagcccat
 381  V  P  E  E  I  F  A  A  D  Y  T  Q  I  D  L  R  V  L  A  H

1201 tactcgggagatgaaaccttgattaaggccttctggcaggggggaagacattcaccggcgc
 401  Y  S  G  D  E  T  L  I  K  A  F  W  Q  G  E  D  I  H  R  R

1261 acggctgcagaaatttttggtatcccgccagaagaagtaactcctgagatgcggcgtatg
 421  T  A  A  E  I  F  G  I  P  P  E  E  V  T  P  E  M  R  R  M

1321 gccaagactataaactttggcattgtttacggcatgagtccttacggtctggcgaaagaa
 441  A  K  T  I  N  F  G  I  V  Y  G  M  S  P  Y  G  L  A  K  E

1381 ctcaaaattggccgccgtgaggccaaggcctttattgagcgctattttgaacgctaccca
 461  L  K  I  G  R  R  E  A  K  A  F  I  E  R  Y  F  E  R  Y  P

1441 ggtgtgaaacgctatatggaacaaatcgtggctgaagcccgagaaaagggctacgtggag
 481  G  V  K  R  Y  M  E  Q  I  V  A  E  A  R  E  K  G  Y  V  E

1501 acccttttcggacgcaaaaggcctcttcctgacatcaatagccctaatcgtacggcgcgc
 501  T  L  F  G  R  K  R  P  L  P  D  I  N  S  P  N  R  T  A  R

1561 gagtttgccgagcgcacggctataaacactcctattcaggggacagccgctgatattatc
 521  E  F  A  E  R  T  A  I  N  T  P  I  Q  G  T  A  A  D  I  I

1621 aagctcgccatgataaaaattcaccggatttttaaagaaaaaggctttgggacaaggatg
 541  K  L  A  M  I  K  I  H  R  I  F  K  E  K  G  F  G  T  R  M

1681 cttcttcaggtgcacgacgaacttctttttgaagtgcctgaaaaagagattgaagaaatc
 561  L  L  Q  V  H  D  E  L  L  F  E  V  P  E  K  E  I  E  E  I

1741 cagccaattgtccgacaaatcatggaaggagtggttgaattgaaggttcctctaaaagta
 581  Q  P  I  V  R  Q  I  M  E  G  V  V  E  L  K  V  P  L  K  V

1801 aacctggcaatagggaaaaattgggcagaggcaaaggcataa      (SEQ ID NO:33)
 601  N  L  A  I  G  K  N  W  A  E  A  K  A  *      (SEQ ID NO:32).
```

The invention further provides an isolated nucleic acid encoding the polypeptide with identity to the *T. indicus* full length DNA polymerase I. The nucleic acid may, for example, have a sequence as shown below (5'-3'):

atggcgcagaaaagcttgtttcctaaaaaattaccatttaaagatgataaagaccccatcttcgttattgacgggagttcttttgt
ttaccgggcttactatgccataagagggcatctatcaaaccgcaaagggctcccaaccaaggcggtctttgggtttacccag
atgcttttaaagcttttgcgtgagatgaaccctgagtatgtggtggtgtgctttgacgccaaagggcctacttttcgccacgag
atgtacaaagaatacaaagccaaccgcccccccatgccagatgatctttccgtccagattccctatatcaaagaggtaacca
gggcctttggagtccctattcttgaaatagaaggctttgaagctgacgatctcatcgccgctattgccactcgtatggaaaga
ccaattgtcatcgttggtggagataaagatttgttcccccttatttcagagaaagttgtcatgtgggaccccatgaaagacgaa
ctgattgacgaaagctggataaagaaacgttttggcattgaacctaaaaagctccttgatgtaagggcccttgccggcgata
gcattgataacgtgccaggggttccgggtattggtgaaaaaacggccctaaggctcataaaagaatacggttcccttgaag
aagtccttaaccatgccgaagaaataaaacaaaagcgcttgcgtgaaaacctcatcaaacacgccggagacgcccttattt
ccaaaaaactggttgagcttgaggccaaagccccaatccccttgagcctgatttttaccgcaaacggccattaaatgccct

aaaactaagggaactcttccttgagcttgaatttaaaaagctcttaaaggaacttccagctactaaaaccctttcgatgaccag
atacgagctggttcttgacccggataaagtaaaagaaattgtagaaaaggccaaaggggccgaagtggtggctattgacct
tgaaagtgatacgaaagaccccatgcgtgggaaaatagtaggggtctcgctttgttttaacccgcccaaagcctattatttcc
cttttagacatgaaggccttgaggcccaaaagcagcttccctgggaggcctttactcatctggccagcctcattgaagaccc
ctcagttaaaaagataggccacaatatcaagtatgacttgattattcttgctcgctacggcgtaactttaaagggccttgaagg
ggataccatgctggcttcgtatctccttgatccaacacgtcgtacccacggccttgatgagctggccgaagaggtcctgggg
cataccatgatttttttacaaggaagtgactaaagaactggccaaaggagagagctttgccagggtccctcttgaaaaggcaa
aagtttacgcctgtgaagacgcccacgttacctatctgctttatcaatatttctggcccaaactcaaagaggaaagcctctgga
aggtctttacggaaattgatcgacctttaatagaagttttggcccacatggaaatggtaggtattaagattgacaccgcctatct
tagaggactttcgcgagaaatggctgaaaagttaaaggagcttgaagaaaaaatttacaccctggctggtgaaaaatttaat
atcaattccagcaaacaactgggccagattttatttgaaaagctaaaactccctacggttaaaaagaccccaaaaaaaacgg
cctattcaacggataacgaagtattagaggaactttctgcggtccacgaacttccgcgtctgatacttgagtatagaactctgg
ctaaactcaaatctacttatgttgatgccctcccgaagatggttaatcctgaaactggtcgtcttcatacttcctttaaccagacg
gttacggccactggaagactttcaagcagtgaccctaatcttcaaaatattcctgtgcgtggtgaagaggggcttaagattcg
ccaggcctttgtgccggaggagatttttgctgccgattacactcagatcgatctgcgagttttagcccattactcgggagatga
aaccttgattaaggccttctggcaggggaagacattcaccggcgcacggctgcagaaatttttggtatcccgccagaaga
agtaactcctgagatgcggcgtatggccaagactataaactttggcattgtttacggcatgagtccttacggtctggcgaaag
aactcaaaattggccgccgtgaggccaaggcctttattgagcgctattttgaacgctacccaggtgtgaaacgctatatgga
acaaatcgtggctgaagcccgagaaaagggctacgtggagacccttttcggacgcaaaaggcctcttcctgacatcaata
gccctaatcgtacggcgcgcgagtttgccgagcgcacggctataaacactcctattcaggggacagccgctgatattatca
agctcgccatgataaaaattcaccggatttttaaagaaaaaggctttgggacaaggatgcttcttcaggtgcatgacgagctt
attttgaagcgcctgaaaaagagattgaagaaatccagccaattgtccgacaaatcatggaaggagtggttgaattgaagg
ttcctctaaaagtaaacctggcaatagggaaaaattgggcagaggcaaaggcataa (SEQ ID NO:4).

The nucleotide of SEQ ID NO:4 encodes the *T. indicus* full length DNA polymerase I of SEQ ID NO:2 as follows:

```
   1 atggctcaaaaaagtttgtttcctaaaaaattaccatttaaagatgataaagaccccatc
   1  M  A  Q  K  S  L  F  P  K  K  L  P  F  K  D  D  K  D  P  I

  61 ttcgttattgacgggagttcttttgtttaccgggcttactatgccataagagggcatcta
  21  F  V  I  D  G  S  S  F  V  Y  R  A  Y  Y  A  I  R  G  H  L

 121 tcaaaccgcaaagggctcccaaccaaggcggtctttgggtttacccagatgcttttaaag
  41  S  N  R  K  G  L  P  T  K  A  V  F  G  F  T  Q  M  L  L  K

 181 cttttgcgtgagatgaaccctgagtatgtggtggtgtgctttgacgccaaagggcctact
  61  L  L  R  E  M  N  P  E  Y  V  V  V  C  F  D  A  K  G  P  T

 241 tttcgccacgagatgtacaaagaatacaaagccaaccgccccccccatgccagatgatctt
  81  F  R  H  E  M  Y  K  E  Y  K  A  N  R  P  P  M  P  D  D  L

 301 tccgtccagattccctatatcaaagaggtaaccagggcctttggagtccctattcttgaa
 101  S  V  Q  I  P  Y  I  K  E  V  T  R  A  F  G  V  P  I  L  E

 361 atagaaggctttgaagctgacgatctcatcgccgctattgccactcgtatggaaagacca
 121  I  E  G  F  E  A  D  D  L  I  A  A  I  A  T  R  M  E  R  P

 421 attgtcatcgttggtggagataaagatttgttccccccttatttcagagaaagttgtcatg
 141  I  V  I  V  G  G  D  K  D  L  F  P  L  I  S  E  K  V  V  M
```

```
 481 tgggaccccatgaaagacgaactgattgacgaaagctggataaagaaacgttttggcatt
 161  W  D  P  M  K  D  E  L  I  D  E  S  W  I  K  K  R  F  G  I

 541 gaacctaaaaagctccttgatgtaagggcccttgccggcgatagcattgataacgtgcca
 181  E  P  K  K  L  L  D  V  R  A  L  A  G  D  S  I  D  N  V  P

 601 ggggttccgggtattggtgaaaaaacggccctaaggctcataaaagaatacggttccctt
 201  G  V  P  G  I  G  E  K  T  A  L  R  L  I  K  E  Y  G  S  L

 661 gaagaagtccttaaccatgccgaagaaataaaacaaaagcgcttgcgtgaaaacctcatc
 221  E  E  V  L  N  H  A  E  E  I  K  Q  K  R  L  R  E  N  L  I

 721 aaacacgccggagacgcccttatttccaaaaaactggttgagcttgaggccaaagcccca
 241  K  H  A  G  D  A  L  I  S  K  K  L  V  E  L  E  A  K  A  P

 781 atccccccttgagcctgatttttaccgcaaacggccattaaatgccctaaaactaagggaa
 261  I  P  L  E  P  D  F  Y  R  K  R  P  L  N  A  L  K  L  R  E

 841 ctcttccttgagcttgaatttaaaaagctcttaaaggaacttccagctactaaaacccttt
 281  L  F  L  E  L  E  F  K  K  L  L  K  E  L  P  A  T  K  T  L

 901 tcgatgaccagatacgagctggttcttgacccggataaagtaaaagaaattgtagaaaag
 301  S  M  T  R  Y  E  L  V  L  D  P  D  K  V  K  E  I  V  E  K

 961 gccaaaggggccgaagtggtggctattgaccttgaaagtgatacgaaagaccccatgcgt
 321  A  K  G  A  E  V  V  A  I  D  L  E  S  D  T  K  D  P  M  R

1021 gggaaaatagtaggggtctcgctttgttttaacccgcccaaagcctattatttccctttt
 341  G  K  I  V  G  V  S  L  C  F  N  P  P  K  A  Y  Y  F  P  F

1081 agacatgaaggccttgaggcccaaaagcagcttccctgggaggcctttactcatctggcc
 361  R  H  E  G  L  E  A  Q  K  Q  L  P  W  E  A  F  T  H  L  A

1141 agcctcattgaagacccctcagttaaaaagataggccacaatatcaagtatgacttgatt
 381  S  L  I  E  D  P  S  V  K  K  I  G  H  N  I  K  Y  D  L  I

1201 attcttgctcgctacggcgtaactttaaagggccttgaaggggataccatgctggcttcg
 401  I  L  A  R  Y  G  V  T  L  K  G  L  E  G  D  T  M  L  A  S

1261 tatctccttgatccaacacgtcgtacccacggccttgatgagctggccgaagaggtcctg
 421  Y  L  L  D  P  T  R  R  T  H  G  L  D  E  L  A  E  E  V  L

1321 gggcataccatgatttttacaaggaagtgactaaagaactggccaaaggagagagctttt
 441  G  H  T  M  I  F  Y  K  E  V  T  K  E  L  A  K  G  E  S  F

1381 gccagggtccctcttgaaaaggcaaaagtttacgcctgtgaagacgcccacgttacctat
 461  A  R  V  P  L  E  K  A  K  V  Y  A  C  E  D  A  H  V  T  Y

1441 ctgctttatcaatatttctggcccaaactcaaagaggaaagcctctggaaggtctttacg
 481  L  L  Y  Q  Y  F  W  P  K  L  K  E  E  S  L  W  K  V  F  T

1501 gaaattgatcgacctttaatagaagttttggcccacatggaaatggtaggtattaagatt
 501  E  I  D  R  P  L  I  E  V  L  A  H  M  E  M  V  G  I  K  I

1561 gacaccgcctatcttagaggactttcgcgagaaatggctgaaaagttaaaggagcttgaa
 521  D  T  A  Y  L  R  G  L  S  R  E  M  A  E  K  L  K  E  L  E

1621 gaaaaaatttacaccctggctggtgaaaaatttaatatcaattccagcaaacaactgggc
 541  E  K  I  Y  T  L  A  G  E  K  F  N  I  N  S  S  K  Q  L  G
```

```
1681 cagattttatttgaaaagctaaaactccctacggttaaaaagaccccaaaaaaaaacggcc
 561  Q  I  L  F  E  K  L  K  L  P  T  V  K  K  T  P  K  K  T  A

1741 tattcaacggataacgaagtattagaggaactttctgcggtccacgaacttccgcgtctg
 581  Y  S  T  D  N  E  V  L  E  E  L  S  A  V  H  E  L  P  R  L

1801 atacttgagtatagaactctggctaaactcaaatctacttatgttgatgccctcccgaag
 601  I  L  E  Y  R  T  L  A  K  L  K  S  T  Y  V  D  A  L  P  K

1861 atggttaatcctgaaactggtcgtcttcatacttcctttaaccagacggttacggccact
 621  M  V  N  P  E  T  G  R  L  H  T  S  F  N  Q  T  V  T  A  T

1921 ggaagactttcaagcagtgaccctaatcttcaaaatattcctgtgcgtggtgaagagggg
 641  G  R  L  S  S  S  D  P  N  L  Q  N  I  P  V  R  G  E  E  G

1981 cttaagattcgccaggcctttgtgccggaggagatttttgctgccgattacactcagatc
 661  L  K  I  R  Q  A  F  V  P  E  E  I  F  A  A  D  Y  T  Q  I

2041 gatctgcgagttttagcccattactcgggagatgaaaccttgattaaggccttctggcag
 681  D  L  R  V  L  A  H  Y  S  G  D  E  T  L  I  K  A  F  W  Q

2101 ggggaagacattcaccggcgcacggctgcagaaatttttggtatcccgccagaagaagta
 701  G  E  D  I  H  R  R  T  A  A  E  I  F  G  I  P  P  E  E  V

2161 actcctgagatgcggcgtatggccaagactataaactttggcattgtttacggcatgagt
 721  T  P  E  M  R  R  M  A  K  T  I  N  F  G  I  V  Y  G  M  S

2221 ccttacggtctggcgaaagaactcaaaattggccgccgtgaggccaaggcctttattgag
 741  P  Y  G  L  A  K  E  L  K  I  G  R  R  E  A  K  A  F  I  E

2281 cgctattttgaacgctacccaggtgtgaaacgctatatggaacaaatcgtggctgaagcc
 761  R  Y  F  E  R  Y  P  G  V  K  R  Y  M  E  Q  I  V  A  E  A

2341 cgagaaaagggctacgtggagacccttttcggacgcaaaaggcctcttcctgacatcaat
 781  R  E  K  G  Y  V  E  T  L  F  G  R  K  R  P  L  P  D  I  N

2401 agccctaatcgtacggcgcgcgagtttgccgagcgcacggctataaacactcctattcag
 801  S  P  N  R  T  A  R  E  F  A  E  R  T  A  I  N  T  P  I  Q

2461 gggacagccgctgatattatcaagctcgccatgataaaaattcaccggatttttaaagaa
 821  G  T  A  A  D  I  I  K  L  A  M  I  K  I  H  R  I  F  K  E

2521 aaaggctttgggacaaggatgcttcttcaggtgcatgacgagcttatttttgaagcgcct
 841  K  G  F  G  T  R  M  L  L  Q  V  H  D  E  L  I  F  E  A  P

2581 gaaaaagagattgaagaaatccagccaattgtccgacaaatcatggaaggagtggttgaa
 861  E  K  E  I  E  E  I  Q  P  I  V  R  Q  I  M  E  G  V  V  E

2641 ttgaaggttcctctaaaagtaaacctggcaatagggaaaaattgggcagaggcaaaggca
 881  L  K  V  P  L  K  V  N  L  A  I  G  K  N  W  A  E  A  K  A

2701 taa     (SEQ ID NO:4)
 901  *      (SEQ ID NO:2).
```

Alternatively, the nucleic acid has the sequence shown below (5'-3'):

atggcgcagaaaagcttgtttcctaaaaaattaccatttaaagatgataaagaccccatcttcgttattgacgggagttcttttgt
ttaccgggcttactatgccataagagggcatctatcaaaccgcaaagggctcccaaccaaggcggtctttgggtttacccag
atgcttttaaagcttttgcgtgagatgaaccctgagtatgtggtggtgtgctttgacgccaaagggcctacttttcgccacgag
atgtacaaagaatacaaagccaaccgcccccccatgccagatgatctttccgtccagattccctatatcaaagaggtaacca
gggcctttggagtccctattcttgaaatagaaggctttgaagctgacgatctcatcgccgctattgccactcgtatggaaaga
ccaattgtcatcgttggtggagataaagatttgttcccccttatttcagagaaagttgtcatgtgggaccccatgaaagacgaa
ctgattgacgaaagctggataaagaaacgttttggcattgaacctaaaaagctccttgatgtaagggcccttgccggcgata
gcattgataacgtgccaggggttccgggtattggtgaaaaaacggccctaaggctcataaaagaatacggttcccttgaag
aagtccttaaccatgccgaagaaataaaacaaaagcgcttgcgtgaaaacctcatcaaacacgccggagacgcccttattt
ccaaaaaactggttgagcttgaggccaaagccccaatcccccttgagcctgatttttaccgcaaacggccattaaatgccct
aaaactaagggaactcttccttgagcttgaatttaaaaagctcttaaaggaacttccagctactaaaaccctttcgtatgacca
gtacgagctggttcttgacccggataaagtaaaagaaattgtagaaaaggccaaaggggccgaagtggtggctattgacct
tgaaagtgatacgaaagaccccatgcgtgggaaaatagtaggggtctcgctttgttttaacccgcccaaagcctattatttcc
cttttagacatgaaggccttgaggcccaaaagcagcttccctgggaggcctttactcatctggccagcctcattgaagaccc
ctcagttaaaaagataggccacaatatcaagtatgacttgattattcttgctcgctacggcgtaactttaaagggccttgaagg
ggataccatgctggcttcgtatctccttgatccaacacgtcgtacccacggccttgatgagctggccgaagaggtcctgggg
cataccatgatttttacaaggaagtgactaaagaactggccaaaggagagagctttgccagggtccctcttgaaaaggcaa
aagtttacgcctgtgaagacgcccacgttacctatctgctttatcaatatttctggcccaaactcaaagaggaaagcctctgga
aggtctttacggaaattgatcgacctttaatagaagttttggcccacatggaaatggtaggtattaagattgacaccgcctatct
tagaggactttcgcgagaaatggctgaaaagttaaaggagcttgaagaaaaaatttacaccctggctggtgaaaaatttaat
atcaattccagcaaacaactgggccagattttatttgaaaagctaaaactccctacggttaaaaagaccccaaaaaaaacgg
cctattcaacggataacgaagtattagaggaactttctgcggtccacgaacttccgcgtctgatacttgagtatagaactctgg
ctaaactcaaatctacttatgttgatgccctcccgaagatggttaatcctgaaactggtcgtcttcatacttcctttaaccagacg
gttacggccactggaagactttcaagcagtgaccctaatcttcaaaatattcctgtgcgtggtgaagaggggcttaagattcg
ccaggcctttgtgccggaggagatttttgctgccgattacactcagatcgatctgcgagttttagcccattactcgggagatga
aaccttgattaaggccttctggcaggggaagacattcaccggcgcacggctgcagaaatttttggtatcccgccagaaga
agtaactcctgagatgcggcgtatggccaagactataaactttggcattgtttacggcatgagtccttacggtctggcgaaag
aactcaaaattggccgccgtgaggccaaggcctttattgagcgctattttgaacgctacccaggtgtgaaacgctatatgga
acaaatcgtggctgaagcccgagaaaagggctacgtggagacccttttcggacgcaaaaggcctcttcctgacatcaata
gccctaatcgtacggcgcgcgagtttgccgagcgcacggctataaacactcctattcaggggacagccgctgatattatca
agctcgccatgataaaaattcaccggatttttaaagaaaaaggctttgggacaaggatgcttcttcaggtgcacgacgaactt
ctttttgaagtgcctgaaaaagagattgaagaaatccagccaattgtccgacaaatcatggaaggagtggttgaattgaaggt
tcctctaaaagtaaacctggcaatagggaaaaattgggcagaggcaaaggcataa (SEQ ID NO: 35).

The nucleotide of SEQ ID NO:35 encodes the *T. indicus* full length DNA polymerase I of SEQ ID NO:34 as follows:

```
    1 atggcgcagaaaagcttgtttcctaaaaaattaccatttaaagatgataaagaccccatc
    1  M  A  Q  K  S  L  F  P  K  K  L  P  F  K  D  D  K  D  P  I

   61 ttcgttattgacgggagttcttttgtttaccgggcttactatgccataagagggcatcta
   21  F  V  I  D  G  S  S  F  V  Y  R  A  Y  Y  A  I  R  G  H  L

  121 tcaaaccgcaaagggctcccaaccaaggcggtctttgggtttacccagatgcttttaaag
   41  S  N  R  K  G  L  P  T  K  A  V  F  G  F  T  Q  M  L  L  K

  181 cttttgcgtgagatgaaccctgagtatgtggtggtgtgctttgacgccaaagggcctact
   61  L  L  R  E  M  N  P  E  Y  V  V  V  C  F  D  A  K  G  P  T
```

```
241 tttcgccacgagatgtacaaagaatacaaagccaaccgccccccccatgccagatgatctt
 81  F  R  H  E  M  Y  K  E  Y  K  A  N  R  P  P  M  P  D  D  L

301 tccgtccagattccctatatcaaagaggtaaccagggcctttggagtccctattcttgaa
101  S  V  Q  I  P  Y  I  K  E  V  T  R  A  F  G  V  P  I  L  E

361 atagaaggctttgaagctgacgatctcatcgccgctattgccactcgtatggaaagacca
121  I  E  G  F  E  A  D  D  L  I  A  A  I  A  T  R  M  E  R  P

421 attgtcatcgttggtggagataaagatttgttccccttatttcagagaaagttgtcatg
141  I  V  I  V  G  G  D  K  D  L  F  P  L  I  S  E  K  V  V  M

481 tgggaccccatgaaagacgaactgattgacgaaagctggataaagaaacgttttggcatt
161  W  D  P  M  K  D  E  L  I  D  E  S  W  I  K  K  R  F  G  I

541 gaacctaaaaagctccttgatgtaagggcccttgccggcgatagcattgataacgtgcca
181  E  P  K  K  L  L  D  V  R  A  L  A  G  D  S  I  D  N  V  P

601 ggggttccgggtattggtgaaaaaacggccctaaggctcataaaagaatacggttccctt
201  G  V  P  G  I  G  E  K  T  A  L  R  L  I  K  E  Y  G  S  L

661 gaagaagtccttaaccatgccgaagaaataaaacaaaagcgcttgcgtgaaaacctcatc
221  E  E  V  L  N  H  A  E  E  I  K  Q  K  R  L  R  E  N  L  I

721 aaacacgccggagacgcccttatttccaaaaaactggttgagcttgaggccaaagcccca
241  K  H  A  G  D  A  L  I  S  K  K  L  V  E  L  E  A  K  A  P

781 atccccccttgagcctgatttttaccgcaaacggccattaaatgccctaaaactaagggaa
261  I  P  L  E  P  D  F  Y  R  K  R  P  L  N  A  L  K  L  R  E

841 ctcttccttgagcttgaatttaaaaagctcttaaaggaacttccagctactaaaacccctt
281  L  F  L  E  L  E  F  K  K  L  L  K  E  L  P  A  T  K  T  L

901 tcgtatgaccagtacgagctggttcttgacccggataaagtaaaagaaattgtagaaaag
301  S  Y  D  Q  Y  E  L  V  L  D  P  D  K  V  K  E  I  V  E  K

961 gccaaaggggccgaagtggtggctattgaccttgaaagtgatacgaaagaccccatgcgt
321  A  K  G  A  E  V  V  A  I  D  L  E  S  D  T  K  D  P  M  R

1021 gggaaaatagtaggggtctcgctttgttttaacccgcccaaagcctattatttccctttt
341  G  K  I  V  G  V  S  L  C  F  N  P  P  K  A  Y  Y  F  P  F

1081 agacatgaaggccttgaggcccaaaagcagcttccctgggaggcctttactcatctggcc
361  R  H  E  G  L  E  A  Q  K  Q  L  P  W  E  A  F  T  H  L  A

1141 agcctcattgaagacccctcagttaaaaagataggccacaatatcaagtatgacttgatt
381  S  L  I  E  D  P  S  V  K  K  I  G  H  N  I  K  Y  D  L  I

1201 attcttgctcgctacggcgtaactttaaagggccttgaaggggataccatgctggcttcg
401  I  L  A  R  Y  G  V  T  L  K  G  L  E  G  D  T  M  L  A  S

1261 tatctccttgatccaacacgtcgtacccacggccttgatgagctggccgaagaggtcctg
421  Y  L  L  D  P  T  R  R  T  H  G  L  D  E  L  A  E  E  V  L

1321 gggcataccatgatttttacaaggaagtgactaaagaactggccaaaggagagagctttt
441  G  H  T  M  I  F  Y  K  E  V  T  K  E  L  A  K  G  E  S  F
```

```
1381 gccagggtccctcttgaaaaggcaaaagtttacgcctgtgaagacgcccacgttacctat
 461  A  R  V  P  L  E  K  A  K  V  Y  A  C  E  D  A  H  V  T  Y

1441 ctgctttatcaatatttctggcccaaactcaaagaggaaagcctctggaaggtctttacg
 481  L  L  Y  Q  Y  F  W  P  K  L  K  E  E  S  L  W  K  V  F  T

1501 gaaattgatcgacctttaatagaagttttggcccacatggaaatggtaggtattaagatt
 501  E  I  D  R  P  L  I  E  V  L  A  H  M  E  M  V  G  I  K  I

1561 gacaccgcctatcttagaggactttcgcgagaaatggctgaaaagttaaaggagcttgaa
 521  D  T  A  Y  L  R  G  L  S  R  E  M  A  E  K  L  K  E  L  E

1621 gaaaaaatttacaccctggctggtgaaaaatttaatatcaattccagcaaacaactgggc
 541  E  K  I  Y  T  L  A  G  E  K  F  N  I  N  S  S  K  Q  L  G

1681 cagattttatttgaaaagctaaaactccctacggttaaaaagaccccaaaaaaaacggcc
 561  Q  I  L  F  E  K  L  K  L  P  T  V  K  K  T  P  K  K  T  A

1741 tattcaacggataacgaagtattagaggaactttctgcggtccacgaacttccgcgtctg
 581  Y  S  T  D  N  E  V  L  E  E  L  S  A  V  H  E  L  P  R  L

1801 atacttgagtatagaactctggctaaactcaaatctacttatgttgatgccctcccgaag
 601  I  L  E  Y  R  T  L  A  K  L  K  S  T  Y  V  D  A  L  P  K

1861 atggttaatcctgaaactggtcgtcttcatacttcctttaaccagacggttacggccact
 621  M  V  N  P  E  T  G  R  L  H  T  S  F  N  Q  T  V  T  A  T

1921 ggaagactttcaagcagtgaccctaatcttcaaaatattcctgtgcgtggtgaagagggg
 641  G  R  L  S  S  S  D  P  N  L  Q  N  I  P  V  R  G  E  E  G

1981 cttaagattcgccaggcctttgtgccggaggagatttttgctgccgattacactcagatc
 661  L  K  I  R  Q  A  F  V  P  E  E  I  F  A  A  D  Y  T  Q  I

2041 gatctgcgagttttagcccattactcgggagatgaaaccttgattaaggccttctggcag
 681  D  L  R  V  L  A  H  Y  S  G  D  E  T  L  I  K  A  F  W  Q

2101 ggggaagacattcaccggcgcacggctgcagaaatttttggtatcccgccagaagaagta
 701  G  E  D  I  H  R  R  T  A  A  E  I  F  G  I  P  P  E  E  V

2161 actcctgagatgcggcgtatggccaagactataaactttggcattgtttacggcatgagt
 721  T  P  E  M  R  R  M  A  K  T  I  N  F  G  I  V  Y  G  M  S

2221 ccttacggtctggcgaaagaactcaaaattggccgccgtgaggccaaggcctttattgag
 741  P  Y  G  L  A  K  E  L  K  I  G  R  R  E  A  K  A  F  I  E

2281 cgctattttgaacgctacccaggtgtgaaacgctatatggaacaaatcgtggctgaagcc
 761  R  Y  F  E  R  Y  P  G  V  K  R  Y  M  E  Q  I  V  A  E  A

2341 cgagaaaagggctacgtggagacccttttcggacgcaaaaggcctcttcctgacatcaat
 781  R  E  K  G  Y  V  E  T  L  F  G  R  K  R  P  L  P  D  I  N

2401 agccctaatcgtacggcgcgcgagtttgccgagcgcacggctataaacactcctattcag
 801  S  P  N  R  T  A  R  E  F  A  E  R  T  A  I  N  T  P  I  Q

2461 gggacagccgctgatattatcaagctcgccatgataaaaattcaccggatttttaaagaa
 821  G  T  A  A  D  I  I  K  L  A  M  I  K  I  H  R  I  F  K  E

2521 aaaggctttgggacaaggatgcttcttcaggtgcacgacgaacttctttttgaagtgcct
 841  K  G  F  G  T  R  M  L  L  Q  V  H  D  E  L  L  F  E  V  P
```

```
2581 gaaaaagagattgaagaaatccagccaattgtccgacaaatcatggaaggagtggttgaa
 861  E  K  E  I  E  E  I  Q  P  I  V  R  Q  I  M  E  G  V  V  E

2641 ttgaaggttcctctaaaagtaaacctggcaatagggaaaaattgggcagaggcaaaggca
 881  L  K  V  P  L  K  V  N  L  A  I  G  K  N  W  A  E  A  K  A

2701 taa  (SEQ ID NO:35)
 901  *   (SEQ ID NO:34).
```

Also encompassed by the invention are variants of the nucleic acids, as defined below.

Further provided is a vector comprising the isolated nucleic acid as described herein.

Additionally provided is a host cell transformed with the nucleic acid or the vector of the invention.

A recombinant polypeptide expression from the host cell is also encompassed by the invention.

In another aspect of the invention there is provided a kit comprising the polypeptide as described herein and/or the composition described herein and/or the isolated nucleic acid as described herein and/or the vector as described herein and/or the host cell as described herein, together with packaging materials therefor. The kit may, for example, comprise components including the polypeptide for carrying out a reaction requiring DNA polymerase activity, such as PCR or LAMP.

The invention further provides a method of amplifying a sequence of a target nucleic acid using a thermocycling reaction, for example PCR, comprising the steps of:

(1) contacting the target nucleic acid with the polypeptide having thermostable DNA polymerase activity as described herein; and

(2) incubating the target nucleic acid with the polypeptide under thermocycling reaction conditions which allow amplification of the target nucleic acid.

Another aspect of the invention encompasses a method of amplifying a sequence of a target nucleic acid using an isothermal reaction, for example LAMP, comprising the steps of:

(1) contacting the target nucleic acid with the polypeptide having thermostable DNA polymerase activity as described herein; and

(2) incubating the target nucleic acid with the polypeptide under isothermal reaction conditions which allow amplification of the target nucleic acid.

The present invention also encompasses structural variants of the polypeptides as defined herein. As used herein, a "variant" means a polypeptide in which the amino acid sequence differs from the base sequence from which it is derived in that one or more amino acids within the sequence are substituted for other amino acids. Amino acid substitutions may be regarded as "conservative" where an amino acid is replaced with a different amino acid with broadly similar properties. Non-conservative substitutions are where amino acids are replaced with amino acids of a different type.

By "conservative substitution" is meant the substitution of an amino acid by another amino acid of the same class, in which the classes are defined as follows:

Class	Amino acid examples
Nonpolar:	A, V, L, I, P, M, F, W
Uncharged polar:	G, S, T, C, Y, N, Q
Acidic:	D, E
Basic:	K, R, H.

As is well known to those skilled in the art, altering the primary structure of a peptide by a conservative substitution may not significantly alter the activity of that peptide because the side-chain of the amino acid which is inserted into the sequence may be able to form similar bonds and contacts as the side chain of the amino acid which has been substituted out. This is so even when the substitution is in a region which is critical in determining the peptides conformation.

Non-conservative substitutions are possible provided that these do not interrupt with the function of the DNA binding domain polypeptides.

Broadly speaking, fewer non-conservative substitutions will be possible without altering the biological activity of the polypeptides. Determination of the effect of any substitution (and, indeed, of any amino acid deletion or insertion) is wholly within the routine capabilities of the skilled person, who can readily determine whether a variant polypeptide retains the thermostable DNA polymerase activity according to the invention. For example, when determining whether a variant of the polypeptide falls within the scope of the invention, the skilled person will determine whether the variant retains enzyme activity (i.e., polymerase activity) at least 60%, preferably at least 70%, more preferably at least 80%, yet more preferably 90%, 95%, 96%, 97%, 98%, 99% or 100% of the non-variant polypeptide. Activity may be measured by, for example, any standard measure such as the number of bases of a template sequence which can be replicated in a given time period.

Suitably, variants may have a sequence which is at least 55% identical, 60% identical, 65% identical, for example at least 70% or 75% identical, such as at least 80%, 85%, 90%, 95%, 96%, 97%, 98% or even 99% identical to the sequence of any of SEQ ID NOs:1, 2, 32 or 34.

For example, the invention encompasses a polypeptide having thermostable DNA polymerase activity and comprising or consisting essentially of an amino acid sequence of SEQ ID NOs: 1, 2, 32 or 34 with up to about one third of the amino acid sequence from the N- or C-terminus having been deleted, or having at least 55% sequence identity to such a sequence. For example, up to about 300 amino acids may be removed from either the N- or C-terminus of SEQ ID NOs:2 or 34; up to about 205 amino acids may be removed from either the N- or C-terminus of SEQ ID NOs:1 or 32.

Using the standard genetic code, further nucleic acids encoding the polypeptides may readily be conceived and manufactured by the skilled person. The nucleic acid may be

DNA or RNA, and where it is a DNA molecule, it may for example comprise a cDNA or genomic DNA.

The invention encompasses variant nucleic acids encoding the polypeptides of the invention. The term "variant" in relation to a nucleic acid sequences means any substitution of, variation of, modification of, replacement of, deletion of, or addition of one or more nucleic acid(s) from or to a polynucleotide sequence providing the resultant polypeptide sequence encoded by the polynucleotide exhibits at least the same properties as the polypeptide encoded by the basic sequence. The term therefore includes allelic variants and also includes a polynucleotide which substantially hybridises to the polynucleotide sequence of the present invention. Such hybridisation may occur at or between low and high stringency conditions. In general terms, low stringency conditions can be defined a hybridisation in which the washing step takes place in a 0.330-0.825 M NaCl buffer solution at a temperature of about 40-48°C below the calculated or actual melting temperature (T_m) of the probe sequence (for example, about ambient laboratory temperature to about 55°C), while high stringency conditions involve a wash in a 0.0165-0.0330 M NaCl buffer solution at a temperature of about 5-10°C below the calculated or actual T_m of the probe(for example, about 65°C). The buffer solution may, for example, be SSC buffer (0.15M NaCl and 0.015M tri-sodium citrate), with the low stringency wash taking place in 3 x SSC buffer and the high stringency wash taking place in 0.1 x SSC buffer. Steps involved in hybridisation of nucleic acid sequences have been described for example in Sambrook et al. (1989; Molecular Cloning, Cold Spring Harbor Laboratory Press, Cold Spring Harbor).

Typically, variants have 55% or more of the nucleotides in common with the nucleic acid sequence of the present invention, more typically 60%, 65%, 70%, 80%, 85%, or even 90%, 95%, 98% or 99% or greater sequence identity.

Variant nucleic acids of the invention may be codon-optimised for expression in a particular host cell.

DNA polymerases and nucleic acids of the invention may be prepared synthetically using conventional synthesisers. Alternatively, they may be produced using recombinant DNA technology or isolated from natural sources followed by any chemical modification, if required. In these cases, a nucleic acid encoding the chimeric protein is incorporated into a suitable expression vector, which is then used to transform a suitable host cell, such as a prokaryotic cell such as *E. coli*. The transformed host cells are cultured and the protein isolated therefrom. Vectors, cells and methods of this type form further aspects of the present invention.

Sequence identity between nucleotide and amino acid sequences can be determined by comparing an alignment of the sequences. When an equivalent position in the compared sequences is occupied by the same amino acid or base, then the molecules are identical at that position. Scoring an alignment as a percentage of identity is a function of the number of identical amino acids or bases at positions shared by the compared sequences. When comparing sequences, optimal alignments may require gaps to be introduced into one or more of the sequences to take into consideration possible insertions and deletions in the sequences. Sequence comparison methods may employ gap penalties so that, for the same number of identical molecules in sequences being compared, a sequence alignment with as few gaps as possible, reflecting higher relatedness between the two compared sequences, will achieve a higher score than one with many gaps. Calculation of maximum percent identity involves the production of an optimal alignment, taking into consideration gap penalties.

In addition to the BLASTP program mentioned above, further suitable computer programs for carrying out sequence comparisons are widely available in the commercial and public sector. Examples include the MatGat program (Campanella et al., 2003, BMC Bioinformatics 4: 29), the Gap program (Needleman & Wunsch, 1970, J. Mol. Biol. 48: 443-453) and the FASTA program (Altschul et al., 1990, J. Mol. Biol. 215: 403-410). MatGAT v2.03 is freely available from the website http://bitincka.com/ledion/matgat/ (accessed on 11 February 2009) and has also been submitted for public distribution to the Indiana University Biology Archive (IUBIO Archive). Gap and FASTA are available as part of the Accelrys GCG Package

Version 11.1 (Accelrys, Cambridge, UK), formerly known as the GCG Wisconsin Package. The FASTA program can alternatively be accessed publically from the European Bioinformatics Institute (http://www.ebi.ac.uk/fasta) (accessed on 11 February 2009) and the University of Virginia (http://fasta.biotech.virginia.edu/fasta_www/cgi or http://fasta.bioch.virginia.edu/fasta_www2/fasta_list2.shtml as accessed on 11 February 2009). FASTA may be used to search a sequence database with a given sequence or to compare two given sequences (see http://fasta.bioch.virginia.edu/fasta_www/cgi/search_frm2.cgi, accessed on 11 February 2009). Typically, default parameters set by the computer programs should be used when comparing sequences. The default parameters may change depending on the type and length of sequences being compared. A sequence comparison using the MatGAT program may use default parameters of Scoring Matrix = Blosum50, First Gap = 16, Extending Gap = 4 for DNA, and Scoring Matrix = Blosum50, First Gap = 12, Extending Gap = 2 for protein. A comparison using the FASTA program may use default parameters of Ktup = 2, Scoring matrix = Blosum50, gap = -10 and ext = -2.

In one aspect of the invention, sequence identity is determined using the MatGAT program v2.03 using default parameters as noted above.

As used herein, a "DNA polymerase" refers to any enzyme that catalyzes polynucleotide synthesis by addition of nucleotide units to a nucleotide chain using a nucleic acid such as DNA as a template. The term includes any variants and recombinant functional derivatives of naturally occurring nucleic acid polymerases, whether derived by genetic modification or chemical modification or other methods known in the art.

As used herein, "thermostable" DNA polymerase activity means DNA polymerase activity which is relatively stable to heat and which functions at high temperatures, for example 45-100°C, preferably 55-100°C, 65-100°C, 75-100°C, 85-100°C or 95-100°C, as compared, for example, to a non-thermostable form of DNA polymerase.

Brief Description of Figures

Particular non-limiting embodiments of the present invention will now be described with reference to the following Figures, in which:

Figure 1 is a diagram illustrating a gene walking method employed in cloning a novel DNA polymerase from *Thermodesulfatator indicus* according to one embodiment of the invention;

Figure 2 is a diagram showing the structure of a new pET24a(+)HIS region used in cloning of the *T. indicus* DNA polymerase;

Figure 3 is an SDS PAGE gel showing expression of Large fragments of the cloned *T. indicus* DNA polymerase. Lane 1 is a size marker, lane 2 is induced control with pET24a(+)HIS vector without insert, lane 3 is 100µl *T. indicus* DNA polymerase, Large fragment with N-terminal HIS tag, lane 4 is 100µl *T. indicus* DNA polymerase, Large fragment without N-terminal HIS tag, lane 5 is 20µl *T. indicus* DNA polymerase, Large fragment with N-terminal HIS tag, lane 6 is 20µl *T. indicus* DNA polymerase, Large fragment without N-terminal HIS tag, lane 7 is 5µl *T. indicus* DNA polymerase, Large fragment with N-terminal HIS tag, lane 8 is 5µl *T. indicus* DNA polymerase, Large fragment without N-terminal HIS tag, lane 9 is 50u *T. indicus* DNA polymerase, Large fragment with N-terminal HIS tag purified via single step chelating sepharose purification, and lane 10 is 12.5u KlenTaq DNA polymerase. Volumes refer to amount of protein loaded from that volume of induced *E. coli* KRX culture;

Figure 4 is an SDS PAGE gel showing expression of full length embodiments of the cloned *T. indicus* DNA polymerase. Lane 1 is a size marker, lane 2 is induced Control with pET24a(+)HIS vector without insert, lane 3 is 100µl *T. indicus* DNA Polymerase, full length with N-terminal HIS tag, lane 4 is 100µl *T. indicus* DNA Polymerase, full length without N-terminal HIS tag, and lane 5 is 25u *Pfu* DNA Polymerase. Volumes refer to amount of protein loaded from that volume of induced *E. coli* KRX culture;

Figure 5 is an agarose gel of PCR reaction samples showing amplification of lambda (λ) DNA using the cloned *T. indicus* DNA polymerase. Lane 1 is a lambda *Eco*R I/*Hin*d III Size Marker, lane 2 is a 500bp, 400bp, 350bp, 275bp, 225bp and 175bp size marker, lane 3 shows amplification product using 1.25u *Taq* DNA polymerase, lane 4 shows 2µl induced *T. indicus* DNA polymerase, Large fragment without N-terminal HIS tag, lane 5 shows 2µl *T. indicus* DNA polymerase, Large fragment with N-terminal HIS tag, lane 6 shows 8µl *T. indicus* DNA polymerase, Large fragment with N-terminal HIS tag and purified via single step chelating sepharose purification, lane 7 shows 10µl *T. indicus* DNA polymerase, full length with N-terminal HIS tag, and lane 8 shows amplification product using induced pET24a(+)HIS vector lacking insert (as negative control). Volumes refer to amount of protein loaded from that volume of induced *E. coli* KRX culture; and

Figure 6 is an agarose gel of LAMP reaction samples showing amplification results using the cloned *T. indicus* DNA polymerase. Lane 1 is a lambda *Eco*R I/*Hin*d III Size Marker, lane 2 is a 500bp, 400bp, 350bp, 275bp, 225bp and 175bp size marker, lane 3 shows amplification product using 8u *Bst* DNA polymerase, Large fragment, lane 4 shows 2µl *T. indicus* DNA polymerase, Large fragment without N-terminal HIS tag, lane 5 shows 2µl *T. indicus* DNA polymerase, Large fragment with N-terminal HIS tag, lane 6 shows 8µl *T. indicus* DNA polymerase, Large fragment with N-terminal HIS tag and purified via single step chelating sepharose purification, lane 7 shows 10µl *T. indicus* DNA polymerase, full length with N-terminal HIS tag, and lane 8 shows amplification product using induced pET24a(+)HIS vector lacking insert (as negative control). Volumes refer to amount of protein loaded from that volume of induced *E. coli* KRX culture.

Examples

Agar-plated cultures of *Thermodesulfatator indicus* were obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (German Collection of Microorganisms and Cell Cultures; Accession No. DSM 15286). As described below, following extraction and amplification of gDNA from the cultures, a gene walking

method was used as outlined below to reach the predicted 5' start and the 3' stop of DNA polymerase A gene ("DNA polA", encoding DNA polymerase I). A Large (or Klenow) fragment of the DNA polymerase I was found to be highly efficient in both PCR and LAMP reactions.

Example 1 *Genomic DNA extraction*

The method for genomic DNA extraction from *T. indicus* cultures was derived from Gotz et al. (2002; Int. J. Syst. Evol. Microbiol. 52: 1349-1359) which is a modification of a method described in Ausubel et al. (1994; Current Protocols in Molecular Biology, Wiley, New York).

Cell pellets were resuspended in 567µl 1x TE buffer (10mM Tris/HCl, pH8.0; 1mM EDTA), 7.5% Chelex 100 (Sigma), 50mM EDTA (pH7.0), 1% (w/v) SDS and 200µg Proteinase K and incubated with slow rotation for 1h at 50°C. Chelex was removed by centrifugation. Then 100µl 5M NaCl and 80µl 10% (w/v) cetyltrimethylammonium bromide in 0.7M NaCl were added to the cell lysate and the sample incubated for 30 mins at 65°C. The DNA was extracted with phenol/chloroform, isopropanol precipitated and the DNA resuspended in water. DNA concentration was estimated on a 1% agarose gel.

Example 2 *Initial screening for DNA polA gene*

The screening method was derived from Shandilya et al. (2004, Extremophiles 8: 243-251).

Using degenerate polA primers PolATF1 and PolATR (see below), a ~570bp fragment was amplified from 10ng *T. indicus* gDNA.

The PolATF1 primer has the sequence:

5' – CATTTTTGCTGCCGATTAywsncarathga – 3' (SEQ ID NO:5); and

the PolATR primer has the sequence:

5' – AACCGCGAAGTTTTTATTyragyagyac – 3' (SEQ ID NO:6).

The PCR reaction mix was as follows:

10x PCR Buffer	10µl
(750mM Tris-HCl, pH8.8, 200mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween-20)	
5mM dNTP's	2µl
5' primer (10pM/µl)	2.5µl
3' primer (10pM/µl)	2.5µl
gDNA	10ng
Taq DNA Polymerase (5u/µl)	0.25µl
Water	To 50µl.

PCR cycling conditions were 4 minute initial denaturation at 94°C followed by 45 cycles of: 10 seconds denaturation at 94°C, 30 seconds annealing at 42°C, 30 second extension at 72°C. Final extension at 72°C for 7mins. 4°C hold.

A ~570bp amplified product was TA cloned (Invitrogen pCR2.1 kit. Cat#K2000-01) and sequenced using M13 Forward (5'-TGT AAA ACG ACG GCC AGT-3')(SEQ ID NO:7) and Reverse (5'-AGC GGA TAA CAA TTT CAC ACA GGA–3')(SEQ ID NO:8) primers on an ABI-3100 DNA sequencer. Sequencing data confirmed the fragment was DNA polymerase A (DNA polA) gene.

Example 3 *DNA polA gene walking*

From the amplification product obtained in Example 2, primers were designed to 'walk along' *T. indicus* gDNA to reach the 5' start and 3' stop of the DNA polA gene.

10ng gDNA was digested individually with 5u of various 6 base pair-cutter restriction endonucleases in 10µl reaction volume and incubated for 3 h at 37°C. 12 individual digest reactions were run, using a unique 6-cutter restriction enzyme (RE) for each. 5µl digested template was then self-ligated using 12.5u T4 DNA Ligase, 1µl 10x ligase buffer in 50µl reaction volume, with an overnight incubation at 16°C.

Self-ligated DNA was then used as template in two rounds of PCR. As illustrated in Figure 1, the first round of PCR employed primers 2 and 3 (see below), while a second round (nested-round) used primers 1 and 4 (see below) to give specificity to amplification.

First round PCR:

The first round PCR reaction mix was as follows:

Self-ligation reaction (~100pg/µl DNA)	2µl
10x PCR Buffer (200mM Tris-HCl, pH8.8, 100mM KCl, 100mM $(NH_4)_2SO_4$, 1% (v/v) Triton X-100, 20mM $MgSO_4$)	5µl
5mM dNTP's	2µl
Primer 2	25pM
Primer 3	25pM
Taq/Pfu (20:1) (5u/µl)	1.25u
Water	To 50µl.

Cycling conditions were 4 minute initial denaturation at 94°C followed by 35 cycles of: 10 seconds denaturation at 94°C, 10 seconds annealing at 55°C, 5 minute extension at 72°C. Final extension at 72°C for 7mins. 4°C hold.

Primer 2 [15286_2_(pos.2085)] has the sequence:
5'– AATCAAGGTTTCATCTCCCG-3' (SEQ ID NO:9); and

Primer 3 [15286_3 _(pos.2453)] has the sequence:
5'– TATTCAGGGGACAGCCGCTG-3' (SEQ ID NO:10).

Second round (nested) PCR:

The second round PCR reaction mix was as follows:

First round PCR reaction	1µl
10x PCR Buffer	5µl
(200mM Tris-HCl, pH8.8, 100mM KCl, 100mM $(NH_4)_2SO_4$, 1% (v/v) Triton X-100, 20mM $MgSO_4$)	
5mM dNTP's	2µl
Primer 1	25pM
Primer 4	25pM
Taq/Pfu (20:1) (5u/µl)	1.25u
Water	To 50µl.

Cycling conditions were 4 minute initial denaturation at 94°C followed by 25 cycles of: 10 seconds denaturation at 94°C, 10 seconds annealing at 55°C, 5 minute extension at 72°C. Final extension at 72°C for 7 minutes. 4°C hold.

Primer 1 [15286_1 _(pos.2063)] has the sequence:
5' – TAATGGGCTAAAACTCGCAG - 3' (SEQ ID NO:11); and

Primer 4 [15286_4 _(pos.2521)] has the sequence:
5' – AAGGCTTTGGGACAAGGATG - 3' (SEQ ID NO:12).

Amplified PCR fragments were ExoSAP treated and sequenced using the nested primers to reveal further DNA polA sequence data from which new gene walking primers could be designed. Two further separate steps of gene walking were required to generate fragments reaching the start and end of the *T. indicus* DNA polA gene.

Further gene walking step 1:

1st round PCR using Primers 2 and 3 (as above), followed by nested PCR using Primers 1 and 4 (as above).

PCR fragments between ~1.5kb and ~2.5kb were obtained from *Hin*d III, *Kpn* I, and *Eco*R V digested/self-ligated reaction templates.

These fragments were sequenced using the nested primers (Primers 1 and 4). Sequencing of fragments reached the C-terminal STOP codon for DNA polA and gave a further ~1100bp of sequence data towards the N-terminal. New gene walking primers were designed to walk towards the N-terminal.

Further gene walking step 2:

1st round PCR using Primers 6 and 7 (see below), followed by nested PCR using Primers 5 and 8 (see below).

Primer 5 [15286_5_(pos.1036)] has the sequence:
5' – TCT CGC TTT GTT TTA ACC C - 3' (SEQ ID NO:13);

Primer 6 [15286_6_(pos.1013)] has the sequence:
5' – CAT GCG TGG GAA AAT AGT A - 3' (SEQ ID NO:14);

Primer 7 [15286_7 _(pos.1008)] has the sequence:
5' – ACT TTA TCC GGG TCA AGA AC - 3' (SEQ ID NO:15); and

Primer 8 [15286_8 _(pos.941)] has the sequence:
5' –TTT CGT ATC ACT TTC AAG GTC - 3' (SEQ ID NO:16).

PCR fragments between ~750bp and 2kb were obtained from *Hin*d III, *Pst* I, and *Kpn* I digested/self-ligated reaction templates.

These fragments were sequenced using the nested primers (Primers 5 and 8). This sequence data showed the fragments reached the N-terminal ATG start codon for DNA polA.

Example 4 *Amplification of full length ("FL") and Large (Klenow) fragment ("LF") DNA polA*

Based on the sequence data derived from the gene walking protocol described in Example 3, a start and stop for the Large (Klenow) fragment could be predicted (based on alignment with known DNA polA sequences, for example the *Taq* KlenTaq fragment), allowing specific primers to be designed to amplify the entire Large fragment gene (~1.7kb).

These specific primers were:

15286_FL_Upper(NdeI)

5'-GTC CAC CAT ATG GCG CAG AAA AGC TTG TTT CCT AAA AAA TTA CCA TTT AAA GAT GA –3' (SEQ ID NO:17);

15286_LF_Upper(NdeI)

5'-CTT GAA CAT ATG GGC CTC TTA AAA GAA CTT CCA GCT AC - 3' (SEQ ID NO:18); and

15286_Lower(SalI)

5' – AGC CCT GTC GAC GGA TCC GCC AGC TTA TGC CTT TGC CTC TGC - 3' (SEQ ID NO:19).

Restriction sites (underlined in the above primer sequences) for *Nde*I or *Sal*I, as noted above, were built into the primers to facilitate cloning into expression vectors.

Gene products were amplified using a high fidelity Phusion DNA polymerase (New England Biolabs).

The PCR reaction mix was as follows:

5x HF Phusion reaction Buffer	20µl
5mM dNTP's	4µl
Upper primer (FL or LF)	25pM
Lower primer	25pM
gDNA	10ng
Phusion DNA Polymerase (2u/µl)	0.5µl
Water	To 100µl.

Cycling conditions were 30 seconds initial denaturation at 98°C followed by 25 cycles of: 3 seconds denaturation at 98°C, 10 seconds annealing at 55°C, 1.5 minute extension at 72°C. Final extension at 72°C for 7 mins. 4°C hold.

Example 5 *pET24a(+)HIS vector construction*

The pET24a(+) vector (Novagen) was modified to add a 6x HIS tag upstream of NdeI site (see Figure 2). The HIS tag was inserted between XbaI and BamHI sites as follows.

An overlapping primer pair, of which an upper primer (*XbaI*) has the sequence:
5' – TTC CCC TCT AGA AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT ATA CTA TG CAC CA – 3' (SEQ ID NO:20), and

a lower primer (*BamHI*) has the sequence:
5' – GAA TTC GGA TCC GCT AGC CAT ATG GTG ATG GTG ATG GTG CAT AGT ATA TCT CCT T – 3' (SEQ ID NO:21).

were amplified by PCR, RE digested and ligated into pET24a(+). The ligation reaction was transformed into *E. coli* TOP10F' (Invitrogen) and plated on Luria Broth plates plus kanamycin. Colonies were screened by PCR and verified by sequencing using T7 sequencing primers:
T7_Promoter: 5'-AAATTAATACGACTCACTATAGGG–3' (SEQ ID NO:22),
T7_Terminator: 5'-GCTAGTTATTGCTCAGCGG–3' (SEQ ID NO:23).

Example 6 *Cloning of full length and Large fragment DNA polA*

PCR products from Example 4 were purified using Promega Wizard purification kit and then RE digested using *Nde* I/*Sal* I. DNA was phenol/chloroform extracted, ethanol-precipitated and resuspended in water. The full length ("FL") and Large fragment ("LF") sequences were then each ligated into pET24a(+) and pET24a(+)HIS, between *Nde* I and *Sal* I, and electroporated into KRX cells (Promega). Colonies were screened by PCR using vector-specific T7 primers.

Example 7 *Expression of full length and Large fragment DNA polymerases*

Recombinant colonies from Example 6 were grown up overnight in 5 ml Luria Broth (including Kanamycin/Chloramphenicol). 50 ml Terrific Broth baffled shake flasks were inoculated by 1/100 dilution of overnight culture. Cultures were grown at 37°C, 275 rpm to OD_{600}~1 then brought down to 24°C and induced with L-rhamnose to 0.1% final concentration, and IPTG to 10mM final concentration. Cultures were incubated for a further 18h at 24°C, 275rpm. 10ml of the culture was then harvested by centrifugation for 10mins at 5,000xg and cells were resuspended in 1ml Lysis buffer (50mM Tris-HCl, pH8.0, 100mM NaCl, 1mM EDTA) and sonicated for 2 bursts of 30s (40v) on ice. Samples were centrifuged at 5,000xg for 5min and heat lysed at 70°C for 20min to denature background *E.coli* proteins. Samples were centrifuged and aliquots of supernatant were size fractionated on 8% SDS-PAGE.

As shown in Figure 3, *T. indicus* Large fragment DNA polymerase I was expressed at the predicted ~70kDa.

Figure 4 shows that *T. indicus* full length DNA polymerase I was expressed at the predicted ~103kDa.

DNA polymerases are known to sometimes run slightly faster than expected on SDS PAGE gels, so that their apparent molecular weight is smaller than predicted.

Example 8 *PCR activity assay*

PCR activity of the samples obtained in Example 7 were tested in a 500bp λDNA PCR assay. *Taq* DNA polymerase (1.25 u) was used as positive control.

The PCR solution contained:

10x PCR Buffer	5µl
(750mM Tris-HCl, pH8.8, 200mM $(NH_4)_2SO_4$, 0.1% (v/v) Tween-20)	
5mM dNTP mix	2µl
Enzyme test sample	1µl
Upper λ primer	25pM
Lower λ primer	25pM
λDNA	1ng
Water	To 50µl.

The Upper λ primer has the sequence:
5'-GATGAGTTCGTGTCCGTACAACTGG-3' (SEQ ID NO:24),
while the Lower primer has the sequence:
5'-GGTTATCGAAATCAGCCACAGCGCC-3' (SEQ ID NO:25).

PCR proceeded with 35 cycles of: 3 seconds denaturation at 94°C, 10 seconds annealing at 55°C, 30 seconds extension at 72°C. Final extension at 72°C for 7mins. 4°C hold.

An aliquot of the reaction products were run out on a 1.5% agarose gel, and the results are shown in Fig. 5. Under the PCR conditions used, the *T. indicus* Large fragment, both with and without an N-terminal HIS tag, showed comparable PCR activity to *Taq* DNA polymerase (lane 3), while the *T. indicus* full length DNA polymerase did not yield detectable PCR product (lane 7). Under the PCR assay conditions used here, *Bst* DNA polymerase did not yield any detectable PCR product (data not shown).

Example 9 *LAMP activity assay*

Samples obtained in Example 7 were also tested for loop-mediated isothermal amplification (LAMP) activity.

LAMP primers (see Nagamine et al., 2002) used were:

Lambda-FIP-LAMP ("FIP")

5'-CAGCCAGCCGCAGCACGTTCGCTCATAGGAGATATGGTAGAGCCGC-3' (SEQ ID NO:26);

Lambda-BIP-LAMP ("BIP")

5'-GAGAGAATTTGTACCACCTCCCACCGGGCACATAGCAGTCCTAGGGA CAGT - 3' (SEQ ID NO:27);

Lambda-F3-LAMP ("F3")

5' – GGCTTGGCTCTGCTAACACGTT - 3' (SEQ ID NO:28);

Lambda-B3-LAMP ("B3")

5'-GGACGTTTGTAATGTCCGCTCC - 3' (SEQ ID NO:29);

Lambda-loopF-LAMP ("loopF")

5'-CTGCATACGACGTGTCT - 3' (SEQ ID NO:30); and

Lambda-loopB-LAMP ("loopB")

5' – ACCATCTATGACTGTACGCC - 3' (SEQ ID NO:31).

LAMP was performed in a total 25µl reaction mixture containing 0.8µM each of FIP and BIP, 0.2µM each of F3 and B3, 0.4µM each of loopF and loopB primers, 1.6mM dNTPs, 1M betaine (Sigma), 2mM $MgSO_4$, 1x Bst buffer (New England Biolabs), 1ng λDNA, and either 8u *Bst* DNA polymerase large fragment (New England Biolabs; positive control) or 1µl test sample (from Example 7), made up to volume with water. The mixture was incubated at 65°C for 1h and an aliquot run out on 1% agarose gel stained with ethidium bromide for detection of amplification.

Results of the LAMP assay are shown in Figure 6. Under the LAMP conditions used, the *T. indicus* Large fragment, both with and without an N-terminal HIS tag, showed comparable PCR activity to *Bst* DNA polymerase Large fragment (lane 3), while the *T. indicus* full length DNA polymerase did not yield detectable LAMP product (lane 7). It is possible that under these LAMP conditions, the full length DNA polymerase has 5'→3' exonuclease activity which destroys any LAMP amplification product. Under the LAMP assay conditions used here, *Taq* DNA polymerase did not yield any detectable LAMP product (data not shown).

Example 10 *Thermostability assay*

Thermostability of the *T. indicus* Large fragment was tested using the 500bp λDNA PCR assay as described above in Example 7. Samples of the induced Large fragment were incubated at 95°C for 0, 2, 4, 6, 8, 10, 15 or 20 min, then used in the 500bp λDNA PCR assay. Under the conditions used, the Large fragment was found to be unaffected by up to 4 min incubation at 95°C, showed reduced PCR activity after 6 min incubation, and was unable to produce detectable PCR product after 8 min incubation (data not shown).

This example demonstrates that the *T. indicus* Large fragment was thermostable for a sufficient duration to be effective in PCR but that prolonged incubation at a denaturation temperature of 95°C affected DNA polymerase activity.

Although the present invention has been described with reference to preferred or exemplary embodiments, those skilled in the art will recognise that various modifications and variations to the same can be accomplished without departing from the spirit and scope of the present invention and that such modifications are clearly contemplated herein. No limitation with respect to the specific embodiments disclosed herein and set forth in the appended claims is intended nor should any be inferred.

All documents cited herein are incorporated by reference in their entirety.

Claims

1. A polypeptide having thermostable DNA polymerase activity and comprising or consisting of an amino acid sequence with at least 55% identity to *Thermodesulfatator indicus* DNA polymerase I Large fragment shown in SEQ ID NO: 1.

2. A polypeptide according to claim 1 which exhibits strand displacement activity.

3. A polypeptide according to either of claim 1 or 2, which is suitable for carrying out an isothermal amplification reaction, such as loop-mediated isothermal amplification (LAMP).

4. A polypeptide according to any preceding claim, which is suitable for carrying out a thermocycling amplification reaction, such as a polymerase chain reaction (PCR).

5. A polypeptide according to any preceding claim wherein the amino acid sequence is SEQ ID NO:32 or an amino acid sequence having at least 55% identity to SEQ ID NO:32.

6. An isolated polypeptide having thermostable DNA polymerase activity and comprising or consisting of an amino acid sequence with at least 55% identity to *T. indicus* DNA polymerase I as shown in SEQ ID NO: 2.

7. A polypeptide according to claim 6 wherein the amino acid sequence is SEQ ID NO:34 or an amino acid sequence having at least 55% identity to SEQ ID NO:34.

8. A polypeptide according to claim 6 or 7, in which the polypeptide is an isolated thermostable DNA polymerase I obtainable from *T. indicus* and having a molecular weight of about 103,000 Daltons, or an enzymatically active fragment thereof.

9. A polypeptide having thermostable DNA polymerase activity and comprising amino acid sequence SEQ ID NO:32.

10. A polypeptide having thermostable DNA polymerase activity and consisting essentially of amino acid sequence SEQ ID NO:32.

11. A polypeptide having thermostable DNA polymerase activity and comprising amino acid sequence SEQ ID NO:34.

12. A polypeptide having thermostable DNA polymerase activity and consisting essentially of amino acid sequence SEQ ID NO:34.

13. A polypeptide according to any of claims 1-9 or 11, further comprising a Cren7 enhancer domain or variant thereof.

14. A composition comprising the polypeptide of any of claims 1-13.

15. An isolated nucleic acid encoding the polypeptide of any of claims 1-4.

16. The nucleic acid according to claim 15, having a nucleotide sequence shown in SEQ ID NO: 3.

17. An isolated nucleic acid encoding the polypeptide according to any of claims 1-5, 9 or 10.

18. The nucleic acid according to claim 17, having a nucleotide sequence shown in SEQ ID NO: 33.

19. An isolated nucleic acid encoding the polypeptide of claim 6.

20. The nucleic acid according to claim 19, having a nucleotide sequence shown in SEQ ID NO: 4.

21. An isolated nucleic acid encoding the polypeptide of claim 6, 7, 11 or 12.

22. The nucleic acid according to claim 21, having a nucleotide sequence shown in SEQ ID NO: 35.

23. An isolated nucleic acid sequence encoding the polypeptide according to claim 13.

24. A vector comprising the isolated nucleic acid of any of claims 15-23.

25. A host cell transformed with the nucleic acid of any of claims 15-23 or the vector of claim 14.

26. A kit comprising the polypeptide of any of claims 1-13, and/or the composition of claim 14, and/or the isolated nucleic acid of any of claims 15-23, and/or the vector of claim 24, and/or the host cell of claim 25, together with packaging materials therefor.

27. A method of amplifying a sequence of a target nucleic acid using a thermocycling reaction, comprising the steps of:

(1) contacting the target nucleic acid with the polypeptide of any of claims 1-5 and/or claims 6-8 and/or any of claims 9-12 and/or claim 13; and

(2) incubating the target nucleic acid with the polypeptide under thermocycling reaction conditions which allow amplification of the target nucleic acid.

28. A method of amplifying a sequence of a target nucleic acid using a thermocycling reaction, comprising the steps of:

(1) contacting the target nucleic acid with the polypeptide of any of claims 9-12 and/or claim 13; and

(2) incubating the target nucleic acid with the polypeptide under thermocycling reaction conditions which allow amplification of the target nucleic acid

29. The method according to claim 27 or 28, in which the thermocycling reaction is PCR.

30. A method of amplifying a sequence of a target nucleic acid using an isothermal reaction, comprising the steps of:

(1) contacting the target nucleic acid with the polypeptide of any of claims 1-5 and/or claims 6-8 and/or any of claims 9-12 and/or claim 13; and

(2) incubating the target nucleic acid with the polypeptide under isothermal reaction conditions which allow amplification of the target nucleic acid.

31. A method of amplifying a sequence of a target nucleic acid using an isothermal reaction, comprising the steps of:

(1) contacting the target nucleic acid with the polypeptide of any of claims 9-12 and/or claim 13; and

(2) incubating the target nucleic acid with the polypeptide under isothermal reaction conditions which allow amplification of the target nucleic acid.

32. The method according to claim 30 or 31, in which the isothermal reaction is LAMP.

33. A DNA polymerase I or an enzymatically active fragment thereof substantially as herein described with reference to the accompanying drawings.

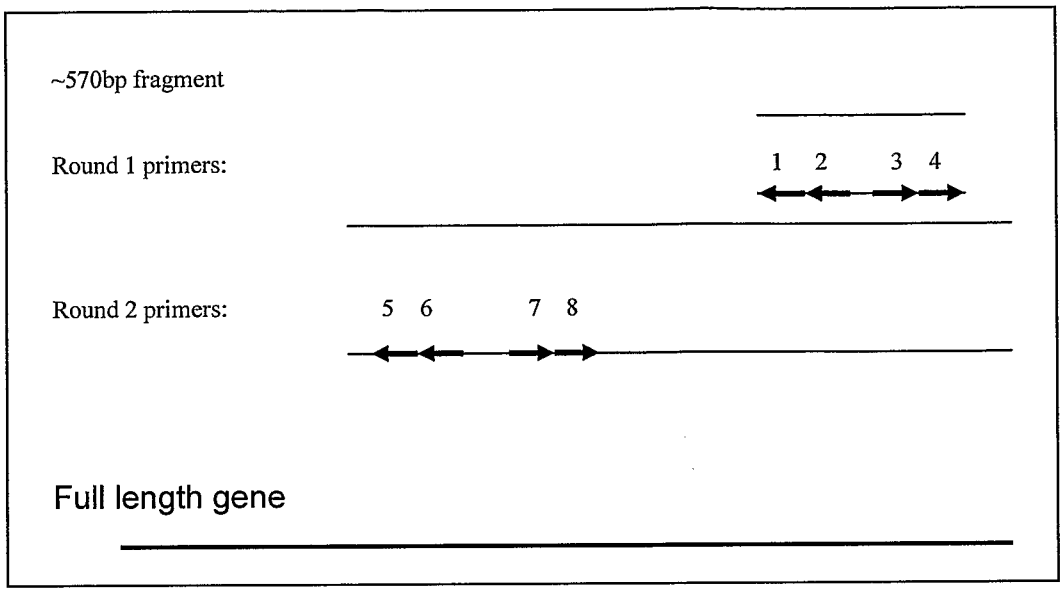

Figure 1

New pET24a(+)HIS region:

Xba I *Nde* I

..**tctaga**ataattttgtttaactttaagaaggagatatact**ATGcaccatcaccatcaccat**atggcta..
(SEQ ID NO:36)

M H H H H H H M
(SEQ ID NO:37)

Figure 2

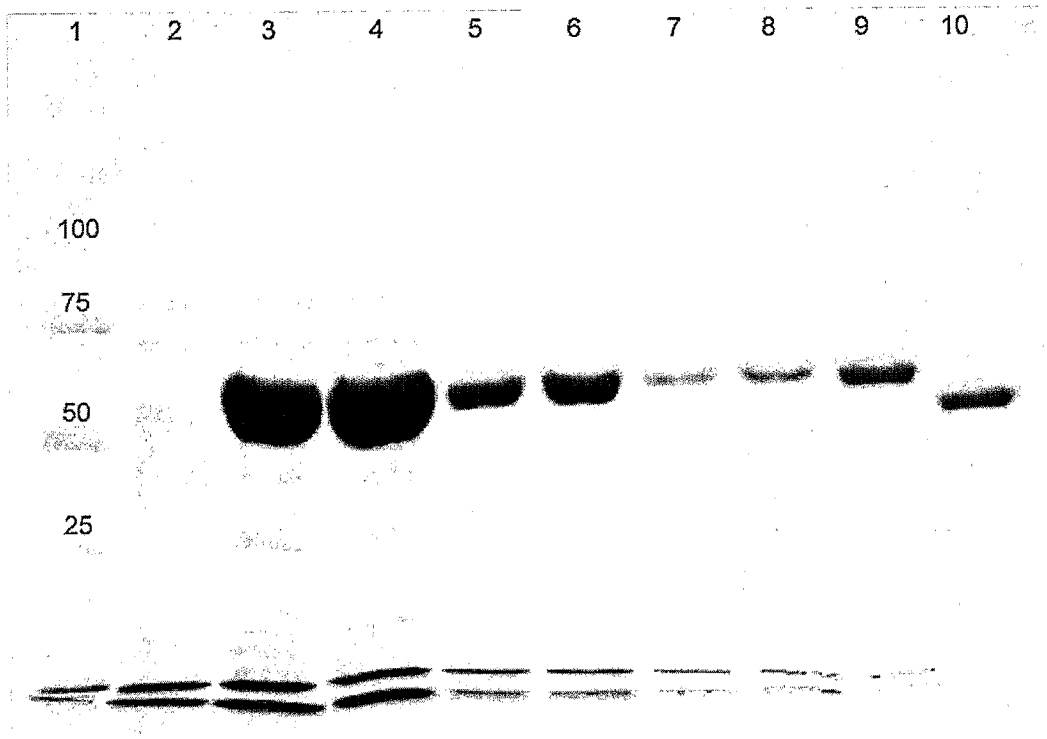

Figure 3

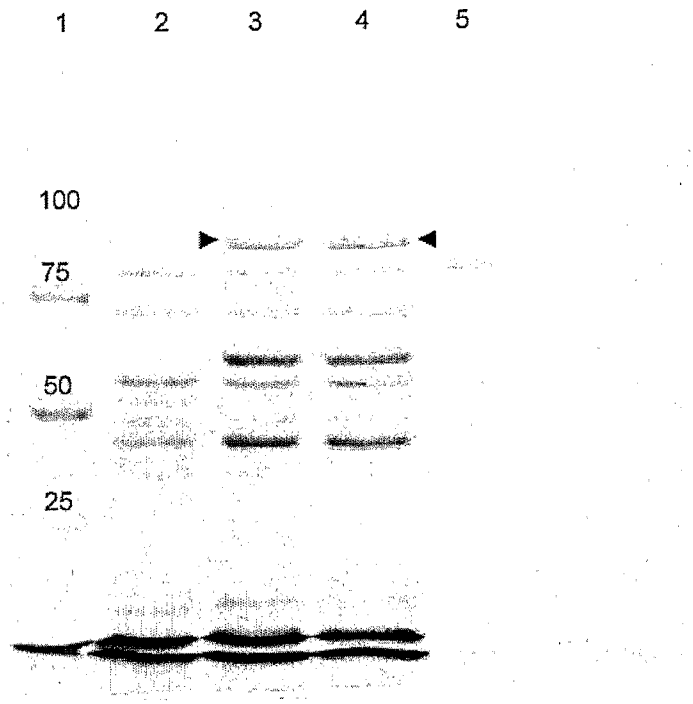

Figure 4

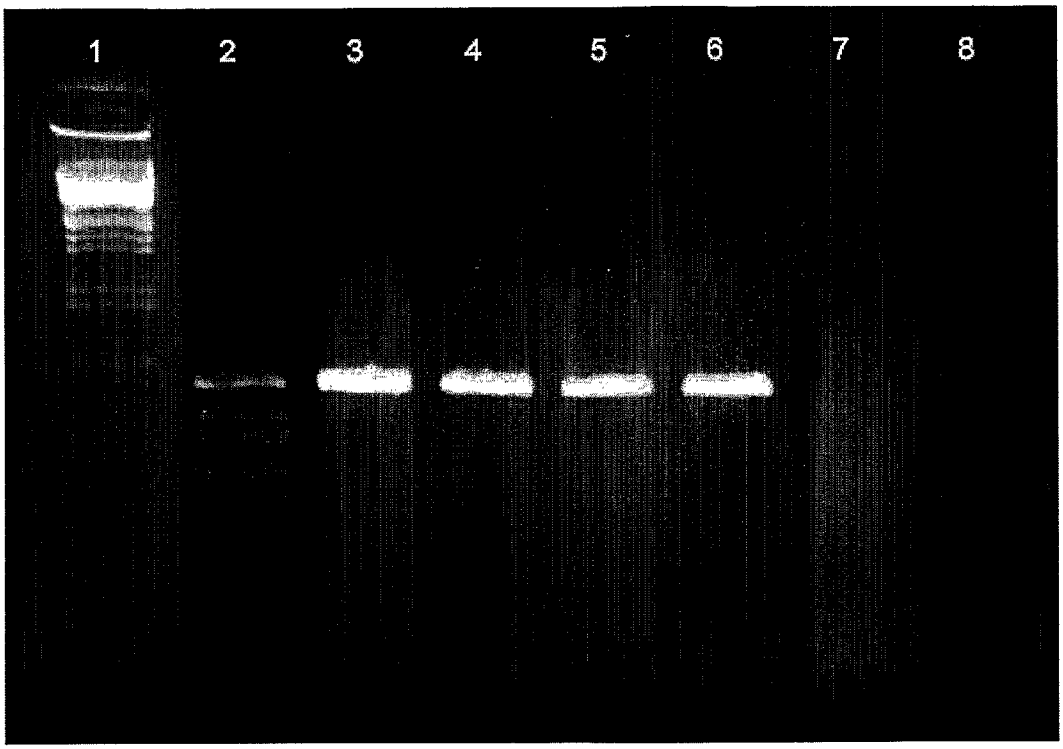

Figure 5

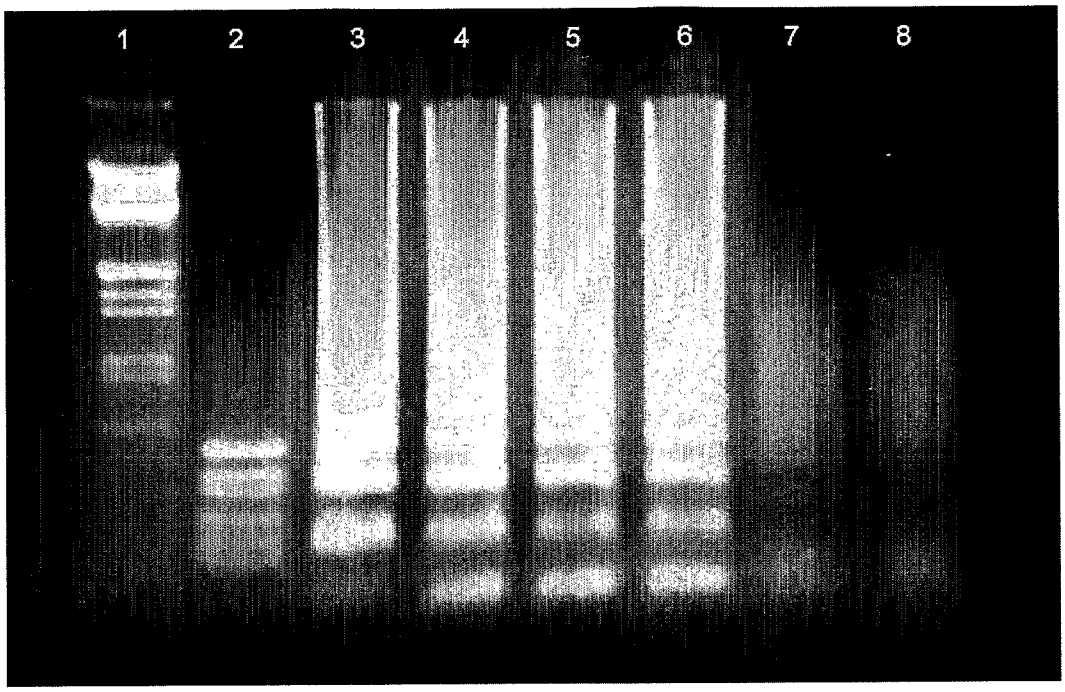

Figure 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2009/000411

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07K14/195 C12N9/12

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07K C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, BIOSIS, Sequence Search, FSTA, WPI Data, EMBASE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2006/030455 A (PROKARIA EHF [IS]; HJORLEIFSDOTTIR SIGRIDUR [IS]; ERNSTSON SVEINN [IS]) 23 March 2006 (2006-03-23) claims 1-8; table 1	
A	MOUSSARD H ET AL: "Thermodesulfatator indicus gen. nov., sp. nov., a novel thermophilic chemolithoautotrophic sulfate-reducing bacterium isolated from the Central Indian Ridge." INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY, vol. 54, no. 1, January 2004 (2004-01), pages 227-233, XP002526275 ISSN: 1466-5026 cited in the application figure 3	

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
4 May 2009	05/08/2009

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Deleu, Laurent

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2009/000411

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 03/048308 A (APPLERA CORP [US]) 12 June 2003 (2003-06-12) page 1 - page 2 page 50 - page 53; example 1	
A	SOUTHWORTH M W ET AL: "CLONING OF THERMOSTABLE DNA POLYMERASES FROM HYPERTHERMOPHILIC MARINE ARCHEA WITH EMPHASIS ON THERMOCOCCUS SP. 9 N-7 AND MUTATIONS AFFECTING 3'-5' EXONUCLEASE ACTIVITY" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE, WASHINGTON, DC.; US, vol. 93, no. 11, 28 May 1996 (1996-05-28), pages 5281-5285, XP000652319 ISSN: 0027-8424 page 5283 - page 5285	
A	GUO LI ET AL: "Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea." NUCLEIC ACIDS RESEARCH MAR 2008, vol. 36, no. 4, March 2008 (2008-03), pages 1129-1137, XP002526277 ISSN: 1362-4962 Published online 20 december 2007. figure 2	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/GB2009/000411

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

see additional sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers allsearchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search reportcovers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

se additional sheet(s)

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-5, 9, 10, 15-18 (complete) and 13, 14, 23-33 (all partially)

A polypeptide having thermostable DNA polymerase activity and comprising or consisting of an amino acid sequence with at least 55% identity to SEQ ID NO: 1 or SEQ ID NO: 32. Said polypeptide which exhibits strand displacement activity. A nucleic acid encoding said polypeptide. A nucleic acid having the nucleotide sequence of SEQ ID NO: 3 or SEQ ID NO: 33. A method of amplifying a sequence of a target nucleic acid using a thermocycling reaction comprising the step of contacting the target nucleic acid with said polypeptide.

---

2. claims: 6-8, 11, 12, 19-22 (complete) and 13, 14, 23-33 (all partially)

A polypeptide having thermostable DNA polymerase activity and comprising or consisting of an amino acid sequence with at least 55% identity to SEQ ID NO: 2 or SEQ ID NO: 34. A nucleic acid encoding said polypeptide. A nucleic acid having the nucleotide sequence of SEQ ID NO: 4 or SEQ ID NO: 35. A method of amplifying a sequence of a target nucleic acid using a thermocycling reaction comprising the step of contacting the target nucleic acid with said polypeptide.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2009/000411

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2006030455 A	23-03-2006	NONE	
WO 03048308 A	12-06-2003	AU 2002346498 A1	17-06-2003