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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 December 2012 (20.12.2012)

WIPO | PCT

(10) International Publication Number
WO 2012/173905 A1

(51) International Patent Classification:
*C12N 9/12* (2006.01) *C12Q 1/68* (2006.01)
*C12N 15/54* (2006.01) *C12N 15/63* (2006.01)
*C12N 15/09* (2006.01) *C07H 21/00* (2006.01)
*C12N 15/10* (2006.01)

(21) International Application Number:
PCT/US2012/041802

(22) International Filing Date:
10 June 2012 (10.06.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/497,524 16 June 2011 (16.06.2011) US

(71) Applicant *(for all designated States except US)*: **THE REGENTS OF THE UNVERSITY OF CALIFORNIA** [US/US]; Office Of Technology Transfer, 1111 Franklin Street, 5th Floor, Oakland, CA 94607 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **BERNICK, David** [US/US]; University California At Santa Cruz, 1156 High Street, CA 95064 (US). **HOLMES, Andrew** [US/US]; University California At Santa Cruz, 1156 High Street, CA 95064 (US). **NIVALA, Jeffrey** [US/US]; University California At Santa Cruz, 1156 High Street, CA 95064 (US).

(74) Agent: **BELL & ASSOCIATES**; 58 West Portal Ave #121, San Francisco, CA 94127 (US).

(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, CL, 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, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, 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, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, 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, RS, SE, SI, SK, SM, 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))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title**: SALT-TOLERANT DNA POLYMERASES

Fig. 1


(57) **Abstract**: Four novel sequences of type B DNA polymerases and variants and analogues thereof useful for applications involving DNA polymerization in high salt conditions.

WO 2012/173905 A1

**IN THE PATENT COOPERATION TREATY**

**Title: Salt-tolerant DNA polymerases**

**Inventors:**
David L. Bernick (US)
Andrew Holmes (US)
Jeffrey M. Nivala (US)

**Address for inventors:** UCSC, 1156 High Street, Santa Cruz, CA 95064, USA

**Assignee**: The Regents of the University of California
Office of Technology Transfer
1111 Franklin Street, 5th floor
Oakland, California 94607 USA

**Attorneys for the applicant**:
BELL & ASSOCIATES
58 West Portal Avenue # 121,
San Francisco, California 94127
info@bell-iplaw.com
PTO Customer No. 039843

**Government sponsorship**
This invention was made with support from a UCSC/SOE start-up grant No. 19900-4041. The government has certain rights in the invention.

**Relation to other applications:**
This PCT application takes priority to and claims the benefit of US provisional application No. 61/497,524 filed 16-JUN-2011 and titled Salt-tolerant DNA polymerases.

[001] **Field of the invention**

[002] DNA polymerases that function at high salt concentrations.

[003] **Background**

[004] DNA polymerase enzymes are naturally-occurring intracellular enzymes, and are used by a cell to replicate a nucleic acid strand using a template molecule to manufacture a complementary nucleic acid strand. The in-vitro use of enzymes having DNA polymerase activity has in recent years become more common in a variety of biochemical applications including cDNA synthesis and DNA sequencing reactions (see Sambrook et al., (2nd ed. Cold Spring Harbor Laboratory Press, 1989) hereby incorporated by reference herein), and amplification of nucleic acids by methods such as the polymerase chain reaction (PCR) (Mullis et al., U.S. Pat. Nos. 4,683,195, 4,683,202, and 4,800,159, hereby incorporated by reference herein) and RNA transcription-mediated amplification methods (e. g., Kacian et al., PCT Publication No. WO91/01384 which enjoys common ownership with the present application and is hereby incorporated by reference herein). Methods such as PCR make use of cycles of primer extension through the use of a DNA polymerase activity, followed by thermal denaturation of the resulting double-stranded nucleic acid in order to provide a new template for another round of primer annealing and extension.

[005] Various scientific and industrial applications exist in which it would be advantageous to use a DNA polymerase that function efficiently at high salt concentrations. In sequencing, GC compressions can be resolved by using high salt concentrations. In nanopore sequencing high salt concentration boosts the signal to noise ratio for ionic-current-based nanopore measurements. Salt tolerant DNA polymerases may be found among members of the extreme halophiles, in which salt tolerance is achieved not by exclusion of monovalent ions from the cytosol, but by adapting intracellular machinery function in elevated salt. As an example of salt tolerance among members of the extreme halophiles, malate dehydrogenase from the archaeal halophile Haloarcula marismortui incorporates a salt-adaptive strategy where the high ionic concentration from the environment is not only tolerated but is incorporated within the protein. Sodium and chloride ions are found incorporated within the molecule itself. When considering viruses that infect extreme halophiles, not only are proteins of the viral capsid exposed directly to the environment, but the proteins of the replication machinery must operate effectively within the elevated salt environment of its archaeal host.

[006] Phi29 is a widely used and commercially successful salt tolerant DNA polymerase from bacillus phage Phi29, however its salt tolerance is limited, and in 1 M KCl there is no detectable binding of the enzyme to the DNA substrate.

[007] It would be desirable to identify DNA polymerases that can function at elevated salt concentrations, for example at concentrations of at least 5%, 10%, 20% or 30% wt/vol salt, such as KCl or NaCl.

[008] **Short description of the invention**

[009] The invention encompasses isolated DNA polymerase comprising an amino acid sequence having at least 60% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6 and 8. Percent sequence identity between the claimed DNA polymerase of the invention and SEQ ID Nos 2, 4, 6 and 8 may, in various embodiments, be, for example, at least 60%, or 70% or 80% or 90% or 95% or 99% or 100% identity.

[0010] The DNA polymerases of the invention catalyze DNA polymerization under ionic conditions of at least 5% and up to 25% KCl (or other monovalent salt) wt/vol, wherein the rate (or average rate) of polymerization is at least 20 bases per minute. Salt concentrations may also be expressed as molarities, and The DNA polymerases of the invention catalyze DNA polymerization in monovalent salt concentrations in excess of 0.75M up to 4M.

[0011] The invention encompasses a biologically active fragment of RD0, RD1, RD2, or RD3, that catalyzes DNA polymerization, where the fragment comprises a polypeptide of at least 20 contiguous amino acid residues of an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6 and 8.

[0012] The invention encompasses an isolated polynucleotide, the polynucleotide having at least 60% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID Nos 1, 3, 5 and 7. Percent sequence identity of the claimed polynucleotides may be at least 50%, in various embodiments, be for example, at least 60%, or 70% or 80% or 90% or 95% or 99% or 100% identity. The isolated polynucleotide may have a sequence selected from the group consisting of SEQ ID Nos 2, 4, 6 and 8.

[0013] The invention further encompasses polynucleotides described cloned into vectors and host cells comprising such vectors. The claimed polynucleotides may be recombinant and/or synthetic.

[0014] The invention includes or variants and analogues of any of the foregoing.
The invention further encompasses a method for DNA synthesis at high salt concentration, comprising: a) providing a DNA polymerase comprising an amino acid sequence having at least 60% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6 and 8, b) contacting a DNA polymerase with a nucleic acid template under conditions of high salt concentration, wherein high salt concentration is defined as ionic conditions of at least 3% salt KCl (or other monovalent salt) wt/vol, c) effecting template dependent synthesis of DNA. In such methods the high salt concentration may comprise

conditions of between 5% and 25% salt KCl (or other monovalent salt) wt/vol.

[0015] The invention further encompasses a method for detecting a target polynucleotide in a sample, said method comprising the steps of a) providing a labelled polynucleotide probe having a sequence that comprises at least 16 contiguous nucleotides of the polynucleotide sequence of RD0, RD1, RD2, or RD3, and b) contacting said probe with a sample putatively containing a target polynucleotide complementary to the probe, c) hybridizing the probe and the target polynucleotide, d) detecting the presence or absence of said hybridization complex.

[0016] Additionally, the invention further encompasses kit comprising a sequencing reagent for DNA sequencing or amplification, the reagent comprising a high salt tolerant DNA polymerase having at least 60% sequence identity to a polypeptide sequence selected from the group consisting of RD0, RD1, RD2, and RD3.

[0017] **Description of the sequence listing**

[0018] The application discloses sequences in standard FASTA format, for both nucleotide and protein, for RD0, RD1, RD2, and RD3 (SEQ IDs 1 to 8). Also disclosed are the sequence for the closest currently known sequences from the salterproviruses His1 and His2 (SEQ IDs 9 to 12). The information recorded in electronic form (if any) submitted (under Rule 13*ter* if appropriate) with this application is identical to the sequence listing as contained in the application as filed. The sequences represented in the accompanying sequence listing are as follows:

SEQ ID No.1. >rd0-nuc = the nucleotide sequence for rd0
SEQ ID No.2. >rd0 = the amino acid sequence for rd0
SEQ ID No.3. >rd1-nuc = the nucleotide sequence for rd1
SEQ ID No.4. >rd1= the amino acid sequence for rd1
SEQ ID No.5. >rd2-nuc = the nucleotide sequence for rd2
SEQ ID No.6. >rd2 = the amino acid sequence for rd2
SEQ ID No.7. >rd3-nuc = the nucleotide sequence for rd3
SEQ ID No.8. >rd3 = the amino acid sequence for rd3
SEQ ID No.9. >His1V_gp12 = the nucleotide sequence for His1V_gp12
SEQ ID No.10. >His1V_gp12 length =717 = the amino acid sequence for His1V_gp12
SEQ ID No.11. >His2V_gp14 = the nucleotide sequence for His2V_gp14
SEQ ID No.12. >His2V_gp14 length=720 = the amino acid sequence for His2V_gp14

[0019] The salterproviruses His1 and His2 sequences are viral sequences, and as such, they will not have a genus-species. These were not discovered by our work and are provided as reference sequences. Michael Dyall-Smith and coworkers. The other 8 sequences are derived from metagenomic sequences found

by sequencing DNA isolated from a pond at location A23. We suspect that these are vial sequences but we cannot know what organism or virus they are from.

[0020] **Description of the figures**

[0021] Fig 1. shows a Sequence comparison between Phi29, his1_gp12, his1_gp14 and rd0, rd1, rd2 and rd3. The original drawing is in color. In the key on the bottom left, the word Exonuclease is in red (first arrow). Palm is in green (second arrow). TRP1 is in blue (third arrow). Fingers is in purple (forth arrow). TRP2 is in black (fifth arrow). And Thumb is in orange (sixth arrow).

[0022] **General Representations Concerning the Disclosure**

[0023] The embodiments disclosed in this specification are exemplary and do not limit the invention. Other embodiments can be utilized and changes can be made. As used in this specification, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a part" includes a plurality of such parts, and so forth. The term "comprises" and grammatical equivalents thereof are used in this specification to mean that, in addition to the features specifically identified, other features are optionally present. Where reference is made in this specification to a method comprising two or more defined steps, the defined steps can be carried out in any order or simultaneously (except where the context excludes that possibility), and the method can optionally include one or more other steps which are carried out before any of the defined steps, between two of the defined steps, or after all the defined steps (except where the context excludes that possibility). Where reference is made herein to "first" and "second" features, this is generally done for identification purposes; unless the context requires otherwise, the first and second features can be the same or different, and reference to a first feature does not mean that a second feature is necessarily present (though it may be present). Where reference is made herein to "a" or "an" feature, this includes the possibility that there are two or more such features.

[0024] This specification incorporates by reference all documents referred to herein and all documents filed concurrently with this specification or filed previously in connection with this application, including but not limited to such documents which are open to public inspection with this specification.

[0025] Where KCl is mentioned, this salt is used as an example only, and 'KCl' may be substituted in all instances for any other salt such as any monovalent salt.

[0026] The information recorded in electronic form (if any) submitted (under Rule 13*ter* if appropriate) with this application is identical to the sequence listing as contained in the application as filed.

[0027] **Definitions**

[0028] The term "amino acid sequence" refers to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these.

[0029] The term "amplification" relates to the production of additional copies of a nucleic acid sequence e.g., using polymerase chain reaction (PCR).

[0030] The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')2, and Fv fragments, which are capable of binding an epitopic determinant.

[0031] The term "similarity" refers to a degree of complementarily. There may be partial similarity or complete similarity. The word "identity" may substitute for the word

[0032] "similarity." A partially complementary sequence that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid is referred to as "substantially similar."

[0033] The phrase "percent identity" as applied to polynucleotide or polypeptide sequences refers to the percentage of residue matches between at least two sequences aligned using a standardized algorithm such as any of the BLAST suite of programs (e.g., blast, blastp, blastx, nucleotide blast and protein blast) using, for example, default parameters. BLAST tools are very commonly used and are available on the NCBI web site.

[0034] A "variant" of a particular polypeptide sequence is defined in this disclosure as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the

[0035] "BLAST 2 Sequences" tool set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 86%, at least 90%, at least 95%, or at least 98% or greater sequence identity over a certain defined length of one of the polypeptides.

[0036] The word "homologue" when used to describe a sequence refers to a sequence that is a variant of another and wherein the two sequences are evolutionarily related. In the present disclosure, when a particular gene or protein is referred to, the term is meant to encompass homologues and orthologues, variants, derivatives, and mutants of such a gene or protein. The present invention is not limited to embodiments employing the exact sequence of any of the disclosed proteins, polypeptides, polynucleotides etc, but encompasses any variant that is related by structure, sequence, function or is derived in any way from the

named protein. For example, the present invention encompasses polypeptides having, for example, at least 30% primary amino acid sequence similarity to a an envelope glycoprotein over a length of at least 100 amino acid residues, or in other embodiments, at least 40%, 50%, 75%, 90% or 99% primary protein sequence similarity.

[0037] "Variants and analogues" of polynucleotides encompass polynucleotides that show structural similarity to the polynucleotide of which it is an analogue or variant. Structural similarity for polynucleotides refers to sequence similarity. A polynucleotide analogue may have, for example, at least 99%, 95%, 90%, 85%, 80%, 70%, 60%, 50%, or at least 40% similarity over the entire length of the original polynucleotide. Often variants that share functional motifs have a good deal less than 40% overall sequence similarity, and yet may still be reasonably described as variants or analogues. Alternatively it may have a similarity of at least 99%, 95%, 90%, 85%, 80%, 70%, 60%, 50%, or at least 40%, 30% or at least 20% similarity over a shorter length, for example over at least 1000 nucleotides, or at least 500, at least 250, at least 150, at least 100, at least 50 or at least 25 polynucleotides. Variants may be derivatives of the polynucleotide of which they are a variant, they may be chemically or biochemically modified and have one or more amino nucleotide substitutions, additions, and/or deletions. Variants may share certain functionally significant motifs with the polynucleotide of which they are a variant. These motifs may encode the portion of a protein that includes the active site of a protein, the portion that is essential to enzymatic activity. Sequence similarities and homologies may be reliably and consistently determined by using any of the well known Basic Local Alignment Search Tool (BLAST) software tools.

[0038] Percent identity between polynucleotide sequences may be determined using the default parameters of the CLSUATL W algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLSUATL W as the "percent similarity" between aligned polynucleotide sequence pairs.

[0039] Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410). The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with

the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such default parameters may be, for example:

*[0040]* *Matrix: BLOSUM62*

*[0041]* *Reward for match: 1*

*[0042]* *Penalty for mismatch: -2*

*[0043]* *Open Gap: 5 and Extension Gap: 2 penalties*

*[0044]* *Gap x drop-off: 50*

*[0045]* *Expect: 10*

*[0046]* *Word Size: 11*

[0047] *Filter: on*

[0048] Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

[0049] Percent identity between polypeptide sequences may be determined using the default parameters of the CLSUATL W algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLSUATL W, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLSUATL W as the "percent similarity" between aligned polypeptide sequence pairs.

[0050] Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

*Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

[0051] Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

[0052] **Detailed description of the invention**

[0053] Described here are four novel sequences of type B DNA polymerases and variants and analogues thereof that can be used in applications involving DNA in high salt conditions. These DNA polymerases are stable and operate under ionic conditions of at least 5% salt (KCl or other monovalent salt) wt/vol, and up to 25% wt/vol. These molecules remain stable at salt concentrations well beyond any currently commercially available DNA polymerase and provide polymerization at a rate (or an average rate) of at least 20 bases per minute.

[0054] The four novel Halophilic DNA polymerases, called RD0, RD1, RD2, and RD3, are described and their polynucleotide and protein sequences are disclosed. The polymerases were isolated from a metagenome of viruses of halophilic archebacteria.

[0055] The high salt tolerance of these DNA polymerases may be very useful for various applications in which high salt concentration is an advantage. For example, the polymerases are useful for sequencing in which they provide better resolution of GC rich compressions. Additionally the polymerases are useful for nanopore sequencing where a high salt concentration will boost the signal to noise ratio for ionic-current-based nanopore measurements.

[0056] The closest homologue to RD0, RD1, RD2, and RD3 appears to be a protein derived the known halophilic salterproviruses viruses His1 (His1V_gp12 YP_529524.1) and His2 (His2V_gp14 YP_529644.1) which demonstrates a 38%-50% identity to RD0, RD1, RD2, and RD3. The functions of these His1 and His2 proteins is unknown, although it may be speculated that they may have DNA polymerase activity.

| Percent ID* | | |
|---|---|---|
| | His1 | His2 |
| RD0 | 38% | 40% |

| | | |
|---|---|---|
| RD1 | 40% | 50% |
| RD2 | 44% | 43% |
| RD3 | 38% | 39% |

[0057] *Identities to His1 / His2 DNA polymerases (Blastp, wordsize 2, Blosum45)

[0058] RD0, RD1, RD2, and RD3 all possess the TPR2 motif, a functional motif found TPR2, a specific subdomain of protein-priming DNA polymerases.

[0059] The commercially available Phi29 DNA polymerase is 18.1%, 18.4%, 20.4% and 19.3% identical at the protein level with RD0, RD1, RD2, and RD3 respectively. Phi29 is a widely used and commercially successful salt tolerant DNA polymerase from bacillus phage Phi29, however its salt tolerance is limited and in 1 M KCl there is no detectable binding of the enzyme to the DNA substrate.

[0060] The DNA polymerases of the invention are stable and exhibit appreciable enzymatic activity under ionic conditions of at least 3% salt (KCl or other monovalent salt) wt/vol, or in other embodiments, at least 5% wt/vol Monovalent salt, or at least 7% wt/vol Monovalent salt, or at least 10% wt/vol Monovalent salt, or at least 15% wt/vol Monovalent salt, or at least 20% wt/vol Monovalent salt, or at least 25% wt/vol Monovalent salt. The salt may be others salts other than KCl, such as NaCl; KCl is given only as an example. The important property is that the DNA polymerases of the invention function under high salt concentrations.

[0061] **Embodiments**

[0062] In one aspect, the invention provides isolated DNA polymerases comprising:

[0063] a) an amino acid sequence selected from SEQ ID Nos 2, 4, 6 and 8,

b) an amino acid sequence having at least 60% (or in other embodiments at least 50%, at least 70%, at least 80%, at least 90% and at least 95%) sequence identity to an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6 and 8,

c) a biologically active fragment of RD0, RD1, RD2, or RD3, where the fragment comprises a polypeptide of at least 20 contiguous amino acid residues of an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6 and 8, or

d) an immunogenic fragment of RD0, RD1, RD2, or RD3, where the fragment comprises a polypeptide of at least 10 (or 15, 20, 25, 30, 35, 40, 50 or 60) contiguous amino acid residues of an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6 and 8, or variants and analogues of any of the foregoing.

[0064] The invention further encompasses an isolated polynucleotide encoding any of the above polypeptides or amino acid sequences and or variants and analogues thereof, including polynucleotides with sequences selected from SEQ ID Nos 1, 3, 5 and 7, and polynucleotides having at least 60% (or in other embodiments at least 70%, at least 80%, at least 90% and at least 95%) sequence identity to an polynucleotide sequence selected from the group consisting of SEQ ID Nos 1, 3, 5 and 7.

[0065] Additionally, the invention provides a recombinant and/or synthetic polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding any of the above polypeptides or amino acid sequences, or to a polynucleotide disclosed herein such as polynucleotides with sequences selected from SEQ ID Nos 1, 3, 5 and 7. The invention also includes recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide having at least 60% (or in other embodiments at least 70%, at least 80%, at least 90% and at least 95%) sequence identity to an polynucleotide sequence selected from the group consisting of SEQ ID Nos 1, 3, 5 and 7.

[0066] The invention also provides a method for producing any of the above polypeptides or amino acid sequences. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

[0067] Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide comprising (1) the polynucleotide sequence of RD0, RD1, RD2, or RD3, (2) a polynucleotide sequence that comprises at least 8 or at least 12 or at least 16 or at least 18 or at least 20 or at least 24 or at least 30 or at least 40 contiguous nucleotides of the polynucleotide sequence of RD0, RD1, RD2, or RD3, (3) a polynucleotide sequence that encodes a high salt tolerant DNA polynucleotide having at least 60% sequence identity to a polynucleotide sequence selected from the group consisting of the polynucleotide sequence of RD0, RD1, RD2, and RD3, (4) a polynucleotide sequence complementary to any of the foregoing. The method comprises a) hybridizing the sample with a probe comprising at least 16 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 30 contiguous nucleotides. In another alternative, the probe comprises at least 60 contiguous nucleotides.

[0068] The invention further provides sequencing reagents for DNA sequencing or amplification, the reagent(s) comprising a high salt tolerant DNA polymerase with an amino acid sequence of RD0, RD1, RD2, or RD3, or a high salt tolerant DNA polynucleotide having at least 60% sequence identity (or in other embodiments at least 70%, at least 80%, at least 90% and at least 95%) to a polypeptide sequence selected from the group consisting of RD0, RD1, RD2, and RD3.

[0069] The invention also encompasses an isolated polynucleotide, the polynucleotide having at least 60% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID Nos 1, 3, 5 and 7, wherein the polynucleotides are recombinant or synthetic. Such polynucleotides may be incorporated into a vector such as an expression vector, and said vector may be transformed into a host cell such as a prokaryotic cell for culture, expression and production of the recombinant polymerase. The invention includes or variants and analogues of any of the foregoing.

[0070] The invention also encompasses a method for DNA synthesis at high salt concentration, comprising: a) providing a DNA polymerase comprising an amino acid sequence having at least 60% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6 and 8, b) contacting a DNA polymerase with a nucleic acid template under conditions of high salt concentration, wherein high salt concentration is defined as ionic conditions of at least 3% salt KCl wt/vol, and c) effecting template dependent synthesis of DNA. Such a method may further comprise contacting a PCR enhancing factor and/or an additive with said DNA polymerase and said nucleic acid template. In such methods the high salt concentration may comprise conditions of between 5% and 25% salt KCl wt/vol.

[0071] The invention also encompasses a kit comprising a sequencing reagent for DNA sequencing or amplification, the reagent comprising a high salt tolerant DNA polymerase having at least 60% sequence identity to a polypeptide sequence selected from the group consisting of RD0, RD1, RD2, and RD3.

[0072] **Further embodiments and examples**

[0073] Identification, isolation, and engineering of high salt tolerant DNA polymerases specifically with a view to nanopore sequencing.

[0074] Rationale: A method for boosting signal to noise for ionic-current-based nanopore measurements is to run the assays in higher ionic strength buffers. Although phi29 DNAP can replicate DNA in 0.6 M KCl, the rate of catalysis is slowed two-fold relative to the rate in 0.3 M KCl, and in 1 M KCl there is no detectable

binding of the enzyme to DNA substrates (Cherf, unpublished observation). Therefore it is desirable to identify and use salt-adapted polymerases that otherwise share structural and functional properties with phi29 DNAP polymerase.

[0075] Among members of the extreme halophiles, salt-tolerance is achieved not by exclusion of monovalent ions from the cytosol, but by adapting intracellular machinery function in elevated salt. As an example of salt tolerance among members of the extreme halophiles, malate dehydrogenase from the archaeal halophile Haloarcula marismortui incorporates a salt-adaptive strategy where the high ionic concentration from the environment is not only tolerated but is incorporated within the protein; sodium and chloride ions are found incorporated within the molecule itself (Richard et al.). When considering viruses that infect extreme halophiles, not only are proteins of the viral capsid exposed directly to the environment, but the proteins of the replication machinery must operate effectively within the elevated salt environment of its archaeal host.

[0076] Purified viral DNA from the halophilic virus His2 is capable of transfecting a wide range of haloarchaeal species (Porter & Dyall-Smith, 2008). Though this may not be environmentally relevant, it does require that proteins of His2 must tolerate the ionic conditions of multiple hosts. The polymerase encoded by His1 and His2 may then not only tolerate high ionic strength but would be expected to properly fold and function in elevated salt.

[0077] **DNA polymerases from salterproviruses His1 and His2.**

[0078] The inventors have reviewed the literature for viruses of halophilic extremophiles and conducted a sequence homology comparison between phi29 DNAP and DNA and protein sequences of halophilic viruses and archaeal extremophiles. Promising candidates include salterprovirus His1 and His2 (Bath et al. 2006; Prangishvili et al. 2006). His1 and His2 were found to contain a putative single protein DNA polymerase with sequence homology to phi29 DNA polymerase including the TPR2 sub-domain of phi29 DNA polymerase (sub-domain responsible for high level of processivity by 29 DNA polymerase, Rodriguez et al. 2005).

[0079] The inventors initially isolated and characterized the DNA polymerase from His1 and His2. Both His1 and His2 viruses and hosts are available through the German Resource Centre for Biological Material (DSMZ). Halophilic archaeal virus alternatives to His1 and His2 were also identified, and include- Hs1, H, Ch1, HF1, HF2, HRPV-1, HHPV-1 and SH1 (Kukkardo, 2008).

[0080] **Expression and purification of proteins from halophiles and haloviruses.**

[0081] Isolation of these proteins can be accomplished using proven expression systems, including E. coli. Seryl-tRNA synthetase from Haloarcula marismortui was successfully cloned and overexpressed as a thioredoxin fusion (Taupin et al.). An alternative expression and purification strategy can be employed for halophilic proteins that takes advantage of the increased solubility of halophilic proteins under elevated salt concentration (salting-in), while proteins of the expression host become insoluble (salting out). Folding, stability and subsequent activity of the halophilic protein can then be improved. A third expression strategy makes use of transformation in a halophilic expression host (Haloferax volcanii) in order to maintain folding of the protein within the 1-4M KCl presented by the intracellular environment of this host (Plosser & Pfeifer).

[0082] **Isolation of novel haloviruses and halovirus DNA.**

[0083] Sampling from extreme environments is limited mainly by safe access to the environment and appropriate authorization permits. In the case of halophilic environments with near-saturating conditions, numerous salt lakes including the Pink lakes of the western United States are known to harbor vast numbers of the Halobacteriaiaceae. Western salt lakes have provided the essential conditions for many halophilic discoveries including Owens lake, the Salton Sea (Swan et al.), Death Valley (Mormile et al.) and the Great salt lake in Utah (Tsai et al., 1995). Additionally, Halorubrum californiense, was recently isolated in a solar salt pond near the southern tip of the San Francisco Bay (Pesenti et al.). Viruses of these Haloarchaeal hosts, the haloviruses, outnumber the cellular population by 10-100 fold (Porter et al.), and in some cases, classic plaque assays are possible, as demonstrated by Dyall-Smith and coworkers (Gunde-Cimerman et al.; Stedman et al.).

## Claims

1. An isolated DNA polymerase comprising an amino acid sequence having at least 60% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6, and 8.

2. The DNA polymerase of claim 1 wherein the an amino acid sequence is selected from the group consisting of SEQ ID Nos 2, 4, 6, and 8.

3. The DNA polymerase of claim 1 that catalyzes DNA polymerization under ionic conditions of at least 3% monovalent salt wt/vol, wherein the average rate of polymerization is at least 20 bases per minute.

4. A biologically active fragment of RD0, RD1, RD2, or RD3, that catalyzes DNA polymerization, where the fragment comprises a polypeptide of at least 20 contiguous amino acid residues of an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6, and 8

5. An isolated polynucleotide, the polynucleotide having at least 60% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID Nos 1, 3, 5 and 7.

6. The isolated polynucleotide of claim 5 having a sequence selected from the group consisting of SEQ ID Nos 1, 3, 5 and 7.

7. A vector comprising the polynucleotide according to claim 5.

8. A host cell comprising the vector according to claim 7.

9. The host cell according to claim 8, wherein the host cell is a prokaryotic cell.

10. The polynucleotide of claim 5, wherein the polynucleotide is recombinant.

11. The polynucleotide of claim 5, wherein the polynucleotide is synthetic.

12. A method for DNA synthesis at high salt concentration, comprising:
a) providing a DNA polymerase comprising an amino acid sequence having at least 60% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID Nos 2, 4, 6, and 8.

b) contacting a DNA polymerase with a nucleic acid template under conditions of high salt concentration, wherein high salt concentration is defined as ionic conditions of at least 3% monovalent salt wt/vol,
c) effecting template dependent synthesis of DNA.

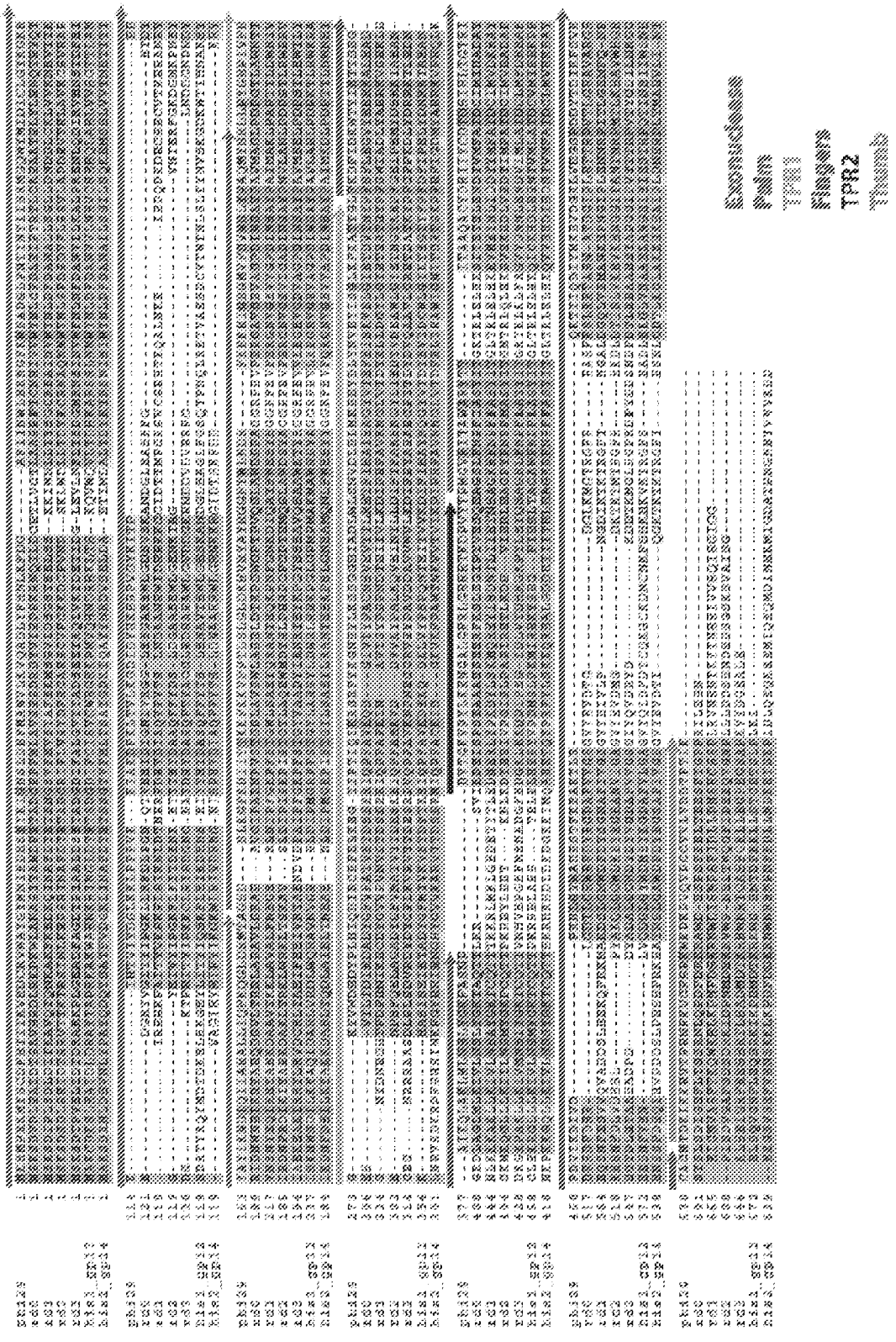
13. The method of claim 12, further comprising contacting a PCR enhancing factor and/or an additive with said DNA polymerase and said nucleic acid template.

14. The method of claim 12, wherein said high salt concentration comprises conditions of between 5% and 25% monovalent salt wt/vol.

15. A method for detecting a target polynucleotide in a sample, said method comprising the steps of a) providing a labelled polynucleotide probe having a sequence that comprises at least 16 contiguous nucleotides of the polynucleotide sequence of RD0, RD1, RD2, or RD3, and b) contacting said probe with a sample putatively containing a target polynucleotide complementary to the probe, c) hybridizing the probe and the target polynucleotide, d) detecting the presence or absence of said hybridization complex.

16. A kit comprising a sequencing reagent for DNA sequencing or amplification, the reagent comprising a high salt tolerant DNA polymerase having at least 60% sequence identity to a polypeptide sequence selected from the group consisting of RD0, RD1, RD2, and RD3.

Fig. 1



**INTERNATIONAL SEARCH REPORT**

International application No.
PCT/US 2012/041802

A. CLASSIFICATION OF SUBJECT MATTER

***C12N 9/12*** *(2006.01)*
***C12N 15/54*** *(2006.01)*
***C12N 15/09*** *(2006.01)*
***C12N 15/10*** *(2006.01)*
***C12Q 1/68*** *(2006.01)*
***C12N 15/63*** *(2006.01)*
***C07H 21/00*** *(2006.01)*

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)

C12N 9/12, 15/54, 15/09, 15/10, C12Q 1/68, C12N 15/63, C07H 21/00

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 practicable, search terms used)

PAJ, Esp@cenet, PCT Online, USPTO DB, WIPO, RUPTO, EAPATIS, PatSearch

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| A | US 6333159 B1 (WASHINGTON UNIVERSITY) 25.12.2001, abstract, SEQ ID NO: 1, 2 | 1-16 |
| A | RU 2198222 C2 (ROSHE DIAGNOSTICS GMBH) 10.02.2003, abstract, claim 1 | 1-16 |
| A | US 5691142 A (THIRD WAVE TECHNOLOGIES, INC.) 25.11.1997, abstract, claim 1 | 1-16 |
| A | US 5968743 A (HITACHI, LTD.) 19.10.1999, claims 16, 17, 18, 19 | 1-16 |

☐ Further documents are listed in the continuation of Box C.

☐ 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 |
|---|---|
| 04 August 2012 (04.08.2012) | 04 October 2012 (04.10.2012) |
| Name and mailing address of the ISA/ FIPS<br>Russia, 123995, Moscow, G-59, GSP-5,<br>Berezhkovskaya nab., 30-1<br><br>Facsimile No. +7 (499) 243-33-37 | Authorized officer<br><br>I. Tereshkina<br><br>Telephone No. (495)531-65-15 |