



EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention of the grant of the patent:
03.07.2024 Bulletin 2024/27

(21) Application number: **16784755.7**

(22) Date of filing: **14.10.2016**

(51) International Patent Classification (IPC):
C12Q 1/6869 ^(2018.01) **C12Q 1/6844** ^(2018.01)

(52) Cooperative Patent Classification (CPC):
(C-Sets available)
C12Q 1/6844; C12Q 1/6869 (Cont.)

(86) International application number:
PCT/US2016/057073

(87) International publication number:
WO 2017/066592 (20.04.2017 Gazette 2017/16)

EXHIBIT #

Metzker 2210

12/03/2025

(54) **METHODS AND KITS FOR HIGHLY MULTIPLEX SINGLE PRIMER EXTENSION**

VERFAHREN UND KITS ZUR HOCHMULTIPLEX-EINZELPRIMER-VERLÄNGERUNG

PROCÉDÉS ET KITS D'EXTENSION D'AMORCE UNIQUE HAUTEMENT MULTIPLEX

(84) Designated Contracting States:
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

(30) Priority: **16.10.2015 US 201562242766 P**

(43) Date of publication of application:
22.08.2018 Bulletin 2018/34

(73) Proprietor: **Qiagen Sciences, LLC**
Germantown, MD 20874 (US)

(72) Inventors:
• **WANG, Yexun**
San Diego, CA 92130 (US)
• **PENG, Quan**
Clarksburg, Maryland 20871 (US)

(74) Representative: **Schiweck Weinzierl Koch**
Patentanwälte Partnerschaft mbB
Ganghoferstraße 68 B
80339 München (DE)

(56) References cited:
WO-A1-2013/175815

- **QUAN PENG ET AL: "Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes", BMC GENOMICS, BIOMED CENTRAL LTD, LONDON, UK, vol. 16, no. 1, 7 August 2015 (2015-08-07) , page 589, XP021224943, ISSN: 1471-2164, DOI: 10.1186/S12864-015-1806-8**
- **JUN WATANABE ET AL: "Use of a competitive probe in assay design for genotyping of the UGT1A1*28 microsatellite polymorphism by the smart amplification process", BIOTECHNIQUES, vol. 43, no. 4, 1 October 2007 (2007-10-01), pages 479-484, XP055180958, ISSN: 0736-6205, DOI: 10.2144/000112563**
- **MITANI YASUMASA ET AL: "RAPID SNP DIAGNOSTICS USING ASYMMETRIC ISOTHERMAL AMPLIFICATION AND A NEW MISMATCH-SUPPRESSION TECHNOLOGY", HHS PUBLIC ACCESS AUTHOR MANUSCRIPT, NATURE PUBLISHING GROUP, GB, vol. 4, no. 3, 1 March 2007 (2007-03-01), pages 257-262, XP009084707, ISSN: 1548-7091, DOI: 10.1038/NMETH1007**
- **KENJI FUKUI ET AL: "Thermostable Mismatch-Recognizing Protein MutS Suppresses Nonspecific Amplification during Polymerase Chain Reaction (PCR)", INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES, vol. 14, no. 3, 21 March 2013 (2013-03-21) , pages 6436-6453, XP055326022, DOI: 10.3390/ijms14036436**

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

Remarks:

The complete document including Reference Table(s) and the Sequence Listing(s) can be downloaded from the EPO website

(52) Cooperative Patent Classification (CPC): (Cont.)

C-Sets

**C12Q 1/6844, C12Q 2521/514, C12Q 2527/127,
C12Q 2533/101, C12Q 2535/122, C12Q 2537/143;
C12Q 1/6869, C12Q 2521/514, C12Q 2527/127,
C12Q 2533/101, C12Q 2535/122, C12Q 2537/143**

Description

CROSS REFERENCE TO RELATED APPLICATION

5 **[0001]** This application claims the benefit under 35 U.S.C. § 119(e) to U.S. Provisional Application No. 62/242,766, filed October 16, 2015.

BACKGROUND

10 Technical Field

[0002] The present disclosure relates to highly multiplex single primer extensions with improved efficiency and specificity and their use in next generation sequencing (NGS).

15 Description of the Related Art

[0003] Targeted sequencing is one of the major applications for NGS. It is a cost efficient way to sequence the most relevant portion of the genome or transcriptomes.

20 **[0004]** Highly multiplex PCR is one of the popular methods to prepare samples for targeted sequencing. Although it is a simple yet powerful tool to enrich a large gene region, it has limited ability to detect large structural variants such as gene fusion or splicing variants because the two primers per amplicon design cannot accommodate large or unknown structural changes between the two predefined primer sites. In addition, highly multiplex PCR is difficult to be scaled up to cover very large regions such as whole transcriptome or exome in a single reactions tube. The many pairs of primers will introduce unwanted artifacts when multiplex scale increases dramatically.

25 **[0005]** A recent alternative approach to enriching target loci is single primer extensions. The single primer design allows the detection of large and unknown structural variants, and improves scalability for covering very large panels. However, because target specificity is provided by only one primer, the overall enrichment specificity is generally much lower than PCR enrichment using two target-specific primers.

30 **[0006]** One way to improve single primer specificity is to conduct the single primer annealing at more stringent conditions (e.g., at high temperature and high salt condition) and only perform polymerase extension after primer binding has reached equilibrium. However, high stringent hybridization conditions typically reduce DNA polymerase activity. In addition, performing polymerase extension after primer binding is inconvenient. Another challenge in enriching a very large region is the limitation of each primer concentration. Because of the physical and chemical constraints, when more different primers are added in one reaction (to cover more regions), the concentration of each primer inevitably has to be lowered in proportion. For example, each primer may need to be at 2 nM when 20,000 primers are used to enrich 500 genes, while each primer may be at 20 nM when 2000 primers are used to enrich 50 genes. In other cases, the concentration is limited by how the primer is synthesized. For example, oligonucleotides prepared by massive parallel synthesis on microchip have much lower yields than by traditional column-based synthesis approach. In both scenarios, lower primer concentrations will lead to lower template annealing efficiency under standard conditions.

40 **[0007]** One way to improve annealing efficiency is to prolong the annealing time to hours, which is not very desirable. While it is possible to change the annealing chemistry to accelerate proper primer-template binding, doing so increases non-specific primer binding and extension by DNA polymerase, leading to overall low specificity.

[0008] Peng et al. (BMC Genomics 16:589, 2015) relates to reducing amplification artifacts in high multiplex PCR amplicon sequencing by using molecular barcodes.

45 **[0009]** Watanabe et al. (Biotechniques 43:479-84, 2007) relates to using a competitive probe in assay design for genotyping a microsatellite polymorphism by the smart amplification process, which process uses a unique asymmetric primer design and Taq MutS.

[0010] Mitani et al. (Nature Methods 4:257-62, 2007) relates to using the same smart amplification process in rapid SNP diagnostics.

50 **[0011]** Fukui et al. (International Journal of Molecular Sciences 14:6436-53, 2013) relates to the use of MutS in suppressing non-specific amplification during PCR.

SUMMARY

55 **[0012]** In one aspect, the present disclosure provides a method for performing a highly multiplex single primer extension reaction, comprising:

- a. extending at least about 1,000 different primers in a single primer extension reaction using a plurality of target

nucleic acids as templates in the presence of (1) a MutS protein, and (2) Mg^{2+} at a minimum concentration of about 6 mM to generate extension products.

[0013] In certain embodiments, at least about 4,000 or at least about 9,000 additional different primers are extended in the single primer extension reaction.

[0014] In certain embodiments, the single primer extension reaction is performed in the absence of a primer capable of specifically annealing to a portion of the extension products.

[0015] In certain embodiments, most of the target nucleic acids comprise a common sequence, and wherein the reaction is performed in the presence of a boosting primer that comprises the common sequence or a portion thereof that is at least about 10 nucleotides in length.

[0016] In certain embodiments, each of the primers other than the boosting primer is present in the single primer extension reaction at a concentration of no more than about 20 nM, preferably at a concentration of about 1 nM to about 10 nM; and/or wherein Mg^{2+} is present in the single primer extension reaction at a concentration of about 6 mM to about 20 mM.

[0017] In certain embodiments, the plurality of target nucleic acids are:

- (i) genomic DNA or amplification products of genomic DNA;
- (ii) cDNA or amplification products of cDNA; or
- (iii) bisulfite treated genomic DNA or its amplified products.

[0018] Also described is that the MutS protein is a *Thermus aquaticus* MutS protein.

[0019] In certain embodiments it is foreseen that (i) the primer annealing and extension specificity of the single primer extension reaction is at least about 80%, and/or

(ii) the average primer annealing and extension efficiency of the single primer extension reaction is at least about 10%.

[0020] In certain embodiments, the method provided herein further comprises:

- b. optionally amplifying the extension product(s) to generate amplified extension product(s), and
- c. sequencing the extension product(s) or the amplified extension products.

[0021] In certain embodiments, the target nucleic acids are partially double-stranded, and comprise a double-stranded target nucleic acid sequence and a single-stranded common sequence at the 5' terminus of each strand of the target nucleic acid sequence, optionally wherein the method prior to step a.,

- (i) fragmenting DNA molecules to generate double-stranded DNA fragments, and
- (ii) ligating the single-stranded common sequence to the 5' terminus of each strand of the double-stranded DNA fragment to provide partially double-stranded target nucleic acids.

[0022] In certain embodiments, each primer other than the boosting primer if present comprises at its 5' terminus a universal sequence that is incapable of specifically annealing to a region of a target nucleic acid.

[0023] In a related aspect, the present disclosure provides a method for sequencing nucleic acids, comprising:

a. providing a plurality of partially double-stranded target nucleic acids each of which comprises a double-stranded target nucleic acid sequence, and a single-stranded common sequence at the 5' terminus of each strand of the target nucleic acid sequence, wherein the single-stranded common sequence is the same among the plurality of the target nucleic acids; optionally wherein step a. comprises:

- a1. fragmenting double-stranded DNA molecules to generate double-stranded DNA fragments, and
- a2. ligating the single-stranded common sequence to the 5' terminus of each strand of the double-stranded DNA fragments to provide the plurality of partially double-stranded nucleic acids;

- b. providing a plurality of primers capable of specifically annealing to portions of the target nucleic acid sequences;
- c. extending the primers using the target nucleic acid sequences as templates in the presence of (1) a MutS protein and (2) Mg^{2+} at a minimum concentration of about 6 mM to obtain extension products; and
- d. sequencing the extension products or their amplification products.

[0024] In certain embodiments, extending the primers in step c. is performed in the presence of a boosting primer comprising the common sequence or a portion thereof that is at least about 10 nucleotides in length.

[0025] Also described is that step a. comprises:

- a1. fragmenting double-stranded DNA molecules to generate double-stranded DNA fragments, and
- a2. ligating the single-stranded common sequence to the 5' terminus of each strand of the double-stranded DNA fragments to provide the plurality of partially double-stranded nucleic acids.

[0026] In certain embodiments, the method comprises: prior to step d., a further step of amplifying the extension products of step c.

[0027] In another aspect, the present application provides a kit for performing a primer extension reaction, comprising:

- (1) a pool of at least about 1,000 different primers, preferably at least about 10,000 different primers,
- (2) a buffer concentrate comprising a MutS protein, dNTPs, and Mg^{2+} , wherein the concentration of Mg^{2+} in a highly multiplex single primer extension reaction mixture generated from diluting the buffer concentrate is at least about 6 mM, and
- (3) a DNA polymerase.

[0028] In certain embodiments, the MutS protein is a *Thermus aquaticus* MutS protein.

BRIEF DESCRIPTION OF THE DRAWINGS

[0029]

Figure 1 is a schematic representation of an embodiment of the present disclosure, showing that single primer extension is used in sequencing.

Figure 2 is a graph showing primer annealing and extension efficiencies at various Mg^{2+} concentrations.

Figure 3 is a graph comparing primer annealing and extension specificities between an extension reaction performed using 2 nM each primer and 12 mM Mg^{2+} and an extension reaction performed using 20 nM each primer and 4 mM Mg^{2+} .

Figure 4 is a graph comparing primer annealing and extension specificities and uniformities (T50) between an extension reaction performed at 12 mM Mg^{2+} without MutS and an extension reaction performed at 12 mM Mg^{2+} with MutS.

DETAILED DESCRIPTION

[0030] Targeted sequence enrichment using single primer extension approach has great advantages over traditional two primer PCR amplicon based enrichment. Because one boundary of the amplified product is not constrained by one predefined primer, it has the ability to detect unexpected structural variations in the sequence, such as fusions and splicing variants. However, due to only one primer being used, single primer extension has traditionally suffered from low enrichment specificity. Another challenge in target enrichment is the low primer annealing efficiency especially when the primer concentration is low to accommodate large gene panels. Existing approaches to promoting primer binding typically inadvertently reduce binding specificity.

[0031] Faced with the challenges of low primer annealing efficiency and specificity, the present inventors have identified a solution to increasing primer annealing extension efficiency while improving primer extension specificity via a combination of Mg^{2+} at a concentration higher than in a typical PCR reaction and a MutS protein. This combination is especially beneficial when a large number of very low concentrations of primers are used in single primer extension process to enrich a large genomics region for sequencing analysis.

[0032] In the following description, any ranges provided herein include all the values in the ranges. It should also be noted that the term "or" is generally employed in its sense including "and/or" (i.e., to mean either one, both, or any combination thereof of the alternatives) unless the content dictates otherwise. Also, as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the content dictates otherwise. The terms "include," "have," "comprise" and their variants are used synonymously and to be construed as non-limiting. The term "about" refers to $\pm 10\%$ of a reference a value. For example, "about 50°C" refers to "50°C + 5°C" (i.e., 50°C $\pm 10\%$ of 50°C).

I. Methods for Performing Single Extension Reactions

[0033] In one aspect, the present disclosure provides a method for performing a highly multiplex single primer extension reaction that comprises extending at least about 1,000 different primers in a single primer extension reaction using a plurality of target nucleic acids as templates in the presence of (1) a MutS protein, and (2) Mg^{2+} at a minimum concentration of about 6 mM to generate extension products.

[0034] The term "single primer extension" or "single primer extension reaction" as used in the present disclosure refers to a reaction in which a primer is extended using a single-stranded target nucleic acid or one strand of a double-stranded target nucleic acid as a template. A single primer extension does not include extension of another primer that anneals to the complementary strand of the single-stranded target nucleic acid or the other strand of the double-stranded target nucleic acid except where a boosting primer is used as described below.

[0035] A "primer" is an oligonucleotide capable of specifically annealing to a target nucleic acid, which allows the extension of the primer using the target nucleic acid as a template. Generally, such a primer or a sufficiently long portion thereof (e.g., at least about 10, 11, 12, 13, 14, or 15 nucleotides long) is substantially or completely complementary to a region of the target nucleic acid to allow specific annealing of the primer to the target nucleic acid. A primer used in a multiplex single primer extension reaction as disclosed herein for targeted nucleic acid enrichment may also be referred to as an "enrichment primer."

[0036] An "oligonucleotide" refers to a short polymer composed of deoxyribonucleotides, ribonucleotides or combinations thereof. Oligonucleotides are generally between about 10 to about 100 nucleotides, preferably about 12 to about 60 nucleotides, in length.

[0037] The terms "complementary" and "complement" and their variants, as used herein, refer to any two nucleic acid sequences or portions thereof that form a hybridized duplex by base pairing. One nucleic acid fragment may be completely complementary to another nucleic acid fragment if all of the nucleotides in the fragment form base pairing with nucleotides in the corresponding antiparallel positions on the other nucleic acid fragment. A primer (or a portion thereof that is at least about 10 nucleotides in length) is "substantially complementary" to a target nucleic acid if at least 90% (e.g., at least 95%, at least 98%, or at least 99%) of the residues of the primer (or the portion thereof) are complementary to corresponding residues in the target nucleic acid over the length of the primer (or the portion thereof).

[0038] In certain embodiments, a primer is substantially or completely complementary to a target nucleic acid over the complete length of the primer. In certain other embodiments, a primer is not substantially complementary to a target nucleic acid over the complete length of the primer. In such embodiments, the primer may comprise a sequence at its 3' end that is substantially or completely complementary to the target nucleic acid so that the primer may specifically anneal to the target nucleic acid, and another sequence at its 5' end that is not substantially complementary to the target nucleic acid sequence.

[0039] A single primer extension reaction may include multiple cycles of annealing of a primer to a single-stranded target nucleic acid or a strand of a double-stranded target nucleic acid, extending the primer, and denaturing the extension product from its complementary template to increase the amount of extension product.

[0040] In certain embodiments, the single primer extension reaction does not comprise any primer capable of specifically annealing to the complementary strand of the single-stranded target nucleic acid or the other strand of the double-stranded target nucleic acid. In such embodiments, multiple cycles of annealing, extending, and denaturing lead to linear amplification of the extension product.

[0041] A primer is capable of specifically annealing to a template nucleic acid if the primer is substantially or completely complementary to the template nucleic acid or a portion thereof that is at least about 10 nucleotides in length.

[0042] In certain other embodiments, the single primer extension reaction comprises a boosting primer. As discussed below, including a boosting primer in a single primer extension reaction that comprises multiple cycles of denaturing, annealing and extending will lead to exponential amplification of the extension product.

[0043] A "multiplex single primer extension reaction" is a single primer extension reaction where a plurality of primers (also referred to as "a primer set" or "a pool of primers") other than a boosting primer as described below are extended using one or more target nucleic acids as templates.

[0044] A "highly multiplex single primer extension reaction" refers to a multiplex single primer extension reaction where at least about 1,000 different primers other than a boosting primer are extended using a plurality of target nucleic acids as templates.

[0045] In certain embodiments, a highly multiplex single primer extension reaction may contain about 1,000 to about 100,000 different primers. For example, a highly multiplex single primer extension may contain at least about 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 4,500, or 5,000 different primers, and/or at most about 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, or 100,000 different primers, including any combination of the above-noted minimum and maximum numbers of different primers such as about 1,000 to about 10,000, about 1,000 to about 50,000, about 1,000 to about 100,000, about 5,000 to about 20,000, about 5,000 to about 50,000, or about 5,000 to about 100,000 different primers.

[0046] In a highly multiplex single primer extension reaction, primers other than a boosting primer (if present) may anneal to different genes, different regions of the same genes, and/or different strands of the same regions of the same genes as long as none of such primers forms a pair with another primer in the same reaction that allows exponentially amplification of a portion of a target nucleic acid delineated by the primer pair. In cases where two primers in a multiplex single primer extension reaction anneal to different portions of the same strand of a target nucleic acid, preferably, the different portions are sufficiently apart from each other so that the extension of one primer does not interfere with the

extension of the other primer.

[0047] In certain embodiments, each primer other than a boosting primer (if present) in a highly multiplex single primer extension reaction comprises at its 5' terminus a universal sequence that is incapable of specifically annealing to a region of a target nucleic acid. Such a universal sequence is useful in designing additional primers for amplifying the products of the highly multiplex single primer extension reaction. Additional description of such a universal sequence is provided in Section II below.

[0048] The primers may be obtained in any methods known in the art for synthesizing oligonucleotides. In certain embodiments, the primers are synthesized on microchips.

[0049] In a multiplex single primer extension reaction, the initial concentration (*i.e.*, the concentration at the start of the single primer extension reaction) of each primer other than a boosting primer may be about 0.01 nM to about 100 nM, preferably about 1 nM to about 40 nM, such as about 2 nM to about 20 nM.

[0050] In certain embodiments, each primer other than a boosting primer in a multiplex single primer extension reaction has an initial concentration of at least about 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, or 0.5 nM and/or at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nM, including any combination of the above-noted minimum and maximum concentrations.

[0051] Target nucleic acids useful as templates in single primer extension reactions disclosed herein may be any nucleic acids of interest, including genomic DNA, cDNA, genomic DNA treated with bisulfite, and the amplification products of the above-noted nucleic acids.

[0052] Target nucleic acids may be single-stranded, double-stranded, or partially double-stranded. In certain embodiments, target nucleic acids are partially double-stranded and comprise a double-stranded target nucleic acid sequence and a single-stranded common sequence at the 5' terminus of each strand of the target nucleic acid sequence. The single-stranded common sequence is the same among the target nucleic acids, and allows the design of a boosting primer as described below to increase or maximize the yield of extension products.

[0053] Target nucleic acids may be isolated from any samples, sources, or organisms of interest that contain nucleic acids, including materials obtained from clinical, forensic, and environmental settings. Preferably, the sample is a biological sample derived from a human, animal, plant, bacteria or fungi. The term "sample" also includes processed samples including preserved, fixed and/or stabilized samples, such as formalin fixed and paraffin-embedded (FFPE samples) and other samples that were treated with cross-linking fixatives such as glutaraldehyde.

[0054] Exemplary samples from which nucleic acids may be prepared include, but are not limited to, blood, swabs, body fluid, tissues including but not limited to, liver, spleen, kidney, lung, intestine, brain, heart, muscle, and pancreas, cell cultures, food samples, plant tissues or samples, as well as lysates, extracts, or materials and fractions obtained from the samples described above, or any cells, microorganisms and viruses that may be present on or in a sample, and the like.

[0055] Isolating target nucleic acids from a sample of interest may be performed by any method known in the art useful for nucleic acid isolation or purification. In addition, many kits for nucleic acid preparation are commercially available and may be used, including QIAamp DNA mini kit, QIAamp FFPE Tissue kit, and PAXgene DNA kit.

[0056] Alternatively, target nucleic acids useful as templates in single primer extension reactions disclosed herein may be obtained by modifying nucleic acids isolated from a sample. For example, in certain embodiments, target nucleic acids may be generated by (i) fragmenting DNA molecules (*e.g.*, genomic DNA) to obtain double-stranded DNA fragments, and (ii) ligating a single-stranded common sequence to the 5' terminus of each stand of the double-stranded DNA fragments.

[0057] As indicated above, in certain embodiments, a single primer extension reaction may further comprise a boosting primer. A boosting primer may be included in a multiplex single primer extension reaction where most (*i.e.*, at least about 90%, such as at least about 95%, about 98% or about 99%) or all of target nucleic acids each comprise a common sequence. The boosting primer comprises the common sequence or a portion thereof that is at least about 10 nucleotides in length, and is thus able to specifically anneal to a nucleic acid strand that comprises the complement of the common sequence and be extended using such a strand as a template in the single primer extension reaction. Thus, the presence of a boosting primer allows exponential amplification of target nucleic acids. However, because it comprises a sequence common to most or all of target nucleic acids, it does not affect specificity of the amplification. An exemplary use of a boosting primer is shown in Figure 1 as described below in detail.

[0058] The initial concentration of a boosting primer may be from about 10 nM to about 10 mM, preferably from about 100 nM to about 2 mM, such as about 200 nM to about 600 nM.

[0059] Single primer extension reactions disclosed herein are performed in the presence of a MutS protein.

[0060] A MutS protein is a protein that binds to heteroduplex DNA containing mispaired or unpaired bases or a small DNA insertion/deletion loops of one to a few (*e.g.*, 2, 3, 4, or 5) nucleotides and has sequence homology with *Escherichia coli* MutS protein (GenBank Accession No. GI: 127556, SEQ ID NO: 1). It includes members of the MutS family described in Eisen, Nucleic Acids Research 26: 4291-4300, 1998, especially those in the MutS-I lineage.

[0061] In certain embodiments, the MutS protein is thermostable, that is, capable of at least partially maintaining (*e.g.*,

at least about 10%, about 20%, about 30%, about 40%, or about 50% of) its activity of binding to heteroduplex DNA containing mispaired or unpaired bases or a small DNA insertion/deletion loop of one to a few nucleotides at a temperature of at least about 50°C, such as at least about 60°C, about 70°C, about 75°C, about 80°C, about 85°C, about 90°C, or about 95°C. Such MutS protein may be of thermophilic or hyperthermophilic bacterial origin. Exemplary thermostable MutS proteins include MutS proteins from *Aquifex aeolicus*, *Aquifex pyrophilus*, *Thermotoga maritima*, *Thermus thermophilus* and *Thermus aquaticus* (see Fukui et al., Int. J. Mol. Sci. 14:6436-6453, 2013; U.S. Patent No. 5,877,280; Biswas and Hsieh, Journal of Biological Chemistry 271 :5040-5048, 1996; and Takamatsu et al., Nucleic Acid Research 24:640-647, 1996).

[0062] In certain preferred embodiments, the MutS protein is the MutS protein from *Thermus aquaticus* (herein referred to as "Taq MutS") (GenBank Accession No. GI: 1203807, SEQ ID NO:2). In related embodiments, the MutS protein is a MutS protein with a sequence identity of at least about 30%, such as at least about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, or about 99%, with Taq MutS.

[0063] In certain preferred embodiments, the MutS protein is the MutS protein from *Thermus thermophilus* (herein referred to as "Tth MutS") (GenBank Accession No. GI: 1871501, SEQ ID NO:3). In related embodiments, the MutS protein is a MutS protein having a sequence identity of at least about 30%, such as at least about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, or about 99%, with Tth MutS.

[0064] In certain preferred embodiments, the MutS protein is the MutS protein from *Aquifex aeolicus* (herein referred to as "Aae MutS") (GenBank Accession No. GI: 2983001, SEQ ID NO:4). In related embodiments, the MutS protein is a MutS protein having a sequence identity of at least about 30%, such as at least about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, or about 99%, with Aae MutS.

[0065] In certain preferred embodiments, the MutS protein is the MutS protein from *Aquifex pyrophilus* (herein referred to as "Apy MutS") (GenBank Accession No. GI: 1619907, SEQ ID NO:5). In related embodiments, the MutS protein is a MutS protein having a sequence identity of at least about 30%, such as at least about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, or about 99%, with Apy MutS.

[0066] In certain preferred embodiments, the MutS protein is the MutS protein from *Thermotoga maritima* (herein referred to as "Tma MutS") (GenBank Accession No. GI: 1619909, SEQ ID NO:6). In related embodiments, the MutS protein is a MutS protein having a sequence identity of at least about 30%, such as at least about 40%, about 45%, about 50%, about 55%, about 60%, about 65%, about 70%, about 75%, about 80%, about 85%, about 90%, about 95%, about 96%, about 97%, about 98%, or about 99%, with Tma MutS.

[0067] For determining sequence identity, the WU-BLAST-2 program (Altschul et al., Methods in Enzymology 266:460-480, 1996) is used. This program uses several search parameters, most of which are set to the default values. The adjustable parameters are set with the following values: overlap span=1, overlap fraction=0.125, word threshold (T)=11. The HSP S and HSP S2 parameters are dynamic values and are established by the program itself depending upon the composition of the particular sequence and composition of the particular database against which the sequence of interest is being searched. A percent sequence identity value is determined by the number of matching identical residues divided by the total number of residues of the "longer" sequence in the aligned region. The "longer" sequence is the one having the most actual residues in the aligned region (gaps introduced by WU-Blast-2 to maximize the alignment score are ignored).

[0068] MutS proteins may be obtained from an organism in which they occur in nature. Alternatively, they can be produced by recombinant methods.

[0069] A MutS protein may be present in a highly multiplex single primer extension reaction at a concentration of about 0.05 to about 10 uM, such as about 0.1 to about 1 uM, or about 0.2 to about 0.5 uM.

[0070] In addition to a MutS protein, single primer extension as disclosed herein is performed in the presence of Mg²⁺ at a minimum concentration of about 6 mM.

[0071] In certain embodiments, the concentration of Mg²⁺ is at least about 6, at least about 7, at least about 8, at least about 9, or at least about 10 mM and/or at most about 15, at most about 16, at most about 17, at most about 18, at most about 19, at most about 20mM, including any combination of the above-noted minimum and maximum concentrations, such as about 6 mM to about 16 mM, about 6 mM to about 20 mM, about 8 mM to about 16 mM, about 8 mM to about 20 mM, about 10 mM to about 16 mM, or about 10 mM to about 20 mM.

[0072] Single primer extension reactions disclosed herein also comprise dNTPs. Typically, dNTPs are each present at an initial concentration (i.e., concentration at the start of the reaction) of about 40 uM to about 1000 uM in the reaction mixtures.

[0073] Single primer extension reactions disclosed herein also comprise a DNA polymerase for primer extensions. Preferably, DNA polymerases are thermostable. Exemplary DNA polymerases include Taq polymerase (from *Thermus*

aquaticus), Tfi polymerase (from *Thermus filiformis*), Bst polymerase (from *Bacillus stearothermophilus*), Pfu polymerase (from *Pyrococcus furiosus*), Tth polymerase (from *Thermus thermophilus*), Pow polymerase (from *Pyrococcus woesei*), Tli polymerase (from *Thermococcus litoralis*), Ultima polymerase (from *Thermotoga maritima*), KOD polymerase (from *Thermococcus kodakaraensis*), Pol I and II polymerases (from *Pyrococcus abyssi*) and Pab (from *Pyrococcus abyssi*), Amplitaq Gold® DNA polymerase (Applied Biosciences), Stoffel fragment of Amplitaq® DNA Polymerase (Roche), KOD polymerase (EMD Biosciences), KOD Hot Start polymerase (EMD Biosciences), Deep Vent™ DNA polymerase (New England Biolabs), Phusion polymerase (New England Biolabs), Klentaq1 polymerase (DNA Polymerase Technology, Inc), Klentaq Long Accuracy polymerase (DNA Polymerase Technology, Inc), Omni KlenTaq™ DNA polymerase (DNA Polymerase Technology, Inc), Omni KlenTaq™ LA DNA polymerase (DNA Polymerase Technology, Inc), Platinum® Taq DNA Polymerase (Invitrogen), Hemo Klentaq™ (New England Biolabs), Platinum® Taq DNA Polymerase High Fidelity (Invitrogen), Platinum® Pfx (Invitrogen), Accuprime™ Pfx (Invitrogen), and Accuprime™ Taq DNA Polymerase High Fidelity (Invitrogen).

[0074] In embodiments where two or more primers are able to anneal to closely located regions of a target nucleic acid, the DNA polymerase preferably does not have strand displacement activity, flap endonuclease or 5'→3' exonuclease activity, with which the polymerase may destroy downstream primer extension products.

[0075] In certain preferred embodiments, the DNA polymerase does not have 3'→5' exonuclease activity.

[0076] As disclosed above, the combination of a relatively high concentration of Mg²⁺ and a MutS improves primer annealing and extension specificity in a highly multiplex single primer extension reaction. In certain embodiments, the primer annealing and extension specificity is at least about 80%, about 85%, about 90% or about 95%.

[0077] The specificity of primer annealing and extension in highly multiplex single primer extension can be measured, after next generation sequencing, by the percentage of reads mapped to the targeted loci.

[0078] Also as disclosed above, the combination of a relatively high concentration of Mg²⁺ and a MutS also improves primer annealing and extension efficiency in a highly multiplex primer extension reaction. In certain embodiments, the average primer annealing and extension efficiency of all of the primers other than a boosting primer (if present) in a highly multiplex single primer extension reaction is at least about 5%, about 10%, about 15%, about 20%, or about 25%.

[0079] The efficiency of primer annealing of an individual enrichment primer in a highly multiplex single primer extension reaction can be measured by measuring the number of DNA copies generated by the individual enrichment primer after primer extension, such as by a qPCR assay. For example, in certain embodiments where each enrichment primer comprises at its 5' terminus a universal sequence that is incapable of specifically annealing to a target nucleic acid, one qPCR primer can be designed to specifically anneal to the universal sequence and another qPCR primer can be designed to have a sequence identical to a portion of the target nucleic acid 5' to the region where the enrichment primer anneals. Such a qPCR assay measures the number of new DNA molecules generated by primer extension. The ratio of new DNA molecules to the input DNA amount represents the percentage of input DNA being annealed and extended by the enrichment primer.

[0080] In certain embodiments, the primer annealing and extension specificity is at least about 80%, about 85%, about 90% or about 95% and the average primer annealing and extension efficiency is at least about 5%, about 10%, about 15%, about 20%, or about 25%, including all possible combinations of the above-noted specificity and efficiency.

[0081] Products of single primer extension reactions disclosed herein may be further amplified and/or analyzed (e.g., sequenced). The further amplification may be performed using any appropriate methods known in the art.

[0082] Sequencing of extension products or amplified extension products may also be performed using any appropriate methods known in the art. In certain embodiments, high throughput sequencing platforms known in the art may be used to sequence nucleic acids prepared from single primer extension products (see, Myllykangas et al., *Bioinformatics for High Throughput Sequencing*, Rodriguez-Ezpeleta et al. (eds.), Springer Science+Business Media, LLC, 2012, pages 11-25). Exemplary high throughput DNA sequencing systems include, but are not limited to, the GS FLX sequencing system originally developed by 454 Life Sciences and later acquired by Roche (Basel, Switzerland), Genome Analyzer developed by Solexa and later acquired by Illumina Inc. (San Diego, CA) (see, Bentley, *Curr Opin Genet Dev* 16:545-52, 2006; Bentley et al., *Nature* 456:53-59, 2008), the SOLiD sequence system by Life Technologies (Foster City, CA) (see, Smith et al., *Nucleic Acid Res* 38: e142, 2010; Valouev et al., *Genome Res* 18:1051-63, 2008), CGA developed by Complete Genomics and acquired by BGI (see, Drmanac et al., *Science* 327:78-81, 2010), PacBio RS sequencing technology developed by Pacific Biosciences (Menlo Park, CA) (see, Eid et al., *Science* 323: 133-8, 2009), and Ion Torrent developed by Life Technologies Corporation (see, U.S. Patent Application Publication Nos. 2009/0026082; 2010/0137143; and 2010/0282617).

[0083] The analysis of the extension products may be used in identifying structural variations (e.g., nucleotide mutations, deletions, insertions) in target nucleic acids when compared with corresponding reference sequences. In addition, sequencing extension products from cDNA is useful in detecting fusions and splicing variants. Furthermore, comparing sequences of extension products from bisulfite treated genomic DNA with those from untreated genomic DNA may be used in determining methylation patterns of the genomic DNA.

II. Using Single Primer Extension in NGS

[0084] In a related aspect, the present disclosure provides a method for using highly multiplex single primer extension reactions described above in high throughput nucleic acid sequencing. The method comprises:

a. providing a plurality of partially double-stranded target nucleic acids each of which comprises a double-stranded target nucleic acid sequence, and a single-stranded common sequence at the 5' terminus of each strand of the target nucleic acid sequence, wherein the single-stranded common sequence is the same among the plurality of the target nucleic acids; optionally wherein step a. comprises:

a1. fragmenting double-stranded DNA molecules to generate double-stranded DNA fragments, and
a2. ligating the single-stranded common sequence to the 5' terminus of each strand of the double-stranded DNA fragments to provide the plurality of partially double-stranded nucleic acids;

b. providing a plurality of primers capable of specifically annealing to portions of the target nucleic acid sequences;
c. extending the primers using the target nucleic acid sequences as templates in the presence of (1) a MutS protein and (2) Mg^{2+} at a minimum concentration of about 6 mM to obtain extension products; and
d. sequencing the extension products or their amplification products. The partially double-stranded target nucleic acids may be prepared from any nucleic acids of interest (e.g., genomic DNA and bisulfite-treated genomic DNA) isolated from any samples, sources, or organisms as described above in Section I.

[0085] In certain embodiments, extending the primers in step c. is performed in the presence of a boosting primer comprising the common sequence or a portion thereof that is at least about 10 nucleotides in length.

[0086] Steps a1 and a2, above, may be performed using any appropriate methods known in the art. For example, step a2 may be performed using an adapter with the 5' portion in single-stranded formation and containing the common sequence and its 3' end portion forming a short double-stranded structure with a complementary oligonucleotide. The double-stranded 3' end can be ligated to the 5' terminus of each strand of the double-stranded DNA fragments in a regular double-strand DNA ligation reaction.

[0087] The single-stranded common sequence comprises at least about 10 nucleotides (e.g., at least about 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 nucleotides) so that a boosting primer may be designed to specifically anneal to the common sequence.

[0088] The plurality of primers in step b., single primer extensions in step c., and sequencing in step b. may be performed as described above in Section I.

[0089] In certain embodiments, the method also comprises: prior to step d., a further step of amplifying the extension products of step c. The amplification may be performed using a primer at least substantially complementary to the common sequence or a portion of the common sequence that is at least about 10 nucleotides in length and another primer capable of specifically annealing to a universal sequence at the 5' ends of the primers that have been extended in the single primer extension reaction.

[0090] An exemplary method is shown in Figure 1. Briefly, in step 1, genomic DNA is first fragmented into smaller fragment (see step 1). In step 2, each fragment is then 5' ligated with a sequencing adapter ("NGSCounter") (an exemplary common sequence). In step 3, target-specific primers each having the same 5' universal sequence are annealed to the fragmented DNA and extended by DNA polymerase to the sequencing adapter ends. This process can be repeated multiple cycles to maximize the yield of single primer extension. An adapter primer ("boosting primer") can also be included to boost the yield of single primer extension. In step 4, the newly synthesized DNA strand, containing the 5' universal sequence and 3' adapter sequence, can be further amplified by a pair of an adapter primer (that comprises the boosting primer sequence or a portion thereof) and a universal primer (that comprises the universal sequence or a portion thereof) to prepare universal PCR library. In step 5, the library is quantified. In step 6, the nucleic acids in the library are sequenced.

III. Kits for Performing Single Primer Extensions

[0091] In another aspect, the present disclosure provides a kit for performing a highly multiplex single primer extension reaction, comprising: (1) a pool of at least 1,000 different primers, preferably at least about 10,000 different primers, (2) a buffer concentrate comprising a MutS protein, dNTPs, and Mg^{2+} , wherein the concentration of Mg^{2+} in a highly multiplex single primer extension reaction mixture generated from diluting the buffer concentrate is at least about 6 mM, and (3) a DNA polymerase.

[0092] The pool of primers in the kit may contain about 1,000 to about 100,000 different primers, preferably at least about 10,000 different primers. For example, a primer pool may contain at least about 1,000, 1,500, 2,000, 2,500, 3,000,

3,500, 4,000, 4,500, or 5,000 different primers, and/or at most about 10,000, 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000, or 100,000 different primers, including any combination of the above-noted minimum and maximum numbers of different primers such as about 1,000 to about 10,000, about 1,000 to about 50,000, about 1,000 to about 100,000, about 5,000 to about 20,000, about 5,000 to about 50,000, or about 5,000 to about 100,000 different primers.

[0093] The primers in the pool may preferably be at the same concentration or a substantially similar concentration (*i.e.*, the highest concentration of a primer is at most twice the lowest concentration of another primer).

[0094] Each primer in the pool may be about 0.01 nM to about 100 nM, preferably about 1 nM to about 40 nM, such as about 2 nM to about 20 nM.

[0095] In certain embodiments, each primer in the pool has a concentration of at least about 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, or 0.5 nM and/or at most about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nM, including any combination of the above-noted minimum and maximum concentrations.

[0096] The buffer concentrate in the kit comprises a MutS protein, dNTPs, and Mg^{2+} , wherein the concentration of Mg^{2+} in a highly multiplex single primer extension reaction mixture generated from diluting the buffer concentrate is at least about 6 mM, such as about 6 mM to about 20 mM.

[0097] In certain embodiments, the concentration of Mg^{2+} in a highly multiplex single primer extension reaction diluted from a buffer concentrate is at least about 6, at least about 7, at least about 8, at least about 9, or at least about 10 mM and/or at most about 15, at most about 16, at most about 17, at most about 18, at most about 19, at most about 20 mM, including any combination of the above-noted minimum and maximum concentrations, such as about 6 mM to about 16 mM, about 6 mM to about 20 mM, about 8 mM to about 16 mM, about 8 mM to about 20 mM, about 10 mM to about 16 mM, or about 10 mM to about 20 mM.

[0098] A MutS protein may be present in a highly multiplex single primer extension reaction diluted from a buffer concentrate at a concentration of about 0.05 to about 10 μ M, such as about 0.1 to about 1 μ M, or about 0.2 to about 0.5 μ M.

[0099] dNTPs are typically each present at a concentration of about 40 to about 1000 μ M in a highly multiplex single primer extension reaction diluted from a buffer concentrate.

[0100] Additional descriptions of various components of the kit, including primers, MutS proteins, dNTPs, DNA polymerases may be found above in section I.

[0101] The kit may further comprise one or more of the following additional components: a single-stranded sequence that may be ligated to the 5' terminus of each strand of double-stranded target nucleic acids as described above in Sections I and II, a boosting primer as described above in Sections I and II, a DNA ligase (*e.g.*, T4 DNA ligase and *E. coli* DNA ligase), a ligation buffer, a primer pair for amplifying single primer extension products (*e.g.*, those described above in Section II), and a sequencing primer.

[0102] In a related aspect, the present disclosure provides use of a kit disclosed herein in performing a highly multiplex primer extension reaction.

[0103] In another related aspect, the present disclosure provides use of a kit disclosed herein in nucleic acid sequencing.

EXAMPLES

EXAMPLE 1

EFFECTS OF Mg^{2+} CONCENTRATIONS ON PRIMER EXTENSION EFFICIENCY AND SPECIFICITY

[0104] Extension reactions of primers (each primer at a 2 nM concentration) were performed at Mg^{2+} concentrations from 2 mM to 20 mM. The primer extension reaction was set up in 1x miScript buffer supplemented with various amount of Mg^{2+} , 2.4 U HotStatTaq, 20 nM single target specific primer with 5' universal sequence and 20 ng genomic DNA. The reaction proceeded for 20 minutes at 95°C, 15 minutes at 55°C, 15 minutes at 65°C and 7 minutes at 72°C. Specific qPCR was used to quantify the amount of extension products.

[0105] The results (Figure 2) show that increasing the concentration of Mg^{2+} improved the primer annealing and extension efficiency.

[0106] DNA enriched using 12 mM Mg^{2+} and over 1300 primers at 2 nM each and DNA enriched using 4 mM Mg^{2+} and the same primers at 20 nM each were sequenced to determine primer annealing and extension specificity.

[0107] The results (Figure 3) show that the higher Mg^{2+} concentration significantly reduced the specificity of primer annealing. While not wishing to be bound by any particular theory, the present inventors believe that under higher Mg^{2+} concentrations, non-specific binding of primers to DNA templates was stabilized along with increased specific annealing.

EXAMPLE 2

EFFECTS OF MUTS ON SPECIFICITY OF PRIMER EXTENSIONS PERFORMED AT RELATIVELY HIGH Mg^{2+} CONCENTRATIONS

[0108] This example demonstrates that MutS improved specificity of highly multiplex targeted enrichment performed at a relatively high Mg^{2+} concentration.

MethodsFragmenting qDNA(sonication method)

[0109] 1 ug genomic DNA (gDNA) was diluted in 100ul elution buffer and sonicated on BIORUPTOR300 (power level low, 10 cycles of 30 seconds on and 30 seconds off). The resulting DNA fragments were purified using QIAGEN's MinElute Column.

Library Construction

[0110] A sequencing library was prepared using the above-generated gDNA fragments according to the End Repair, A-Addition, Adapter Ligation and Clean up and Size Selection of DNA protocols as described in QIAGEN GeneRead Library Prep for Illumina Kit. The library was quantified and diluted to 20ng/ul.

Primer Extension

[0111] The primer extension reactions were set up as follows:

Primer Extension		20 nM Primer Setup	2 nM Primer setup	
		Volume (ul)	Volume (ul)	
GeneRead V2 buffer (5x)		8	8	
dNTP (2 mM each)		4	4	
DNA library (from previous step, 20 ng/ul)		1	1	
Taq		1.6	1.6	
SPE Primer pool (76 nM)	20 nM final	10.5	1.05	2 nM final
Illumina_F primer (10 uM)		1.6	1.6	
MgCl ₂ (80mM)		0	4	
Taq MutS		0	3	
H ₂ O		13.3	15.75	
Total		40	40	

[0112] The final Mg^{2+} concentrations of 20 nM primer setup and 2 nM primer setup were 4 and 12 mM, respectively.

[0113] Primer extension reactions were performed in a thermal cycler as follows: 95°C for 15 minutes, 5 cycles of 95°C for 15 seconds and 68°C for 10 minutes, 72°C for 5 minutes, and 4°C hold.

[0114] 60 ul H₂O was added to each of the primer extension mixtures to bring the volume to 100 ul. The extension products were purified by 2 rounds Ampure beads purification (1.2X beads ratio) and eluted in 15ul Buffer EB.

Universal PCR

[0115] Primer extension products were further amplified in a universal PCR reaction as follows:

Universal qPCR		Volume (ul)
GeneRead V2 buffer (5x)		5
dNTP (2 mM each)		2.5
Template (from previous step)		11.5
Taq		1
Universal primer (IL2) (4 uM)		2.5

(continued)

	Universal qPCR	Volume (ul)
5	IL1_ID(#)-RS2 primer (4 uM)	2.5
	Total	25

[0116] The reaction was performed in a thermal cycler as follows: 95°C for 15 minutes, 26 cycles of 95°C for 15 seconds and 65°C for 2 minutes, and 4°C hold.

10 [0117] The amplification products were purified via one round Ampure beads purification (1.2X beads ratio) and eluted in 20 ul buffer EB. 5 ul of the purified amplification products were checked with a 1.2% Agarose gel.

Library Quantification and Sequencing

15 [0118] Quantification of the amplified library and sequencing were performed according to the manufacturer's instructions.

Results

20 [0119] The results (Figure 4) show that adding Taq MutS in primer extension reactions together with 12 mM Mg²⁺ significantly improve primer annealing and extension specificity at no cost of primer annealing and extension efficiency as indicated by consistently good sequence uniformity.

[0120] T50 is an indicator of sequence uniformity. It is defined as the percentage of total sequence throughput used by regions sequenced at below 50% of the mean read depth of all target regions. In a perfect uniform scenario, the T50 value is 50.

25 [0121] While not wishing to be bound by any particular theory, the present inventors hypothesize that non-specific primer-template hybrids formed during primer extension performed at a high Mg²⁺ concentration presumably contain mismatches and bulges in the middle of the primer sequences, but perfectly anneal at the 3' termini of the primers so that DNA polymerase can extend from such termini. MutS specifically binds to those bulges created by sequence mismatches and thus block a DNA polymerase from binding to and extending the non-specific primer-template duplex.
30 The specific primer-template duplex would have complementary sequences and no bulge in the structure so that MutS would not bind tightly. Thus, extension of primers that specifically bind to their templates will not be affected.

Claims

- 35 1. A method for performing a highly multiplex single primer extension reaction, comprising:
- 40 a. extending at least about 1,000 different primers in a single primer extension reaction using a plurality of target nucleic acids as templates in the presence of (1) a MutS protein, and (2) Mg²⁺ at a minimum concentration of about 6 mM to generate extension products.
2. The method of claim 1, wherein at least about 4,000 or at least about 9,000 additional different primers are extended in the single primer extension reaction.
- 45 3. The method of claim 1 or 2, wherein the single primer extension reaction is performed in the absence of a primer capable of specifically annealing to a portion of the extension products.
- 50 4. The method of any one of claims 1 to 3, wherein most of the target nucleic acids comprise a common sequence, and wherein the reaction is performed in the presence of a boosting primer that comprises the common sequence or a portion thereof that is at least about 10 nucleotides in length.
- 55 5. The method of any one of claims 1 to 4, wherein each of the primers other than the boosting primer is present in the single primer extension reaction at a concentration of no more than about 20 nM, preferably at a concentration of about 1 nM to about 10 nM; and/or wherein Mg²⁺ is present in the single primer extension reaction at a concentration of about 6 mM to about 20 mM.
6. The method of any one of claims 1 to 5, wherein the plurality of target nucleic acids are:

- (i) genomic DNA or amplification products of genomic DNA;
- (ii) cDNA or amplification products of cDNA; or
- (iii) bisulfite treated genomic DNA or its amplified products.

5 **7.** The method of any one of claims 1 to 6, wherein

- (i) the primer annealing and extension specificity of the single primer extension reaction is at least about 80%, and/or
- 10 (ii) the average primer annealing and extension efficiency of the single primer extension reaction is at least about 10%.

8. The method of any one of claims 1 to 7, further comprising:

- 15 b. optionally amplifying the extension product(s) to generate amplified extension product(s), and
- c. sequencing the extension product(s) or the amplified extension products.

9. The method of any one of claims 1 to 8, wherein the target nucleic acids are partially double-stranded, and comprise a double-stranded target nucleic acid sequence and a single-stranded common sequence at the 5' terminus of each strand of the target nucleic acid sequence, optionally wherein the method comprises, prior to step a.,

- 20 (i) fragmenting DNA molecules to generate double-stranded DNA fragments, and
- (ii) ligating the single-stranded common sequence to the 5' terminus of each strand of the double-stranded DNA fragment to provide partially double-stranded target nucleic acids.

25 **10.** The method of any one of claims 1 to 9, wherein each primer other than the boosting primer if present comprises at its 5' terminus a universal sequence that is incapable of specifically annealing to a region of a target nucleic acid.

11. A method for sequencing nucleic acids, comprising:

30 a. providing a plurality of partially double-stranded target nucleic acids each of which comprises a double-stranded target nucleic acid sequence, and a single-stranded common sequence at the 5' terminus of each strand of the target nucleic acid sequence, wherein the single-stranded common sequence is the same among the plurality of the target nucleic acids; optionally wherein step a. comprises:

- 35 a1. fragmenting double-stranded DNA molecules to generate double-stranded DNA fragments, and
- a2. ligating the single-stranded common sequence to the 5' terminus of each strand of the double-stranded DNA fragments to provide the plurality of partially double-stranded nucleic acids;

- 40 b. providing a plurality of primers capable of specifically annealing to portions of the target nucleic acid sequences;
- c. extending the primers using the target nucleic acid sequences as templates in the presence of (1) a MutS protein and (2) Mg²⁺ at a minimum concentration of about 6 mM to obtain extension products; and
- d. sequencing the extension products or their amplification products.

45 **12.** The method of claim 11, wherein extending the primers in step c. is performed in the presence of a boosting primer comprising the common sequence or a portion thereof that is at least about 10 nucleotides in length.

13. The method of claim 11 or 12, comprising prior to step d., a further step of amplifying the extension products of step c.

50 **14.** A kit for performing a primer extension reaction, comprising:

- (1) a pool of at least about 1,000 different primers, preferably at least about 10,000 different primers,
- (2) a buffer concentrate comprising a MutS protein, dNTPs, and Mg²⁺, wherein the concentration of Mg²⁺ in a highly multiplex single primer extension reaction mixture generated from diluting the buffer concentrate is at least about 6 mM, and
- 55 (3) a DNA polymerase.

15. The method of any one of claims 1 to 6 or the kit of claim 14, wherein the MutS protein is a *Thermus aquaticus* MutS protein.

Patentansprüche

1. Ein Verfahren zum Durchführen einer hochmultiplexen Einzelprimer-Verlängerungsreaktion, umfassend:
 - a. Verlängern von mindestens etwa 1.000 verschiedenen Primern in einer einzigen Primer-Verlängerungsreaktion unter Verwendung einer Vielzahl von Zielnukleinsäuren als Vorlagen in Gegenwart von (1) einem MutS-Protein und (2) Mg^{2+} in einer Mindestkonzentration von etwa 6 mM, um Verlängerungsprodukte zu erzeugen.
2. Das Verfahren nach Anspruch 1, wobei mindestens etwa 4.000 oder mindestens etwa 9.000 zusätzliche unterschiedliche Primer in der Einzelprimer-Verlängerungsreaktion verlängert werden.
3. Das Verfahren nach Anspruch 1 oder 2, wobei die Einzelprimer-Verlängerungsreaktion in Abwesenheit eines Primers durchgeführt wird, der in der Lage ist, spezifisch an einen Teil der Verlängerungsprodukte zu binden (annealing).
4. Das Verfahren nach einem der Ansprüche 1 bis 3, wobei die meisten der Zielnukleinsäuren eine gemeinsame Sequenz umfassen, und wobei die Reaktion in Gegenwart eines Boosting-Primers durchgeführt wird, der die gemeinsame Sequenz oder einen Teil davon umfasst, der mindestens etwa 10 Nukleotide lang ist.
5. Das Verfahren nach einem der Ansprüche 1 bis 4, wobei jeder andere Primer als der Boosting-Primer in der Einzelprimer-Verlängerungsreaktion in einer Konzentration von nicht mehr als etwa 20 nM, vorzugsweise in einer Konzentration von etwa 1 nM bis etwa 10 nM, vorliegt; und/oder wobei Mg^{2+} in der Einzelprimer-Verlängerungsreaktion in einer Konzentration von etwa 6 mM bis etwa 20 mM vorliegt.
6. Das Verfahren nach einem der Ansprüche 1 bis 5, wobei die Vielzahl der Zielnukleinsäuren sind:
 - i) genomische DNA oder Amplifikationsprodukte genomischer DNA;
 - ii) cDNA oder Amplifikationsprodukte von cDNA; oder
 - iii) mit Bisulfit behandelte genomische DNA oder ihre amplifizierten Produkte.
7. Das Verfahren nach einem der Ansprüche 1 bis 6, wobei
 - (i) die Primer-Annealing- und Verlängerungsspezifität der Einzelprimer-Verlängerungsreaktion mindestens etwa 80 % beträgt, und/oder
 - (ii) die durchschnittliche Primer-Annealing- und Verlängerungseffizienz der einzelnen Primer-Verlängerungsreaktion mindestens etwa 10% beträgt.
8. Das Verfahren nach einem der Ansprüche 1 bis 7, weiter umfassend:
 - b. optional Verstärken des/der Verlängerungsprodukt(e), um ein amplifiziertes Verlängerungsprodukt(e) zu erzeugen, und
 - c. Sequenzieren des/der Verlängerungsprodukt(e) oder der amplifizierten Verlängerungsprodukte.
9. Das Verfahren nach einem der Ansprüche 1 bis 8, wobei die Zielnukleinsäuren teilweise doppelsträngig sind und eine doppelsträngige Zielnukleinsäuresequenz und eine einzelsträngige gemeinsame Sequenz am 5'-Terminus jedes Strangs der Zielnukleinsäuresequenz umfassen, optional wobei das Verfahren vor Schritt a., umfasst
 - i) Fragmentieren von DNA-Molekülen, um doppelsträngige DNA-Fragmente zu erzeugen, und
 - ii) Ligieren der einzelsträngigen gemeinsamen Sequenz an den 5'-Terminus jedes Strangs des doppelsträngigen DNA-Fragments, um teilweise doppelsträngige Zielnukleinsäuren bereitzustellen.
10. Das Verfahren nach einem der Ansprüche 1 bis 9, wobei jeder andere Primer als der Boosting-Primer, falls vorhanden, an seinem 5'-Ende eine universelle Sequenz umfasst, die nicht in der Lage ist, spezifisch in einen Bereich einer Zielnukleinsäure zu binden (annealing).
11. Ein Verfahren zum Sequenzieren von Nukleinsäuren, umfassend:
 - a. Bereitstellen einer Vielzahl von teilweise doppelsträngigen Zielnukleinsäuren, wobei jede eine doppelsträngige Zielnukleinsäuresequenz umfasst, und eine einzelsträngige gemeinsame Sequenz am 5'-Terminus jedes

Strangs der Zielnukleinsäuresequenz, wobei die einzelsträngige gemeinsame Sequenz gleich ist unter der Vielzahl von Zielnukleinsäuren; optional wobei Schritt a. umfasst:

- 5 a1. Fragmentieren doppelsträngiger DNA-Moleküle, um doppelsträngige DNA-Fragmente zu erzeugen, und
a2. Ligieren der einzelsträngigen gemeinsamen Sequenz an den 5'-Terminus jedes Strangs der doppelsträngigen DNA-Fragmente, um die Vielzahl von teilweise doppelsträngigen Nukleinsäuren bereitzustellen;
- b. Bereitstellen einer Vielzahl von Primern, die in der Lage sind, spezifisch an Abschnitte der Zielnukleinsäuresequenzen zu binden (annealing);
- 10 c. Verlängern der Primer unter Verwendung der Zielnukleinsäuresequenzen als Templates in Gegenwart von (1) einem MutS-Protein und (2) Mg^{2+} in einer Mindestkonzentration von etwa 6 mM, um Verlängerungsprodukte zu erhalten; und
- d. Sequenzierung der Verlängerungsprodukte oder ihrer Amplifikationsprodukte.
- 15 **12.** Das Verfahren nach Anspruch 11, wobei das Verlängern der Primer in Schritt c. in Gegenwart eines Boosting-Primers durchgeführt wird, der die gemeinsame Sequenz oder einen Teil davon umfasst, der mindestens etwa 10 Nukleotide lang ist.
- 13.** Das Verfahren nach Anspruch 11 oder 12, umfassend vor dem Schritt d. einen weiteren Schritt des Amplifizierens der Verlängerungsprodukte des Schritts c.
- 20 **14.** Ein Kit zum Durchführen einer Primer-Verlängerungsreaktion, umfassend:
 - (1) einen Pool von mindestens etwa 1.000 verschiedenen Primern, vorzugsweise mindestens etwa 10.000 verschiedenen Primern,
 - (2) ein Pufferkonzentrat, das ein MutS-Protein, dNTPs und Mg^{2+} umfasst, wobei die Konzentration von Mg^{2+} in einem hochmultiplex Einzelprimer-Verlängerungsreaktionsgemisch, das durch Verdünnen des Pufferkonzentrats erzeugt wird, mindestens etwa 6 mM beträgt, und
 - (3) eine DNA-Polymerase.
- 30 **15.** Das Verfahren nach einem der Ansprüche 1 bis 6 oder das Kit nach Anspruch 14, wobei das MutS-Protein ein *Thermus aquaticus* MutS-Protein ist.

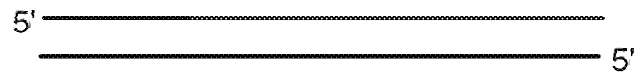
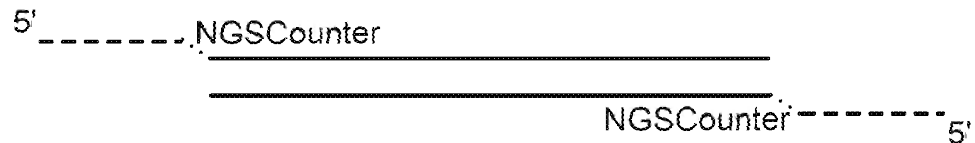
35 Revendications

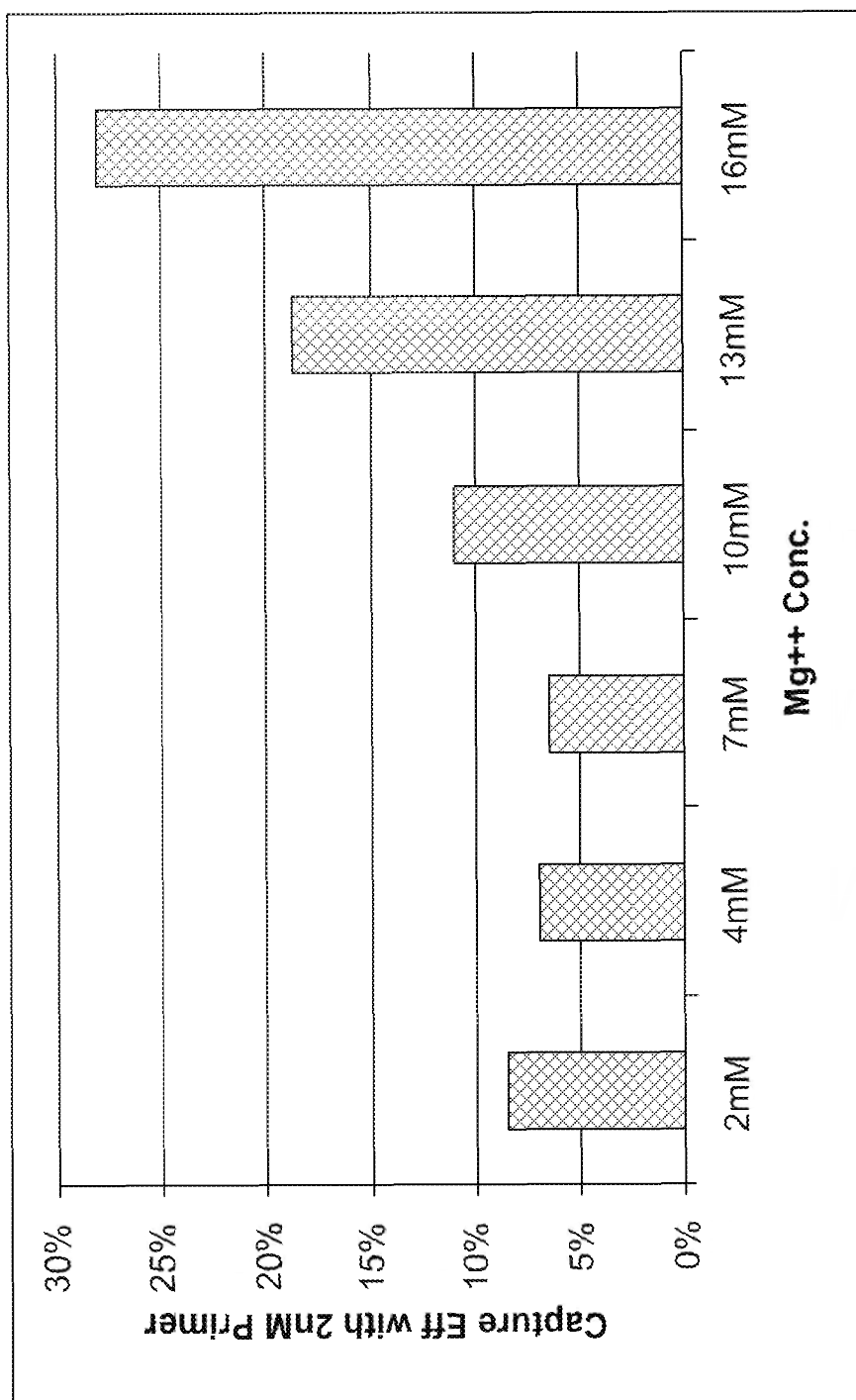
1. Procédé pour effectuer une réaction d'extension d'amorce unique hautement multiplex, comprenant :
 - 40 a. l'extension d'au moins environ 1 000 amorces différentes dans une réaction d'extension d'amorce unique en utilisant une pluralité d'acides nucléiques cibles comme matrices en présence (1) d'une protéine MutS et (2) de Mg^{2+} à une concentration minimale d'environ 6 mM pour générer des produits d'extension.
2. Procédé selon la revendication 1, dans lequel au moins environ 4 000 ou au moins environ 9 000 amorces différentes supplémentaires sont étendues dans la réaction d'extension d'amorce unique.
- 45 3. Procédé selon la revendication 1 ou 2, dans lequel la réaction d'extension d'amorce unique est réalisée en l'absence d'une amorce capable de s'hybrider spécifiquement à une partie des produits d'extension.
- 50 4. Procédé selon l'une quelconque des revendications 1 à 3, dans lequel la plupart des acides nucléiques cibles comprennent une séquence commune, et dans lequel la réaction est effectuée en présence d'une amorce de renforcement qui comprend la séquence commune ou une partie de celle-ci qui est d'une longueur d'au moins 10 nucléotides environ.
- 55 5. Procédé selon l'une quelconque des revendications 1 à 4, dans lequel chacune des amorces autres que l'amorce de renforcement est présente dans la réaction d'extension d'amorce unique à une concentration ne dépassant pas environ 20 nM, de préférence à une concentration d'environ 1 nM à environ 10 nM ; et/ou dans lequel Mg^{2+} est présent dans la réaction d'extension d'amorce unique à une concentration d'environ 6 mM à environ 20 mM.

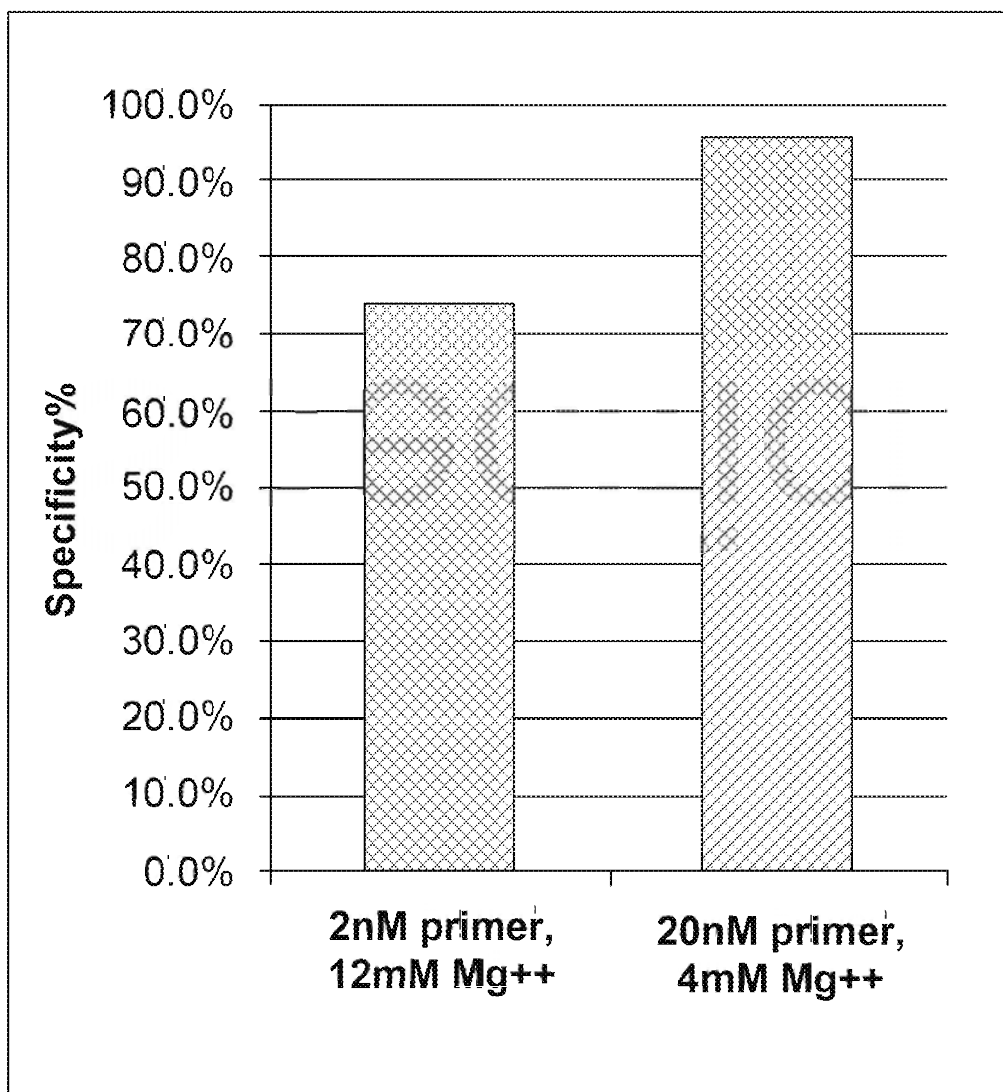
6. Procédé selon l'une quelconque des revendications 1 à 5, dans lequel la pluralité d'acides nucléiques cibles sont :
- (i) de l'ADN génomique ou des produits d'amplification d'ADN génomique ;
 - (ii) de l'ADNc ou des produits d'amplification d'ADNc ; ou
 - (iii) de l'ADN génomique traité au bisulfite ou ses produits amplifiés.
7. Procédé selon l'une quelconque des revendications 1 à 6, dans lequel
- (i) la spécificité d'hybridation et d'extension d'amorce de la réaction d'extension d'amorce unique est d'au moins 80 % environ, et/ou
 - (ii) l'efficacité moyenne d'hybridation et d'extension d'amorce de la réaction d'extension d'amorce unique est d'au moins 10 % environ.
8. Procédé selon l'une quelconque des revendications 1 à 7, comprenant en outre :
- b. éventuellement l'amplification du ou des produit(s) d'extension pour générer un ou des produit(s) d'extension amplifié(s), et
 - c. le séquençage du ou des produit(s) d'extension ou des produits d'extension amplifiés.
9. Procédé selon l'une quelconque des revendications 1 à 8, dans lequel les acides nucléiques cibles sont partiellement double brin, et comprennent une séquence d'acide nucléique cible double brin et une séquence commune simple brin à l'extrémité 5' de chaque brin de la séquence d'acide nucléique cible, éventuellement dans lequel le procédé comprend, avant l'étape a.,
- (i) la fragmentation de molécules d'ADN pour générer des fragments d'ADN double brin, et
 - (ii) la ligature de la séquence commune simple brin à l'extrémité 5' de chaque brin du fragment d'ADN double brin pour obtenir des acides nucléiques cibles partiellement double brin.
10. Procédé selon l'une quelconque des revendications 1 à 9, dans lequel chaque amorce autre que l'amorce de renforcement, si elle est présente, comprend à son extrémité 5' une séquence universelle incapable de s'hybrider spécifiquement à une région d'un acide nucléique cible.
11. Procédé de séquençage d'acides nucléiques, comprenant :
- a. la fourniture d'une pluralité d'acides nucléiques cibles partiellement double brin comprenant chacun une séquence d'acide nucléique cible double brin et une séquence commune simple brin à l'extrémité 5' de chaque brin de la séquence d'acide nucléique cible, la séquence commune simple brin étant la même dans la pluralité d'acides nucléiques cibles ; éventuellement dans lequel l'étape a. comprend :
 - a1. la fragmentation de molécules d'ADN double brin pour générer des fragments d'ADN double brin, et
 - a2. la ligature de la séquence commune simple brin à l'extrémité 5' de chaque brin des fragments d'ADN double brin pour fournir la pluralité d'acides nucléiques partiellement double brin ;
 - b. la fourniture d'une pluralité d'amorces capables de s'hybrider spécifiquement à des parties des séquences d'acide nucléique cibles ;
 - c. l'extension des amorces en utilisant les séquences d'acides nucléiques cibles comme matrices en présence (1) d'une protéine MutS et (2) de Mg^{2+} à une concentration minimale d'environ 6 mM pour obtenir des produits d'extension ; et
 - d. le séquençage des produits d'extension ou de leurs produits d'amplification.
12. Procédé selon la revendication 11, dans lequel l'extension des amorces à l'étape c. est effectuée en présence d'une amorce de renforcement comprenant la séquence commune ou une partie de celle-ci qui est d'une longueur d'au moins 10 nucléotides environ.
13. Procédé selon la revendication 11 ou 12, comprenant avant l'étape d. une étape supplémentaire d'amplification des produits d'extension de l'étape c.
14. Kit pour la réalisation d'une réaction d'extension d'amorce, comprenant :

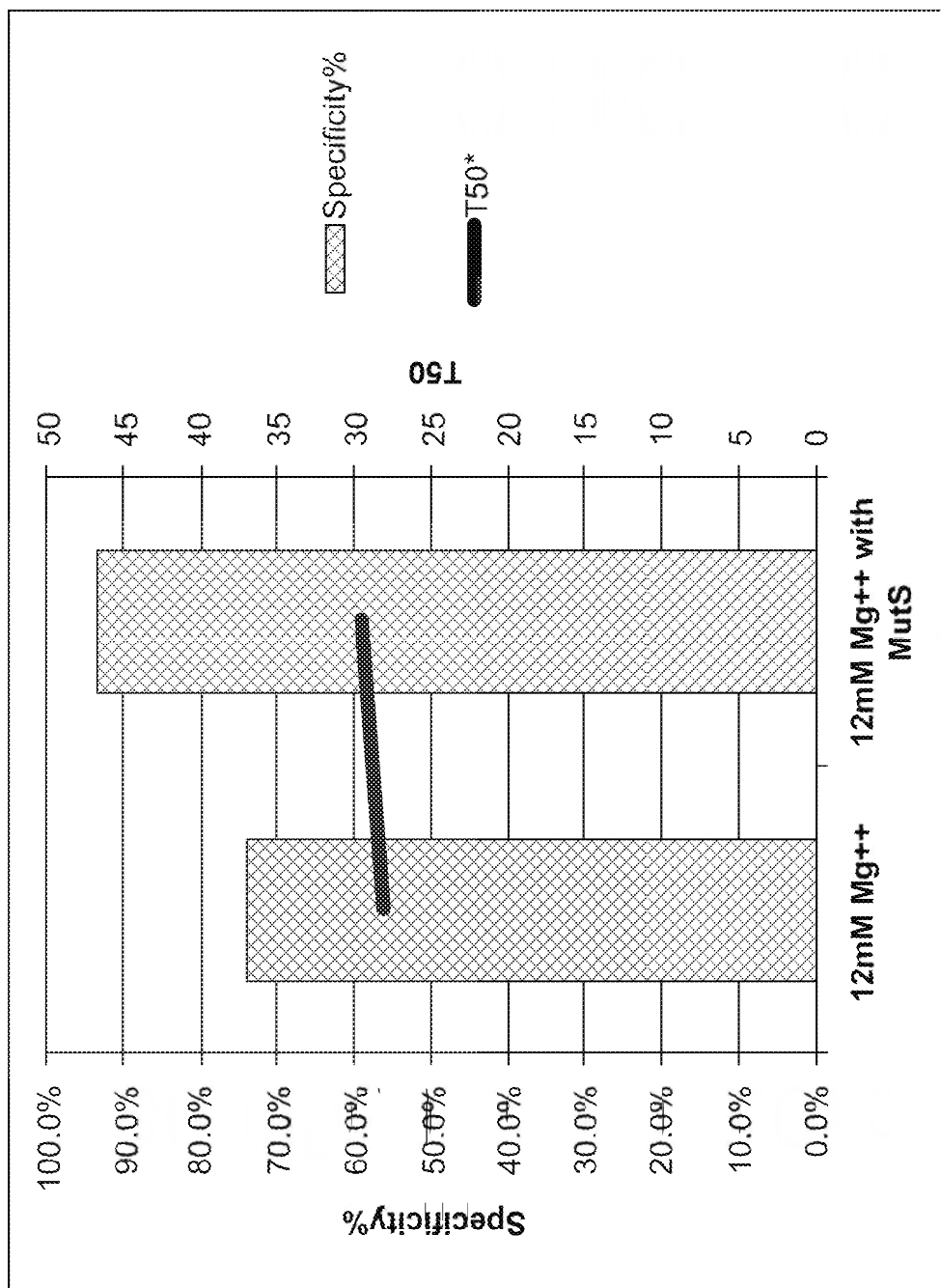
- (1) un ensemble d'au moins environ 1 000 amorces différentes, de préférence au moins environ 10 000 amorces différentes,
- (2) un concentré de tampon comprenant une protéine MutS, des dNTP et du Mg^{2+} , la concentration de Mg^{2+} dans un mélange réactionnel d'extension d'amorce unique hautement multiplex généré par la dilution du concentré de tampon étant d'au moins environ 6 mM, et
- (3) une ADN polymérase.

15. Procédé selon l'une quelconque des revendications 1 à 6 ou kit selon la revendication 14, dans lequel la protéine MutS est une protéine MutS de *Thermus aquaticus*.

1. DNA random fragmentation**2. Library construction with NGSCounter****3. Single primer extension****4. Universal PCR library****5. Library quantification****6. Sequencing****FIG. 1**

**FIG. 2**

**FIG. 3**

**FIG. 4**

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 62242766 [0001]
- US 5877280 A [0061]
- US 20090026082 [0082]
- US 20100137143 A [0082]
- US 20100282617 A [0082]

Non-patent literature cited in the description

- **PENG et al.** *BMC Genomics*, 2015, vol. 16, 589 [0008]
- **WATANABE et al.** *Biotechniques*, 2007, vol. 43, 479-84 [0009]
- **MITANI et al.** *Nature Methods*, 2007, vol. 4, 257-62 [0010]
- **FUKUI et al.** *International Journal of Molecular Sciences*, 2013, vol. 14, 6436-53 [0011]
- **EISEN.** *Nucleic Acids Research*, 1998, vol. 26, 4291-4300 [0060]
- **FUKUI et al.** *Int. J. Mol. Sci.*, 2013, vol. 14, 6436-6453 [0061]
- **BISWAS ; HSIEH.** *Journal of Biological Chemistry*, 1996, vol. 271, 5040-5048 [0061]
- **TAKAMATSU et al.** *Nucleic Acid Research*, 1996, vol. 24, 640-647 [0061]
- **ALTSCHUL et al.** *Methods in Enzymology*, 1996, vol. 266, 460-480 [0067]
- **MYLLYKANGAS et al.** *Bioinformatics for High Throughput Sequencing*. Springer Science+Business Media, LLC, 2012, 11-25 [0082]
- **BENTLEY.** *Curr Opin Genet Dev*, 2006, vol. 16, 545-52 [0082]
- **BENTLEY et al.** *Nature*, 2008, vol. 456, 53-59 [0082]
- **SMITH et al.** *Nucleic Acid Res*, 2010, vol. 38, e142 [0082]
- **VALOUEV et al.** *Genome Res*, 2008, vol. 18, 1051-63 [0082]
- **DRMANAC et al.** *Science*, 2010, vol. 327, 78-81 [0082]
- **EID et al.** *Science*, 2009, vol. 323, 133-8 [0082]