

B-FAMILY DNA POLYMERASE VARIANT AND KIT COMPRISING THE SAME

CROSS REFERENCE

This application claims priority to, and the benefit of, U.S. Provisional Application
5 No. US63/249,813, filed on September 29, 2021, the content thereof is incorporated by
reference herein.

FIELD OF THE INVENTION

The present invention relates to, among other things, B-family DNA polymerase
variants and kits comprising the same for use particularly in the context of nucleic acid
10 sequencing by synthesis.

SEQUENCE LISTING

The present application is being filed along with a Sequence Listing in electronic
format. The Sequence Listing is provided as a file entitled 2022-06-28-PCT-SEQ.txt, created
on June 28, 2022, which is 126 kb in size. The information in the electronic format of
15 Sequence Listing is incorporated herein by reference in its entirety.

BACKGROUND OF THE INVENTION

Next-generation sequencing (NGS) technologies have revolutionized modern
biological and biomedical research. The engines responsible for this innovation are DNA
polymerases, which catalyze the biochemical reaction for deriving template sequence
20 information. As one of the widely adopted NGS platforms, the DNA sequencing-by-synthesis
(SBS) technique is a unique, polymerase-dependent approach for that the consecutive
sequencing reaction thereof concurrently generates a newly synthesized DNA strand as a
result.

The general SBS approach involves (i) incorporation of nucleotide analogues bearing
25 fluorescent reporters, (ii) identification of the incorporated nucleotide by its fluorescent
emissions, and (iii) cleavage of the fluorophore, along with the reinitiation of the polymerase
reaction for stepwise sequence determination. In contrast to conventional Sanger sequencing,
the fluorescently labeled DNA fragments of different sizes are generated in a single reaction
and, thus, required DNA separation by electrophoresis followed with a fluorescent detection.
30 The SBS approach is more robust than Sanger's dideoxy-sequencing method because

individual DNA fragments, or clusters, for sequence determination are immobilized on a high-density array. Hence, the SBS fluorescent signals from each DNA fragments or clusters can be directly detected without the need for prior DNA separations. Furthermore, by using massive parallel microarray chip, the SBS approach can easily be scaled up to increase overall sequencing output and greatly reduces the cost of DNA sequencing.

Nevertheless, the nucleotides used in the SBS approach are variedly modified. These nucleotide analogues carry either a larger chemical blocking group at the 3'-hydroxyl (3'-OH) position of deoxyribose or a bulky fluorescent molecule on the nucleobase. These large, chemical modifications on the nucleotide are not well-tolerated by naturally occurring nucleic acid polymerase. The active-site pocket of nucleic acid polymerase is pre-arranged in a proper geometry for accommodating a correct and matched canonical nucleotide having a normal 3'-OH group and a nucleobase. The elimination, or any substitution, of the 3'-OH group with a bulky, chemical group on the nucleotide, such as 2', 3'-dideoxynucleotide (ddNTP) and 3'-O-azidomethyl-dNTP, respectively, significantly alters the nucleotide configuration within the active-site pocket of nucleic acid polymerase and reduces the nucleotide binding affinity and overall DNA synthesis efficiency of nucleic acid polymerase. Likewise, any modifications on the nucleobase or 5'-triphosphate group of the nucleotide, disrupts the interactions between the nucleotide and active-site residues of nucleic acid polymerase and leads to a poor utilization of these modified nucleotides for nucleic acid synthesis by the polymerase.

In view of that a series of nucleotide modifications are constantly created for rapidly changing DNA polymerase-based sequencing technologies, it is crucial to look for, design or evolve compatible enzymes for ever-changing DNA sequencing chemistries, such as US 11104888B2 disclosed a variant of archaeal 9°N DNA polymerase (a B-family Pol), which has been found to incorporate reversible dye-terminator nucleotides well.

SUMMARY OF THE INVENTION

Owing to the diverse properties of modified nucleotide analogues, the naturally occurring nucleic acid polymerases cannot readily utilize these nucleotide analogues as a substrate for the template-directed nucleic acid synthesis. Thus, the tailor-made, modified nucleic acid polymerase is a prerequisite for exerting the utilities of these non-canonical nucleotides for a variety of nucleic acid sequencing-by-synthesis applications.

The inventor has discovered the novel positions/regions in the amino acid sequences of B-family DNA polymerases that play crucial parts in enhancing the substrate affinity of said polymerases for modified nucleotides, thereby improving the nucleotide-incorporation efficiency in the DNA sequencing-by-synthesis method.

5 Accordingly, in one aspect, the invention provides a B-family DNA polymerase variant comprising: a motif Exo I, a motif Exo II, a motif Exo III, a motif A, a motif B, and a motif C corresponding respectively to positions 349 to 364, 450 to 476, 590 to 608, 706 to 730, 843 to 855, and 940 to 956 of a consensus sequence (SEQ ID NO:1); at least one amino acid substitution at a position residing in a motif selected from the group consisting of the
10 motif Exo I, the motif Exo II, and the motif Exo III; and at least one amino acid substitution at a positions residing in a motif selected from the group consisting of the motif A, the motif B, and the motif C.

In one embodiment, the B-family DNA polymerase variant is modified from a wild-type B-family DNA polymerase having an amino acid sequence selected from the group
15 consisting of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17.

In one embodiment, the wild-type B-family DNA polymerase is *Thermococcus gorgonarius* DNA polymerase (Tgo), *Thermococcus kodakarensis* DNA polymerase (Kod1), *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N), *Pyrococcus furiosus* DNA polymerase (Pfu), *Thermococcus litoralis* DNA polymerase (Vent), *Methanococcus maripaludis* DNA polymerase (Mma), *Methanosarcina acetivorans* DNA polymerase (Mac),
20 human DNA polymerase delta catalytic p125 subunit (hPOLD), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD), *Pyrobaculum islandicum* DNA polymerase, (Pis) *Sulfolobus solfataricus* DNA polymerase (Sso), *Pseudomonas aeruginosa* DNA polymerase II (Pae), *Escherichia coli* DNA polymerase II (Eco), Bacteriophage
25 (*Escherichia* phage) RB69 DNA polymerase (RB69), Bacteriophage (*Escherichia* phage) T4 DNA polymerase (T4), or Bacteriophage (*Bacillus* phage) Phi29 DNA polymerase (Phi29).

In one embodiment, the amino acid of corresponding to position 354 of SEQ ID NO:1 is any amino acid other than D; the amino acid corresponding to position 355 of SEQ ID NO:1 is any amino acid other than I; the amino acid corresponding to position 356 of SEQ ID
30 NO:1 is any amino acid other than E; the amino acid corresponding to position 715 of SEQ ID

NO:1 is any amino acid other than L; the amino acid corresponding to position 716 of SEQ ID NO:1 is any amino acid other than Y; and the amino acid of corresponding to position 717 of SEQ ID NO:1 is any amino acid other than P.

In one embodiment, the amino acid corresponding to position 354 of SEQ ID NO:1 is
5 any amino acid other than D; the amino acid corresponding to position 355 of SEQ ID NO:1 is any amino acid other than I; the amino acid corresponding to position 356 of SEQ ID NO:1 is any amino acid other than E; the amino acid corresponding to position 715 of SEQ ID NO:1 is any amino acid other than L; the amino acid corresponding to position 716 of SEQ ID NO:1 is any amino acid other than Y; the amino acid corresponding to position 717 of SEQ ID
10 NO:1 is any amino acid other than P; and the amino acid corresponding to position 854 of SEQ ID NO:1 is any amino acid other than A.

In one embodiment, the B-family DNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO:3; and wherein: D141 of SEQ ID NO:3 corresponding to position
15 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:3 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L408 of SEQ ID NO:3 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y409 of SEQ ID NO:3 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and P410 of SEQ ID NO:3 corresponding to position 717 of SEQ ID NO:1 is substituted with A,
20 C, G, S or T.

In one embodiment, the B-family DNA polymerase variant is derived from *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N) having a wild-type amino acid sequence of SEQ ID NO:4; and wherein: D141 of SEQ ID NO:4 corresponding to position
25 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:4 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L408 of SEQ ID NO:4 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y409 of SEQ ID NO:4 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and P410 of SEQ ID NO:4 corresponding to position 717 of SEQ ID NO:1 is substituted with A,
C, G, S or T.

30 In one embodiment, the B-family DNA polymerase variant is derived from

Pyrococcus furiosus DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO:5; and wherein: D141 of SEQ ID NO:5 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:5 corresponding to position 356 of SEQ ID NO:1 is substituted with A; I409 of SEQ ID NO:5 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y410 of SEQ ID NO:5 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and P411 of SEQ ID NO:5 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

In one embodiment, the B-family DNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO:6; and wherein: D141 of SEQ ID NO:6 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:6 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L411 of SEQ ID NO:6 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y412 of SEQ ID NO:6 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and P413 of SEQ ID NO:6 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

In one embodiment, the B-family DNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO:8; and wherein: D198 of SEQ ID NO:8 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E200 of SEQ ID NO:8 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L485 of SEQ ID NO:8 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y486 of SEQ ID NO:8 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and P487 of SEQ ID NO:8 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

In one embodiment, the B-family DNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO:11; and wherein: D171 of SEQ ID NO:11 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E173 of SEQ ID NO:11 corresponding to position 356 of SEQ ID NO:1 is substituted with A; M426 of SEQ ID NO:11 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y427 of SEQ ID NO:11 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and

P428 of SEQ ID NO:11 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

In one embodiment, the B-family DNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO:12; and wherein: D231 of SEQ ID NO:12 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E233 of SEQ ID NO:12 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L518 of SEQ ID NO:12 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y519 of SEQ ID NO:12 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; an P520 of SEQ ID NO:12 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

In one embodiment, the B-family DNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO:3; and wherein: D141 of SEQ ID NO:3 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:3 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L408 of SEQ ID NO:3 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y409 of SEQ ID NO:3 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; P410 of SEQ ID NO:3 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and A485 of SEQ ID NO:3 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

In one embodiment, the B-family DNA polymerase variant is derived from *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N) having a wild-type amino acid sequence of SEQ ID NO:4; and wherein: D141 of SEQ ID NO:4 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:4 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L408 of SEQ ID NO:4 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y409 of SEQ ID NO:4 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; P410 of SEQ ID NO:4 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and A485 of SEQ ID NO:4 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

In one embodiment, the B-family DNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO:5; and wherein: D141 of SEQ ID NO:5 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:5 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L409 of SEQ ID NO:5 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y410 of SEQ ID NO:5 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; P411 of SEQ ID NO:5 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and A486 of SEQ ID NO:5 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

In one embodiment, the B-family DNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO:6; and wherein: D141 of SEQ ID NO:6 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:6 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L411 of SEQ ID NO:6 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y412 of SEQ ID NO:6 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; P413 of SEQ ID NO:6 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and A488 of SEQ ID NO:6 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

In one embodiment, the B-family DNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO:8; and wherein: D198 of SEQ ID NO:8 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E200 of SEQ ID NO:8 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L485 of SEQ ID NO:8 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y486 of SEQ ID NO:8 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; P487 of SEQ ID NO:8 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and A565 of SEQ ID NO:8 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

In one embodiment, the B-family DNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO:11; and wherein: D171 of SEQ ID NO:11 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E173 of SEQ ID NO:11 corresponding to position 356 of SEQ ID NO:1 is substituted with A; M426 of SEQ ID NO:11 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y427 of SEQ ID NO:11 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; P428 of SEQ ID NO:11 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and A508 of SEQ ID NO:11 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

In one embodiment, the B-family DNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO:12; and wherein: D231 of SEQ ID NO:12 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E233 of SEQ ID NO:12 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L518 of SEQ ID NO:12 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y; Y519 of SEQ ID NO:12 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; P520 of SEQ ID NO:12 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and A601 of SEQ ID NO:12 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

The present invention further provides a kit for performing a sequencing-by-synthesis reaction, comprising: a B-family DNA polymerase variant as described above, a primer, and nucleotide analogues under conditions suitable for incorporation into the primer, thereby determining the nucleobase complementary to the nucleotide analogues incorporated in the primer.

Accordingly, the present invention relates to the specific B-family DNA polymerase variants that exhibit an improved incorporation of nucleotide analogues for synthesizing polynucleotides and sequencing the associated nucleic acid template. More particularly, the DNA sequencing-by-synthesis method can be efficiently performed by said B-family DNA polymerase variants with a normal nucleic acid template and reversible dye-terminator

nucleotides to precisely determine the sequence of associated nucleic acid template.

BRIEF DESCRIPTION OF THE DRAWINGS

In order to make the above and other objects, features, advantages and embodiments of the present invention more obvious and understandable, the drawings are described as follows:

Figure 1 illustrates a schematic diagram of a general SBS approach.

Figure 2 shows an amino acid sequence alignment of various wild-type B-family DNA polymerases (PolB) related to the present invention and their consensus sequence.

Figure 3 depicts a general structure of a dye-labelled nucleotide analogue with a removable 3'-blocking moiety.

Figure 4 depicts the structures of 3'-O-N₃-dNTP.

Figure 5A to 5D, respectively, depict the structures of N₃-dATP-linker-Atto532, N₃-dCTP-linker-Atto700, N₃-dGTP-linker-Cy5 and N₃-dTTP-linker-ROX.

Figure 6 shows the result of the reaction described in Example 4.

Figure 7 shows the result of the reaction described in Example 6.

Figure 8 shows the result of the reaction described in Example 7.

Figure 9 shows the result of the reaction described in Example 8.

Figure 10 shows the result of the reaction described in Example 9.

DETAILED DESCRIPTION OF THE INVENTION

TERMINOLOGY

The terms used in this specification are generally within the scope of the present invention and the specific context of each term has its usual meaning in related fields. The specific terms used to describe the present invention in this specification will be described below or elsewhere in this specification, so as to help people in the industry understand the relevant description of the present invention. The same term has the same scope and meaning in the same context. In addition, there is more than one way to express the same thing; therefore, the terms discussed in this article may be replaced by alternative terms and synonyms, and whether a term is specified or discussed in this article does not have any special meaning. This article provides synonyms for certain terms, but the use of one or more synonyms does not mean that other synonyms are excluded.

As used herein, unless the context clearly indicates otherwise, "a" and "the" can also be interpreted as plural. In addition, in this specification and the scope of the attached patent application, unless otherwise stated in the context, "middle" and "inner" include "located in"; and unless otherwise stated in the context, the direction of the tip of the projectile was defined
5 as "upper" or "lower". Furthermore, titles and subtitles may be attached to the description for easy reading, but these titles do not affect the scope of the present invention.

As used herein, an "amino acid" refers to any monomer unit that can be incorporated into a peptide, polypeptide, or protein. As used herein, the term "amino acid" includes the following twenty natural or genetically encoded alpha-amino acids: alanine (Ala or A),
10 arginine (Arg or R), asparagine (Asn or N), aspartic acid (Asp or D), cysteine (Cys or C), glutamine (Gln or Q), glutamic acid (Glu or E), glycine (Gly or G), histidine (His or H), isoleucine (Ile or I), leucine (Leu or L), lysine (Lys or K), methionine (Met or M), phenylalanine (Phe or F), proline (Pro or P), serine (Ser or S), threonine (Thr or T), tryptophan (Trp or W), tyrosine (Tyr or Y), and valine (Val or V). In cases where "X"
15 residues are undefined, these should be defined as "any amino acid".

As used herein, the term "nucleic acid" or "polynucleotide" refers to a polymer that can be corresponded to a ribose nucleic acid (RNA) or deoxyribose nucleic acid (DNA) polymer, or an analog thereof. This includes polymers of nucleotides such as RNA and DNA, as well as synthetic forms, modified (e.g., chemically or biochemically modified) forms
20 thereof, and mixed polymers (e.g., including both RNA and DNA subunits). Exemplary modifications include methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (e.g., methyl phosphonates, phosphotriesters, phosphoamidates, carbamates, and the like), pendentmoieties (e.g., polypeptides), intercalators (e.g., acridine, psoralen, and the like),
25 chelators, alkylators, and modified linkages (e.g., alpha anomeric nucleic acids and the like). Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Typically, the nucleotide monomers are linked via phosphodiester bonds, although synthetic forms of nucleic acids can comprise other linkages (e.g., peptide nucleic acids as described in Nielsen
30 et al. (*Science* 254:1497-1500, 1991). A nucleic acid can be or can include, e.g., a

chromosome or chromosomal segment, a vector (e.g., an expression vector), an expression cassette, a naked DNA or RNA polymer, the product of a polymerase chain reaction (PCR), an oligonucleotide, a probe, and a primer. A nucleic acid can be, e.g., single-stranded, double-stranded, or triple-stranded and is not limited to any particular length. Unless otherwise
5 indicated, a particular nucleic acid sequence optionally comprises or encodes complementary sequences, in addition to any sequence explicitly indicated.

As used herein, the term "DNA sequencing by synthesis (SBS)" refers to a polymerase-dependent, next-generation sequencing (NGS) approach. As illustrated in Fig.1, the SBS approach substantially involves the following processes: (i) incorporation of
10 nucleotide analogues bearing labels (e.g., fluorescent reporters); optionally, the labeled nucleotides include a reversible termination property that terminates further primer extension once a nucleotide has been added to a primer; (ii) identification of the incorporated nucleotide by its labels (e.g., fluorescent emissions); and (iii) cleavage of the labels (e.g., fluorophore and/or the terminator), along with the reinitiation of the polymerase reaction for step-wise
15 sequence determination. Some SBS embodiments include detection of a proton released upon incorporation of a nucleotide into an extension product. For example, sequencing based on detection of released protons can use an electrical detector and associated techniques that are commercially available from Ion Torrent (Guilford, CT, a Life Technologies subsidiary) or sequencing methods and systems described in US Pat. App. Pub. Nos. 2009/0026082 A1;
20 2009/0127589 A1; 2010/0137143 A1; or 2010/0282617 A1, each of which is incorporated herein by reference.

As used herein, the term "template" is a polynucleotide, or a polynucleotide mimic, that contains the desired or unknown target nucleotide sequence. In some instances, the terms "target sequence", "template polynucleotide", "target nucleic acid", "target polynucleotide",
25 "nucleic acid template", "template sequence", and variations thereof, are used interchangeably. Specifically, the term "template" refers to a strand of nucleic acid on which a complementary copy is synthesized from nucleotides or nucleotide analogues through the replication of a template-dependent or template-directed nucleic acid polymerase. Within a nucleic acid duplex, the template strand is, by the convention definition, depicted and
30 described as the "bottom" strand. Similarly, the non-template strand is often depicted and

described as the "top" strand. The "template" strand may also be referred to as the "sense" or "plus" strand and the non-template strand as the "antisense" or "minus" strand.

As used herein, the term "primer" refers to a short single-stranded oligonucleotide, a polynucleotide, or a modified nucleic acid analogue used by nucleic acid polymerase as an initiator to elongate a nucleic acid chain.

As used herein, the terms "nucleotide incorporation", "analogue incorporation", "incorporating nucleotide" and "incorporating analogue" are known to those skilled in the art and used to describe a process or reaction for nucleic acid synthesis. Thus, as used herein, the term "incorporation" is known to flexibly refer to add one, or more nucleotides, or any specified nucleic acid precursors to the 3'-hydroxyl terminus of a nucleic acid primer. For example, the nucleoside triphosphate, such as deoxyguanosine triphosphate (dGTP), is a substrate, or a precursor, for DNA synthesis by DNA polymerase. Once the dGTP is incorporated into the elongated DNA strand, it becomes a deoxyguanosine monophosphate (dGMP) moiety of the newly synthesized DNA. In other words, when a dGTP nucleotide is converted into a dGMP moiety of DNA, the person skilled in the art may say that one dGTP is incorporated into the DNA.

As used herein, the term "nucleotide analogue" is known to those of skills in the art to describe the chemically modified nucleotides or artificial nucleotides, which are structural mimics of canonical nucleotides. These nucleotide analogues can serve as substrates for nucleic acid polymerases to synthesize nucleic acid. A nucleotide analogue may have one or more altered components of a nucleotide (e.g., the phosphate backbone, pentose sugar, and nucleobase), which changes the structure and configuration of a nucleotide and affects its interactions with other nucleobases and the nucleic acid polymerases. For example, a nucleotide analogue having an altered nucleobase may confer alternative base-pairing and base-stacking properties in the DNA or RNA. Furthermore, by way of example, the modification at the base may generate various nucleosides such as inosine, methyl-5-deoxycytidine, deoxyuridine, dimethylamino-5-deoxyuridine, diamino-2,6-purine or bromo-5-deoxyuridine, and any other analogues which permits hybridization. In other exemplary aspects, modifications may take place at the level of sugar moiety (for example, replacement of a deoxyribose by an analogue), and/or at the level of the phosphate group (for example,

boronate, alkylphosphonate, or phosphorothioate derivatives). A nucleotide analogue monomer may have a phosphate group selected from a monophosphate, a diphosphate, a triphosphate, a tetraphosphate, a pentaphosphate, and a hexaphosphate. Other examples of nucleotide analogues also include nucleotides having a removable blocking moiety. Examples of the removable blocking moiety include, but are not limited to, a 3'-O-blocking moiety, a base blocking moiety, and a combination thereof. Examples of the 3'-O-blocking moiety include, but are not limited to, O-N₃, O-azidomethyl, O-amino, O-allyl, O-phenoxyacetyl, O-methoxyacetyl, O-acetyl, O-(p-toluene)sulfonate, O-phosphate, O-nitrate, O-[4-methoxy]-tetrahydrothiopyranyl, O-tetrahydrothiopyranyl, O-[5-methyl]-tetrahydrofuranlyl, O-[2-methyl, 4-methoxy]-tetrahydropyranyl, O-[5-methyl]-tetrahydropyranyl, and O-tetrahydrothiofuranlyl, O-2-nitrobenzyl, O-methyl, and O-acyl. Examples of the base blocking moiety may be a reversible dye-terminator. Examples of the reversible dye-terminator include, but are not limited to, a reversible dye-terminator of Illumina MiSeq, a reversible dye-terminator of Illumina HiSeq, a reversible dye-terminator of Illumina Genome Analyzer IIX, a reversible dye-terminator of Helicos Biosciences Heliscope, and a reversible dye-terminator of LaserGen's Lightning Terminators.

As used herein, "B-family DNA polymerases (PolBs)" refers to the most common template-dependent nucleic acid polymerases or replicases in all domains of life and many DNA viruses. Like most nucleic acid polymerases, natural PolBs require a duplex primer-template DNA with a free 3'-hydroxyl (3'-OH) group at the primer terminus, all four nucleoside triphosphates (dATP, dTTP, dCTP, and dGTP), and catalytic, divalent cations (Mg²⁺ or Mn²⁺, etc.) for catalyzing the nucleotidyl transferase reaction of adding nucleotides to the 3'-OH terminus of a primer. The PolB enzymes, such as bacterial Pol II and archaeal B-family DNA polymerases, are replicative and repair polymerases that inherently contain a catalytic polymerase domain and a 3'→5' exonucleolytic, or proofreading, domain for removing the mis-incorporated nucleotide from the growing primer strand during nucleic acid replication. The term "3'→5' exonucleolytic domain" refers to a region of the amino acid sequence of a polymerase, which exerts the nucleic acid degradation activity from the 3'-terminus of the primer or polynucleotide chain. Coordinately, the term "catalytic polymerase domain" refers to a region of the amino acid sequence of a polymerase, which exerts the

catalytic DNA/RNA polymerase activity for adding nucleotides to the 3'-terminus of a primer or polynucleotide chain.

All known structures of PolB catalytic polymerase domain resemble the shape of human right hand, where the key functional regions are characterized as fingers, palm, and thumb subdomains, respectively. The most conserved region is the palm subdomain, which contains the essential residues for catalysis. The protein sequence-alignment among various B-family DNA polymerases from different kingdoms of life and DNA viruses reveals that the PolB polymerases generally harbor six semi-conserved or conserved motifs (I–VI) for their essential exonuclease and polymerase functions. The first three sequence-motifs (Exo I, Exo II, Exo III) are in the 3'→5' exonucleolytic domain, while the other three motifs (designated as Motif A, B, and C, respectively) reside in the polymerase domain (Hopfner et al, Proc. Natl. Acad. Sci. USA 96, 3600–3605, 1999).

As used herein, the term "mutant" in the context of DNA polymerases of the present invention, means a polypeptide, typically recombinant, that comprises one or more amino acid substitutions relative to a corresponding, functional DNA polymerase.

As used herein, in the context of DNA polymerase variants, "corresponding to another sequence" (e.g., regions, fragments, nucleotide or amino acid positions, or the like) is based on the convention of numbering according to nucleotide or amino acid position number and then aligning the sequences in a manner that maximizes the percentage of sequence identity. An amino acid "corresponding to position X of specific sequence" refers to an amino acid in a polypeptide of interest that aligns with the equivalent amino acid of a specified sequence. Generally, as described herein, the amino acid corresponding to a position of a polymerase can be determined using an alignment algorithm such as BLAST. Because not all positions within a given "corresponding region" need to be identical, non-matching positions within a corresponding region may be regarded as "corresponding positions". Accordingly, as used herein, referral to an "amino acid position corresponding to amino acid position X of a specified DNA polymerase" refers to equivalent positions, based on the alignment, in other DNA polymerases and structural homologues and families.

As used herein, the term "semi-conserved" used herein refers to the segment of polymerase that has a similar property of amino acid residue or an identical amino acid

residue in the homologous position of different PolBs from various sources. The term "conserved" means the segment of polymerase having the same amino acid residue in the homologous position of different PolBs from various sources.

As used herein, the term "consensus sequence of SEQ ID NO: 1" used herein refers to a reference sequence comprising the conserved amino acids of cross-species B-family DNA polymerase. The consensus sequence of SEQ ID NO: 1 is a virtual sequence and is generated by aligning the following 16 wild type B-family DNA polymerases to obtain the conserved amino acids: *Thermococcus gorgonarius* DNA polymerase (Tgo), *Thermococcus kodakarensis* DNA polymerase (Kod1), *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N), *Pyrococcus furiosus* DNA polymerase (Pfu), *Thermococcus litoralis* DNA polymerase (Vent), *Methanococcus maripaludis* DNA polymerase (Mma), *Methanosarcina acetivorans* DNA polymerase (Mac), human DNA polymerase delta catalytic p125 subunit (hPOLD), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD), *Pyrobaculum islandicum* DNA polymerase (Pis), *Sulfolobus solfataricus* DNA polymerase (Sso), *Pseudomonas aeruginosa* DNA polymerase II (Pae), *Escherichia coli* DNA polymerase II (Eco), Bacteriophage (*Escherichia* phage) RB69 DNA polymerase (RB69), Bacteriophage (*Escherichia* phage) T4 DNA polymerase (T4), and Bacteriophage (*Bacillus* phage) Phi29 DNA polymerase (Phi29). The positions of motifs Exo I, Exo II, Exo III, Motif A, B, and C are defined by the inventor using the consensus sequence of SEQ ID NO: 1 of the present invention; therefore, it shall be noted that the positions of these motifs defined in the present invention are not totally the same as those described in the prior art.

OBJECTIVES

In general, to be considered and ultimately applied for a SBS approach, the DNA polymerase should possess the following properties: (i) The polymerase must be a DNA-dependent DNA polymerase; in other words, the polymerase requires a DNA template for replication. (ii) The polymerase should rapidly incorporate nucleotides; DNA polymerases with a faster nucleotide incorporation rate can catalyze DNA synthesis much more efficiently. (iii) The polymerase must have a high replicative fidelity to minimize systematic errors; in other words, the polymerase should be able to accurately read DNA template sequence information and faithfully incorporate the correct, matched nucleotides along the DNA

template. (iv) The polymerase should possess a long, intrinsic, and replicative processivity. As used herein, the processivity of DNA polymerase is defined as the number of dNTP or the nucleotide analogue incorporated during complex formation with a primer/template (P/T) DNA. (v) The polymerase should function as a monomer, for that it can be produced and
5 further modified more easily (Chen, C.Y. DNA Polymerases Drive DNA Sequencing-by-Synthesis Technologies: Both Past and Present. *Front. Microbiol.* 2014;5:305. doi: 10.3389/fmicb.2014.00305).

Therefore, one objective of the subject application is to provide B-family DNA polymerase (PolB) variants based on the above rationale, so that the PolB variants have an
10 improved incorporation efficiency of nucleotide analogues for synthesizing polynucleotides and sequencing the associated nucleic acid template. The aforesaid enzymatic nucleic acid synthesis and sequencing can be efficiently performed by these PolB variants in the presence of a nucleic acid template and nucleotide analogues. More particularly, the aforesaid enzymatic nucleic acid synthesis and sequencing can be efficiently performed by these PolB
15 variants with a normal nucleic acid template and reversible dye-terminator nucleotides to precisely determine the sequence of this associated nucleic acid template. For instance, the PolB variants may depend on a normal, unmodified nucleic acid template to direct synthesis and incorporation of modified nucleotide analogues to the 3'-terminus of primer. The incorporated nucleotide carrying a distinct fluorescent reporter can be used to determine the
20 DNA sequence information of the nucleic acid template.

The inventor has discovered that the PolB variants have a broader substrate specificity, which means the altered PolB variants can utilize not only naturally occurring nucleotides, but also varieties of modified nucleotides for the template-dependent nucleic acid synthesis. Therefore, these PolB variants may improve the nucleotide-incorporation efficiency in the
25 DNA sequencing-by-synthesis method to determine the sequence of nucleic acids.

B-FAMILY DNA POLYMERASE VARIANT

Therefore, provided herein are altered polymerase, which is described based on the amino acid sequence of the consensus sequence of SEQ ID NO: 1. An altered polymerase includes substitution mutations at one or more residues when compared to the consensus
30 sequence of SEQ ID NO: 1. A substitution mutation can be at the same position or a

functionally equivalent position compared to the consensus sequence of SEQ ID NO: 1. The skilled person will readily appreciate that an altered polymerase described herein is not naturally occurring. Therefore, an altered polymerase described herein is based on the consensus sequence of SEQ ID NO: 1 and further includes substitution mutations at one or
5 more residues. In one embodiment, a substitution mutation is at a position functionally equivalent to an amino acid of the consensus sequence of SEQ ID NO: 1. "Functionally equivalent" means that the altered polymerase has the amino acid substitution at the amino acid position in the consensus sequence of SEQ ID NO: 1 that has the same functional role in both the consensus sequence and the altered polymerase.

10 In general, functionally equivalent substitution mutations in two or more different polymerases occur at homologous amino acid positions in the amino acid sequences of the polymerases. Hence, "functionally equivalent" also encompasses mutations that are "positionally equivalent" or "homologous" to a given mutation, regardless of whether the particular function of the mutated amino acid is known. It is possible to identify the locations
15 of functionally equivalent and positionally equivalent amino acid residues in the amino acid sequences of two or more different polymerases on the basis of sequence alignment and/or molecular modelling. For instance, the amino acid sequence-alignment of 16 wild-type B-family DNA polymerases to identify positionally equivalent and/or functionally equivalent residues is set forth in FIG. 2. Thus, for the exemplary residue 141 of the Tgo, Kod1, 9°N,
20 Pfu, and Vent DNA polymerases and the residue 198 of the Mac DNA polymerase are functionally equivalent and positionally equivalent. Likewise, for the exemplary residue 143 of the Tgo, Kod1, 9°N, Pfu, and Vent DNA polymerases and the residue 200 of the Mac polymerase are functionally equivalent and positionally equivalent. The skilled person can easily identify functionally equivalent residues in other DNA polymerases.

25 In accordance with some embodiments of the present invention, the provided B-family DNA polymerase variant comprising: a motif Exo I, a motif Exo II, a motif Exo III, a motif A, a motif B, and a motif C corresponding respectively to positions 349 to 364, 450 to 476, 590 to 608, 706 to 730, 843 to 855, and 940 to 956 of the consensus sequence of SEQ ID NO:1; at least one amino acid substitution (one or more amino acid substitutions, or a
30 combination of amino acid substitutions) at positions residing in the motif Exo I, the motif

Exo II, and the motif Exo III; and at least one amino acid substitution (one or more amino acid substitutions, or a combination of amino acid substitutions), at positions residing in the motif A, the motif B, and the motif C.

In accordance with some embodiments of the present invention, the B-family DNA polymerase variant is modified from a wild-type B-family DNA polymerase having an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17, which are derived from the wild-type B-family DNA polymerase including *Thermococcus gorgonarius* DNA polymerase (Tgo), *Thermococcus kodakarensis* DNA polymerase (Kod1), *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N), *Pyrococcus furiosus* DNA polymerase (Pfu), *Thermococcus litoralis* DNA polymerase (Vent), *Methanococcus maripaludis* DNA polymerase (Mma), *Methanosarcina acetivorans* DNA polymerase (Mac), human DNA polymerase delta catalytic p125 subunit (hPOLD), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD), *Pyrobaculum islandicum* DNA polymerase (Pis), *Sulfolobus solfataricus* DNA polymerase (Sso), *Pseudomonas aeruginosa* DNA polymerase II (Pae), *Escherichia coli* DNA polymerase II (Eco), Bacteriophage (*Escherichia*) phage RB69 DNA polymerase (RB69), Bacteriophage (*Escherichia* phage) T4 DNA polymerase (T4), or Bacteriophage (*Bacillus* phage) Phi29 DNA polymerase (Phi29), respectively.

In accordance with some embodiments of the present invention, the amino acid corresponding to position 354 of SEQ ID NO:1 is any amino acid other than D; for example, said amino acid can be substituted with N, Q, S, T, Y, A, C, G, I, L, M, F, P, W, V, R, H, K or E. The amino acid corresponding to position 355 of SEQ ID NO:1 is any amino acid other than I; for example, said amino acid can be substituted with N, Q, S, T, Y, A, C, G, L, M, F, P, W, R, H, K, D or E. The amino acid corresponding to position 356 of SEQ ID NO:1 is any amino acid other than E; for example, said amino acid can be substituted with N, Q, S, T, Y, A, C, G, I, L, M, F, P, W, V, R, H, K or D. The amino acid corresponding to position 715 of SEQ ID NO:1 is any amino acid other than L or M; for example, said amino acid can be substituted with N, Q, S, T, Y, A, C, G, I, F, P, W, V, R, H, K, D or E. The amino acid corresponding to position 716 of SEQ ID NO:1 is any amino acid other than Y; for example, said amino acid can be substituted with N, Q, S, T, A, C, G, I, L, M, F, P, W, V, R, H, K, D or

E. The amino acid of corresponding to position 717 of SEQ ID NO:1 is any amino acid other than P; for example, said amino acid can be substituted with N, Q, S, T, Y, A, C, G, I, L, M, F, W, V, R, H, K, D or E. In a preferred embodiment of the present invention, the amino acid corresponding to position 854 of SEQ ID NO:1 is any amino acid other than A; for example, said amino acid can be substituted with N, Q, S, T, Y, C, G, I, L, M, F, P, W, V, R, H, K, D or E.

In accordance with some embodiments of the present invention, the B-family DNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO:3; and wherein: D141 of SEQ ID NO:3 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:3 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L408 of SEQ ID NO:3 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y, preferably A, Q or Y, more preferably Y; Y409 of SEQ ID NO:3 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I, preferably A; and P410 of SEQ ID NO:3 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, F, G, S or T, preferably F or G, more preferably G. In certain embodiments, A485 of SEQ ID NO:3 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L, preferably E or L, more preferably L.

In accordance with some embodiments of the present invention, the B-family DNA polymerase variant is derived from *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N) having a wild-type amino acid sequence of SEQ ID NO:4; and wherein: D141 of SEQ ID NO:4 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:4 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L408 of SEQ ID NO:4 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y, preferably A, Q or Y, more preferably Y; Y409 of SEQ ID NO:4 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I, preferably A; and P410 of SEQ ID NO:4 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, F, G, or S or T, preferably F or G, more preferably G. In certain embodiments, A485 of SEQ ID NO:4 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L, preferably E or L, more preferably L.

In accordance with some embodiments of the present invention, the B-family DNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO:5; and wherein: D141 of SEQ ID NO:5 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:5 corresponding to position 356 of SEQ ID NO:1 is substituted with A; I409 of SEQ ID NO:5 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y, preferably A, Q or Y, more preferably Y; Y410 of SEQ ID NO:5 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I, preferably A; and P411 of SEQ ID NO:5 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, F, S or T, preferably F or G, more preferably G. In certain embodiments, A486 of SEQ ID NO:5 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L, preferably E or L, more preferably L.

In accordance with some embodiments of the present invention, the B-family DNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO:6; and wherein: D141 of SEQ ID NO:6 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E143 of SEQ ID NO:6 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L411 of SEQ ID NO:6 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y, preferably A, Q or Y, more preferably Y; Y412 of SEQ ID NO:6 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I, preferably A; and P413 of SEQ ID NO:6 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, F, G, S or T, preferably F or G, more preferably G. In certain embodiments, A488 of SEQ ID NO:6 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L, preferably E or L, more preferably L.

In accordance with some embodiments of the present invention, the B-family DNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO:8; and wherein: D198 of SEQ ID NO:8 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E200 of SEQ ID NO:8 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L485 of SEQ ID NO:8 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H

or Y, preferably A, Q or Y, more preferably Y; Y486 of SEQ ID NO:8 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I, preferably A; and P487 of SEQ ID NO:8 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, F, G, S or T, preferably F or G, more preferably G. In certain embodiments, A565 of SEQ ID NO:8 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L, preferably E or L, more preferably L.

In accordance with some embodiments of the present invention, the B-family DNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO:11; and wherein: D171 of SEQ ID NO:11 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E173 of SEQ ID NO:11 corresponding to position 356 of SEQ ID NO:1 is substituted with A; M426 of SEQ ID NO:11 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y, preferably A, Q or Y, more preferably Y; Y427 of SEQ ID NO:11 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I, preferably A; and P428 of SEQ ID NO:11 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, F, G, S or T, preferably F or G, more preferably G. In certain embodiments, A508 of SEQ ID NO:11 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L, preferably E or L, more preferably L.

In accordance with some embodiments of the present invention, the B-family DNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO:12; and wherein: D231 of SEQ ID NO:12 corresponding to position 354 of SEQ ID NO:1 is substituted with A; E233 of SEQ ID NO:12 corresponding to position 356 of SEQ ID NO:1 is substituted with A; L518 of SEQ ID NO:12 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y, preferably A, Q or Y, more preferably Y; Y519 of SEQ ID NO:12 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I, preferably A; an P520 of SEQ ID NO:12 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, F, G, S or T, preferably F or G, more preferably G. In certain embodiments, A601 of SEQ ID NO:12 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L, preferably E or L, more preferably L.

MUTATING POLYMERASES

Various types of mutagenesis techniques are optionally used in the present disclosure, e.g., to modify polymerases to produce the variants of the subject application, or using random or semi-random mutational approaches. In general, any available mutagenesis procedure can be used for making polymerase mutants. Such mutagenesis procedures optionally include selection of mutant nucleic acids and polypeptides for one or more activity of interest. Procedures that can be used include, but are not limited to: the site-directed point mutagenesis, random point mutagenesis, *in vitro* or *in vivo* homologous recombination (DNA shuffling and combinatorial overlap PCR), mutagenesis using uracil containing templates, oligonucleotide-directed mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using gapped duplex DNA, point mismatch repair, mutagenesis using repair-deficient host strains, restriction-selection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, degenerate PCR, double-strand break repair, and many others known to skilled person.

NUCLEIC ACIDS ENCODING THE DNA POLYMERASE VARIANT

Nucleic acids encoding the DNA polymerase variant of disclosed herein are also a feature of embodiments presented herein. A particular amino acid can be encoded by multiple codons, and certain translation systems (e.g., prokaryotic or eukaryotic cells) often exhibit codon bias, e.g., different organisms often prefer one of the several synonymous codons that encode the same amino acid. As such, nucleic acids presented herein are optionally "codon optimized," meaning that the nucleic acids are synthesized to include codons that are preferred by the particular translation system being employed to express the polymerase. For example, when it is desirable to express the polymerase in a bacterial cell (or even a particular strain of bacteria), the nucleic acid can be synthesized to include codons most frequently found in the genome of that bacterial cell, for efficient expression of the polymerase. A similar strategy can be employed when it is desirable to express the polymerase in a eukaryotic cell, e.g., the nucleic acid can include codons preferred by that eukaryotic cell.

The nucleic acid molecules described herein may also, advantageously, be included in a suitable expression vector to express the polymerase proteins encoded therefrom in a suitable host. Incorporation of cloned DNA into a suitable expression vector for subsequent

transformation of said cell and subsequent selection of the transformed cells is well known to those skilled in the art. Such an expression vector includes a vector having a nucleic acid according to the embodiments presented herein operably linked to regulatory sequences, such as promoter regions, that are capable of effecting expression of said DNA fragments. Such
5 vectors may be transformed into a suitable host cell to provide for the expression of the B-family DNA polymerase variants according to the embodiments presented herein.

The nucleic acid molecule may encode a mature protein or a protein having a prosequence, including that encoding a leader sequence on the preprotein which is then cleaved by the host cell to form a mature protein. The vectors may be, for example, plasmid,
10 virus or phage vectors provided with an origin of replication, and optionally a promoter for the expression of said nucleotide and optionally a regulator of the promoter. The vectors may contain one or more selectable markers, such as, for example, an antibiotic resistance gene.

COMPOSITIONS INCLUDING THE B-FAMILY DNA POLYMERASE VARIANT

15 The present disclosure also provides compositions that include the B-family DNA polymerase variant described herein. The composition can include other components in addition to the B-family DNA polymerase variant. For example, the composition can include a buffer, a nucleotide solution, or a combination thereof. The nucleotide solution can include nucleotides, such as nucleotides that are labelled, synthetic, modified, or a combination
20 thereof. In one embodiment, a composition includes target nucleic acids, such as a library of target nucleic acids.

KIT FOR PERFORMING SBS REACTION

The present invention further provides a kit for performing a sequencing-by-synthesis reaction, comprising: at least one B-family DNA polymerase variant as described above, a
25 primer, a polynucleotide template, and nucleotide analogues under conditions suitable for incorporation into the primer, thereby determining the templating nucleobase complementary to the incorporated nucleotide at the primer terminus. Optionally, other reagents such as buffers and solutions required for the B-family DNA polymerase variant and nucleotide solution are also included. Instructions for use of the assembled or packaged components are
30 also typically included.

In certain embodiments, the nucleotide solution includes labelled nucleotides. In certain embodiments, the nucleotides are synthetic nucleotides. In certain embodiments, the nucleotides are chemically modified nucleotides. In certain embodiments, a modified nucleotide has an altered chemical group at the 3' sugar hydroxyl such that the substituent is larger in size than the naturally occurring 3' hydroxyl group. In certain embodiments, the modified nucleotides include a modified nucleotide or nucleoside molecule that includes a purine or pyrimidine base and a ribose or deoxyribose sugar moiety having a removable 3'-blocking group covalently attached thereto. In certain embodiments, the modified nucleotides are fluorescently labelled to allow direct detection. In certain embodiments, the modified nucleotides include a nucleotide or nucleoside having a base attached to a detectable label via a cleavable linker. In certain embodiments, the detectable label includes a fluorescent label.

METHOD OF USE

The present invention further provides a method for incorporating nucleotide analogues into DNA comprising allowing the following components to interact: (i) B-family DNA polymerase variant according to any of the above embodiments, (ii) a DNA template; and (iii) a nucleotide solution. In certain embodiments, the DNA template comprises a clustered array on a glass or silicon wafer. Thus, a multiplexing, parallel nucleic acid synthesis and sequencing can be performed to determine the sequences of various nucleic acids. The PolB variant driven and array-based sequencing-by-synthesis method can increase the throughput of DNA sequencing and drive down the overall costs of DNA sequencing.

In certain embodiments, the PolB variants described herein can be used to incorporate the nucleotide conjugates or analogue covalently linked with an enzyme, an antibody, a chemical group, such as a biotin, a desthiobiotin, or a fluorophore on the base, phosphate moiety, or pentose sugar of nucleotide, to the 3'-terminus of the primer or nucleic acids initiator in a template-dependent synthesis manner, respectively.

The incorporation of these nucleotide conjugates or analogues into the nucleic acids by PolB variants during the nucleic acid synthesis concurrently add the desired component, such as an enzyme, an antibody, or a chemical group to the newly synthesized nucleic acids in a sequence-dependent manner. Common components used to label or generate nucleic acid probes and conjugates are known in the art, which include, but are not limited to, nucleotide

analogues, modified linkers, such as a biotin, a thiol, an azido or an amine group, fluorophores, enzymes, and antibodies.

Alternatively, in other embodiments, the post-synthetic modifications of nucleic acids can be achieved by covalently or non-covalently coupling with an enzyme, an antibody, a chemical group, or a fluorophore via a modified linker on the base, phosphate moiety, or pentose sugar of synthetic nucleotide. As a result, the desired component can be covalently or non-covalently associated with the specific base, which is added to the 3'-terminus of newly synthesized polynucleotides. The base-specific conjugate can be used as a reporter to determine the sequence of nucleic acid template.

Still, in some embodiments, the PolB variants described herein can be used in the sequencing-by-synthesis method to incorporate the 3'-modified reversible dye-terminator (RDT) nucleotide, which bears a cleavable chemical substitution at the 3'-hydroxyl group (3'-OH) on the nucleotide sugar moiety and a removable fluorophore molecule on the nucleotide base with a distinct fluorescent spectrum. The template-directed synthesis and individual RDT-nucleotide incorporation by the PolB variants can be used as a fluorescent readout for A, T, C, or G base, respectively, to determine the sequence of associated nucleic acids template.

EXAMPLES

In this section, the contents of the present invention will be described in detail through the following examples. These examples are for illustration only, and those skilled in the art can easily think of various modifications and changes. As such, various embodiments of the present invention will be described in detail below, while the invention is not limited to said various embodiments listed in this specification.

Example 1

Protein Sequence Alignment of Various B-family DNA Polymerases and Determination of Consensus Sequences and Variants

Fig.2 shows the amino acid sequence alignment of 16 wild-type B-family DNA polymerases (PolB) and the outcome of aligned consensus sequence is listed in the bottom (SEQ ID NO:1). The 16 wild-type B-family DNA polymerases being aligned are *Thermococcus gorgonarius* DNA polymerase (Tgo, SEQ ID NO:2), *Thermococcus kodakarensis* DNA polymerase (Kod1, SEQ ID NO:3), *Thermococcus sp.* (strain 9°N-7) DNA

polymerase (9°N, SEQ ID NO:4), *Pyrococcus furiosus* DNA polymerase (Pfu, SEQ ID NO:5), *Thermococcus litoralis* DNA polymerase (Vent, SEQ ID NO:6), *Methanococcus maripaludis* DNA polymerase (Mma, SEQ ID NO:7), *Methanosarcina acetivorans* DNA polymerase (Mac, SEQ ID NO:8), human DNA polymerase delta catalytic p125 subunit (hPOLD, SEQ ID NO:9), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD, SEQ ID NO:10), *Pyrobaculum islandicum* DNA polymerase, (Pis, SEQ ID NO:11) *Sulfolobus solfataricus* DNA polymerase (Sso, SEQ ID NO:12), *Pseudomonas aeruginosa* DNA polymerase II (Pae, SEQ ID NO:13), *Escherichia. coli* DNA polymerase II (Eco, SEQ ID NO:14), Bacteriophage RB69 DNA polymerase (RB69, SEQ ID NO:15),
10 Bacteriophage T4 DNA polymerase (T4, SEQ ID NO:16), and Bacteriophage Phi29 DNA polymerase (Phi29, SEQ ID NO:17).

As shown in FIG. 2, various regions of wild-type PolBs are highly conserved while other regions are more variable. Those of skill in the art will immediately recognize and understand that mutations in addition to those specifically identified and discussed herein may
15 be also made in the variable regions of wild-type PolBs without altering, or without substantially altering, the polymerase activity of the mutated enzyme. Likewise, conservative mutations at conserved residues/positions of PolB may be made without altering, or substantially altering, the polymerase activity of the mutated enzyme. Enzyme engineering based on comparative structure analysis with other functionally related enzymes or homologs
20 is a common and useful technique in the molecular biology field that allows a person of skill to reasonably predict the effect of a given mutation on the catalytic activity of the enzyme. Using the structural data and known physical properties of amino acids, those of skill in the art can mutate enzymes, such as the DNA polymerases encompassed by the present invention, without altering, or without substantially altering, the essential, intrinsic characteristics of the
25 enzymes.

Besides, the motif Exo I, motif Exo II, motif Exo III, motif A, motif B, and motif C regions corresponding respectively to positions 349 to 364, 450 to 476, 590 to 608, 706 to 730, 843 to 855, and 940 to 956 of the consensus sequence of SEQ ID NO:1 are marked in FIG. 2. More specifically, the polymerase variant in the present invention is based on
30 substitution mutations at one or more residues correspondingly residing in said motifs.

Example 2

Creating B-family DNA polymerase variants

Gene synthesis approach and mutagenesis technique are adapted to create exemplary PolB variants according to the properties of conserved/consensus amino acids in the conserved and semi-conserved regions of selective PolBs, which are disclosed herein. For instance, the site-directed mutagenesis approach, which is well-known approach to those skilled in the art, is conducted to change the amino acid residues in the conserved motif Exo I, motif Exo II, motif Exo III, motif A, motif B, and motif C regions of an exemplary wild-type PolB, respectively.

In some embodiments, the procedure for obtaining DNA polymerase variants is generally divided into three steps: Step 1: Gene synthesis of DNA polymerase and its exonuclease-deficient (Exo⁻) mutant, Step 2: Construction of DNA polymerase variant in the desired region, and Step 3: Expression and purification of DNA polymerase and its variants. As described in more detail below, the techniques used in said procedure are well-known to those skilled in the art.

In Step 1, the codon-optimized gene fragment encoding the wild-type, intron-free DNA polymerase is synthesized by Genomics BioSci & Tech Co. (New Taipei City, Taiwan). The 3' to 5' exonuclease-deficient DNA polymerase (designated as Exo⁻) is also provided by the same vendor. By "Exo⁻" as used herein, it means that wild-type PolBs has combinatory mutations at the positions corresponding to D354 of SEQ ID NO: 1, which is substituted with an Alanine (D354A) residue, and E356 of SEQ ID NO: 1, which is also substituted with an Alanine residue (E356A), respectively.

In Step 2, the synthetic wild-type and 3' to 5' exonuclease-deficient DNA polymerase gene are respectively subcloned into the pET28b vector using the NdeI and NotI sites. The sequences of recombinant plasmids are confirmed by DNA sequencing. To create the polymerase variant at the desired motif region of the PolB protein backbone, the site-directed mutagenesis, is conducted. Briefly, the site-directed mutagenesis PCR was performed with the recombinant plasmids using the Q5 Site-directed Mutagenesis Kit from New England Biolabs (Ipswich, MA) to introduce the amino acid substitution. The products are first analyzed by 1% agarose gel to confirm the amplicon size and the rest of PCR reaction mixture is then treated

with DpnI at 37°C for an hour. The mixture is further incubated at 70°C for 10 mins to inactivate the DpnI function. The DpnI-treated PCR reaction mixture is then purified by the Qiagen's QIAquick PCR Purification Kit (Whatman, MA). The purified DNA fragment is treated with the mixture of T4 PNK and T4 DNA ligase. The re-circularized PCR-amplified
5 DNA is transformed back into the *E. coli* cells. The plasmid DNA was later extracted from the *E. coli* cells using the Qiagen Plasmid Mini Kit (Whatman, MA). The mutagenic sequences of the polymerase variants at the desired motif region, or regions, are confirmed by DNA sequencing.

In Step 3, *E. coli* Acella cells harboring the plasmid DNA carrying specific polymerase
10 variant gene are grown in 2L of LB medium supplemented with 0.5% glucose and 50 µg/ml carbenicillin at 37°C. When the cell density reaches an absorbance value at OD_{600nm} around 0.6~0.8, an 1mM of isopropyl β-D-1-thiogalactopyranoside (IPTG) is added to induce protein expression. Cells are grown for additional 4 hours at 37°C and then harvested by centrifugation at 4°C for 10 min at 7,000 ×g. Cell pellets are resuspended with buffer A [50
15 mM Tris-HCl (pH 7.5), 300 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5% (v/v) glycerol] containing 1mM benzamidine hydrochloride. Cell lysis is achieved by incubation with 50 mg lysozyme on ice for 1 hour followed by sonication. The cell lysate is clarified by centrifugation at 18,000 ×g for 25 min at 4°C. The clarified crude cell extract is incubated at 70°C for 30 minutes and then cooled down at 4°C. The heat-treated cell extract is further
20 clarified by centrifugation at 18,000 ×g for 25 minutes at 4°C. After centrifugation, the supernatant is diluted with buffer A without NaCl and loaded onto a HiTrap Heparin column (Cytiva Life Sciences, Marlborough, MA, USA) pre-equilibrated in buffer A in the ÄKTA pure chromatography system (Cytiva Life Sciences, Marlborough, MA, USA). The protein is eluted with the linear 100 mM to 1 M NaCl gradient using the buffer B [50 mM Tris-HCl (pH
25 8.0), 1 M NaCl, 0.5 mM EDTA, 1 mM DTT, 5% (v/v) glycerol]. Column fractions are analyzed by 10% SDS-PAGE. Fractions containing desired protein are pooled and dialyzed against the storage buffer [50 mM Tris-HCl (pH 7.5), 250 mM NaCl, 0.5 mM EDTA, 1 mM DTT, 5 % (v/v) glycerol] at 4°C overnight. The dialyzed protein fraction pool containing the target protein is concentrated using an Amicon filter unit (MW cut-off 50,000). The
30 concentrated protein pool is aliquoted and stored at -20°C. Each mutant polymerase variant

was purified with the same procedures as described above. The final protein concentration is determined by the Bradford reaction (Bradford, 1976) using the Bio-Rad Protein Assay (Hercules, CA) with bovine serum albumin as a standard.

Example 3

5 Template-dependent DNA Synthesis Assay

The PolB variants provided herein are tested for DNA sequencing-by-synthesis (SBS) approach. To further evaluate the activities (incorporation of nucleotide analogues) of the PolB variants, modified nucleotides and duplex primer/template (P/T)-DNA are used herein.

In Example 3, the modified nucleotides are nucleotide analogues with a removable
10 blocking moiety (Fig. 3); more specifically, the removable blocking moiety is a 3'-O-azidomethyl group. More specifically, the modified nucleotides are 3'-O-azidomethyl-dNTP including 3'-O-azidomethyl-dATP, 3'-O-azidomethyl-dCTP, 3'-O-azidomethyl-dGTP, 3'-O-azidomethyl-dTTP (as shown in Fig.4). Furthermore, said modified nucleotides are linked
15 with various forms of dyes, such as N₃-dATP-linker-Atto532 (Fig. 5A), N₃-dCTP-linker-Atto700 (Fig. 5B), N₃-dGTP-linker-Cy5 (Fig.5C) and N₃-dTTP-linker-ROX (Fig. 5D).

In Example 3, the following synthetic oligonucleotides were used as a duplex primer and template (P/T)-DNA to determine the template-dependent DNA synthesis activity of PolB variants:

Cy5-38-mer primer:

20 5'-GCTTGCACAAGTTCGTTCAATGATACGGCGACCACCGA-3' (SEQ ID NO: 18);

The oligonucleotide is labeled with a fluorescent Cyanine5 (Cy5) dye at the 5'-end. The corresponding 45-mer template:

25 5'-CGTTCNTCGGTGGTCGCCGTATCATTGAACGAACTTGTGCAAGC-3' (SEQ ID NO: 19); wherein the letter "N" represents either A, T, C, or G.

FAM-38-mer primer:

20 5'-GCTTGCACAAGTTCGTTCAATGATACGGCGACCACCGA-3' (SEQ ID NO: 20)

FAM-19-mer primer: (SEQ ID NO: 21)

30 5'-GCTTGCACAGGGCCTCGAC-3'

Example 4

Activity of exonuclease-deficient PolB variant
on incorporating the 3'-modified nucleotides

In this example, the activities of exonuclease-deficient PolB variants (Kod1, Pfu, Vent, Mac, Sso) on incorporating 3'-modified nucleotides are compared to that of Pol812; more specifically, said 3'-modified nucleotides are 3'-O-azidomethyl nucleotides. As a polymerase of reference, Pol812, a commercial variant of 9°N polymerase is used.

In procedure, the duplex primer-template (P/T) DNA was formed by annealing the FAM-labeled 19-mer primer with the complementary 29-mer DNA template at a molar ratio of 1:1.5 in the 1x TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA] containing 100 mM NaCl. The DNA annealing reaction was performed in the Bio-Rad Thermal Cycler (Hercules, CA) by first heating up the sample mixture to 98°C for 3 minute and then gradually cooling it down (5°C/30 seconds) to 4°C.

The template-dependent DNA synthesis assay was performed in the reaction mixture (10 µl) containing a 50 nM of duplex P/T-DNA substrate and 200 nM of the polymerase in the ThermoPol buffer [20 mM Tris-HCl (pH 8.8), 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, and 0.1% Triton X-100]. The DNA synthesis reaction was initiated by adding 10 µM of 3'-O-azidomethyl-dATP. The reaction mixtures were incubated at 55°C. The reactions were terminated after 2 minutes incubation by adding an equal volume (10 µl) of 2x quench solution [95% deionized formamide and 25 mM EDTA]. The sample mixtures were denatured at 95°C for 10 min and analyzed by 20% polyacrylamide gel electrophoresis containing 8M urea (Urea-PAGE). The DNA synthesis reaction products were then visualized by imaging the gel on the Amersham Typhoon Laser Scanner (Cytiva Life Sciences, Marlborough, MA, USA).

The result of the reaction described above is shown in Fig. 6. In Fig. 6, "S" stands for "duplex P/T-DNA substrate", and the activity of nucleotide incorporation can be inferred from the extent to which the samples shift upward. As such, the result has indicated that none of the nucleotide-incorporation activity of the exemplary PolBs with exonuclease deficiency is comparable to that of Pol812 enzyme. In view of this, it can be concluded that the exonuclease deficiency is not sufficient to endow the PolB enzymes with the activity of

incorporating the 3'-O-azidomethyl-dNTP.

Example 5

Activity of exemplary PolB variants on incorporating the 3'-modified nucleotides

In this example, the exonuclease-deficient PolB (Kod1, Pfu, Vent, Mac, Sso, and 9°N) variants are further substituted with different amino acids in varied regions, and the activity on incorporating 3'-O-azidomethyl-dATP thereof are then evaluated. The procedure applied for evaluating said incorporation is the same with that described above.

The results of the activity performance in correspondence to the PolB variants with varied amino acid forms of substitutions are shown in Table. 1. In Table. 1, “+” stands for “the duplex P/T-DNA substrate is extended with one base”; “+/-” stands for “less than 10% of the duplex P/T-DNA substrate is extended”; and “-” stands for “the duplex P/T-DNA substrate without the nucleotide incorporation”. Based on Table.1, it can be inferred that only when combined with additional amino acid substitutions (either only in Motif A or in both Motifs A and B), the exonuclease-deficient PolB variants can incorporate 3'-O-azidomethyl-dATP. Besides, according to M04 in the Table.1, the PolB variant cannot incorporate 3'-O-azidomethyl-dATP when it is equally substituted in Motif A region as Pol812; However, said PolB variant exhibits the activity of incorporating the 3'-O-azidomethyl-dNTP with amino acid substitutions of L485Y+Y486A+P487G.

Table.1

Wild-type Enzymes	Variant No.	Modification Substitutions	Activity
Pfu	P01	Exo ⁻	-
	P02	Exo ⁻ +L409Y+Y410A+P411G	+
	P03	Exo ⁻ +L409Y+Y410A+P411G+A486L	+
Kod1	K01	Exo ⁻	-
	K02	Exo ⁻ +L408Y+Y409A+P410G	+
	K03	Exo ⁻ +L408Y+Y409A+P410G+A485L	+
Vent	V01	Exo ⁻	-
	V02	Exo ⁻ +L411Y+Y412A+P413G	+
	V03	Exo ⁻ +L411Y+Y412A+P413G+A488L	+
	V04	Exo ⁻ +L411Y+Y412A+P413P+A488L	+
Mac	M01	Exo ⁻	-
	M02	Exo ⁻ +L485Y+Y486A+P487G	+
	M03	Exo ⁻ + L485A+Y486A+P487I+A565L	+
	M04	Exo ⁻ + L485A+Y486A+P487I	-

Sso	S01	Exo ⁻	-
	S02	Exo ⁻ +L518Y+Y519A+P520G	+
	S03	Exo ⁻ +L518Y+Y519A+P520G+A601L	+
	S04	Exo ⁻ +L518A+Y519A+P520I+A601L	+
9°N	N01	Exo ⁻ +M129A+C223S	-
	Pol812	Exo ⁻ +L408A+Y409A+P410I+A485V +M129A+C223S	+
	N02	Exo ⁻ +L408Y+Y409A+P410G+A485V +M129A+C223S	+
Pis	I01	Exo ⁻ +M426Y+Y427A+P428G+A508L	+
	I02	Exo ⁻ +M426Y+Y427A+P428P+A508L	+

Example 6

Efficiency and fidelity of exemplary PolB variants
on incorporating modified nucleotides

5 In this section, the efficiency and fidelity of exemplary PolB variants introduced above on incorporating 3'-O-azidomethyl nucleotides are compared to that of Pol812. As representatives of the exemplary PolB variants, variant V03 (Vent with Exo⁻+L411Y+Y412A+P413G+A488L) and M03 (Mac with Exo⁻+L485Y+Y486A+P487G+A565L) are evaluated herein.

10 In the experimental procedure, the duplex primer-template (P/T) DNA was formed by annealing the FAM-labeled 19-mer primer with the complementary 29-mer DNA template at a molar ratio of 1:1.5 in the 1x TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA] containing 100 mM NaCl. The DNA annealing reaction was performed in the Bio-Rad Thermal Cycler (Hercules, CA) by first heating up the sample mixture to 98°C for 3 minute
15 and then gradually cooling it down (5°C/30 seconds) to 4°C.

The template-dependent DNA synthesis assay was performed in the reaction mixture (10 µl) containing a 50 nM of duplex P/T-DNA substrate and 200 nM of the polymerase in the ThermoPol buffer [20 mM Tris-HCl (pH 8.8), 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, and 0.1% Triton X-100]. The DNA synthesis reaction was initiated by adding 10 µM
20 of 3'-O-azidomethyl-dATP. The reaction mixtures were incubated at 55°C. The reactions were terminated after 0.5, 1, 2 and 5 minutes, respectively, incubation by adding an equal volume (10 µl) of 2x quench solution [95% formamide and 25 mM EDTA]. The sample mixtures were denatured at 95°C for 10 min and analyzed by 20% polyacrylamide gel electrophoresis

containing 8M urea (Urea-PAGE). The DNA synthesis reaction products were then visualized by imaging the gel on the Amersham Typhoon Laser Scanner (Cytiva Life Sciences, Marlborough, MA, USA).

The result of the reaction described above is shown in Fig. 7. In Fig. 7, “S” stands for “duplex P/T-DNA substrate”, and the activity of incorporation can be inferred from the extent to which the samples shift upward. As such, the result has indicated that the efficiency and fidelity of both V03 and M03 on incorporating 3'-O-azidomethyl-dATP are comparable to that of Pol812. The other data (not shown) obtained using 3'-O-azidomethyl-dTTP, 3'-O-azidomethyl-dCTP, and 3'-O-azidomethyl-dGTP, respectively, are all consistent with the result above. In view of this, it can be inferred that the combination of specific amino acid substitutions in Motif A and/or Motif B endow PolB variants with improved efficiency and fidelity of incorporating the 3'-O-N₃ nucleotides.

Example 7

Activity of exemplary PolB variants on incorporating the reversible dye terminator nucleotides

In this example, the activities of exemplary PolB variants on incorporating the reversible dye-terminator nucleotides (Cy5-N₃-dGTP) are compared to that of Pol812. As representatives of the exemplary PolB variants, exonuclease-deficient PolB variants (Kod1 with Exo⁻, Pfu with Exo⁻, Vent with Exo⁻), variant P03 (Pfu with Exo⁻ +L409Y+Y410A+P411G+A486L), variant K03 (Kod1 with Exo⁻ +L408Y+Y409A+P410G+A485L), variant V04 (Vent with Exo⁻ +L411Y+Y412A+P413P+A488L) and variant V03 (Vent with Exo⁻ +L411Y+Y412A+P413G+A488L) are evaluated herein.

In the experimental procedure, the duplex primer-template (P/T) DNA was formed by annealing the FAM-labeled 38-mer primer with the complementary 45-mer DNA template at a molar ratio of 1:1.5 in the 1x TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA] containing 100 mM NaCl. The DNA annealing reaction was performed in the Bio-Rad Thermal Cycler (Hercules, CA) by first heating up the sample mixture to 98°C for 3 minute and then gradually cooling it down (5°C/30 seconds) to 4°C.

The template-dependent DNA synthesis assay was performed in the reaction mixture

(10 µl) containing a 50 nM of duplex P/T-DNA substrate and 200 nM of the polymerase in the ThermoPol buffer [20 mM Tris-HCl (pH 8.8), 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, and 0.1% Triton X-100]. The DNA synthesis reaction was initiated by adding 10 µM of Cy5-N₃-dGTP. The reaction mixtures were incubated at 55°C. The reactions were

5 terminated at 5 minutes incubation by adding an equal volume (10 µl) of 2x quench solution [95% formamide and 25 mM EDTA]. The sample mixtures were denatured at 95°C for 10 min and analyzed by 20% polyacrylamide gel electrophoresis containing 8M urea (Urea-PAGE). The DNA synthesis reaction products were then visualized by imaging the gel on the Amersham Typhoon Laser Scanner (Cytiva Life Sciences, Marlborough, MA, USA).

10 The result of the reaction described above is shown in Fig. 8. In Fig. 8, “C” stands for “Control (Pol812)”; “S” stands for “duplex P/T-DNA substrate”, “1” stands for “Pfu with Exo⁻”; “2” stands for “Kod1 with Exo⁻”; “3” stands for “Vent with Exo⁻”; “4” stands for “P03”; “5” stands for “K03”; “6” stands for “V04”; and “7” stands for “V03”. The result shown in Fig.8 has indicated that the activity of the exonuclease-deficient PolB variants

15 (Kod1 with Exo⁻, Pfu with Exo⁻, Vent with Exo⁻) are not comparable with Pol812, while variants with combined substitutions can all exhibit improved performances and are comparable to that of Pol812. In view of this, it can be inferred that the combination of specific amino acid substitutions in Motif A and/or Motif B endow PolB variants with improved activity of incorporating the reversible dye- terminator nucleotides (Cy5-N₃-dGTP).

20 Example 8

Activity of exemplary PolB variants on incorporating reversible dye-terminator nucleotides

In this example, the activities of other exemplary PolB variants on incorporating another reversible dye-terminator nucleotides (Cy5-N₃-dGTP) are compared to that of Pol812.

25 As representatives of the exemplary PolB variants, variant I01(Pis with Exo⁻+M426Y+Y427A+P428G+A508L) and variant I02 (Pis with Exo⁻+M426Y+Y427A+P428P+A508L) are evaluated herein.

In the experimental procedure, the duplex primer-template (P/T) DNA was formed by annealing the FAM-labeled 38-mer primer with the complementary 45-mer DNA template at

30 a molar ratio of 1:1.5 in the 1x TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA]

containing 100 mM NaCl. The DNA annealing reaction was performed in the Bio-Rad Thermal Cycler (Hercules, CA) by first heating up the sample mixture to 98°C for 3 minute and then gradually cooling it down (5°C/30 seconds) to 4°C.

The template-dependent DNA synthesis assay was performed in the reaction mixture (10 µl) containing a 50 nM of duplex P/T-DNA substrate and 200 nM of the polymerase in the ThermoPol buffer [20 mM Tris-HCl (pH 8.8), 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, and 0.1% Triton X-100]. The DNA synthesis reaction was initiated by adding 10 µM of Cy5-N₃-dGTP. The reaction mixtures were incubated at 55°C. The reactions were terminated at 5 minutes incubation by adding an equal volume (10 µl) of 2x quench solution [95% formamide and 25 mM EDTA]. The sample mixtures were denatured at 95°C for 10 min and analyzed by 20% polyacrylamide gel electrophoresis containing 8M urea (Urea-PAGE). The DNA synthesis reaction products were then visualized by imaging the gel on the Amersham Typhoon Laser Scanner (Cytiva Life Sciences, Marlborough, MA, USA).

The result of the reaction described above is shown in Fig. 9. In Fig. 9, “S” stands for “duplex P/T-DNA substrate”, “1” stands for “Pol812”; “2” stands for “I01”; and “3” stands for “I02”. The result shown in Fig.9 has indicated that variants with combined Motif A and Motif B substitutions can exhibit improved performances that are comparable to that of Pol812, and the activity of I02 is even better than that of Pol812. In view of this, it can be inferred that the combination of specific amino acid substitutions in Motif A and/or Motif B may endow PolB variants with improved activity of incorporating the reversible dye-terminator nucleotides (Cy5-N₃-dGTP).

Example 9

Activity of exemplary PolB variants on incorporating the reversible dye-terminator nucleotide

In this example, the activities of various exemplary PolB variants on incorporating the reversible dye-terminator nucleotides (N₃-dNTP-Linker-Dye) are evaluated. The representatives of the exemplary PolB variants examined herein are as follows:

variant 1: Pol812 (strain 9°N);

variant 2: 9°N with Exo⁻+L408Y+Y409A+P410G+A485V;

variant 3: 9°N with Exo⁻+L408Y+Y409A+P410G+A485L;

variant 4: Vent with Exo⁻+L411Y+Y412A+P413G+A488L;

variant 5: Vent with Exo⁻+L411Y+Y412A+P413G+A488L+K480M;

variant 6: Pfu with Exo⁻+L409Y+Y410A+P411G+A486L;

variant 7: Pfu with Exo⁻+L409Y+Y410A+P411G+A486L+I478M;

- 5 The PolB variants described above are examined with the N₃-dTTP-Linker-ROX (structure thereof is shown in Fig.5D), N₃-dCTP-Linker-Atto700 (structure thereof is shown in Fig.5B)

10 In the experimental procedure, the duplex primer-template (P/T) DNA was formed by annealing the FAM-labeled 38-mer primer/ Cy5-labeled 38-mer primer with the complementary 45-mer DNA template at a molar ratio of 1:1.5 in the 1x TE buffer [10 mM Tris-HCl (pH 8.0) and 1 mM EDTA] containing 100 mM NaCl. The DNA annealing reaction was performed in the Bio-Rad Thermal Cycler (Hercules, CA) by first heating up the sample mixture to 98°C for 3 minute and then gradually cooling it down (5°C/30 seconds) to 4°C.

15 The template-dependent DNA synthesis assay was performed in the reaction mixture (10 µl) containing a 50 nM of duplex P/T-DNA substrate and 200 nM of the polymerase in the ThermoPol buffer [20 mM Tris-HCl (pH 8.8), 10 mM (NH₄)₂SO₄, 10 mM KCl, 2 mM MgSO₄, and 0.1% Triton X-100]. The DNA synthesis reaction was initiated by adding 10 µM of N₃-dTTP-Linker-ROX/ N₃-dCTP-Linker-Atto700. The reaction mixtures were incubated at 55°C. The reactions were terminated at 5 minutes incubation by adding an equal volume (10 µl) of 2x quench solution [95% formamide and 25 mM EDTA]. The sample mixtures were denatured at 95°C for 10 min and analyzed by 20% polyacrylamide gel electrophoresis containing 8M urea (Urea-PAGE). The DNA synthesis reaction products were then visualized by imaging the gel on the Amersham Typhoon Laser Scanner (Cytiva Life Sciences, Marlborough, MA, USA).

25 The result of the reaction described above is shown in Fig. 10. In Fig. 10, “S” stands for “duplex P/T-DNA substrate”, “1” to “7” stand for “variant1” to “variant7”. In view of the result shown in Fig.10, the PolB variants from strain 9°N with “YAG” substitution in the Motif A combined with the A485 substitutions without M129A and C223S can exhibit comparable performances on incorporating the N₃-dNTP-Linker-Dye with that of Pol812, which comprises “AAI” substitution in the Motif A with A485, M129A and C223S

30

substitutions. Besides, the result shown in Fig.10 has indicated that such combination of specific substitutions in the Motif A and/or Motif B may endow PolB variants from different strains with improved activity of incorporating the N₃-dTTP-Linker-ROX or N₃-dCTP-Linker-Atto700 as well. Besides, the other data (not shown) obtained through using FAM-labeled 38-mer primer/ N₃-dGTP-Linker-Cy5 and Cy5-labeled 38-mer primer/ N₃-dATP-Linker-Atto532 are both consistent with the result above.

In summary, the B-family DNA polymerase variants provided herein can exhibit improved incorporation of nucleotide analogues for synthesizing polynucleotides and sequencing the associated nucleic acid template. The specific embodiments of the present invention have been disclosed, but it is not intended to limit the present invention. Those with ordinary knowledge in the technical field to which the present invention belongs are capable of understanding. And in the case of deviating from the principle and spirit of the present invention, various changes and modifications can be made to it, so the scope of protection of the present invention should be based on those defined in the scope of the accompanying patent application.

CLAIMS

What is claimed is:

1. A B-family DNA polymerase variant comprising:
 - a motif Exo I, a motif Exo II, a motif Exo III, a motif A, a motif B, and a motif C
 - 5 corresponding respectively to positions 349 to 364, 450 to 476, 590 to 608, 706 to 730, 843 to 855, and 940 to 956 of a consensus sequence (SEQ ID NO:1);
 - at least one amino acid substitution at a position in a motif selected from the group consisting of the motif Exo I, the motif Exo II, and the motif Exo III; and
 - at least one amino acid substitution at a position in a motif selected from the group
 - 10 consisting of the motif A, the motif B, and the motif C.
2. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is modified from a wild-type B-family DNA polymerase having an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 and 17.
- 15 3. The B-family DNA polymerase variant of Claim 2, wherein the wild-type B-family DNA polymerase is *Thermococcus gorgonarius* DNA polymerase (Tgo), *Thermococcus kodakarensis* DNA polymerase (Kod1), *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N), *Pyrococcus furiosus* DNA polymerase (Pfu), *Thermococcus litoralis* DNA polymerase (Vent), *Methanococcus maripaludis* DNA polymerase (Mma), *Methanosarcina*
- 20 *acetivorans* DNA polymerase (Mac), human DNA polymerase delta catalytic p125 subunit (hPOLD), *Saccharomyces cerevisiae* DNA polymerase delta catalytic subunit (ScePOLD), *Pyrobaculum islandicum* DNA polymerase, (Pis) *Sulfolobus solfataricus* DNA polymerase (Sso), *Pseudomonas aeruginosa* DNA polymerase II (Pae), *Escherichia coli* DNA polymerase II (Eco), Bacteriophage (*Escherichia* phage) RB69 DNA polymerase (RB69),
- 25 Bacteriophage (*Escherichia* phage) T4 DNA polymerase (T4), or Bacteriophage (*Bacillus* phage) Phi29 DNA polymerase (Phi29).
4. The B-family DNA polymerase variant of Claim 1, wherein:
 - i. the amino acid corresponding to position 354 of SEQ ID NO:1 is any amino acid other than D;
 - 30 ii. the amino acid corresponding to position 355 of SEQ ID NO:1 is any amino

acid other than I;

iii. the amino acid corresponding to position 356 of SEQ ID NO:1 is any amino acid other than E;

iv. the amino acid corresponding to position 715 of SEQ ID NO:1 is any amino acid other than L;

v. the amino acid corresponding to position 716 of SEQ ID NO:1 is any amino acid other than Y; and

vi. the amino acid corresponding to position 717 of SEQ ID NO:1 is any amino acid other than P.

5 10 5. The B-family DNA polymerase variant of Claim 1, wherein:

i. the amino acid corresponding to position 354 of SEQ ID NO:1 is any amino acid other than D;

ii. the amino acid corresponding to position 355 of SEQ ID NO:1 is any amino acid other than I;

15 iii. the amino acid corresponding to position 356 of SEQ ID NO:1 is any amino acid other than E;

iv. the amino acid corresponding to position 715 of SEQ ID NO:1 is any amino acid other than L;

v. the amino acid corresponding to position 716 of SEQ ID NO:1 is any amino acid other than Y;

20 vi. the amino acid corresponding to position 717 of SEQ ID NO:1 is any amino acid other than P; and

vii. the amino acid corresponding to position 854 of SEQ ID NO:1 is any amino acid other than A.

25 6. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO:3; and wherein:

i. D141 of SEQ ID NO:3 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

30 ii. E143 of SEQ ID NO:3 corresponding to position 356 of SEQ ID NO:1 is

substituted with A;

iii. L408 of SEQ ID NO:3 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

iv. Y409 of SEQ ID NO:3 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and

v. P410 of SEQ ID NO:3 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

7. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N) having a wild-type amino acid sequence of SEQ ID NO:4; and wherein:

i. D141 of SEQ ID NO:4 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E143 of SEQ ID NO:4 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. L408 of SEQ ID NO:4 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

iv. Y409 of SEQ ID NO:4 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and

v. P410 of SEQ ID NO:4 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

8. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO:5; and wherein:

i. D141 of SEQ ID NO:5 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E143 of SEQ ID NO:5 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. I409 of SEQ ID NO:5 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

iv. Y410 of SEQ ID NO:5 corresponding to position 716 of SEQ ID NO:1 is

substituted with A, C, D, F, G, H or I; and

v. P411 of SEQ ID NO:5 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

9. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO:6; and wherein:

i. D141 of SEQ ID NO:6 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E143 of SEQ ID NO:6 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. L411 of SEQ ID NO:6 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

iv. Y412 of SEQ ID NO:6 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and

v. P413 of SEQ ID NO:6 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

10. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO:8; and wherein:

i. D198 of SEQ ID NO:8 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E200 of SEQ ID NO:8 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. L485 of SEQ ID NO:8 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

iv. Y486 of SEQ ID NO:8 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and

v. P487 of SEQ ID NO:8 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

11. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA

polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO:11; and wherein:

- i. D171 of SEQ ID NO:11 corresponding to position 354 of SEQ ID NO:1 is substituted with A;
- 5 ii. E173 of SEQ ID NO:11 corresponding to position 356 of SEQ ID NO:1 is substituted with A;
- iii. M426 of SEQ ID NO:11 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;
- iv. Y427 of SEQ ID NO:11 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and
- 10 v. P428 of SEQ ID NO:11 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.

12. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO:12; and wherein:

- i. D231 of SEQ ID NO:12 corresponding to position 354 of SEQ ID NO:1 is substituted with A;
- ii. E233 of SEQ ID NO:12 corresponding to position 356 of SEQ ID NO:1 is substituted with A;
- 20 iii. L518 of SEQ ID NO:12 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;
- iv. Y519 of SEQ ID NO:12 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I; and
- v. P520 of SEQ ID NO:12 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T.
- 25

13. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Thermococcus kodakarensis* DNA polymerase (Kod1) having a wild-type amino acid sequence of SEQ ID NO:3; and wherein:

- i. D141 of SEQ ID NO:3 corresponding to position 354 of SEQ ID NO:1 is substituted with A;
- 30

ii. E143 of SEQ ID NO:3 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. L408 of SEQ ID NO:3 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

5 iv. Y409 of SEQ ID NO:3 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I;

v. P410 of SEQ ID NO:3 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and

10 vi. A485 of SEQ ID NO:3 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

14. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Thermococcus sp.* (strain 9°N-7) DNA polymerase (9°N) having a wild-type amino acid sequence of SEQ ID NO:4; and wherein:

15 i. D141 of SEQ ID NO:4 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E143 of SEQ ID NO:4 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. L408 of SEQ ID NO:4 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

20 iv. Y409 of SEQ ID NO:4 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I;

v. P410 of SEQ ID NO:4 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and

25 vi. A485 of SEQ ID NO:4 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

15. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Pyrococcus furiosus* DNA polymerase (Pfu) having a wild-type amino acid sequence of SEQ ID NO:5; and wherein:

30 i. D141 of SEQ ID NO:5 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E143 of SEQ ID NO:5 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. L409 of SEQ ID NO:5 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

5 iv. Y410 of SEQ ID NO:5 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I;

v. P411 of SEQ ID NO:5 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and

10 vi. A486 of SEQ ID NO:5 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

16. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Thermococcus litoralis* DNA polymerase (Vent) having a wild-type amino acid sequence of SEQ ID NO:6; and wherein:

15 i. D141 of SEQ ID NO:6 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E143 of SEQ ID NO:6 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. L411 of SEQ ID NO:6 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

20 iv. Y412 of SEQ ID NO:6 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I;

v. P413 of SEQ ID NO:6 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and

25 vi. A488 of SEQ ID NO:6 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

17. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Methanosarcina acetivorans* DNA polymerase (Mac) having a wild-type amino acid sequence of SEQ ID NO:8; and wherein:

30 i. D198 of SEQ ID NO:8 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E200 of SEQ ID NO:8 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. L485 of SEQ ID NO:8 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

5 iv. Y486 of SEQ ID NO:8 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I;

v. P487 of SEQ ID NO:8 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and

10 vi. A565 of SEQ ID NO:8 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

18. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Pyrobaculum islandicum* DNA polymerase (Pis) having a wild-type amino acid sequence of SEQ ID NO:11; and wherein:

15 i. D171 of SEQ ID NO:11 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

ii. E173 of SEQ ID NO:11 corresponding to position 356 of SEQ ID NO:1 is substituted with A;

iii. M426 of SEQ ID NO:11 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;

20 iv. Y427 of SEQ ID NO:11 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I;

v. P428 of SEQ ID NO:11 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and

25 vi. A508 of SEQ ID NO:11 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.

19. The B-family DNA polymerase variant of Claim 1, wherein the B-family DNA polymerase variant is derived from *Sulfolobus solfataricus* DNA polymerase (Sso) having a wild-type amino acid sequence of SEQ ID NO:12; and wherein:

30 i. D231 of SEQ ID NO:12 corresponding to position 354 of SEQ ID NO:1 is substituted with A;

- ii. E233 of SEQ ID NO:12 corresponding to position 356 of SEQ ID NO:1 is substituted with A;
- iii. L518 of SEQ ID NO:12 corresponding to position 715 of SEQ ID NO:1 is substituted with A, F, I, M, Q, S, H or Y;
- 5 iv. Y519 of SEQ ID NO:12 corresponding to position 716 of SEQ ID NO:1 is substituted with A, C, D, F, G, H or I;
- v. P520 of SEQ ID NO:12 corresponding to position 717 of SEQ ID NO:1 is substituted with A, C, G, S or T; and
- 10 vi. A601 of SEQ ID NO:12 corresponding to position 854 of SEQ ID NO:1 is substituted with C, D, E, F, or L.
20. A kit for performing a sequencing-by-synthesis reaction, comprising: a B-family DNA polymerase variant as defined in any one of claim 1-19, a primer, and nucleotide analogues under conditions suitable for incorporation into the primer.