

CLAIMS

1. A method for introducing mutations into at least one target nucleic acid molecule comprising:
 - a. providing at least one sample comprising at least one target nucleic acid molecule; and
 - b. amplifying the at least one target nucleic acid molecule using a low bias DNA polymerase.
2. Use of a low bias DNA polymerase in a method for introducing mutations into at least one target nucleic acid molecule.
3. The use of claim 2, wherein the method for introducing mutations into at least one target nucleic acid molecule comprises:
 - a. providing at least one sample comprising at least one target nucleic acid molecule; and
 - b. amplifying the at least one target nucleic acid molecule using a low bias DNA polymerase.
4. The method or use of any one of the preceding claims, wherein the mutations are substitution mutations.
5. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase mutates adenine, thymine, guanine, and cytosine nucleotides in the at least one target nucleic acid molecule at a rate ratio of 0.5-1.5:0.5-1.5:0.5-1.5:0.5-1.5, 0.6-1.4:0.6-1.4:0.6-1.4:0.6-1.4, 0.7-1.3:0.7-1.3:0.7-1.3:0.7-1.3, 0.8-1.2:0.8-1.2:0.8-1.2:0.8-1.2, or around 1:1:1:1 respectively.
6. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase mutates adenine, thymine, guanine, and cytosine nucleotides in the at least one target nucleic acid molecule at a rate ratio of 0.7-1.3:0.7-1.3:0.7-1.3:0.7-1.3 respectively.

7. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase mutates between 1% and 15%, between 2% and 10%, or around 8% of the nucleotides in the at least one target nucleic acid molecule.
8. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase mutates between 0% and 3%, or between 0% and 2% of the nucleotides in the at least one target nucleic acid molecule per round of replication.
9. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase incorporates nucleotide analogs into the at least one target nucleic acid molecule.
10. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase mutates adenine, thymine, guanine, and/or cytosine in the at least one target nucleic acid molecule using a nucleotide analog.
11. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase replaces guanine, cytosine, adenine, and/or thymine with a nucleotide analog.
12. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase introduces guanine or adenine nucleotides using a nucleotide analog at a rate ratio of 0.5-1.5:0.5-1.5, 0.6-1.4:0.6-1.4, 0.7-1.3:0.7-1.3, 0.8-1.2:0.8-1.2, or around 1:1 respectively.
13. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase introduces guanine or adenine nucleotides using a nucleotide analog at a rate ratio of 0.7-1.3:0.7-1.3 respectively.
14. The method or use of any one of claims 9-13, wherein the method comprises a step of amplifying the at least one target nucleic acid molecule using a low bias DNA polymerase, the step of amplifying the at least one target nucleic

acid molecule using a low bias DNA polymerase is carried out in the presence of the nucleotide analog, and the step of amplifying the at least one target nucleic acid molecule provides at least one target nucleic acid molecule comprising the nucleotide analog.

15. The method or use of any one of claims 9-14, wherein the nucleotide analog is dPTP.
16. The method or use of claim 15, wherein the low bias DNA polymerase introduces guanine to adenine substitution mutations, cytosine to thymine substitution mutations, adenine to guanine substitution mutations, and thymine to cytosine substitution mutations.
17. The method or use of claim 16, wherein the low bias DNA polymerase introduces guanine to adenine substitution mutations, cytosine to thymine substitution mutations, adenine to guanine substitution mutations, and thymine to cytosine substitution mutations at a rate ratio of 0.5-1.5:0.5-1.5:0.5-1.5:0.5-1.5, 0.6-1.4:0.6-1.4:0.6-1.4:0.6-1.4, 0.7-1.3:0.7-1.3:0.7-1.3:0.7-1.3, 0.8-1.2:0.8-1.2:0.8-1.2:0.8-1.2, or around 1:1:1:1 respectively.
18. The method or use of claim 16 or 17, wherein the low bias DNA polymerase introduces guanine to adenine substitution mutations, cytosine to thymine substitution mutations, adenine to guanine substitution mutations, and thymine to cytosine substitution mutations at a rate ratio of 0.7-1.3:0.7-1.3:0.7-1.3:0.7-1.3 respectively.
19. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase is a high fidelity DNA polymerase.
20. The method or use of claim 19, wherein, in the absence of nucleotide analogs, the high fidelity DNA polymerase introduces less than 0.01%, less than 0.0015%, less than 0.001%, between 0% and 0.0015%, or between 0% and 0.001% mutations per round of replication.

21. The method or use of claim 14 or 15, wherein the method comprises a further step of amplifying the at least one target nucleic acid molecule comprising nucleotide analogs in the absence of nucleotide analogs.
22. The method or use of claim 21, wherein the step of amplifying the at least one target nucleic acid molecule comprising nucleotide analogs in the absence of nucleotide analogs is carried out using the low bias DNA polymerase.
23. The method or use of any one of the preceding claims, wherein the method provides a mutated at least one target nucleic acid molecule and the method further comprises a further step of amplifying the mutated at least one target nucleic acid molecule using the low bias DNA polymerase.
24. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase has low template amplification bias.
25. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase comprises a proof-reading domain and/or a processivity enhancing domain.
26. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase comprises a fragment of at least 400, at least 500, at least 600, at least 700, or at least 750 contiguous amino acids of:
 - a. a sequence of SEQ ID NO. 2;
 - b. a sequence at least 95%, at least 98%, or at least 99% identical to SEQ ID NO. 2;
 - c. a sequence of SEQ ID NO. 4;
 - d. a sequence at least 95%, at least 98%, or at least 99% identical to SEQ ID NO. 4;
 - e. a sequence of SEQ ID NO. 6;

- f. a sequence at least 95%, at least 98%, or at least 99% identical to SEQ ID NO. 6;
 - g. a sequence of SEQ ID NO. 7; or
 - h. a sequence at least 95%, at least 98%, or at least 99% identical to SEQ ID NO. 7.
- 27. The method or use of claim 26, wherein the low bias DNA polymerase comprises:
 - a. a sequence of SEQ ID NO. 2;
 - b. a sequence at least 95%, at least 98%, or at least 99% identical to SEQ ID NO. 2;
 - c. a sequence of SEQ ID NO. 4;
 - d. a sequence at least 95%, at least 98%, or at least 99% identical to SEQ ID NO. 4;
 - e. a sequence of SEQ ID NO. 6;
 - f. a sequence at least 95%, at least 98%, or at least 99% identical to SEQ ID NO. 6;
 - g. a sequence of SEQ ID NO. 7; or
 - h. a sequence at least 95%, at least 98%, or at least 99% identical to SEQ ID NO. 7.
- 28. The method or use of claim 27, wherein the low bias DNA polymerase comprises a sequence at least 98% identical to SEQ ID NO. 2.
- 29. The method or use of claim 27, wherein the low bias DNA polymerase comprises a sequence at least 98% identical to SEQ ID NO. 4.
- 30. The method or use of claim 27, wherein the low bias DNA polymerase comprises a sequence at least 98% identical to SEQ ID NO. 6.
- 31. The method or use of claim 27, wherein the low bias DNA polymerase comprises a sequence at least 98% identical to SEQ ID NO. 7.

32. The method or use of any one of the preceding claims, wherein the low bias DNA polymerase is a thermococcal polymerase, or derivative thereof.
33. The method or use of claim 32, wherein the low bias DNA polymerase is a thermococcal polymerase.
34. The method or use of claim 32 or 33, wherein the thermococcal polymerase is derived from a thermococcal strain selected from the group consisting of *T.kodakarensis*, *T.siculi*, *T.celer* and *T.sp* KS-1.
35. The method or use of any one of the preceding claims, further comprising introducing barcodes into the at least one target nucleic acid molecule.
36. The method or use of any one of the preceding claims, further comprising introducing sample tags into the at least one target nucleic acid molecule.
37. The method or use of claim 36, wherein a group of sample tags is used and target nucleic acid molecules from different samples are labelled with different sample tags from the group.
38. The method or use of claim 37, wherein each sample tag differs from substantially all other sample tags in the group by at least 1 low probability mutation difference or at least 3 high probability mutation differences.
39. The method or use of claim 38, wherein each sample tag differs from substantially all other sample tags in the group by at least 3 low probability mutation differences.
40. The method or use of claim 38 or 39, wherein each sample tag differs from substantially all other sample tags in the group by between 3 and 25, or between 3 and 10 low probability mutation differences.
41. The method or use of any one of claims 38-40, wherein the low probability mutation is a transversion mutation or an indel mutation.

42. The method or use of any one of claims 37-41, wherein each sample tag differs from substantially all other sample tags in the group by at least 5 high probability mutation differences.
43. The method or use of claim 42, wherein each sample tag differs from substantially all other sample tags in the group by between 5 and 25, or between 5 and 10 high probability mutation differences.
44. The method or use of any one of claims 38-43, wherein the high probability mutation is a transition mutation.
45. The method or use of any one of claims 37-44, wherein the group of sample tags is obtainable from the method of any one of claims 71-75.
46. The method or use of any one of the preceding claims, further comprising introducing adapters into each of the at least one target nucleic acid molecule.
47. The method or use of claim 46, comprising introducing a first adapter at the 3' end of the at least one target nucleic acid molecule and a second adapter at the 5' end of the at least one target nucleic acid molecule, wherein the first adapter and the second adapter can anneal to one another.
48. The method or use of claim 47, wherein the at least one target nucleic acid molecule is amplified using primers that are identical to one another and complementary to a portion of the first adapter.
49. The method or use of claim 47 or 48, wherein the first adapter is complementary to a nucleic acid molecule that is at least 80%, at least 90%, at least 95%, at least 99%, or 100% identical to the second adapter.
50. The method or use of claim 48 or 49, wherein the primers comprise a second primer binding site, and the method comprises amplifying the at least one target nucleic acid molecule using the primers, removing the primers and

further amplifying the at least one target nucleic acid molecule using a second set of primers that anneal to the second primer binding site.

51. The method or use of any one of the preceding claims, wherein the method further comprises introducing barcodes, sample tags and adapters into each of the target nucleic acid molecules.
52. The method or use of any one of the preceding claims, wherein the barcodes, sample tags and/or adapters are introducing by tagmentation or by shearing and ligation.
53. The method or use of any one of any one of the preceding claims, wherein the at least one target nucleic acid molecule is greater than 1 kbp, greater than 1.5 kbp, greater than 2 kbp, greater than 4 kbp, greater than 5kbp, greater than 7 kbp, or greater than 8 kbp.
54. A method for determining a sequence of at least one target nucleic acid molecule comprising the method for introducing mutations of any one of claims 1 or 3-53.
55. The method of claim 54, comprising steps of:
 - a. performing the method of any one of claims 1 or 3-53 to provide at least one mutated target nucleic acid molecule;
 - b. sequencing regions of the at least one mutated target nucleic acid molecule to provide mutated sequence reads; and
 - c. assembling a sequence for at least a portion of the at least one target nucleic acid molecule using the mutated sequence reads.
56. The method of claim 54, comprising steps of:
 - a. performing the method of any one of claims 1 or 3-53 to provide at least one mutated target nucleic acid molecule;
 - b. fragmenting and/or amplifying the at least one mutated target nucleic acid molecule to provide at least one fragmented and/or amplified mutated target nucleic acid molecule;

- c. sequencing regions of the at least one fragmented and/or amplified mutated target nucleic acid molecule to provide mutated sequence reads; and
- d. assembling a sequence for at least a portion of the at least one target nucleic acid molecule using the mutated sequence reads.

57. A method for engineering a protein comprising the method for introducing mutations of any one of claims 1 or 3-53.
58. The method of claim 57, comprising steps of:
- a. performing the method of any one of claims 1 or 3-53 to provide at least one mutated target nucleic acid molecule;
 - b. inserting the at least one mutated target nucleic acid molecule into a vector; and
 - c. expressing a protein encoded by the at least one mutated target nucleic acid molecule.
59. The method of claim 58 comprising steps of:
- a. providing at least one sample comprising at least one target nucleic acid molecule; and
 - b. amplifying the at least one target nucleic acid molecule using a low bias DNA polymerase in the presence of a nucleotide analog to provide at least one target nucleic acid molecule comprising the nucleotide analog;
 - c. amplifying the at least one target nucleic acid molecule comprising the nucleotide analog in the absence of the nucleotide analog to provide at least one mutated target nucleic acid molecule;
 - d. inserting the at least one mutated target nucleic acid molecule into a vector; and
 - e. expressing a protein encoded by the at least one mutated target nucleic acid molecule.

60. The method of claim 58 or 59, wherein the method further comprises a step of testing the activity or assessing the structure of the protein encoded by the at least one mutated target nucleic acid molecule.
61. The method of any one of claims 58-60, wherein the vector is a plasmid, a virus, a cosmid, or an artificial chromosome.
62. The method of any one of claims 58-61, wherein the step of expressing a protein encoded by the at least one mutated target nucleic acid molecule is achieved by transforming bacterial cells, transfecting eukaryotic cells or transducing eukaryotic cells with the vector.
63. A group of sample tags, wherein each sample tag differs from substantially all other sample tags in the group by at least one low probability mutation difference or at least three high probability mutation differences.
64. The group of sample tags of claim 63, wherein each sample tag differs from substantially all other sample tags in the group by at least 3 low probability mutation differences.
65. The group of sample tags of claim 63 or 64, wherein each sample tag differs from substantially all other sample tags in the group by between 3 and 25, or between 3 and 10 low probability mutation differences.
66. The group of sample tags of any one of claims 63-65, wherein the low probability mutation is a transversion mutation or an indel mutation.
67. The group of sample tags of any one of claims 63-66, wherein each sample tag differs from substantially all other sample tags in the group by at least 5 high probability mutation differences.
68. The group of sample tags of any one of claims 63-67, wherein each sample tag differs from substantially all other sample tags in the group by between 5 and 25, or between 5 and 10 high probability mutation differences.

69. The group of sample tags of any one of claims 63-68, wherein the high probability mutation is a transition mutation.
70. The group of sample tags of any one of claims 63-69, wherein each sample tag is at least 8 nucleotides, at least 10 nucleotides, at least 12 nucleotides, between 8 and 50 nucleotides, between 10 and 50 nucleotides, or between 10 and 50 nucleotides in length.
71. A method for designing a group of sample tags suitable for use in a method for introducing mutations into at least one target nucleic acid molecule comprising:
- a. analysing the method for introducing mutations into at least one target nucleic acid molecule and determining the average number of low probability mutations that take place during the method for introducing mutations into at least one target nucleic acid molecule; and
 - b. determining sequences for a group of sample tags wherein each sample tag differs from substantially all sample tags in the group by more low probability differences than the average number of low probability mutations that take place during the method for introducing mutations into at least one target nucleic acid molecule.
72. The method of claim 71, further comprising:
- a. (i) analysing the method for introducing mutations into at least one target nucleic acid molecule and determining the average number of high probability mutations that take place during the method for introducing mutations into at least one target nucleic acid molecule; and
 - (ii) determining sequences for a group of sample tags wherein each sample tag differs from substantially all sample tags in the group by more high probability differences than the average number of high probability mutations that take place during the method for introducing mutations into at least one target nucleic acid molecule.

73. The method of claim 71 or 72, wherein the low probability mutation is a transversion mutation or an indel mutation.
74. The method of any one of claims 72-73, wherein the high probability mutation is a transition mutation.
75. The method of any one of claims 71-74, which is a computer implemented method.
76. The method or use of any one of the preceding claims, wherein the step of amplifying the at least one target nucleic acid molecule using a low bias DNA polymerase is carried out using dNTPs at unequal concentrations.
77. The method of any one of the preceding claims, wherein:
- (i) the method comprises a further step of amplifying the at least one target nucleic acid molecule comprising nucleotide analogs in the absence of nucleotide analogs and the further step of amplifying the at least one target nucleic acid molecule comprising nucleotide analogs in the absence of nucleotide analogs is carried out using dNTPs at unequal concentrations; or
 - (ii) the method provides a mutated at least one target nucleic acid molecule, the method comprises a further step of amplifying the mutated at least one target nucleic acid molecule using the low bias DNA polymerase and the further step of amplifying the mutated at least one target nucleic acid molecule using the low bias DNA polymerase is carried out using dNTPs at unequal concentrations.
78. A method for introducing mutations into at least one target nucleic acid molecule comprising:
- a. providing at least one sample comprising at least one target nucleic acid molecule; and
 - b. introducing mutations into the at least one target nucleic acid molecule by amplifying the at least one target nucleic acid molecule using a DNA

polymerase to provide a mutated at least one target nucleic acid molecule,
wherein step b. is carried out using dNTPs at unequal concentrations.

79. The method of claim 78, wherein step b. is carried out in the presence of a nucleotide analog.
80. The method of claim 79, wherein the nucleotide analog is dPTP.
81. The method of claim 79 or 80, wherein the method comprises a further step c. of amplifying the mutated at least one target nucleic acid molecule in the absence of the nucleotide analog.
82. The method of claim 81, wherein step c. is carried out using dNTPs at unequal concentrations.
83. The method of any one of claims 76-82, wherein dNTPs at unequal concentrations are used to alter the profile of mutations that are introduced.
84. The method of claim 83, wherein dNTPs at unequal concentrations are used to reduce bias in the profile of mutations that are introduced.
85. The method of any one of the preceding claims, wherein the method is a method for introducing mutations in a low bias mutation profile.
86. The method of any one of claims 76-85, wherein the dNTPs at unequal concentrations comprise dATP, dCTP, dTTP and dGTP and one or two of dATP, dCTP, dTTP or dGTP are at a lower concentration compared to other dNTPs.
87. The method of any one of claims 76-86, wherein using dNTPs at unequal concentrations comprises a step of identifying a dNTP whose level should be increased or decreased in order to reduce bias in the profile of mutations that are introduced.

88. The method of any one of claims 76-87, wherein the dNTPs at unequal concentrations comprise dTTP at a lower concentration than other dNTPs.
89. The method of claim 88, wherein the dNTPs at unequal concentrations comprise dTTP at a concentration less than 75%, less than 70%, less than 60%, less than 55%, between 25% and 75%, between 25% and 70, between 25% and 60%, or around 50% of the concentration of dATP, dCTP or dGTP.
90. The method of claim 89, wherein the dNTPs at unequal concentrations comprise dTTP at a concentration less than 75%, less than 70%, less than 60%, less than 55%, between 25% and 75%, between 25% and 70, between 25% and 60%, or around 50% of the concentration of dCTP.
91. The method of claim 90, wherein the dNTPs at unequal concentrations comprise dTTP at a concentration less than 60% of the concentration of dCTP.
92. The method of claim 87, wherein the dNTPs at unequal concentrations comprise dTTP at a concentration between 25% and 60% of the concentration of dCTP.
93. The method of any one of claims 77, or 81-92, wherein the step of amplifying the at least one target nucleic acid molecule comprising nucleotide analogs in the absence of nucleotide analogs or amplifying the mutated at least one target nucleic acid molecule in the absence of the nucleotide analog is carried out using dNTPs at unequal concentrations.
94. The method of claim 93, wherein the dNTPs at unequal concentrations comprises dATP at a lower concentration compared to other dNTPs.

95. The method of claim 94, wherein the dNTPs at unequal concentrations comprises dATP at a concentration less than 75%, less than 70%, less than 60%, less than 55%, between 25% and 75%, between 25% and 70, between 25% and 60%, or around 50% of the concentration of dTTP, dCTP or dGTP.
96. The method of claim 95, wherein the dNTPs at unequal concentrations of dNTPs comprises dATP at a concentration less than 75%, less than 70%, less than 60%, less than 55%, between 25% and 75%, between 25% and 70, between 25% and 60%, or around 50% of the concentration of dGTP.
97. The method of claim 96, wherein the dNTPs at unequal concentrations comprises dATP at a concentration less than 60% of the concentration of dGTP.
98. The method of claim 96 or 97, wherein the dNTPs at unequal concentrations comprises dATP at a concentration between 25% and 60% of the concentration of dGTP.
99. A group of sample tags obtainable by the method of any one of claims 71-74.
100. A computer readable medium configured to perform the method of any one of claims 71-74.
101. A method for preferentially amplifying target nucleic acid molecules that are larger than 1 kbp in length comprising:
- a. providing at least one sample comprising target nucleic acid molecules;
 - b. introducing a first adapter at the 3' end of target nucleic acid molecules and a second adapter at the 5' end of target nucleic acid molecules; and
 - c. amplifying the target nucleic acid molecules using primers that are complementary to a portion of the first adapter,
- wherein the first adapter and the second adapter can anneal to one another.

102. The method of claim 101, wherein the primers are identical to one another.
103. The method of claim 101 or claim 102, wherein the first adapter is complementary to a nucleic acid molecule that is at least 80%, at least 90%, at least 95%, at least 99%, or 100% identical to the second adapter.
104. The method of any one of claims 101-103, wherein the method is a method for preferentially amplifying target nucleic acid molecules that are larger than 1.5 kbp in length.
105. The method of any one of any one of claims 101-104, further comprising a step of sequencing the target nucleic acid molecules.