

ENGINEERED TRANSPOSASES AND METHODS OF USE THEREOF**CROSS-REFERENCE TO RELATED APPLICATIONS**

- 5 This application claims the benefit of U.S. Provisional Application No. 63/622,732, filed January 19, 2024, and U.S. Provisional Application No. 63/623,064, filed January 19, 2024, both of which are herein incorporated by reference, in their entireties.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

- 10 The instant application contains a Sequence Listing which has been submitted electronically in XML format and is hereby incorporated by reference in its entirety. Said XML copy, created on January 17, 2025, is named "51178-017WO2_Sequence_Listing_1_17_25", and is 12,833,483 bytes in size.

TECHNICAL FIELD

- 15 The present invention provides engineered transposase polypeptides useful in the preparation of next generation sequencing (NGS) libraries, as well as compositions and methods of utilizing these engineered polypeptides.

BACKGROUND

- 20 Transposases are enzymes known to facilitate the transposition of fragments of DNA (transposons) from a donor strand to a target strand at sites flanked by short inverted-repeat sequences or end sequences (ESs). The retroviral integrase superfamily of proteins includes a number of transposases, including Tn5, an IS4 transposase. The transposition mechanism is initiated by the binding of two transposases to donor DNA ESs to form the dimeric synaptic complex or transposome, which
25 catalyzes cleavage of donor DNA. The transposome then transfers the transposon to the target strand (Gradman, R.J., & Reznikoff, W.S. (2008). Tn 5 Synaptic Complex Formation : Role of Transposase Residue W 450.). Two divalent metal ions (Mg^{2+} or Mn^{2+}) bind to the active site and are critical for catalysis (Lovell, S., Goryshin, I., Reznikoff, W. *et al.* Two-metal active site binding of a Tn5 transposase synaptic complex. Nat. Struct. Mol. Biol. 9:278–281 (2002).)
30 In nature, transposases have limited activity. However, several mutations in Tn5 are known to enhance transposase activity, including L372P, E54K, E110K, E345K, and P242A/G (Reznikoff, W.S. (2003), Tn5 as a model for understanding DNA transposition. Molecular Microbiology, 47: 1199-1206.) Additionally, a 19 bp ES known as mosaic end (MEs) is associated with hyperactivity in Tn5 transposases.
35 Next generation sequencing (NGS) has facilitated sequencing of large DNA libraries using limited samples. However, sample preparation is time-consuming and resource-intensive, necessitating development of more efficient protocols. Tn5 transposase has been used to reduce costs and improve efficiency in preparing samples for large-scale sequencing projects (Hennig BP, Velten L, Racke I, Tu CS, Thoms M, Rybin V, Besir H, Remans K, Steinmetz LM. Large-Scale Low-Cost NGS Library Preparation

Using a Robust Tn5 Purification and Tagmentation Protocol. G3 (Bethesda). 2018 Jan 4;8(1):79-89.) In one example, a hyperactive Tn5 with L372P and E54K mutations was used with MEs in a tagmentation protocol to simultaneously fragment sample DNA and bind sequencing adapters for NGS library preparation (Picelli S, Björklund AK, Reinius B, Sagasser S, Winberg G, Sandberg R. Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. Genome Res. 2014 Dec;24(12):2033-40.)

Improvements to these methods have been reported (Hennig et. al)., and various commercial sample preparation methods now incorporate Tn5 transposase tagmentation. In one instance, a Tn5 transposase is used to create normalized libraries suitable for NGS without the need for independent adjustment or parallelized normalization of input samples or resulting libraries (PCT/US2016/013753).

However, existing transposase NGS prep methods are limited primarily by their high cost. A transposase with increased DNA fragmentation or cleavage and increased adapter loading or ligation would expand the utility of existing transposase NGS library preparation methods. Transposases with other improved properties, including increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased binding to a target polynucleotide or adapter polynucleotide, and/or decreased sequence insertion bias as compared to another wild-type or engineered transposase, are also highly desirable.

SUMMARY

The present invention provides engineered transposase polypeptides useful in preparation of next generation sequencing libraries, as well as compositions and methods of utilizing these engineered polypeptides. The transposases of the present invention are derived from a wild-type (WT) polypeptide described by UniProt accession number A0A3L8PTF2 a predicted IS4 family transposase from *Parashewanella curva* or from a modified hyperactive Tn5 transposase that contains E54K and L372P amino acid substitutions and is described by NCBI Protein accession number HQ908071.1. A C-terminal 6-histidine tagged version of the transposases (SEQ ID NO: 2 or SEQ ID NO: 5704, respectively) recognize the 19 bp Tn5 mosaic end (ME) transposon donor recognition sequence. Beneficial properties of these engineered transposases include increased binding to adapters or donor DNA to form a transposome, increased DNA cleavage, increased DNA fragmentation, increased activity at elevated temperatures, increased soluble expression, increased ligation of the adapters or donor DNA to the target DNA, decreased inhibition and/or decreased sequence insertion bias, as compared to a wild-type transposase or a reference transposase.

In some embodiments, the present invention provides an engineered transposase polypeptide comprising an amino acid sequence having at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity to a reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, optionally comprising at least one substitution or one substitution set at one or more positions, wherein the positions are numbered with reference to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, and wherein the engineered

transposase polypeptide has increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias as compared to another wild-type transposase or engineered transposase. These engineered transposase polypeptides with one or more amino acid substitutions or substitution sets are described, below, in the detailed description of the invention.

In some embodiments, the present invention provides engineered transposases comprising a polypeptide sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, or a functional fragment thereof, wherein the engineered transposase comprises at least one substitution or substitution set in its polypeptide sequence, and wherein the amino acid positions of the polypeptide sequence are numbered with reference to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

In some additional embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 10, 13, 24, 35, 39, 44, 46, 51, 54, 55, 62/361, 72, 75, 82, 89, 92, 107, 128, 137, 164, 167, 177/359, 191, 192, 193, 194, 201, 217, 281, 289, 291, 291/340, 305, 309, 310, 317, 331, 337, 340, 346, 352, 353, 357, 358, 359, 362, 371, 396, 414/435, 415, 428, 437, 443, 447, and 448, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 4, 5, 6, 8, 10, 10/317, 10/317/358, 10/317/359, 10/346, 10/346/359, 10/358, 10/359, 12, 12/373, 13, 15, 16, 18, 19, 26, 30, 30/154, 31, 34, 35, 37, 38, 68, 70, 73, 74, 93, 106, 118, 126, 134, 135, 143, 146, 154/453, 163, 182, 207/286, 208, 226, 237, 246, 247, 259, 262, 263, 264, 280, 282, 286, 287, 288, 289, 291, 292, 293, 298, 305, 309, 310, 313, 317, 317/346, 317/359, 317/445, 320, 346, 351, 355, 356, 358, 359, 360, 368, 369, 410, 411, 422, 426, 427, 428, 446, 448, 453, and 454, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 18/73/74/259, 18/73/286/289/411, 18/74/118/259/286/289/411, 18/74/259, 18/74/259/286/289, 18/74/259/286/411, 18/118/259, 18/259/286/289/411, 18/259/289/411, 18/289/411, 73/74/259/286/289, 73/74/259/289, 73/259/286, 74/259, 74/259/289, 74/286/289/411, 259, 259/286/289, and 289, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 156.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 16/19/355/427, 16/37/146/286/287/289/310/313/355, 16/146/226/287/289/355, 19/37/146/286/287/310/355, 19/135/286/287/289/355, 19/182/355, 37/70/226/310/313/355,

37/74/310/355/427, 37/135/146/182/226/286/287/355/368, 37/146/226/289/427, 37/146/287/289/427, 38, 41, 42, 43, 44, 45, 47, 48, 50, 53, 54, 55, 70/74/226/286/287/289/368/427, 70/74/355/368, 70/135/146/150/182/286/289/313/427, 74/310/313/355, 74/310/355, 74/355/427, 101, 103, 117, 135/226/286/289/355/368, 135/355, 146, 146/150/287/289/310/355, 146/182/355, 146/226/286/355, 146/286/287/289/355/427, 146/289, 146/427, 150/226/310/313, 182/226/355, 182/355, 244, 247, 249, 251, 287/355/427, 300, 317/411, 355, 393, 397, 398, 399, 401, 405, 407, 408, 409, 411, 413, 426, 429, 431, 439, 444, 446, 450, 453, 454, 457, 459, 460, 464, 466, 467, 469, 470, 471, 472, 474, 475, 476, and 478, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 302.

10 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 16, 16/313, 16/409/411/427, 37/150/408, 37/313/317/408/411, 37/317/405/408, 37/317/405/426, 37/405/409/427, 150, 150/313/405/411, 150/405/406, 150/408, 150/408/411, 226/405, 317/460, and 408/426, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 338.

15 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 12, 13, 18, 19, 22, 26, 31/262, 34, 60, 79, 80, 81, 82, 92, 106, 107, 113, 115, 132, 134, 136, 140, 143, 152, 160, 163, 167, 168, 175, 179, 183, 195, 199, 201, 202, 215, 216, 218, 220, 223, 226, 233, 20 235, 236, 237, 243, 255, 259, 261, 262, 266, 268, 270, 273, 282, 290, 292, 293, 309, 319, 320, 326, 343, 346, 365, 370, 374, 380, 382, 386, 389, 390, 408, 412, 414, 415, 424, and 442, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 2, 12, 13, 15, 16, 18, 19, 22, 26, 30, 31/262, 34, 51, 60, 69, 79, 82, 84, 87, 88, 106, 107, 113, 114, 115, 128, 132, 134, 135, 136, 140, 143, 146, 150, 152, 160, 163, 167, 168, 175, 179, 183, 195, 198, 199, 201, 215, 217, 218, 219, 220, 223, 224, 226, 227, 231, 233, 235, 236, 243, 254, 255, 259, 261, 266, 267, 268, 270, 272, 273, 275, 277, 279, 282, 287, 289, 290, 291, 292, 293, 304, 307, 309, 319, 320, 343, 344, 346, 355, 365, 370, 373, 380, 382, 386, 389, 390, 408, 412, 414, 415, 418, 423, 424, and 468, 30 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 44/45/47/408/453, 44/45/47/453, 44/45/47/453/454, 44/45/103/453/457, 44/45/453/454/457, 35 44/47/103/393/408/453/454, 44/47/103/453/454/457, 44/47/393/408/453/457, 44/47/393/453/457, 44/103/408/453/454, 45/47/103/393/453/454/457, 45/47/393/453/454, 45/47/408/423/453/457, 45/47/453, 45/103/393/453/454, 45/453/454, 47/103/408/453/454, 47/393/408/453/454/457, 47/393/453/454, 47/408/453/457, 53/54/244/450, 103/393/408/453/457, 244/450, 408/453/454, 446/450/464, and 453/457, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 38/45/53, 38/45/53/54/476, 38/45/53/244, 38/45/53/470/472, 38/45/54/244, 38/45/54/470/474/478, 38/45/472/476/478, 38/45/476, 38/53/54, 38/53/54/244, 38/53/54/244/472/478, 5 38/53/472, 38/54/244/470/474/476/478, 38/244/470/472/474/478, 38/478, 45/53/54/244/470/474/476/478, 45/53/54/470/474/476, 45/53/54/472/478, 45/54/244/474/478, 53, 53/54/470/476/478, 53/244/474/478, 54, 54/244, 54/244/472/476, 54/472, 244, and 474/478, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

In some further embodiments, the engineered transposase comprises a polypeptide sequence 10 comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 22/38/45/107/259/317/412, 22/38/45/259/317/412, 22/107, 22/107/317/412, 22/134, 22/134/220, 22/134/220/317, 22/134/317/412, 22/317, 38, 38/45/81/143/244, 38/53/243, 38/81, 38/81/244, 38/107/134/220, 38/132, 38/220/317, 45/81, 45/81/243, 45/81/243/343, 45/143/243/270/343, 45/244/270, 53/115/163/243/270/343, 53/143/243/270/343, 53/270, 64/134/220/317/412, 81/132/243, 15 107, 107/134/317, 115/244/270, 134/259, 143/243/270, 243, 243/270, 244, 244/270, 259, 270, and 317, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions 20 selected from 22/38/45/107/259/317/412, 22/38/45/259/317/412, 22/38/64/134/220, 22/45/220/259, 22/64/107/259/317/412, 22/64/134/220, 22/107, 22/134, 22/134/317, 22/134/317/412, 22/220, 22/220/412/415, 22/259/317, 22/259/317/412, 22/317, 38, 38/45/81/143/244, 38/45/143/270, 38/53/243, 38/81, 38/81/143/243/270, 38/81/183/243/270, 38/81/243, 38/81/244, 38/107/134/220, 38/132, 38/220/317, 38/270/343, 45/81, 45/81/132/243/270, 45/81/243, 45/143/243/270/343, 45/244/270, 25 53/143/243/270/343, 53/244/270, 53/270, 64, 64/107/134, 64/107/317, 64/134, 64/134/220/317/412, 81/132, 81/132/243, 107, 107/134/317, 107/412, 115/244/270, 143/243/270, 220/259, 243, 243/270, 244/270, 255, 259, and 270, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

In some further embodiments, the engineered transposase comprises a polypeptide sequence 30 comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 143, 143/175/272/273/415, 152/168/216/237, 152/168/423, 160, 160/201/272, 160/201/273/415, 160/226, 160/244/272/273/415, 160/415, 168, 168/187/216/255/423, 168/216/237, 168/255/423, 168/290/423, 187, 187/423, 199/201/226, 199/201/415, 201, 216, 216/423, 226/244/415, 226/415, 237, 237/290, 244, 255, 272, 290, 290/423, 415, and 423, and/or any combinations thereof, 35 wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 143, 143/160/226/415, 143/175/272/273/415, 152/168/216/237, 152/168/423, 160, 160/201/272, 160/201/273/415, 160/226, 160/244/272/273/415, 160/415, 168, 168/187/216/255/423, 40 168/216/237, 168/216/237/255/343, 168/255/423, 168/290/423, 187, 199/201/226, 199/201/415, 201,

216, 216/423, 226/244/415, 226/415, 237, 237/290, 244, 255, 255/343, 272, 290, 290/423, 415, and 423, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

5 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 72, 90, 91, 92, 93, 95, 122, 129, 141, 142, 143, 144, 145, 150, 151, 151/155, 152, 154, 155, 157, 160, 161, 161/313, 164, 165, 168, 180, 181, 184, 193, 197, 204, 205, 205/472, 206, 299, and 320, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

10 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 90, 92, 122, 129, 141, 143, 144, 145, 150, 151, 151/155, 152, 154, 155, 155/265, 157, 160, 161, 164, 165, 168, 180, 181, 184, 193, 197, 204, 205, 205/472, 206, 299, and 320, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

15 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 129, 129/160, 129/160/164, 129/160/184, 129/164, 129/164/184, 129/205, 146/152/244, 150/155/226, 150/155/244, 152, 154, 154/160/184, 154/164, 154/205, 155, 160, 160/164, 160/164/184, 20 In 160/184, 164, 164/204, 164/205, 204, 205, 226, 226/244, and 244, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1218.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 129, 129/143, 129/154/164, 129/154/164/205, 129/154/205, 129/160, 129/160/164, 25 129/160/205, 129/164/205, 143, 143/150/155/226/244, 143/150/155/244, 143/150/226, 143/150/226/244/343, 143/150/226/343, 143/150/244, 143/150/343, 143/152, 143/152/226, 143/152/226/244, 143/152/226/343, 143/152/343, 143/155, 143/155/226/244, 143/155/226/244/343, 143/155/343, 143/226, 143/226/244/343, 143/226/343, 143/244, 143/244/343, 143/343, 146/152/244, 150/155/226, 150/155/244, 150/155/244/343, 150/226/244/343, 152, 152/174/244/343, 152/226, 30 152/226/343, 152/343, 154, 154/160/184, 154/164, 154/164/204, 154/164/204/205, 154/205, 155, 160, 160/164, 160/164/184, 164, 205, 226, 226/244, and 244, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1218.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions 35 selected from 98/129/157/168, 129/157/168, 151, 151/152/255, 151/152/273, 151/255, 151/343, 157, 157/164, 157/164/168, 157/168, 164, 164/255, 168, 183, 255, 255/343, and 343, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions 40 selected from 129, 129/157, 129/157/168, 129/164, 129/164/168/183, 129/164/183, 143, 143/151,

143/151/152, 143/151/164, 143/164, 143/164/255, 143/164/343, 143/255, 143/343, 151, 151/152, 151/152/164/343, 151/152/343, 151/255, 151/255/343, 151/343, 152, 152/255, 157, 157/164/168, 157/168, 164, 164/168, 164/255, 164/447, 168, 183, 255, 255/343, 298/343, and 343, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO:

5 1580.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 69, 69/164, 72, 73, 91, 92/164, 95, 96, 97, 98, 110, 112, 113, 121, 121/164, 122, 122/164, 124, 126, 127, 139, 140, 141, 144, 144/164, 145, 146, 147, 148, 150, 150/164, 151, 151/164, 153,
10 153/164, 155, 156, 157, 159, 160, 161, 163, 164, 164/172, 164/173, 164/182, 164/188, 164/191, 164/192, 164/195, 164/197, 164/199, 164/202, 164/205/249, 164/209, 165, 166, 167, 169, 170, 171, 172, 173, 177, 184, 187, 188, 190, 191, 192, 193, 194, 195, 197, 197/266, 198, 199, 200, 202, 203, 204, 205, 208, 209, 210, 321, 324, 331, and 349, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

15 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 69, 69/164, 72, 73, 91, 92/164, 95, 97, 98, 121/164, 126, 127, 139, 140, 142, 144, 144/164, 145, 146, 146/164, 148, 150, 150/164, 151, 153, 153/164, 155, 156, 157, 160, 163, 164/173, 164/188, 164/191, 164/192, 164/195, 164/199, 164/205/249, 164/209, 165, 166, 167, 169, 171, 172, 173, 177, 182,
20 187, 188, 190, 191, 192, 194, 195, 197/266, 198, 199, 200, 202, 203, 204, 205, 208, 209, 210, 324, 331, and 349, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions
25 selected from 73, 91, 93, 95, 97, 98, 110, 112, 122, 122/164, 124, 127, 139, 144, 145, 146, 146/164, 147, 148, 150, 150/164, 151, 151/164, 153, 153/164, 155, 156, 157, 159, 160, 161, 162, 163, 164, 164/182, 164/188, 164/191, 164/192, 164/195, 164/199, 164/205/249, 164/209, 165, 166, 167, 169, 171, 172, 177, 184, 187, 188, 190, 191, 192, 193, 195, 197/266, 198, 199, 200, 204, 205, 206, 207, 208, 209, 210, and 324, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to
30 SEQ ID NO: 1580.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 69, 73, 91, 127, 140, 144, 145, 146, 146/164, 150, 150/164, 151, 153, 155, 156, 157, 160, 163, 164, 164/172, 164/188, 164/191, 164/192, 164/195, 164/199, 164/209, 165, 167, 169, 171, 172, 182,
35 183, 188, 191, 192, 193, 195, 196, 198, 199, 200, 204, 205, 208, 210, and 324, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 38/143/168/220/290/317/423, 38/152/154/164/168/220/290/317/423,
40 38/152/168/220/290/317/343/423, 38/152/168/220/290/317/423, 38/154/164/168/220/290/317/423,

38/168/220/290/317/423, and 38/220/317, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions
5 selected from 73/143/169/191/193, 73/156, 73/156/192/435, 73/156/210, 73/163, 73/163/210, 73/192/193, 127, 127/160/172/198, 127/172, 127/198, 146, 146/151/188, 146/324, 151/171/324, 151/324, 156/169, 156/192, 157/171, 163, 163/169, 163/169/191/192/193, 165/171/324, 165/188/324, 165/324, 169, 171/188/324, 171/324, 172, 188, 192, 192/193, 198, 210, and 324 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2368.

10 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 73, 73/163/191/192/193, 127/160/172, 127/160/172/198, 144, 160, 160/172, 163/191/192, and 192 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2368. In some embodiments, the engineered transposase comprises the polypeptide
15 sequence of SEQ ID NO: 2396. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 2396, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 2396, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed.

20 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 8, 13, 17, 18, 20/146/157/167, 26, 34/38, 38/39/45, 67, 70, 73, 73/150/155/169, 73/150/171/192, 73/155, 73/155/169/199, 73/155/171, 73/192, 73/199, 78, 82, 84, 86, 86/384, 89, 90, 91, 95, 102, 110, 111, 113, 114, 115, 122/127, 126/127, 127/131/165, 127/133, 143, 143/146, 143/146/153,
25 143/146/157, 143/146/167, 143/146/167/188, 143/146/188, 143/153/167/188/246, 143/156/157/167/195, 143/156/188, 143/188, 143/188/195, 143/195, 144, 145, 145/151, 146, 146/152, 146/153/157, 146/153/157/167, 146/153/167, 146/157, 146/157/167, 146/167, 150, 150/155/169, 150/155/169/192, 150/155/192, 150/169, 150/192, 150/199, 151/152/154, 152/154, 152/154/155/160, 153/157/167, 155/171, 155/192, 156/270, 160/371, 167/188, 168/172, 169, 184, 190, 192/198, 198, 206, 208, 209, 210,
30 214/220, 215/220, 216/220, 217/220, 219/220, 220, 220/222, 220/225, 220/226, 231, 235, 242, 257, 259, 261, 262, 267, 268, 269, 270, 272, 274, 275, 276, 277, 278, 282, 283/290, 284/290, 287/290, 288/290, 289/290, 290, 290/291, 290/292, 290/293, 290/294, 304, 305, 305/373, 306, 307, 309, 317, 329, 333, 334, 339, 341, 344, 345, 346, 352, 353, 355, 358, 366, 368, 370, 372, 373, 375, 376, 378, 379, 380, 381, 382, 383, 384, 385, 386/393, 387/393, 389/393, 390/393, 391/393, 392/393, 393/394, 414, 415, 420/423,
35 422/423, 423, 423/425, 432, 440, 441, 443, 469, and 472 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2426.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 8, 18, 25, 29, 30, 67, 73, 73/150/155/169, 73/150/155/171, 73/150/171/192, 73/171, 73/192,
40 73/199, 74, 82, 86, 86/384, 113, 115, 118, 127/128, 143, 150, 150/155/169, 150/155/169/192,

150/155/171, 150/155/171/199, 150/155/192, 150/155/199, 150/192, 150/199, 155/192, 168/171/172, 195/198, 197/198, 198/201, 198/203, 198/205, 235, 252, 256, 257, 262, 263, 267/343, 268, 269, 270, 272, 273, 276, 277, 288/290, 289/290, 290, 290/291, 290/292, 290/294, 298, 304, 305, 307, 333, 346, 355, 366, 370, 386/393, 389/393, 393/395, 440, and 472 and/or any combinations thereof, wherein the
 5 amino acid positions are numbered with reference to SEQ ID NO: 2426.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 66, 67, 70, 101, 102, 103, 104, 106, 107, 108, 113, 114, 115, 117, 118, 143, 152, 154, 235, 236/251, 237, 238, 242, 243, 246, 247, 249, 250, 251, 256/352, 257, 259, 285, 286, 287, 288, 289, 290,
 10 292, 293, 294, 295, 296, 322, 323, 326, 329, 330, 333, 339, 340, 343, 344, 345, and 352. and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2540.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions
 15 selected from 18, 18/113/270, 18/113/270/329/370, 18/113/270/370/384, 18/113/270/370/412, 18/113/329/384, 18/259/370/384, 18/270, 18/270/370, 18/270/370/384, 18/370, 18/370/384, 73, 73/115, 73/115/184/333, 73/115/184/341, 73/115/256, 73/184/256, 73/184/273/333, 73/184/333, 73/256, 73/256/375, 73/273, 73/273/333/375, 73/273/375, 73/333, 73/333/375, 73/375, 73/379, 77/233/238/376, 77/233/334/376, 77/262/376, 77/276, 77/334/376, 82, 82/127, 82/127/144/235/290, 82/127/144/266,
 20 82/127/144/266/290, 82/127/144/378, 82/127/235, 82/127/290, 82/144/266/290/378, 82/144/266/378, 82/235/266, 82/235/266/378/383, 82/235/290, 82/235/378, 82/266, 82/266/378, 82/378, 113/270/370/384, 115/256/333/375, 115/256/375, 115/273/375, 115/375, 127/144/290, 127/149/150, 127/266/290/378, 127/378/383, 144/235/290, 144/266/290, 233/334, 233/376, 235, 235/290, 256/333, 262, 266/378, 270/370, 273, 276/352, 290, 290/378, 333, 358, 370, 375, and 378 and/or any combinations thereof,
 25 wherein the amino acid positions are numbered with reference to SEQ ID NO: 2540.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 18, 18/113, 18/113/252, 18/113/252/336/384, 18/113/252/370, 18/113/270,
 18/113/270/370/384, 18/259/370/384, 18/270, 18/270/370, 18/270/370/384, 18/370, 18/370/384, 73,
 30 73/115/184/341, 73/115/341, 73/131/256/273/333/341/375, 73/184/256, 73/184/256/341, 73/184/256/375, 73/184/273/333, 73/184/333, 73/184/341/375, 73/256, 73/256/273, 73/256/273/375, 73/256/333/375, 73/256/375, 73/273, 73/273/333/341, 73/273/333/375, 73/273/375, 73/333, 73/333/341/375, 73/333/375, 73/341, 73/375, 73/379, 77, 77/118, 77/118/143/262/276/334/376, 77/118/143/334/352/376, 77/118/143/376, 77/118/233, 77/118/233/352/376, 77/118/262/334/376, 77/118/334, 77/143/262/276,
 35 77/143/262/334, 77/143/376, 77/233/238/376, 77/233/262, 77/233/276/376, 77/233/334/376, 77/262/376, 77/276, 77/334/376, 77/352, 82, 82/127, 82/127/144/235/290, 82/127/144/266, 82/127/144/266/290, 82/127/144/378, 82/127/235, 82/127/235/266/358, 82/127/290, 82/127/290/358, 82/127/290/378, 82/144/266/290/378, 82/144/266/378, 82/150/235/266, 82/235/266, 82/235/266/290, 82/235/266/378/383, 82/235/290, 82/235/378, 82/266, 82/266/378, 82/290, 82/378, 113, 113/252, 113/252/270/370,
 40 113/270/370/384, 113/336, 118/233/276/334, 118/334, 127/235/266, 127/266/290/378, 127/378/383, 143,

143/276, 143/276/334, 144/235/290, 144/266/290, 144/290/358, 233/334, 233/376, 235,
 235/266/290/358/383, 235/358, 252/329/336/370, 252/384, 256/333, 256/375, 262, 262/276, 262/334,
 266/290, 266/378, 270/370, 273, 276/352, 290, 290/378, 333, 334, 334/376, 336/370, 341, 358, 370, 375,
 and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference
 5 to SEQ ID NO: 2540.

In some further embodiments, the engineered transposase comprises a polypeptide sequence
 comprising at least one substitution or substitution set comprising substitutions at amino acid positions
 selected from 18, 18/77/118/169/233/238/256/266/270, 18/77/150/256/376, 18/118,
 18/118/143/169/238/256/266/333, 18/118/150/169/184/238/256/270/378, 18/118/169, 18/118/233/238,
 10 18/143, 18/143/256, 18/150/169, 18/150/266, 18/169/184/290/333/376, 18/169/233/333, 18/169/392,
 18/184/270, 18/233/256/334, 18/256/266/290/378, 77/113/118/256/266/290, 77/118/290, 77/169/270/290,
 77/233, 77/233/256, 77/238, 113/118/169/184/256/376, 113/169, 113/233, 118, 118/143/238,
 118/233/238/256/378, 118/233/238/270, 143/169/233/270, 143/233/266/290/376, 143/238/270/290/378,
 143/256/334, 143/290, 169, 169/256/266, 183/270/333, 233/238, 233/238/376, 233/392, 238/256,
 15 256/290/303, 270, 290, 376, and 378 and/or any combinations thereof, wherein the amino acid positions
 are numbered with reference to SEQ ID NO: 3402.

In some further embodiments, the engineered transposase comprises a polypeptide sequence
 comprising at least one substitution or substitution set comprising substitutions at amino acid positions
 selected from 18, 18/77/118/169/266/334/376, 18/77/150/256/376, 18/77/256/290/334, 18/113/238,
 20 18/118, 18/118/143/169/238/256/266/333, 18/118/143/169/376, 18/118/169, 18/118/233/238, 18/118/266,
 18/143, 18/143/169/233/238/358, 18/143/233/238/290/376, 18/143/256, 18/143/290, 18/143/333/334/378,
 18/150/169, 18/150/266, 18/169/184/290/333/376, 18/169/233/333, 18/169/238/270/334/376, 18/169/392,
 18/233/256/334, 18/233/266/290/376, 18/256/266/290/378, 18/266/270/378, 18/378,
 77/113/121/143/233/334, 77/118/169/233/266/270, 77/118/238/376, 77/118/290, 77/169/270/290, 77/233,
 25 77/238, 113/238/376, 118, 118/143/238, 118/143/392, 118/169/184/233/266/270/333,
 118/233/238/256/378, 118/256/290/333/334/376/378, 118/256/334/376, 118/266/290, 143/169/233/270,
 143/233/266/290/376, 143/238/270/290/378, 143/256/334, 143/290, 143/364, 150/233/333, 169,
 169/256/266, 183/270/333, 233/238, 233/238/376, 233/392, 238/256, 238/266/270/378, 256/266/334,
 266/270/376, 266/333, and 378 and/or any combinations thereof, wherein the amino acid positions are
 30 numbered with reference to SEQ ID NO: 3402.

In some further embodiments, the engineered transposase comprises a polypeptide sequence
 comprising at least one substitution or substitution set comprising substitutions at amino acid positions
 selected from 77/156, 77/156/270, 77/184/244, 77/184/333, 77/244/270, 77/270, 143,
 143/169/238/257/344, 143/169/255, 143/169/255/257, 143/169/343, 143/255, 143/255/257, 143/255/344,
 35 143/257, 156, 156/270, 169/238, 169/238/255/257/344, 169/257/344, 169/344, 184, 184/244/324, 238,
 238/255/344, 244/270, 257/344, and 270 and/or any combinations thereof, wherein the amino acid
 positions are numbered with reference to SEQ ID NO: 3662.

In some further embodiments, the engineered transposase comprises a polypeptide sequence
 comprising at least one substitution or substitution set comprising substitutions at amino acid positions
 40 selected from 77, 77/156/184/244/324, 77/184/270, 143, 143/238, 143/238/255, 143/238/255/343,

143/238/255/392, 143/238/344, 143/255, 143/257/392, 143/344, 169, 169/238/255/344, 169/238/344, 169/255, 169/343, 169/392, 238/344, 255, 255/257, 255/257/392, 257, and 344 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3662.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 20, 22, 26, 42, 43, 44, 45, 50, 51, 52, 53, 54, 55, 60, 62, 64, 65, 95, 97, 99, 103, 104, 111, 112, 115, 115/143/169, 115/169/343, 115/343, 116, 116/343, 143, 143/169, 154/439, 169, 169/324, 169/324/343, 169/343, 186, 234, 236, 238, 238/246, 239, 239/244, 241, 242, 243, 244, 250, 251, 254, 255, 260, 283, 284, 295, 296, 298/326, 319, 322, 323, 324, 324/343, 326, 328, 329, 333, 334, 338, 340, 341, 343, 344, 347, 366, 434, 439, 442, 443, 444, 445, 446, and 447 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3788.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 26, 26/236, 26/236/260/333, 26/236/295, 26/260, 26/260/295, 26/260/295/333, 26/295/319, 26/333, 64, 64/244, 64/251/260/347, 64/260/295/333, 64/260/347, 103/115/255, 103/115/343, 103/115/343/344, 103/143/255/284/343/344, 103/169/343/344, 103/255/284, 103/343/344, 115, 115/143/255, 115/143/255/343, 115/143/284, 115/143/343, 115/143/344, 115/169, 115/169/343, 115/255, 115/255/344, 115/284, 115/284/343/344, 115/284/344, 115/343, 115/343/344, 115/344, 143, 143/169/255/284/343/344, 143/169/343/344, 143/255, 143/255/284/343/344, 143/255/343, 143/255/343/344, 143/255/344, 143/343/344, 169, 169/255, 169/255/284, 169/255/343, 169/284, 169/343, 169/344, 236, 236/244/333, 236/260, 236/347, 244, 244/251, 244/251/319, 244/260, 244/295, 244/347, 251, 251/260/295/347, 251/333, 255, 255/284, 255/343, 255/343/344, 255/344, 259, 260, 260/295, 260/333, 284, 284/343/344, 284/344, 295, 295/333, 333, 343, 343/344, 344, and 347 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4150.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 115, 115/117, 115/117/118, 115/117/118/152, 115/117/118/169, 115/117/118/246, 115/117/152, 115/117/152/169, 115/117/152/169/246, 115/117/152/246, 115/117/169, 115/117/246, 115/118, 115/118/152, 115/118/152/169, 115/118/152/169/246, 115/118/169, 115/118/246, 115/152, 115/152/169, 115/152/169/246, 115/152/246, 115/152/246/374, 115/169, 115/169/246, 115/246, 117, 117/118, 117/118/152, 117/118/152/169, 117/118/152/246, 117/118/169, 117/118/246, 117/152, 117/152/169, 117/152/246/288, 117/169, 117/246, 118/246, 152, 152/169, 152/169/246, 152/246, 169, 169/246, and 246 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4278. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4642. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4642, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4642, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 73, 101, 114, 115, 118, 143, 145, 150, 151, 152, 153, 154, 156, 157, 160, 161, 163, 166, 167, 168, 169, 171, 172, 190, 191, 192, 195, 213, 214, 218, 219, 223, 225, 233, 233/343, 234, 235, 236, 237, 238, 239, 239/250, 243, 244, 249, 251, 254, 255, 257, 258, 259, 261, 262, 262/283, 277/292, 283, 284, 286, 289, 292, 295, and 296 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4278.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 93, 113, 113/114, 113/115, 113/143, 115/118, 115/177, 118/143/163/168/195/255/315, 118/143/255, 118/163/168/214/255/315, 118/168/214/315, 118/255, 124, 126, 129, 134, 143/154/255, 143/156/168/216/255, 143/163, 143/163/255, 143/168, 143/168/169/255/315, 143/255, 148, 156/214/315, 159, 163, 163/168, 163/168/169, 163/168/169/255/315, 163/168/169/315, 163/168/195/214/255/315, 163/214, 163/255, 163/255/315, 168/315, 186, 194, 197, 206, 207, 208, 210, 213, 255/315, 277, 293, 294, 315, 317, 320, 322, 328, 329, 330, 333, 352, 366, 392, and 445 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4642. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4872. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4872, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4872, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4954. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4954, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4954, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 127, 128, 129, 140, 180, 182, 194, 277, 293, 294, 299, 300, 317, 319, 320, 324, 333, 356, 366, and 392 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4642.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 113, 117, 118, 151, 191, 196, and 205 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4984.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 114/115/143, 114/115/143/233, 114/115/233, 114/115/343, 114/143, 114/143/151/233/238/261, 114/143/213/261/343, 114/143/233, 114/143/233/238, 114/151/261/343, 114/213/238, 114/343, 115, 115/143, 115/143/151/154/213/261, 115/143/151/154/233/343,

115/143/151/233/261, 115/143/151/233/343, 115/143/154/213/233/238/261, 115/143/213, 115/143/233, 115/143/343, 115/151/154/233, 115/151/261, 115/213/233, 115/213/233/238, 115/213/261, 115/233/238, 115/238, 115/343, 115/348, 143, 143/151/154, 143/151/154/213, 143/154, 143/213/233, 143/213/233/238/261/343, 143/213/261/343, 143/233/261, 143/261, 151/154, 151/154/213/343, 151/154/343, 151/213, 151/213/233, 213/233/343, and 233/238/343 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4984. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5186. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5186, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5186, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5354. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5354, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5354, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 117, 117/118/143/168/169/392, 117/118/143/392, 117/118/168/169, 117/118/168/169/392, 117/143/168/169/170/392, 117/143/168/169/392, 117/168/169, 117/168/169/392, 117/168/392, 117/169, 117/392, 118/143, 118/143/168/169, 118/143/169, 118/143/392, 118/168, 118/168/392, 118/169/392, 118/392, 143/168/169/392, 143/169, 143/392, 168/169, 168/169/392, 168/392, 169/392, and 392 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5196.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 134, 134/182/194/294, 134/182/277/293, 134/182/293/320, 134/182/294, 134/194/277, 134/194/277/317, 134/194/293, 134/194/293/294, 134/194/293/317, 134/194/294, 134/194/317, 134/194/317/320, 134/317, 182, 182/194, 182/194/293/320, 182/293/294, 194, 194/277/293/317, 194/277/317/320, 194/277/320, 194/293/294/317, 194/293/320, 277, 293, 317, 317/320, and 320 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5286.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 12/134, 13/134, 13/134/210, 22/134/152, 22/134/235/344, 64/93/134/235, 64/134/344/347, 77/134, 79/134/210, 84/134, 86/134, 87/134, 89/134, 93/134/152/235/344, 93/134/235/344, 93/134/244/344, 93/134/344, 131, 134/152/344, 134/175, 134/210/382, 134/210/402, 134/210/456, 134/210/478, 134/344, 134/377, 134/378, 134/380, 134/381, 134/382, 134/384, 134/386, 134/387, 134/388, 134/390, 134/391, 134/393, 134/395, 134/398, 134/401, 134/402, 134/406, 134/407, 134/421, 134/456, 134/457, 134/459, 134/460, 134/464, 134/467, 134/468, 134/470, 134/471, 134/473,

134/473/474, 134/474, and 134/475 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5412.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions
 5 selected from 73/127/134/160/172/198/290, 73/127/134/160/198/333, 73/127/134/160/210/266, 73/127/134/172/192/266/333, 73/127/134/172/198, 73/127/134/210/290, 73/127/134/210/333, 73/127/134/290/333, 73/127/134/333, 73/134, 73/134/160/172/192/290, 73/134/160/172/198/333, 73/134/160/172/290, 73/134/160/266, 73/134/160/333, 73/134/172, 73/134/172/210/333, 73/134/172/266/333, 73/134/198/210, 73/134/198/210/266, 73/134/210/333, 73/134/333, 73/134/366,
 10 127/134, 127/134/160/172/192/333, 127/134/160/172/210/266/290, 127/134/172/198/266/333, 127/134/192/210, 127/134/198/210/290, 127/134/210/266/290, 127/134/210/333, 134, 134/160/172/333, 134/160/198/266/290, 134/172/192/290/333, 134/192/198, 134/198, 134/198/266/333, 134/210, 134/210/290, and 134/333 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5412. In some embodiments, the engineered transposase
 15 comprises the polypeptide sequence of SEQ ID NO: 5500. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5500, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5500, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed. In some embodiments, the engineered transposase
 20 comprises the polypeptide sequence of SEQ ID NO: 5520. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5520, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5520, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed.

25 In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from 13, 13/178/381, 13/178/381/473, 13/178/457, 13/381, 13/381/457, 13/402/457, 13/457, 178, 178/381/457, 178/381/457/473, 178/402/457, 178/473, 381, 381/393/457, 381/457/473, 381/473, and 393/457 and/or any combinations thereof, wherein the amino acid positions are numbered with reference
 30 to SEQ ID NO: 5422.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions selected from the reference sequence is SEQ ID NO: 5704 and the at least one substitution comprises substitutions at amino acid positions selected from 48, 116, 118, 120, 130, 137, 153, 157, 163, 175, 181,
 35 187, 192, 213, 246, 263, 273, 296, 318, 334, 341, 394, 412, 424, 454, and 458 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5704. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5738. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5738, wherein the C-terminal histidine tag is replaced or removed. In some embodiments, the engineered

transposase comprises the polypeptide sequence of SEQ ID NO: 5738, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed.

In some further embodiments, the engineered transposase comprises a polypeptide sequence comprising at least one substitution or substitution set comprising substitutions at amino acid positions
5 selected from

38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/194/198/220/243/260/266/270/295/317/333/392/423,

38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/392/423,

10 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/392/423/457,

38/73/113/115/117/118/127/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/392/423,

15 38/73/113/115/118/127/152/154/160/162/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423,

38/73/113/115/118/127/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/423,

20 38/73/113/115/118/127/152/154/160/164/168/172/184/189/192/198/220/243/260/266/270/295/317/333/423,

38/73/113/115/118/127/152/154/160/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423,

38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423,

38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/243/266/270/317/333/423,

25 38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/266/270/317/333/423, and

38/113/127/152/154/160/164/168/172/198/220/290/317/423,

and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5702. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5702, wherein the C-terminal histidine tag is replaced or removed. In some
30 embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5702, wherein the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof is replaced or removed.

In some additional embodiments, the present invention provides engineered transposases capable of binding to adapters or donor DNA to form a transposome, cleaving or fragmenting target DNA,
35 and/or ligating the adapters or donor DNA to the target DNA.

In some further embodiments, the present invention provides engineered transposases have at least one improved property, as compared to a wild-type or reference transposase. In some embodiments, the improved property is selected from increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble

expression, decreased inhibition, increased adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias.

In still some additional embodiments, the present invention provides engineered transposases comprise increased polynucleotide fragmentation or cleavage activity as compared to a wild-type or
5 reference transposase.

In yet some additional embodiments, the engineered transposases comprise increased thermal stability as compared to a wild-type or reference transposase. In some further embodiments, the engineered transposase comprises increased activity at elevated temperatures as compared to a wild-type or reference transposase.

10 In some further embodiments, the engineered transposases comprise increased adapter loading or adapter ligation (i.e. tagging) as compared to a wild-type or reference transposase.

In some additional embodiments, the engineered transposases comprise increased binding to a target polynucleotide, as compared to a wild-type or reference transposase.

15 In yet some further embodiments, the engineered transposases comprise decreased sequence insertion bias, as compared to a wild-type or reference transposase.

In some additional embodiments, the engineered polypeptide comprises an amino acid sequence with at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or greater sequence identity to any polypeptide sequence set forth in SEQ ID NOs: 2-2368, 2387-5694, and 5706-5756.

20 In some additional embodiments, the engineered transposase comprises a polypeptide sequence selected from the polypeptide sequences between SEQ ID NOs: 2-2368, 2388-5694, and 5706-5756.

In yet some additional embodiments, the present invention provides purified engineered transposases.

The present invention also provides compositions comprising at least one engineered transposase provided herein.

25 The present invention also provides engineered polynucleotide sequences encoding at least one engineered transposase provided herein.

In some embodiments, the present invention provides engineered polynucleotide sequences comprising at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 1, 3, 155, 301, 337, 605, 971, 1081, 1217, 1579, 1597,
30 1599, 1603, 1719, 2367, 2425, 2539, 3401, 3661, 3787, 4149, 4277, 4641, 4983, 5195, 5285, 5411, 5421, and/or 5703-5755, and/or or a functional fragment thereof, wherein said polynucleotide sequence encodes an engineered polypeptide comprising at least one substitution at one or more amino acid positions. In some embodiments, a functional fragment of an engineered polynucleotide sequence does not contain the segment of nucleotides encoding the C-terminal peptide spacer GGSG (SEQ ID NO:
35 5777) and/or the 6-histidine tag.

In some further embodiments, the engineered polynucleotide sequences encode at least one engineered transposase comprising a sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788,
40 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

In some additional embodiments, the engineered polynucleotide sequence comprises SEQ ID NOs: 1, 3, 155, 301, 337, 605, 971, 1081, 1217, 1579, 1597, 1599, 1603, 1719, 2367, 2425, 2539, 3401, 3661, 3787, 4149, 4277, 4641, 4983, 5195, 5285, 5411, 5421, and/or 5705-5755.

In yet some additional embodiments, the engineered polynucleotide sequence is operably linked
5 to a control sequence.

In still some other embodiments, the engineered polynucleotide sequence is codon-optimized.

The present invention also provides an engineered polynucleotide encoding at least one engineered polypeptide described in the above paragraphs. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NOs: 1-2367, SEQ ID NOs:
10 2385-5693, and SEQ ID NOs: 5705-5755. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NO: 2395. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NO: 4641. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NO: 4871. In some embodiments, the engineered polynucleotide comprises the polynucleotide
15 sequences set forth in SEQ ID NO: 4953. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NO: 5185. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NO: 5353. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NO: 5499. In some embodiments, the engineered polynucleotide comprises the polynucleotide
20 sequences set forth in SEQ ID NO: 5519. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NO: 5701. In some embodiments, the engineered polynucleotide comprises the polynucleotide sequences set forth in SEQ ID NO: 5737. In some embodiments, the engineered polynucleotide comprises one of the foregoing polynucleotide sequence or a polynucleotide sequence encoding an engineered polypeptide described herein (e.g., an
25 engineered transposase), wherein the segment of nucleotides encoding the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) and/or the 6-histidine tag is modified or removed. In some embodiments, the segment of nucleotides encoding the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) is modified or replaced to encode a different C-terminal peptide spacer. In some embodiments, the segment of nucleotides encoding the C-terminal 6-histidine tag is modified or replaced to encode a different tag. In
30 some embodiments, the segment of nucleotides encoding the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) and/or the 6-histidine tag is removed.

The present invention further provides vectors comprising at least one engineered polynucleotide described above. In some embodiments, the vectors further comprise at least one control sequence.

The present invention also provides host cells comprising the vectors provided herein. In some
35 embodiments, the host cell produces at least one engineered polypeptide provided herein.

The present invention further provides methods of producing an engineered transposase polypeptide, comprising the steps of culturing the host cell provided herein under conditions such that the engineered polynucleotide is expressed and the engineered polypeptide is produced. In some
40 embodiments, the methods further comprise the step of recovering the engineered polypeptide. In some embodiments, the methods further comprise the step of purifying at least one engineered transposase.

In some embodiments, the invention provides, a method for generating a mixture of altered nucleic acids comprising exposing a target nucleic acid to an engineered transposase under conditions and for a time sufficient for the engineered transposase to carry out a transposition event, wherein the engineered transposase comprises a polypeptide sequence comprising at least 85% sequence identity to a reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, or a functional fragment thereof, wherein the engineered transposase comprises at least one substitution or substitution set in its polypeptide sequence relative to the reference sequence.

In some embodiments, the altered nucleic acids are fragmented from the target nucleic acid.

In some embodiments, the invention provides, a method of preparing a nucleic acid library comprising:

(a) contacting a target nucleic acid with an engineered transposase under conditions and for a time sufficient for the engineered transposase to carry out a transposition event, wherein the engineered transposase comprises a polypeptide sequence comprising at least 85% sequence identity to a reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, or a functional fragment thereof, wherein the engineered transposase comprises at least one substitution or substitution set in its polypeptide sequence relative to the reference sequences;

(b) fragmenting the target nucleic acid into a plurality of nucleic acid fragments;

(c) selecting a subset of nucleic acid fragments from the plurality of nucleic acid fragments; and

(d) amplifying the subset of nucleic acid fragments to generate the nucleic acid library.

In some embodiments, the invention provides, a method of sequencing a target nucleic acid comprising:

(a) contacting the target nucleic acid with an engineered transposase under conditions and for a time sufficient for the engineered transposase to carry out a transposition event, wherein the engineered transposase comprises a polypeptide sequence comprising at least 85% sequence identity to a reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, or a functional fragment thereof, wherein the engineered transposase comprises at least one substitution or substitution set in its polypeptide sequence relative to the reference sequence;

(b) fragmenting the target nucleic acid into a plurality of nucleic acid fragments;

(c) selecting a subset of nucleic acid fragments from the plurality of nucleic acid fragments;

(d) amplifying the subset of nucleic acid fragments to generate a nucleic acid library; and

(e) sequencing the nucleic acid library.

In some embodiments, the method further includes adding a polynucleotide to the 5' and/or 3' end of each nucleic acid fragment of the plurality of nucleic acid fragments.

In some embodiments, the engineered transposase is any one of the engineered transposases of the foregoing embodiments. In some embodiments, the engineered transposase comprises the

polypeptide sequence of SEQ ID NO: 2396. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4642. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4872. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 4954. In some

5 embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5186. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5354. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5500. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5520. In some embodiments, the engineered transposase comprises the polypeptide

10 sequence of SEQ ID NO: 5702. In some embodiments, the engineered transposase comprises the polypeptide sequence of SEQ ID NO: 5738. In some embodiments, the polypeptide sequence of the engineered transposase is modified to replace or remove the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) and/or the 6-histidine tag. In some embodiments, the polypeptide sequence of the engineered transposase is modified to replace the C-terminal peptide spacer GGSG (SEQ ID NO: 5777) with a

15 different peptide spacer. In some embodiments, the polypeptide sequence of the engineered transposase is modified to replace the 6-histidine tag with a different tag. In some embodiments, the polypeptide sequence of the engineered transposase is modified to remove the C-terminal histidine tag and the peptide spacer GGSG (SEQ ID NO: 5777) or a portion thereof.

 In some embodiments, the engineered transposase is purified.

20 In some embodiments, fragmenting the target nucleic acid comprises tagmentation or random sheering and adapter ligation. In some embodiments, fragmenting the target nucleic acid comprises tagmentation.

 In some embodiments, each altered nucleic acid of the mixture of altered nucleic acids comprises an identifiable sequence tag (IST).

25 In some embodiments, each nucleic acid fragment of the subset of nucleic acid fragments comprises an identifiable sequence tag (IST). In some embodiments, the IST is between 6 and 30 nucleotides in length (e.g., 6-10, 10-15, 15-20, 20-25, 25-30, such as 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 nucleotides in length).

 In some embodiments, each altered nucleic acid of the mixture of altered nucleic acids is between

30 30 and 30000 nucleotides in length (e.g., 30-30000, 30-25000, 30-20000, 30-15000, 30-10000, 30-7500, 30-5000, 30-4000, 30-3000, 30-2500, 30-2000, 30-1500, 30-1000, 30-900, 30-800, 30-700, 30-600, 30-500, 30-400, 30-300, 30-250, 30-200, 30-150, 30-100, 30-50, 100-500, 100-1000, 100-1500, 100-2500, 100-5000, 100-10000, 100-20000, 100-30000, 500-1000, 500-2500, 500-5000, 500-10000, 500-20000, 500-30000, 1000-2500, 1000-5000, 1000-10000, 1000-20000, 1000-30000, 5000-10000, 5000-20000,

35 5000-30000, 10000-20000, 20000-30000, 2500-5000, 7500-10000, or 15000-25000 nucleotides in length).

 In some embodiments, each nucleic acid fragment of the plurality of nucleic acid fragments is between 30 and 30000 nucleotides in length (e.g., 30-30000, 30-25000, 30-20000, 30-15000, 30-10000, 30-7500, 30-5000, 30-4000, 30-3000, 30-2500, 30-2000, 30-1500, 30-1000, 30-900, 30-800, 30-700, 30-600, 30-500, 30-400, 30-300, 30-250, 30-200, 30-150, 30-100, 30-50, 100-500, 100-1000, 100-1500, 100-

40 600, 30-500, 30-400, 30-300, 30-250, 30-200, 30-150, 30-100, 30-50, 100-500, 100-1000, 100-1500, 100-

2500, 100-5000, 100-10000, 100-20000, 100-30000, 500-1000, 500-2500, 500-5000, 500-10000, 500-20000, 500-30000, 1000-2500, 1000-5000, 1000-10000, 1000-20000, 1000-30000, 5000-10000, 5000-20000, 5000-30000, 10000-20000, 20000-30000, 2500-5000, 7500-10000, or 15000-25000 nucleotides in length).

5 In some embodiments, the transposition event occurs in reaction conditions additionally comprising a terminal deoxynucleotide transferase, dNTP, and/or buffer components suitable for the addition of deoxynucleotides to the 3' terminus of each of the plurality of nucleic acid fragments.

In some embodiments, the transposition event occurs in reaction conditions additionally comprising a DNA ligase and/or buffer components suitable for a ligation reaction.

10 In some embodiments, each altered nucleic acid of the mixture of altered nucleic acids further comprises a sample tag and/or a unique molecular identifier (UMI). In some embodiments, each nucleic acid fragment of the subset of nucleic acid fragments further comprises a sample tag and/or a unique molecular identifier (UMI).

15 In some embodiments, amplification of the subset of nucleic acid fragments is performed through polymerase chain reaction (PCR), multiple displacement amplification (MDA), ligase chain reaction (LCR), loop mediated isothermal amplification (LAMP), rolling circle amplification (RCA), or strand displacement amplification (SDA).

In some embodiments, the sequencing comprises next-generation sequencing (NGS).

20 In some embodiments, the sequencing comprises sequencing by synthesis, sequencing by ligation, or nanopore sequencing. In some embodiments, the sequencing by synthesis comprises Illumina™ dye sequencing, single-molecule real-time (SMRT™) sequencing, or pyrosequencing. In some embodiments, the sequencing by ligation comprises polony-based sequencing or SOLiD™ sequencing.

In some embodiments, the target nucleic acid comprises genomic DNA or cDNAs from a single cell. In some embodiments, the target nucleic acid comprises nucleic acids from a plurality of haplotypes.

25 In some embodiments, the target nucleic acid is crosslinked via histones or chromatin from single or multiple cells. In some embodiments, the target nucleic acid has been condensed or optionally treated with one or more condensing agents.

BRIEF DESCRIPTION OF THE DRAWINGS

30 **Figure 1** depicts active transposase variants, bound to appropriate adapter sequences, that are capable of binding target sequences, cleaving the target strand, and ligating the adapter oligonucleotides to the target fragment.

Figure 2 provides a scheme depicting the capillary electrophoresis (CE) assay used in the Examples to assess the activity of the engineered transposase variants.

35 **Figure 3** is a plot of fragment sizes of NGS libraries prepared from 50 ng input DNA sample. Libraries prepared with a transposase that has the polypeptide sequence set forth in SEQ ID NO: 5704 or a commercially available Tn5 transposase are indicated.

Figures 4A through 4D are plots showing the insertion bias of the indicated transposases as determined by the calculated root mean square error (RMSE) of the first 25 bases of the insert.

Figures 5A and 5B are gels showing the results of the enzyme activity assay comparing engineered transposase (SequenceK; **Figure 5A**) with a commercial transposase (hyperactive Tn5) for their ability to convert circular plasmid to linear form. 2.5 pmol transposase was used in the comparison. Plasmids are present in the gel as the higher length bars, i.e. >3000 bp, while the linear form fragments are shown as the smear $\leq$ 3000 bp.

Figure 6 is a plot of fragment sizes of NGS libraries prepared from 50 ng – 200 ng input DNA sample.

Figure 7 is a plot showing the estimated NGS library size generated from 100 ng of human NA12878 genomic DNA using different engineered transposases (SequenceE, SequenceF, SequenceG, SequenceH, SequenceK, and SequenceL) and commercial hyperactive Tn5 transposases (Hyp-Tn5).

Figure 8 is a plot showing GC coverage bias of NGS libraries generated from 100 ng of human NA12878 genomic DNA using different engineered transposases (SequenceE; SequenceK; SequenceF; SequenceG; SequenceH; and SequenceL) and commercial hyperactive Tn5 transposases.

Figure 9 is a plot showing fold-coverage of NGS libraries generated from 100 ng of human NA12878 genomic DNA using different engineered transposases (SequenceE; SequenceK; SequenceF; SequenceG; SequenceH; and SequenceL) and commercial hyperactive Tn5 transposases.

Figures 10A through 10D are plots showing the per base sequence content of the transposition sites of the different commercial transposases and engineered transposases. Each nucleotide (A, T, G, or C) is represented by a single line trace, representing the percentage of reads having that nucleotide at the corresponding position relative to the read start site (transposition site) of the DNA molecules in the sequenced library.

Figure 11 are plots showing GC coverage bias of NGS libraries prepared by ExpressPlex using hyperactive Tn5 transposase and the SequenceH engineered transposase from *Staphylococcus epidermidis*, *Bacillus subtilis*, *Escherichia coli*, and *Pseudomonas aeruginosa*.

Figure 12 is a plot showing the coverage profiles of NGS libraries prepared by ExpressPlex of lambda amplicon 18 DNA samples using a commercial and engineered (SequenceE, SequenceH, and SequenceK) transposases. Data is shown as the raw coverage profile as well as the summary statistic in the form of the coefficient of variation (CV) of the raw coverage profile.

Figure 13 is a plot showing the coverage profiles of NGS libraries prepared by ExpressPlex of pUC19 DNA samples using commercial and engineered (SequenceE, SequenceH, and SequenceK) transposases. Data is shown as the raw coverage profile as well as the summary statistic in the form of the coefficient of variation (CV) of the raw coverage profile.

DESCRIPTION OF THE INVENTION

Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Generally, the nomenclature used herein and the laboratory procedures of cell culture, molecular genetics, microbiology, organic chemistry, analytical chemistry and nucleic acid chemistry described below are those well-known and commonly employed in the art. Such techniques are well-known and described in numerous texts and reference works well known to those of skill in the art. Standard

techniques, or modifications thereof, are used for chemical syntheses and chemical analyses. All patents, patent applications, articles and publications mentioned herein, both *supra* and *infra*, are hereby expressly incorporated herein by reference.

Although any suitable methods and materials similar or equivalent to those described herein find use in the practice of the present invention, some methods and materials are described herein. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art. Accordingly, the terms defined immediately below are more fully described by reference to the invention as a whole.

It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the present invention. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described. Numeric ranges are inclusive of the numbers defining the range. Thus, every numerical range disclosed herein is intended to encompass every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. It is also intended that every maximum (or minimum) numerical limitation disclosed herein includes every lower (or higher) numerical limitation, as if such lower (or higher) numerical limitations were expressly written herein.

As used in this specification and the appended claims, the singular forms “a”, “an” and “the” include plural referents unless the context clearly indicates otherwise. Thus, for example, reference to “a polypeptide” includes more than one polypeptide. Similarly, “comprise,” “comprises,” “comprising” “include,” “includes,” and “including” are interchangeable and not intended to be limiting.

It is to be understood that where descriptions of various embodiments use the term “comprising,” those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language “consisting essentially of” or “consisting of.” It is to be further understood that where descriptions of various embodiments use the term “optional” or “optionally” the subsequently described event or circumstance may or may not occur, and that the description includes instances where the event or circumstance occurs and instances in which it does not. It is to be understood that both the foregoing general description, and the following detailed description are exemplary and explanatory only and are not restrictive of this disclosure. The section headings used herein are for organizational purposes only and not to be construed as limiting the subject matter described.

Abbreviations:

The abbreviations used for the genetically encoded amino acids are conventional and are as follows:

Amino Acid	Three-Letter	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R

Amino Acid	Three-Letter	One-Letter Abbreviation
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

When the three-letter abbreviations are used, unless specifically preceded by an “L” or a “D” or clear from the context in which the abbreviation is used, the amino acid may be in either the L- or D-configuration about α -carbon (C_α). For example, whereas “Ala” designates alanine without specifying the configuration about the α -carbon, “D-Ala” and “L-Ala” designate D-alanine and L-alanine, respectively.

When the one-letter abbreviations are used, upper case letters designate amino acids in the L-configuration about the α -carbon and lower case letters designate amino acids in the D-configuration about the α -carbon. For example, “A” designates L-alanine and “a” designates D-alanine. When polypeptide sequences are presented as a string of one-letter or three-letter abbreviations (or mixtures thereof), the sequences are presented in the amino (N) to carboxy (C) direction in accordance with common convention.

The abbreviations used for the genetically encoding nucleosides are conventional and are as follows: adenosine (A); guanosine (G); cytidine (C); thymidine (T); and uridine (U). These abbreviations are also used interchangeably for nucleosides and nucleotides (nucleosides with one or more phosphate groups). Unless specifically delineated, the abbreviated nucleosides or nucleotides may be either ribonucleosides (or ribonucleotides) or 2'-deoxyribonucleosides (or 2'-deoxyribonucleotides). The nucleosides or nucleotides may also be modified at the 3' position. The nucleosides or nucleotides may be specified as being either ribonucleosides (or ribonucleotides) or 2'-deoxyribonucleosides (or 2'-deoxyribonucleotides) on an individual basis or on an aggregate basis. When nucleic acid sequences are

presented as a string of one-letter abbreviations, the sequences are presented in the 5' to 3' direction in accordance with common convention, and the phosphates are not indicated.

Definitions:

5 In reference to the present invention, the technical and scientific terms used in the descriptions herein will have the meanings commonly understood by one of ordinary skill in the art, unless specifically defined otherwise. Accordingly, the following terms are intended to have the following meanings.

“EC” number refers to the Enzyme Nomenclature of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). The IUBMB biochemical
10 classification is a numerical classification system for enzymes based on the chemical reactions they catalyze.

“ATCC” refers to the American Type Culture Collection whose biorepository collection includes genes and strains.

“NCBI” refers to National Center for Biological Information and the sequence databases provided
15 therein.

“Protein,” “polypeptide,” and “peptide” are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation, phosphorylation, lipidation, myristilation, ubiquitination, etc.). Included within this definition are D- and L-amino acids, and mixtures of D- and L-amino acids, as well as polymers
20 comprising D- and L-amino acids, and mixtures of D- and L-amino acids.

“Amino acids” are referred to herein by either their commonly known three-letter symbols or by the one-letter symbols recommended by IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single letter codes.

As used herein, “polynucleotide,” “oligonucleotide,” and “nucleic acid” are used interchangeably
25 herein and refer to two or more nucleosides or nucleotides that are covalently linked together. The polynucleotide may be wholly comprised of ribonucleotides (i.e., RNA), wholly comprised of 2' deoxyribonucleotides (i.e., DNA), wholly comprised of other synthetic nucleotides or comprised of mixtures of synthetic, ribo- and/or 2' deoxyribonucleotides. The polynucleotides may also include modified nucleotides with substitutions, including 2' substitutions (e.g., 2'-fluoro, 2'-O-methyl, 2'-O-methoxyethyl,
30 locked or constrained ethyl modifications, and others known to those skilled in the art). Nucleosides will be linked together via standard phosphodiester linkages or via one or more non-standard linkages, including but not limited to phosphorothioate linkages. The polynucleotide may be single-stranded or double-stranded or may include both single-stranded regions and double-stranded regions. Moreover, while a polynucleotide will typically be composed of the naturally occurring encoding nucleobases (i.e., adenine,
35 guanine, uracil, thymine and cytosine), it may include one or more modified and/or synthetic nucleobases, such as, for example, inosine, xanthine, hypoxanthine, etc. In some embodiments, such modified or synthetic nucleobases are nucleobases encoding amino-acid sequences. Nucleobases that are modified or synthetic may comprise any known or hypothetical or future discovered modification or structure that would be recognized by one of skill in the art as a modified or synthetic nucleobase. Examples of modified
40 nucleobases or nucleotides include, but are not limited to, diaminopurine, 5-fluorouracil, 5-bromouracil, 5-

chlorouracil, 5-iodouracil, hypoxanthine, xantine, 4-acetylcytosine, 5-(carboxyhydroxymethyl)uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, 5-methyl-2-thiouracil, and 3-(3-amino-3-N-2-carboxypropyl) uracil. A polynucleotide may contain one or more chemical modifications, including modifications on the 5' end, the 3' end, or internally. Examples of chemical modifications include, but are not limited to, addition of functional groups (e.g., biotins, amino modifiers, alkynes, thiol modifiers, or azides), fluorophores (e.g. quantum dots or organic dyes), spacers (e.g. C3 spacer, dSpacer, photo-cleavable spacers), modified bases, or modified backbones. Similarly, the terms "polynucleotide," "oligonucleotide," and "nucleic acid" are intended to comprise any modified or synthetic structure that is now known or discovered in the future that would be recognized by one of skill in the art as being or having the function of a "polynucleotide," "oligonucleotide," or "nucleic acid." An example of a modified or synthetic structure having the function of a "polynucleotide," "oligonucleotide," or "nucleic acid" is PNA or peptide nucleic acid.

"Coding sequence" refers to that portion of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

"Naturally-occurring" or "wild-type" refers to the form found in nature. For example, a naturally occurring or wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

As used herein, "recombinant," "engineered," and "non-naturally occurring" when used with reference to a cell, nucleic acid, or polypeptide, refer to a material, or a material corresponding to the natural or native form of the material, that has been modified in a manner that would not otherwise exist in nature. In some embodiments, the cell, nucleic acid or polypeptide is identical to a naturally occurring cell, nucleic acid or polypeptide, but is produced or derived from synthetic materials and/or by manipulation using recombinant techniques. Non-limiting examples include, among others, recombinant cells expressing genes that are not found within the native (non-recombinant) form of the cell or expressed native genes that are otherwise expressed at a different level.

"Percentage of sequence identity" and "percentage homology" are used interchangeably herein to refer to comparisons among polynucleotides or polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (i.e., gaps) as compared to the reference sequence for optimal alignment of the two sequences. The percentage may be calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result

by 100 to yield the percentage of sequence identity. Alternatively, the percentage may be calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Those of skill in the art appreciate that there are many established algorithms available to align two sequences. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman (Smith and Waterman, *Adv. Appl. Math.*, 2:482 [1981]), by the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, *J. Mol. Biol.*, 48:443 [1970]), by the search for similarity method of Pearson and Lipman (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 [1988]), by computerized implementations of these algorithms (*e.g.*, GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package), or by visual inspection, as known in the art. Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity include, but are not limited to the BLAST and BLAST 2.0 algorithms, which are described by Altschul et al. (See, Altschul et al., *J. Mol. Biol.*, 215: 403-410 [1990]; and Altschul et al., *Nucl. Acids Res.*, 3389-3402 [1977], respectively). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length *W* in the query sequence, which either match or satisfy some positive-valued threshold score *T* when aligned with a word of the same length in a database sequence. *T* is referred to as, the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters *M* (reward score for a pair of matching residues; always >0) and *N* (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity *X* from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters *W*, *T*, and *X* determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (*W*) of 11, an expectation (*E*) of 10, *M*=5, *N*=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (*W*) of 3, an expectation (*E*) of 10, and the BLOSUM62 scoring matrix (See, Henikoff and Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 [1989]). Exemplary determination of sequence alignment and % sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using default parameters provided.

"Reference sequence" refers to a defined sequence used as a basis for a sequence comparison. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, or the full length of the

nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a sequence (*i.e.*, a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptide are typically performed by comparing sequences of the two

5 polynucleotides or polypeptides over a “comparison window” to identify and compare local regions of sequence similarity. In some embodiments, a “reference sequence” can be based on a primary amino acid sequence, where the reference sequence is a sequence that can have one or more changes in the primary sequence. For instance, a “reference sequence based on SEQ ID NO:4 having at the residue corresponding to X14 a valine” or X14V refers to a reference sequence in which the corresponding

10 residue at X14 in SEQ ID NO:4, which is a tyrosine, has been changed to valine.

“Comparison window” refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acids residues wherein a sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less as compared to

15 the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The comparison window can be longer than 20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

As used herein, “substantial identity” refers to a polynucleotide or polypeptide sequence that has at least 80 percent sequence identity, at least 85 percent identity, at least between 89 to 95 percent

20 sequence identity, or more usually, at least 99 percent sequence identity as compared to a reference sequence over a comparison window of at least 20 residue positions, frequently over a window of at least 30-50 residues, wherein the percentage of sequence identity is calculated by comparing the reference sequence to a sequence that includes deletions or additions which total 20% or less of the reference sequence over the window of comparison. In some specific embodiments applied to polypeptides, the

25 term “substantial identity” means that two polypeptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap weights, share at least 80% sequence identity, preferably at least 89% sequence identity, at least 95% sequence identity or more (*e.g.*, 99% sequence identity). In some embodiments, residue positions that are not identical in sequences being compared differ by conservative amino acid substitutions.

“Corresponding to,” “reference to,” and “relative to” when used in the context of the numbering of a given amino acid or polynucleotide sequence refer to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the

35 given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered transposase, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned.

“Amino acid difference” or “residue difference” refers to a change in the amino acid residue at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in a reference sequence. The positions of amino acid differences generally are referred to herein as “X_n,” where n refers to the corresponding position in the reference sequence upon which the residue difference is based. For example, a “residue difference at position X₂₅ as compared to SEQ ID NO: 2” refers to a change of the amino acid residue at the polypeptide position corresponding to position 25 of SEQ ID NO:2. Thus, if the reference polypeptide of SEQ ID NO: 2 has a valine at position 25, then a “residue difference at position X₂₅ as compared to SEQ ID NO:2” an amino acid substitution of any residue other than valine at the position of the polypeptide corresponding to position 25 of SEQ ID NO: 2. In most instances herein, the specific amino acid residue difference at a position is indicated as “X_nY” where “X_n” specified the corresponding position as described above, and “Y” is the single letter identifier of the amino acid found in the engineered polypeptide (i.e., the different residue than in the reference polypeptide). In some embodiments, more than one amino acid can appear in a specified residue position (i.e., the alternative amino acids can be listed in the form X_nY/Z, where Y and Z represent alternate amino acid residues). In some instances (e.g., in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5, 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3) the present invention also provides specific amino acid differences denoted by the conventional notation “A_nB”, where A is the single letter identifier of the residue in the reference sequence, “n” is the number of the residue position in the reference sequence, and B is the single letter identifier of the residue substitution in the sequence of the engineered polypeptide. Furthermore, in some instances, a polypeptide of the present invention can include one or more amino acid residue differences relative to a reference sequence, which is indicated by a list of the specified positions where changes are made relative to the reference sequence. In some additional embodiments, the present invention provides engineered polypeptide sequences comprising both conservative and non-conservative amino acid substitutions.

As used herein, “conservative amino acid substitution” refers to a substitution of a residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar defined class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain is substituted with another aliphatic amino acid (e.g., alanine, valine, leucine, and isoleucine); an amino acid with a hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain (e.g., serine and threonine); an amino acid having an aromatic side chain is substituted with another amino acid having an aromatic side chain (e.g., phenylalanine, tyrosine, tryptophan, and histidine); an amino acid with a basic side chain is substituted with another amino acid with a basic side chain (e.g., lysine and arginine); an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain (e.g., aspartic acid or glutamic acid); and/or a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively. Exemplary conservative substitutions are provided in **Table 1**.

Table 1. Conservative Amino Acid Substitution Examples	
Residue	Possible Conservative Substitutions
A, L, V, I	Other aliphatic (A, L, V, I) Other non-polar (A, L, V, I, G, M)
G, M	Other non-polar (A, L, V, I, G, M)
D, E	Other acidic (D, E)
K, R	Other basic (K, R)
N, Q, S, T	Other polar
H, Y, W, F	Other aromatic (H, Y, W, F)
C, P	None

“Non-conservative substitution” refers to substitution of an amino acid in the polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (*e.g.*, proline for glycine), (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

“Deletion” refers to modification to the polypeptide by removal of one or more amino acids from the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the reference enzyme while retaining enzymatic activity and/or retaining the improved properties of an engineered transposase enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide. In various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

“Insertion” refers to modification to the polypeptide by addition of one or more amino acids from the reference polypeptide. In some embodiments, the improved engineered transposase enzymes comprise insertions of one or more amino acids to the naturally occurring polypeptide as well as insertions of one or more amino acids to other improved transposase polypeptides. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

“Fragment” as used herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence. Fragments can be at least 14 amino acids long, at least 20 amino acids long, at least 50 amino acids long or longer, and up to 70%, 80%, 90%, 95%, 98%, and 99% of the full-length transposase polypeptide, for example the polypeptide of SEQ ID NO: 2 or a transposase provided in the polypeptide sequences of SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

“Isolated polypeptide” refers to a polypeptide which is substantially separated from other contaminants that naturally accompany it, *e.g.*, protein, lipids, and polynucleotides. The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (*e.g.*, host cell or *in vitro* synthesis). The engineered transposase enzymes may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the engineered transposase enzyme can be an isolated polypeptide.

“Substantially pure polypeptide” refers to a composition in which the polypeptide species is the predominant species present (*i.e.*, on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition) and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular species present by mole or % weight. Generally, a substantially pure transposase composition will comprise about 60 % or more, about 70% or more, about 80% or more, about 90% or more, about 95% or more, and about 98% or more of all macromolecular species by mole or % weight present in the composition. In some embodiments, the object species is purified to essential homogeneity (*i.e.*, contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated engineered transposase polypeptide is a substantially pure polypeptide composition.

As used herein, a “target nucleic acid” refers to any nucleic acid (*e.g.*, RNA or DNA) of interest that is selected for amplification or analysis (*e.g.*, sequencing) using a composition (*e.g.*, sequencing oligonucleotides or barcoding oligonucleotides) or method of the invention. In some instances, RNA may be converted to cDNA prior to being used in the methods described herein. In some instances, target nucleic acid includes genomic DNA or cDNAs from a single cell. In some instances, target nucleic acid includes nucleic acids from a plurality of haplotypes. In some instances, target nucleic acid is crosslinked via histones or chromatin from single or multiple cells. In some instances, target nucleic acid has been condensed or optionally treated with one or more condensing agents.

As used herein, “improved enzyme property” refers to at least one improved property of an enzyme. In some embodiments, the present invention provides engineered transposase polypeptides that exhibit an improvement in any enzyme property as compared to a reference transposase polypeptide and/or a wild-type transposase polypeptide, and/or another engineered transposase polypeptide. Thus, the level of “improvement” can be determined and compared between various transposase polypeptides, including wild-type, as well as engineered transposases. Improved properties include, but are not limited, to such properties as increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased adapter loading (donor binding) or adapter ligation (*i.e.* tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias as compared to another wild-type transposase or engineered transposase.

“Sequence insertion bias,” “target bias,” “sequence bias,” or related terms used herein refer to the degree to which a transposome cleaves target nucleotide sequences in a non-random manner based on

the composition of the sequence of the polynucleotide. For example, a transposase with G/C target bias will have an increased rate of binding to G/C rich sequences, as compared to other sequences. As detailed in the Examples below, sequence insertion bias can be approximated or quantified using several methods. One method evaluates bias over the first 15 nucleotides of NGS reads putatively comprising the region of the target bound by the transposome construct. Bias plots generated from NGS data represent this bias as a function of the distance from the insertion site. Overall relative bias for each variant can be calculated as the root-mean-squared deviation (RMSD) of the observed vs. expected representation of all four nucleotides at each position. Conversely, "insertion promiscuity" and "promiscuity" refer to the degree to which a transposome binds to target nucleotide sequences in a random manner, without regard to the sequence composition of the polynucleotide.

"Next generation sequencing" refers to modern high-throughput or massively-parallel sequencing techniques that allow sequencing of many polynucleotide sequences at the same time, as compared to traditional sequencing methods, including the Sanger method, that allowed sequencing of only one polynucleotide at a time. Next generation sequencing methods have facilitated whole genome sequencing and enabled genomics and related fields by requiring less polynucleotide sample and having greater accuracy.

"Adapter," as used herein, refers to a donor polynucleotide that contains the end sequences for transposase binding, as well as other optional application-specific elements. These may include binding sites for sequencing primers, library amplification primers, sample indexes, or other functional donor polynucleotide sequence. The adapter is loaded onto or bound to the transposase, which dimerizes to form the "transposome".

"Barcode," as used herein, refers to a unique oligonucleotide sequence that may allow the corresponding oligonucleotide to be identified. In some instances, the nucleic acid sequence may be located at a specific position in a longer nucleic acid sequence.

"Complement" or "complementary," as used herein in reference to a sequence, refers to the sequence of a first nucleic acid in relation to that of a second nucleic acid, wherein when the first and second nucleic acids are aligned antiparallel (5' end of the first nucleic acid matched to the 3' end of the second nucleic acid, and vice versa) to each other, the nucleotide bases at each position in their sequences will have complementary structures following a lock-and-key principle (i.e., A will be paired with U or T and G will be paired with C). Complementary sequences may include mismatches of up to one third of nucleotide bases. For example, two sequences that are nine bases in length may have mismatches of at most 3, at most 2, or at most 1, or at most 0 nucleotide bases, and remain complementary to one another.

"Transposition" or a "transposition event," as used herein, is the cleavage of a target dsDNA (or hybrid DNA/RNA) sequence and ligation of the pre-loaded adapter (or another sequence containing transposon end sequences) from the transposome onto the target dsDNA at the site of cleavage. In cells, transposition typically refers to the relocation of a segment of DNA to another locus in the genome. In vitro, transposition may result in the fragmentation of a target DNA sample with the simultaneous addition of the adapter sequences to the fragmented DNA ends, i.e., tagmentation.

“Synaptic complex” or “transposase synaptic complex,” as used herein, refers to a protein-nucleic acid complex including one or more transposases and one or more oligonucleotides. In some instances, the one or more oligonucleotides of the synaptic complex are inserted into a nucleic acid sequence of a nucleic acid sample by transposase activity. In some instances, the synaptic complex may include two
5 transposases and two oligonucleotides. In some instances, the insertion of the two oligonucleotides into the nucleic acid sequence of the nucleic acid sample results in fragmentation of the nucleic acid sequence at the site of insertion. In some instances, the transposases may be Tn5 transposases. In some instances, the oligonucleotides may be adapter sequences. In some instances, the synaptic complex is pre-assembled.

10 “Increased enzymatic activity” refers to an improved property of the transposase polypeptides, which can be represented by an increase in specific activity (*e.g.*, product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (*e.g.*, percent conversion of starting amount of substrate to product in a specified time period using a specified amount of transposase) as compared to the reference transposase enzyme. Exemplary methods to determine enzyme activity are
15 provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m , V_{max} or k_{cat} , changes of which can lead to increased enzymatic activity. Improvements in enzyme activity can be from about 1.2 times the enzymatic activity of the corresponding wild-type enzyme, to as much as 2 times, 5 times, 10 times, 20 times, 25 times, 50 times or more enzymatic activity than the naturally occurring or another engineered transposase from which the
20 transposase polypeptides were derived. Transposase activity can be measured by any one of standard assays, such as by monitoring changes in properties of substrates, cofactors, or products. In some embodiments, the amount of products generated can be measured by Liquid Chromatography-Mass Spectrometry (LC-MS), HPLC, capillary electrophoresis or other methods, as known in the art. Comparisons of enzyme activities are made using a defined preparation of enzyme, a defined assay
25 under a set condition, and one or more defined substrates, as further described in detail herein. Generally, when lysates are compared, the numbers of cells and the amount of protein assayed are determined as well as use of identical expression systems and identical host cells to minimize variations in amount of enzyme produced by the host cells and present in the lysates.

30 “Conversion” refers to the enzymatic conversion of the substrate(s) to the corresponding product(s). “Percent conversion” refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, the “enzymatic activity” or “activity” of a transposase polypeptide can be expressed as “percent conversion” of the substrate to the product.

35 “Thermostable” refers to a polypeptide that maintains similar activity (more than 60% to 80% for example) after exposure to elevated temperatures (*e.g.*, 40-80 °C) for a period of time (*e.g.*, 0.5-24 hrs) compared to the wild-type enzyme exposed to the same elevated temperature.

40 “Solvent stable” refers to a polypeptide that maintains similar activity (more than *e.g.*, 60% to 80%) after exposure to varying concentrations (*e.g.*, 5-99%) of solvent (ethanol, isopropyl alcohol, dimethylsulfoxide (DMSO), tetrahydrofuran, 2-methyltetrahydrofuran, acetone, toluene, butyl acetate, methyl tert-butyl ether, etc.) for a period of time (*e.g.*, 0.5-24 hrs) compared to the wild-type enzyme exposed to the same concentration of the same solvent.

“Thermo- and solvent stable” refers to a polypeptide that is both thermostable and solvent stable.

The term “stringent hybridization conditions” is used herein to refer to conditions under which nucleic acid hybrids are stable. As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (T_m) of the hybrids. In general, the stability of a hybrid is a function of ion strength, temperature, G/C content, and the presence of chaotropic agents. The T_m values for polynucleotides can be calculated using known methods for predicting melting temperatures (See e.g., Baldino et al., Meth. Enzymol., 168:761-777 [1989]; Bolton et al., Proc. Natl. Acad. Sci. USA 48:1390 [1962]; Bresslauer et al., Proc. Natl. Acad. Sci. USA 83:8893-8897 [1986]; Freier et al., Proc. Natl. Acad. Sci. USA 83:9373-9377 [1986]; Kierzek et al., Biochem., 25:7840-7846 [1986]; Rychlik et al., 1990, Nucl. Acids Res., 18:6409-6412 [1990] (erratum, Nucl. Acids Res., 19:698 [1991]); Sambrook et al., *supra*; Suggs et al., 1981, in Developmental Biology Using Purified Genes, Brown et al. [eds.], pp. 683-693, Academic Press, Cambridge, MA [1981]; and Wetmur, Crit. Rev. Biochem. Mol. Biol., 26:227-259 [1991]). In some embodiments, the polynucleotide encodes the polypeptide disclosed herein and hybridizes under defined conditions, such as moderately stringent or highly stringent conditions, to the complement of a sequence encoding an engineered transposase enzyme of the present invention.

“Hybridization stringency” relates to hybridization conditions, such as washing conditions, in the hybridization of nucleic acids. Generally, hybridization reactions are performed under conditions of lower stringency, followed by washes of varying but higher stringency. The term “moderately stringent hybridization” refers to conditions that permit target-DNA to bind a complementary nucleic acid that has about 60% identity, preferably about 75% identity, about 85% identity to the target DNA, with greater than about 90% identity to target-polynucleotide. Exemplary moderately stringent conditions are conditions equivalent to hybridization in 50% formamide, 5× Denhart’s solution, 5×SSPE, 0.2% SDS at 42 °C, followed by washing in 0.2×SSPE, 0.2% SDS, at 42 °C. “High stringency hybridization” refers generally to conditions that are about 10 °C or less from the thermal melting temperature T_m as determined under the solution condition for a defined polynucleotide sequence. In some embodiments, a high stringency condition refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65 °C (*i.e.*, if a hybrid is not stable in 0.018M NaCl at 65 °C, it will not be stable under high stringency conditions, as contemplated herein). High stringency conditions can be provided, for example, by hybridization in conditions equivalent to 50% formamide, 5× Denhart’s solution, 5×SSPE, 0.2% SDS at 42 °C, followed by washing in 0.1×SSPE, and 0.1% SDS at 65 °C. Another high stringency condition is hybridizing in conditions equivalent to hybridizing in 5X SSC containing 0.1% (w:v) SDS at 65 °C and washing in 0.1x SSC containing 0.1% SDS at 65 °C. Other high stringency hybridization conditions, as well as moderately stringent conditions, are described in the references cited above.

“Heterologous” polynucleotide refers to any polynucleotide that is introduced into a host cell by laboratory techniques, and includes polynucleotides that are removed from a host cell, subjected to laboratory manipulation, and then reintroduced into a host cell.

“Codon optimized” refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is efficiently expressed in the organism of interest. Although the genetic code is degenerate in that most amino acids are

represented by several codons, called “synonyms” or “synonymous” codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the transposase enzymes may be codon optimized for optimal production from the host organism selected for expression.

As used herein, “preferred, optimal, high codon usage bias codons” refers interchangeably to codons that are used at higher frequency in the protein coding regions than other codons that code for the same amino acid. The preferred codons may be determined in relation to codon usage in a single gene, a set of genes of common function or origin, highly expressed genes, the codon frequency in the aggregate protein coding regions of the whole organism, codon frequency in the aggregate protein coding regions of related organisms, or combinations thereof. Codons whose frequency increases with the level of gene expression are typically optimal codons for expression. A variety of methods are known for determining the codon frequency (e.g., codon usage, relative synonymous codon usage) and codon preference in specific organisms, including multivariate analysis, for example, using cluster analysis or correspondence analysis, and the effective number of codons used in a gene (See e.g., GCG CodonPreference, Genetics Computer Group Wisconsin Package; CodonW, Peden, University of Nottingham; McInerney, *Bioinform.*, 14:372-73 [1998]; Stenico et al., *Nucl. Acids Res.*, 22:437-46 [1994]; Wright, *Gene* 87:23-29 [1990]). Codon usage tables are available for many different organisms (See e.g., Wada et al., *Nucl. Acids Res.*, 20:2111-2118 [1992]; Nakamura et al., *Nucl. Acids Res.*, 28:292 [2000]; Duret, et al., *supra*; Henaut and Danchin, in *Escherichia coli* and *Salmonella*, Neidhardt, et al. (eds.), ASM Press, Washington D.C., p. 2047-2066 [1996]). The data source for obtaining codon usage may rely on any available nucleotide sequence capable of coding for a protein. These data sets include nucleic acid sequences actually known to encode expressed proteins (e.g., complete protein coding sequences-CDS), expressed sequence tags (ESTs), or predicted coding regions of genomic sequences (See e.g., Mount, *Bioinformatics: Sequence and Genome Analysis*, Chapter 8, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. [2001]; Uberbacher, *Meth. Enzymol.*, 266:259-281 [1996]; and Tiwari et al., *Comput. Appl. Biosci.*, 13:263-270 [1997]).

“Control sequence” is defined herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present invention. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter, signal peptide sequence, and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

“Operably linked” is defined herein as a configuration in which a control sequence is appropriately placed (*i.e.*, in a functional relationship) at a position relative to a polynucleotide of interest such that the control sequence directs or regulates the expression of the polynucleotide and/or polypeptide of interest.

“Promoter sequence” refers to a nucleic acid sequence that is recognized by a host cell for expression of a polynucleotide of interest, such as a coding sequence. The promoter sequence contains transcriptional control sequences, which mediate the expression of a polynucleotide of interest. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

“Suitable reaction conditions” refer to those conditions in the biocatalytic reaction solution (*e.g.*, ranges of enzyme loading, substrate loading, cofactor loading, temperature, pH, buffers, co-solvents, etc.) under which a transposase polypeptide of the present invention is capable of converting one or more substrate compounds to a product compound (*e.g.*, ligation or annealing of an adapter or donor polynucleotide and a target polynucleotide to form an annealed or ligated polynucleotide consisting of the adapter and the optionally fragmented target polynucleotide). Exemplary “suitable reaction conditions” are provided in the present invention and illustrated by the Examples.

“Composition” refers to a mixture or combination of one or more substances, wherein each substance or component of the composition retains its individual properties. As used herein, a biocatalytic composition refers to a combination of one or more substances useful for biocatalysis.

“Loading”, such as in “compound loading” or “enzyme loading” or “cofactor loading” refers to the concentration or amount of a component in a reaction mixture at the start of the reaction. “Adapter loading” refers to binding of the transposase to the adapter polynucleotide.

“Substrate” in the context of a biocatalyst mediated process refers to the compound or molecule acted on by the biocatalyst. For example, a transposase biocatalyst used in the synthesis processes disclosed herein acts on an adapter or donor polynucleotide and a target polynucleotide.

“Product” in the context of a biocatalyst mediated process refers to the compound or molecule resulting from the action of the biocatalyst. For example, an exemplary product for a transposase biocatalyst used in a process disclosed herein is an annealed or ligated polynucleotide consisting of the adapter and the optionally fragmented target polynucleotide, as depicted in **Figures 1 and 2**.

“Alkyl” refers to saturated hydrocarbon groups of from 1 to 18 carbon atoms inclusively, either straight chained or branched, more preferably from 1 to 8 carbon atoms inclusively, and most preferably 1 to 6 carbon atoms inclusively. An alkyl with a specified number of carbon atoms is denoted in parenthesis (*e.g.*, (C₁-C₆)alkyl refers to an alkyl of 1 to 6 carbon atoms).

“Alkenyl” refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one double bond but optionally containing more than one double bond.

“Alkynyl” refers to hydrocarbon groups of from 2 to 12 carbon atoms inclusively, either straight or branched containing at least one triple bond but optionally containing more than one triple bond, and additionally optionally containing one or more double bonded moieties.

“Heteroalkyl”, “heteroalkenyl”, and “heteroalkynyl”, refer respectively, to alkyl, alkenyl and alkynyl as defined herein in which one or more of the carbon atoms are each independently replaced with the same or different heteroatoms or heteroatomic groups. Heteroatoms and/or heteroatomic groups which can replace the carbon atoms include, but are not limited to -O-, -S-, -S-O-, -NR^a-, -PH-, -S(O)-, -S(O)₂-, -

S(O) NR^g-, -S(O)₂NR^g, and the like, including combinations thereof, where each R^g is independently selected from hydrogen, alkyl, cycloalkyl, heterocycloalkyl, aryl, and heteroaryl.

“Amino” refers to the group -NH₂. Substituted amino refers to the group -NHR^h, NR^hR^h, and NR^hR^hR^h, where each R^h is independently selected from substituted or unsubstituted alkyl, cycloalkyl, cycloheteroalkyl, alkoxy, aryl, heteroaryl, heteroarylalkyl, acyl, alkoxycarbonyl, sulfanyl, sulfinyl, sulfonyl, and the like. Typical amino groups include, but are limited to, dimethylamino, diethylamino, trimethylammonium, triethylammonium, methylsulfonylamino, furanyl-oxy-sulfamino, and the like.

“Aminoalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with one or more amino groups, including substituted amino groups.

“Aminocarbonyl” refers to -C(O)NH₂. Substituted aminocarbonyl refers to -C(O)NR^hR^h, where the amino group NR^hR^h is as defined herein.

“Oxy” refers to a divalent group -O-, which may have various substituents to form different oxy groups, including ethers and esters.

“Alkoxy” or “alkyloxy” are used interchangeably herein to refer to the group -OR^z, wherein R^z is an alkyl group, including optionally substituted alkyl groups.

“Carboxy” refers to -COOH.

“Carbonyl” refers to -C(O)-, which may have a variety of substituents to form different carbonyl groups including acids, acid halides, aldehydes, amides, esters, and ketones.

“Carboxyalkyl” refers to an alkyl in which one or more of the hydrogen atoms are replaced with one or more carboxy groups.

“Aminocarbonylalkyl” refers to an alkyl substituted with an aminocarbonyl group, as defined herein.

“Halogen” or “halo” refers to fluoro, chloro, bromo and iodo.

“Haloalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with a halogen. Thus, the term “haloalkyl” is meant to include monohaloalkyls, dihaloalkyls, trihaloalkyls, etc. up to perhaloalkyls. For example, the expression “(C₁ - C₂) haloalkyl” includes 1-fluoromethyl, difluoromethyl, trifluoromethyl, 1-fluoroethyl, 1,1-difluoroethyl, 1,2-difluoroethyl, 1,1,1 trifluoroethyl, perfluoroethyl, etc.

“Hydroxy” refers to -OH.

“Hydroxyalkyl” refers to an alkyl group in which one or more of the hydrogen atoms are replaced with one or more hydroxy groups.

“Thiol” or “sulfanyl” refers to -SH. Substituted thiol or sulfanyl refers to -S-R^h, where R^h is an alkyl, aryl or other suitable substituent.

“Sulfonyl” refers to -SO₂-. Substituted sulfonyl refers to -SO₂-R^h, where R^h is an alkyl, aryl or other suitable substituent.

“Alkylsulfonyl” refers to -SO₂-R^z, where R^z is an alkyl, which can be optionally substituted. Typical alkylsulfonyl groups include, but are not limited to, methylsulfonyl, ethylsulfonyl, n-propylsulfonyl, and the like.

“Phosphate” as used herein refers to a functional group comprised of an orthophosphate ion (phosphorous atom covalently linked to four oxygen atoms). The orthophosphate ion is commonly found with one or more hydrogen atoms or organic groups.

“Phosphorylated” as used herein refers to the addition or presence of one or more phosphoryl groups (phosphorous atom covalently linked to the three oxygen atoms).

“Optionally substituted” as used herein with respect to the foregoing chemical groups means that positions of the chemical group occupied by hydrogen can be substituted with another atom (unless otherwise specified) exemplified by, but not limited to carbon, oxygen, nitrogen, or sulfur, or a chemical group, exemplified by, but not limited to, hydroxy, oxo, nitro, methoxy, ethoxy, alkoxy, substituted alkoxy, trifluoromethoxy, haloalkoxy, fluoro, chloro, bromo, iodo, halo, methyl, ethyl, propyl, butyl, alkyl, alkenyl, alkynyl, substituted alkyl, trifluoromethyl, haloalkyl, hydroxyalkyl, alkoxyalkyl, thio, alkylthio, acyl, carboxy, alkoxycarbonyl, carboxamido, substituted carboxamido, alkylsulfonyl, alkylsulfinyl, alkylsulfonylamino, sulfonamido, substituted sulfonamido, cyano, amino, substituted amino, alkylamino, dialkylamino, aminoalkyl, acylamino, amidino, amidoximo, hydroxamoyl, phenyl, aryl, substituted aryl, aryloxy, arylalkyl, arylalkenyl, arylalkynyl, pyridyl, imidazolyl, heteroaryl, substituted heteroaryl, heteroaryloxy, heteroarylalkyl, heteroarylalkenyl, heteroarylalkynyl, cyclopropyl, cyclobutyl, cyclopentyl, cyclohexyl, cycloalkyl, cycloalkenyl, cycloalkylalkyl, substituted cycloalkyl, cycloalkyloxy, pyrrolidinyl, piperidinyl, morpholino, heterocycle, (heterocycle)oxy, and (heterocycle)alkyl; where preferred heteroatoms are oxygen, nitrogen, and sulfur. Additionally, where open valences exist on these substitute chemical groups they can be further substituted with alkyl, cycloalkyl, aryl, heteroaryl, and/or heterocycle groups, that where these open valences exist on carbon they can be further substituted by halogen and by oxygen-, nitrogen-, or sulfur-bonded substituents, and where multiple such open valences exist, these groups can be joined to form a ring, either by direct formation of a bond or by formation of bonds to a new heteroatom, preferably oxygen, nitrogen, or sulfur. It is further contemplated that the above substitutions can be made provided that replacing the hydrogen with the substituent does not introduce unacceptable instability to the molecules of the present invention, and is otherwise chemically reasonable. One of ordinary skill in the art would understand that with respect to any chemical group described as optionally substituted, only sterically practical and/or synthetically feasible chemical groups are meant to be included. “Optionally substituted” as used herein refers to all subsequent modifiers in a term or series of chemical groups. For example, in the term “optionally substituted arylalkyl,” the “alkyl” portion and the “aryl” portion of the molecule may or may not be substituted, and for the series “optionally substituted alkyl, cycloalkyl, aryl and heteroaryl,” the alkyl, cycloalkyl, aryl, and heteroaryl groups, independently of the others, may or may not be substituted.

“Reaction” as used herein refers to a process in which one or more substances or compounds or substrates is converted into one or more different substances, compounds, or processes.

Preparation of Next Generation Sequencing Libraries Using Engineered Transposases

Transposases with improved activity in cleaving or fragmenting polynucleotides and ligating or tagging polynucleotides are necessary to overcome the limitations of existing NGS library preparation methods. Improved transposases with increased polynucleotide fragmentation or cleavage activity,

increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias compared to another wild-type or engineered transposase find utility in library preparation for NGS, but are not limited to these uses. The improved transposases of the present disclosure find use in any application involving cleaving or fragmenting double-stranded polynucleotides and/or ligating adapters or tagging polynucleotides, including applications where a donor polynucleotide is transferred into a target polynucleotide.

The present invention provides novel transposases that have improved activity in the fragmentation or cleavage of double stranded DNA and the ligation of adapters or tagging of the DNA fragments. The transposases of the present disclosure have increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias as compared to another wild-type transposase or engineered transposase. The engineered polypeptides of the present disclosure are variants of SEQ ID NO: 2, a chimeric protein based an IS4 transposase from *Parashewanella curva* or SEQ ID NO: 5704, a modified hyperactive Tn5 transposase, and designed to recognize the 19 bp Tn5 mosaic end (ME) transposon donor recognition sequence. These engineered transposases are capable of improved preparation of NGS libraries from polynucleotide samples.

In nature, transposases facilitate the transposition of fragments of DNA (transposons) from a donor strand to a host strand at sites flanked by short inverted-repeat sequences or end sequences (ESs). The transposition mechanism is initiated by the binding of two transposases to donor DNA ESs to form the dimeric synaptic complex, which catalyzes cleavage of the donor DNA. The transposome (dimeric transposase/DNA complex) then binds to the target DNA to transfer the transposon DNA to the target DNA.

Transposase-mediated NGS library preparation, depicted in **Figure 1**, is similar to naturally-occurring transposition, except that the donor DNA or polynucleotide consists of two, separate adapter sequences rather than one continuous loop of donor DNA or polynucleotide. These adapter sequences are bound to transposases in a loading step that replaces the initial donor DNA or polynucleotide binding and cleavage step. Following adapter loading, the dimerized transposome proceeds to bind the target DNA or polynucleotide sample. Cleavage of the target polynucleotide and ligation of the adapter sequences results in fragmentation of the target DNA or polynucleotide and ligation of sequencing adapters to the target fragments.

The resulting DNA or polynucleotide fragments with ligated adapters comprise the next generation sequencing library. An ideal library preparation method would generate random fragments in the polynucleotide sequence, without bias toward the sequence composition of the target site. Compared to traditional library preparation methods, transposase-mediated fragmentation has lower sequence insertion bias. However, certain transposases are known to have sequence insertion bias. As an example, Tn5 has a slight G/C insertion bias.

In some embodiments, the engineered transposases of the present disclosure have reduced sequence insertion bias or increased sequence promiscuity as compared to a wild-type or reference transposase.

5 In addition to containing the ME or ES sequences for target binding, adapters may comprise sequencing primers or other sequences used to identify or select fragments or groups of fragments from pooled samples. The adapters may encode primer binding sites for next generation sequencing library amplification, binding sites for sequencing primers, sample indexes, or other functional sequences related to a specific application. In some embodiments, the engineered transposases of the present disclosure have increased adapter loading or binding, as compared to a wild-type or reference transposase.

10 In some embodiments, further processing steps are used for library preparation, including gap filling, normalization, pooling of tagged fragment samples, and/or ligation of additional adapters to create dual-tagged libraries. A variety of parameters may be manipulated to create sequencing libraries with various characteristics. For example, varying the concentration of transposase and adapters can change the fragment length. Copy number and sample coverage may also be influenced or controlled using
15 methods known to those of skill in the art.

As described further herein, the engineered transposase polypeptides of the current disclosure exhibit one of more improved properties in the transposase-mediated DNA fragmentation and adapter ligation process depicted in **Figure 1**.

20 In some embodiments, the present invention provides an engineered transposase polypeptide comprising an amino acid sequence having at least 60% sequence identity to an amino acid reference sequence of SEQ ID NO: 2 or SEQ ID NO: 5704 and further comprising one or more amino acid residue differences as compared to the reference amino acid sequence, wherein the engineered transposase polypeptide has increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased
25 adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias as compared to another wild-type transposase or engineered transposase.

30 In particular, the engineered transposase polypeptides of the present disclosure have been engineered for efficient fragmentation or cleavage of DNA and the ligation of adapters in the process described above.

A variety of suitable reaction conditions are known to those skilled in the art, as detailed below and in the Examples.

Engineered Transposase Polypeptides

35 The present invention provides engineered transposase polypeptides useful in preparation of next generation sequencing libraries, as well as compositions and methods of utilizing these engineered polypeptides.

The present invention provides transposase polypeptides, polynucleotides encoding the polypeptides, methods of preparing the polypeptides, and methods for using the polypeptides. Where the

description relates to polypeptides, it is to be understood that it can describe the polynucleotides encoding the polypeptides.

Suitable reaction conditions under which the above-described improved properties of the engineered polypeptides carry out the desired reaction can be determined with respect to concentrations or amounts of polypeptide, substrate, co-substrate, buffer, solvent, pH, conditions including temperature and reaction time, and/or conditions with the polypeptide immobilized on a solid support, as further described below and in the Examples.

In some embodiments, exemplary engineered transposases comprise an amino acid sequence that has one or more residue differences as compared to SEQ ID NO: 2 at the residue positions indicated in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5, 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1. In some embodiments, exemplary engineered transposases comprise an amino acid sequence that has one or more residue differences as compared to SEQ ID NO: 5704 at the residue positions indicated in Tables 36.1 and 36.3.

The structure and function information for the exemplary engineered polypeptides of the present invention are based on the fragmentation or cleavage of double stranded DNA and the ligation or tagging of the DNA fragments, the results of which are shown below in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5, 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3, as further described in the Examples. The odd numbered sequence identifiers (i.e., SEQ ID NOs) in these Tables refer to the nucleotide sequence encoding the amino acid sequence provided by the even numbered SEQ ID NOs in these Tables. Exemplary sequences are provided in the electronic sequence listing file accompanying this invention, which is hereby incorporated by reference herein. The amino acid residue differences are based on comparison to the reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, as indicated.

Proteins based on an IS4 transposase from *Parashewanella curva* or a Tn5 transposase with the ability to recognize the 19 bp Tn5 mosaic end (ME) transposon donor recognition sequence (SEQ ID NO: 2 or SEQ ID NO: 5704, respectively) were designed *in silico*, verified to have transposase activity, and selected for evolution. The transposase polypeptides of the present disclosure are engineered variants of SEQ ID NO: 2 or SEQ ID NO: 5704, which also comprise a C-terminal 6-histidine tag following a GGSG peptide spacer (SEQ ID NO: 5777).

The polypeptides of the present disclosure have residue differences that result in improved properties, including increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias as compared to another wild-type transposase or engineered transposase.

The activity of each engineered transposase relative to the reference polypeptide of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704 was determined as conversion of the substrates described in the Examples herein. In some embodiments, a shake flask purified enzyme (SFP) is used to assess the properties of the engineered transposases, the results of which are provided in the Examples.

In some embodiments, the specific enzyme properties are associated with the residues differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704 at the residue positions indicated herein. In some embodiments, residue differences affecting polypeptide expression can be used to increase expression of the engineered transposases.

In light of the guidance provided herein, it is further contemplated that any of the exemplary engineered polypeptides comprising the polypeptide sequences of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704 find use as the starting amino acid sequence for synthesizing other transposase polypeptides, for example by subsequent rounds of evolution that incorporate new combinations of various amino acid differences from other polypeptides in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5, 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3, and other residue positions described herein. Further improvements may be generated by including amino acid differences at residue positions that had been maintained as unchanged throughout earlier rounds of evolution.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from 10, 13, 24, 35, 39, 44, 46, 51, 54, 55, 62/361, 72, 75, 82, 89, 92, 107, 128, 137, 164, 167, 177/359, 191, 192, 193, 194, 201, 217, 281, 289, 291, 291/340, 305, 309, 310, 317, 331, 337, 340, 346, 352, 353, 357, 358, 359, 362, 371, 396, 414/435, 415, 428, 437, 443, 447, and 448. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from 10A, 13E, 24L, 35L, 39P, 44P, 46A, 51E, 54T, 55K, 62D/361H, 72L, 75H, 82R, 89V, 92L, 107G, 128A, 137L, 164E, 167R, 177I/359A, 191Y, 192E, 193L, 194F, 201G, 217G, 281I, 289P, 291G, 291Y/340L, 305V, 309E, 310D, 317W, 331L, 337I, 340L, 346D, 352L, 353A, 357I, 358M, 358V, 359A, 362I, 371L, 396I, 414D/435L, 415P, 428I, 437R, 443P, 447V, and 448L. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences as compared to SEQ ID NO: 2, selected from S10A, H13E, R24L, M35L, S39P, T44P, S46A, K51E, Q54T, E55K,

R62D/R361H, R72L, G75H, L82R, I89V, I92L, D107G, L128A, I137L, D164E, H167R, M177I/G359A, Q191Y, S192E, Y193L, L194F, N201G, S217G, V281I, K289P, D291G, D291Y/Q340L, I305V, S309E, E310D, I317W, W331L, A337I, Q340L, R346D, V352L, S353A, F357I, I358M, I358V, G359A, L362I, P371L, F396I, N414D/M435L, G415P, L428I, S437R, A443P, A447V, and I448L.

- 5 In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4 and one or more residue differences as compared to SEQ ID NO: 4, selected from 4, 5, 6, 8, 10, 10/317, 10/317/358, 10/317/359, 10/346, 10/346/359, 10/358, 10/359, 12, 12/373, 13,
- 10 15, 16, 18, 19, 26, 30, 30/154, 31, 34, 35, 37, 38, 68, 70, 73, 74, 93, 106, 118, 126, 134, 135, 143, 146, 154/453, 163, 182, 207/286, 208, 226, 237, 246, 247, 259, 262, 263, 264, 280, 282, 286, 287, 288, 289, 291, 292, 293, 298, 305, 309, 310, 313, 317, 317/346, 317/359, 317/445, 320, 346, 351, 355, 356, 358, 359, 360, 368, 369, 410, 411, 422, 426, 427, 428, 446, 448, 453, and 454. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%,
- 15 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4 and one or more residue differences as compared to SEQ ID NO: 4, selected from 4H, 5R, 6A, 8D, 10A, 10A/317W, 10A/317W/358V, 10A/317W/359A, 10A/346D, 10A/346D/359A, 10A/358V, 10A/359A, 12S, 12S/373C, 13V, 15S, 16S, 18A, 19L, 26A, 30N, 30N/154V, 31V, 34Q, 35L, 37K, 38Y, 68A, 70A, 73K, 74A, 93E, 106E, 118S, 126V, 134F, 135R, 143E, 146M,
- 20 154P/453E, 163L, 182A, 207C/286I, 208S, 226K, 237S, 246K, 247R, 259S, 262L, 263R, 264S, 280A, 282L, 286I, 287N, 288P, 289P, 291G, 292E, 293T, 298L, 305V, 309A, 310Q, 313R, 317W, 317W/346D, 317W/359A, 317W/445L, 320H, 346D, 351M, 355L, 356S, 358V, 359A, 360V, 368S, 369F, 410G, 411K, 422Q, 426M, 427A, 428I, 446G, 448L, 453E, and 454A. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%,
- 25 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4 and one or more residue differences as compared to SEQ ID NO: 4, selected from F4H, N5R, T6A, Q8D, S10A, S10A/I317W, S10A/I317W/I358V, S10A/I317W/G359A, S10A/R346D, S10A/R346D/G359A, S10A/I358V, S10A/G359A, E12S, E12S/Y373C, H13V, E15S, K16S, N18A, T19L, K26A, E30N, E30N/A154V, T31V, D34Q, M35L, R37K, C38Y, P68A, V70A, C73K, S74A,
- 30 D93E, S106E, A118S, T126V, T134F, K135R, D143E, C146M, A154P/D453E, A163L, S182A, R207C/T286I, A208S, E226K, D237S, S246K, K247R, K259S, V262L, K263R, V264S, V280A, Y282L, T286I, T287N, S288P, K289P, D291G, M292E, E293T, V298L, I305V, S309A, E310Q, H313R, I317W, I317W/R346D, I317W/G359A, I317W/W445L, A320H, R346D, A351M, C355L, A356S, I358V, G359A, I360V, A368S, I369F, P410G, R411K, K422Q, Q426M, S427A, L428I, E446G, I448L, D453E, and T454A.
- 35 In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 156 and one or more residue differences as compared to SEQ ID NO: 156, selected from 18/73/74/259, 18/73/286/289/411, 18/74/118/259/286/289/411, 18/74/259,
- 40 18/74/259/286/289, 18/74/259/286/411, 18/118/259, 18/259/286/289/411, 18/259/289/411, 18/289/411,

73/74/259/286/289, 73/74/259/289, 73/259/286, 74/259, 74/259/289, 74/286/289/411, 259, 259/286/289, and 289. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 156 and one or

5 more residue differences as compared to SEQ ID NO: 156, selected from 18A/73K/74A/259S, 18A/73K/286I/289P/411K, 18A/74A/118S/259S/286I/289P/411K, 18A/74A/259S, 18A/74A/259S/286I/289P, 18A/74A/259S/286I/411K, 18A/118S/259S, 18A/259S/286I/289P/411K, 18A/259S/289P/411K, 18A/289P/411K, 73K/74A/259S/286I/289P, 73K/74A/259S/289P, 73K/259S/286I, 74A/259S, 74A/259S/289P, 74A/286I/289P/411K, 259S, 259S/286I/289P, and 289P. In some

10 embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 156 and one or more residue differences as compared to SEQ ID NO: 156, selected from N18A/C73K/S74A/K259S, N18A/C73K/T286I/K289P/R411K, N18A/S74A/A118S/K259S/T286I/K289P/R411K, N18A/S74A/K259S, N18A/S74A/K259S/T286I/K289P, N18A/S74A/K259S/T286I/R411K, N18A/A118S/K259S, N18A/K259S/T286I/K289P/R411K, N18A/K259S/K289P/R411K, N18A/K289P/R411K, C73K/S74A/K259S/T286I/K289P, C73K/S74A/K259S/K289P, C73K/K259S/T286I, S74A/K259S, S74A/K259S/K289P, S74A/T286I/K289P/R411K, K259S, K259S/T286I/K289P, and K289P.

In some embodiments, the engineered transposase polypeptide with one or more improved

20 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 302 and one or more residue differences as compared to SEQ ID NO: 302, selected from 16/19/355/427, 16/37/146/286/287/289/310/313/355, 16/146/226/287/289/355, 19/37/146/286/287/310/355, 19/135/286/287/289/355, 19/182/355, 37/70/226/310/313/355, 37/74/310/355/427, 37/135/146/182/226/286/287/355/368, 37/146/226/289/427, 37/146/287/289/427, 38, 41, 42, 43, 44, 45, 47, 48, 50, 53, 54, 55, 70/74/226/286/287/289/368/427, 70/74/355/368, 70/135/146/150/182/286/289/313/427, 74/310/313/355, 74/310/355, 74/355/427, 101, 103, 117, 135/226/286/289/355/368, 135/355, 146, 146/150/287/289/310/355, 146/182/355, 146/226/286/355, 146/286/287/289/355/427, 146/289, 146/427, 150/226/310/313, 182/226/355, 182/355, 244, 247, 249, 251, 287/355/427, 300, 317/411, 355, 393, 397, 398, 399, 401, 405, 407, 408, 409, 411, 413, 426, 429, 431, 439, 444, 446, 450, 453, 454, 457, 459, 460, 464, 466, 467, 469, 470, 471, 472, 474, 475, 476, and 478. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 302 and one or more

35 residue differences as compared to SEQ ID NO: 302, selected from 16S/19L/355L/427A, 16S/37K/146M/286I/287N/289P/310Q/313R/355L, 16S/146M/226K/287N/289P/355L, 19L/37K/146M/286I/287N/310Q/355L, 19L/135R/286I/287N/289P/355L, 19L/182A/355L, 37K/70A/226K/310Q/313R/355L, 37K/74S/310Q/355L/427A, 37K/135R/146M/182A/226K/286I/287N/355L/368S, 37K/146M/226K/289P/427A, 37K/146M/287N/289P/427A, 38F, 38N, 41R, 42G, 43L, 43V, 44G, 44S, 44V, 45M, 45V, 47M, 47S, 48D,

40

50D, 53D, 53L, 54H, 54S, 54V, 55M, 70A/74S/226K/286I/287N/289P/368S/427A, 70A/74S/355L/368S, 70A/135R/146M/150H/182A/286I/289P/313R/427A, 74S/310Q/313R/355L, 74S/310Q/355L, 74S/355L/427A, 101S, 103M, 103R, 117R, 135R/226K/286I/289P/355L/368S, 135R/355L, 146M, 146M/150H/287N/289P/310Q/355L, 146M/182A/355L, 146M/226K/286I/355L,

5 146M/286I/287N/289P/355L/427A, 146M/289P, 146M/427A, 150H/226K/310Q/313R, 182A/226K/355L, 182A/355L, 244I, 244T, 247L, 249F, 249I, 249L, 249S, 249T, 249V, 249Y, 251L, 287N/355L/427A, 300F, 317R/411N, 355L, 393A, 397S, 398V, 399V, 401V, 405F, 405V, 407L, 408R, 409G, 411D, 413Y, 426A, 426H, 426M, 426S, 426V, 426W, 429G, 431M, 439G, 444G, 446K, 446Q, 450L, 450M, 450T, 453G, 453H, 453R, 453T, 454L, 454R, 457A, 459M, 460T, 464L, 464M, 466R, 467G, 469F, 470M, 470N, 470R,

10 470V, 470Y, 471A, 471S, 472R, 474R, 474S, 475P, 476D, 476L, 476S, 476T, 478C, 478G, 478L, and 478V. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 302 and one or more residue differences as compared to SEQ ID NO: 302, selected from K16S/T19L/C355L/S427A,

15 K16S/R37K/C146M/T286I/T287N/K289P/E310Q/H313R/C355L, K16S/C146M/E226K/T287N/K289P/C355L, T19L/R37K/C146M/T286I/T287N/E310Q/C355L, T19L/K135R/T286I/T287N/K289P/C355L, T19L/S182A/C355L, R37K/V70A/E226K/E310Q/H313R/C355L, R37K/A74S/E310Q/C355L/S427A, R37K/K135R/C146M/S182A/E226K/T286I/T287N/C355L/A368S, R37K/C146M/E226K/K289P/S427A, R37K/C146M/T287N/K289P/S427A, C38F, C38N, K41R, S42G,

20 I43L, I43V, T44G, T44S, T44V, I45M, I45V, C47M, C47S, Q48D, S50D, M53D, M53L, Q54H, Q54S, Q54V, E55M, V70A/A74S/E226K/T286I/T287N/K289P/A368S/S427A, V70A/A74S/C355L/A368S, V70A/K135R/C146M/N150H/S182A/T286I/K289P/H313R/S427A, A74S/E310Q/H313R/C355L, A74S/E310Q/C355L, A74S/C355L/S427A, K101S, Q103M, Q103R, K117R, K135R/E226K/T286I/K289P/C355L/A368S, K135R/C355L, C146M,

25 C146M/N150H/T287N/K289P/E310Q/C355L, C146M/S182A/C355L, C146M/E226K/T286I/C355L, C146M/T286I/T287N/K289P/C355L/S427A, C146M/K289P, C146M/S427A, N150H/E226K/E310Q/H313R, S182A/E226K/C355L, S182A/C355L, V244I, V244T, K247L, K249F, K249I, K249L, K249S, K249T, K249V, K249Y, K251L, T287N/C355L/S427A, L300F, W317R/R411N, C355L, E393A, D397S, E398V, D399V, W401V, M405F, M405V, M407L, K408R, K409G, R411D,

30 W413Y, Q426A, Q426H, Q426M, Q426S, Q426V, Q426W, A429G, L431M, R439G, S444G, E446K, E446Q, E450L, E450M, E450T, D453G, D453H, D453R, D453T, T454L, T454R, S457A, V459M, V460T, V464L, V464M, K466R, E467G, L469F, A470M, A470N, A470R, A470V, A470Y, D471A, D471S, G472R, L474R, L474S, L475P, G476D, G476L, G476S, G476T, S478C, S478G, S478L, and S478V.

In some embodiments, the engineered transposase polypeptide with one or more improved

35 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 338 and one or more residue differences as compared to SEQ ID NO: 338, selected from 16, 16/313, 16/409/411/427, 37/150/408, 37/313/317/408/411, 37/317/405/408, 37/317/405/426, 37/405/409/427, 150, 150/313/405/411, 150/405/406, 150/408, 150/408/411, 226/405,

40 317/460, and 408/426. In some embodiments, the engineered transposase polypeptide comprises an

amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 338 and one or more residue differences as compared to SEQ ID NO: 338, selected from 16S, 16S/313R, 16S/409D/411N/427M, 37K/150N/408G, 37K/313R/317R/408G/411N, 37K/317R/405V/408G, 5 37K/317R/405V/426M, 37K/405V/409D/427M, 150N, 150N/313R/405V/411N, 150N/405V/406S, 150N/408G, 150N/408G/411N, 226K/405V, 317R/460L, and 408G/426M. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 338 and one or more residue differences as compared to 10 SEQ ID NO: 338, selected from K16S, K16S/H313R, K16S/K409D/R411N/S427M, R37K/H150N/K408G, R37K/H313R/W317R/K408G/R411N, R37K/W317R/M405V/K408G, R37K/W317R/M405V/Q426M, R37K/M405V/K409D/S427M, H150N, H150N/H313R/M405V/R411N, H150N/M405V/K406S, H150N/K408G, H150N/K408G/R411N, E226K/M405V, W317R/V460L, and K408G/Q426M.

In some embodiments, the engineered transposase polypeptide with one or more improved 15 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 606 and one or more residue differences as compared to SEQ ID NO: 606, selected from 12, 13, 18, 19, 22, 26, 31/262, 34, 60, 79, 80, 81, 82, 92, 106, 107, 113, 115, 132, 134, 136, 140, 143, 152, 160, 163, 167, 168, 175, 179, 183, 195, 199, 201, 202, 215, 216, 218, 220, 223, 226, 233, 20 235, 236, 237, 243, 255, 259, 261, 262, 266, 268, 270, 273, 282, 290, 292, 293, 309, 319, 320, 326, 343, 346, 365, 370, 374, 380, 382, 386, 389, 390, 408, 412, 414, 415, 424, and 442. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 606 and one or more residue differences as 25 compared to SEQ ID NO: 606, selected from 12M, 13I, 18Q, 19L, 22S, 22T, 26R, 31M/262A, 34G, 60L, 79Y, 80T, 81T, 82Q, 92L, 92T, 106Q, 107G, 107H, 107R, 113K, 115R, 132G, 132R, 134E, 134G, 136L, 140C, 140L, 140V, 143G, 143R, 143V, 152S, 160Q, 160T, 163L, 163T, 167R, 168Q, 175T, 179G, 183L, 195G, 199R, 199T, 199W, 201G, 201Q, 202M, 215R, 215V, 216E, 218G, 220R, 223R, 223V, 226L, 226Q, 233R, 235V, 236T, 237G, 237S, 243L, 243M, 255R, 259A, 261E, 262L, 266P, 268A, 270R, 273K, 30 273P, 273T, 282G, 290A, 290R, 290T, 292W, 293V, 309G, 319R, 320R, 320S, 326A, 326L, 343G, 343R, 346G, 346N, 365I, 370F, 374L, 380R, 382A, 382R, 386D, 389G, 390R, 408A, 408R, 412A, 412H, 412I, 412R, 414R, 415S, 424V, and 442K. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ 35 ID NO: 606 and one or more residue differences as compared to SEQ ID NO: 606, selected from E12M, H13I, A18Q, T19L, P22S, P22T, K26R, T31M/V262A, D34G, F60L, T79Y, A80T, V81T, L82Q, I92L, I92T, S106Q, D107G, D107H, D107R, S113K, K115R, F132G, F132R, T134E, T134G, T136L, I140C, I140L, I140V, D143G, D143R, D143V, E152S, G160Q, G160T, A163L, A163T, H167R, F168Q, D175T, R179G, V183L, Q195G, S199R, S199T, S199W, N201G, N201Q, E202M, L215R, L215V, D216E, E218G, P220R, 40 E223R, E223V, E226L, E226Q, A233R, T235V, I236T, D237G, D237S, V243L, V243M, A255R, S259A,

S261E, V262L, A266P, T268A, K270R, D273K, D273P, D273T, Y282G, K290A, K290R, K290T, M292W, E293V, S309G, T319R, A320R, A320S, E326A, E326L, E343G, E343R, R346G, R346N, L365I, T370F, Y374L, L380R, D382A, D382R, E386D, S389G, Q390R, G408A, G408R, G412A, G412H, G412I, G412R, N414R, G415S, A424V, and I442K.

- 5 In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 606 and one or more residue differences as compared to SEQ ID NO: 606, selected from 2, 12, 13, 15, 16, 18, 19, 22, 26, 30, 31/262, 34, 51, 60, 69, 79, 82, 84, 87, 88, 106, 107,
- 10 113, 114, 115, 128, 132, 134, 135, 136, 140, 143, 146, 150, 152, 160, 163, 167, 168, 175, 179, 183, 195, 198, 199, 201, 215, 217, 218, 219, 220, 223, 224, 226, 227, 231, 233, 235, 236, 243, 254, 255, 259, 261, 266, 267, 268, 270, 272, 273, 275, 277, 279, 282, 287, 289, 290, 291, 292, 293, 304, 307, 309, 319, 320, 343, 344, 346, 355, 365, 370, 373, 380, 382, 386, 389, 390, 408, 412, 414, 415, 418, 423, 424, and 468.
- In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence
- 15 having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 606 and one or more residue differences as compared to SEQ ID NO: 606, selected from 2A, 12L, 12M, 13I, 13N, 13Y, 15R, 16S, 18Q, 19L, 22S, 22T, 26R, 30H, 31M/262A, 34G, 51R, 60L, 69A, 69L, 79Y, 82Q, 84R, 87T, 88L, 106Q, 107G, 107H, 107R, 113E, 114P, 115S, 128F, 132G, 132R, 134E, 134G, 134L, 135T, 136L, 140C,
- 20 140L, 140V, 143G, 143R, 143V, 146K, 150R, 152S, 160Q, 160T, 163L, 163T, 167R, 168Q, 175T, 175V, 179G, 183L, 195G, 195S, 198S, 199R, 199T, 199W, 201G, 201Q, 215P, 215V, 217F, 217G, 218G, 219S, 220R, 223R, 223V, 224A, 224S, 226L, 227Y, 231V, 233R, 235A, 235V, 236T, 243L, 243M, 254F, 255R, 259A, 261E, 266P, 267P, 268A, 270A, 270R, 272L, 272P, 273K, 273P, 273T, 275L, 277K, 279G, 282G, 287L, 287R, 289T, 290T, 291S, 292W, 293V, 304R, 307L, 309G, 319R, 320L, 320N, 320R, 320S, 343G,
- 25 343R, 344V, 346G, 346N, 355G, 365I, 370F, 373R, 380G, 380R, 382R, 386D, 386Q, 389G, 390R, 408A, 412A, 412H, 412I, 412R, 414R, 415S, 418S, 423D, 424V, and 468T. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 606 and one or more residue differences as compared to
- 30 SEQ ID NO: 606, selected from S2A, E12L, E12M, H13I, H13N, H13Y, E15R, K16S, A18Q, T19L, P22S, P22T, K26R, E30H, T31M/V262A, D34G, K51R, F60L, E69A, E69L, T79Y, L82Q, Q84R, P87T, E88L, S106Q, D107G, D107H, D107R, S113E, V114P, K115S, L128F, F132G, F132R, T134E, T134G, T134L, K135T, T136L, I140C, I140L, I140V, D143G, D143R, D143V, M146K, H150R, E152S, G160Q, G160T, A163L, A163T, H167R, F168Q, D175T, D175V, R179G, V183L, Q195G, Q195S, Q198S, S199R, S199T,
- 35 S199W, N201G, N201Q, L215P, L215V, S217F, S217G, E218G, L219S, P220R, E223R, E223V, Y224A, Y224S, E226L, S227Y, L231V, A233R, T235A, T235V, I236T, V243L, V243M, P254F, A255R, S259A, S261E, A266P, V267P, T268A, K270A, K270R, K272L, K272P, D273K, D273P, D273T, M275L, P277K, N279G, Y282G, N287L, N287R, P289T, K290T, D291S, M292W, E293V, P304R, T307L, S309G, T319R, A320L, A320N, A320R, A320S, E343G, E343R, E344V, R346G, R346N, L355G, L365I, T370F, Y373R,

L380G, L380R, D382R, E386D, E386Q, S389G, Q390R, G408A, G412A, G412H, G412I, G412R, N414R, G415S, V418S, W423D, A424V, and M468T.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 606 and one or more residue differences as compared to SEQ ID NO: 606, selected from 44/45/47/408/453, 44/45/47/453, 44/45/47/453/454, 44/45/103/453/457, 44/45/453/454/457, 44/47/103/393/408/453/454, 44/47/103/453/454/457, 44/47/393/408/453/457, 44/47/393/453/457, 44/103/408/453/454, 45/47/103/393/453/454/457, 45/47/393/453/454, 45/47/408/423/453/457, 45/47/453, 45/103/393/453/454, 45/453/454, 47/103/408/453/454, 47/393/408/453/454/457, 47/393/453/454, 47/408/453/457, 53/54/244/450, 103/393/408/453/457, 244/450, 408/453/454, 446/450/464, and 453/457.

In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 606 and one or more residue differences as compared to SEQ ID NO: 606, selected from 44G/45V/47M/408R/453R, 44G/45V/47M/453H, 44G/45V/47M/453R/454R, 44G/47M/393A/453R/457G, 44S/45V/103M/453R/457G, 44S/45V/453T/454R/457G, 44S/47M/103M/393A/408R/453R/454R, 44S/47M/103M/453R/454R/457G, 44S/47S/393A/408R/453R/457G, 44S/103M/408R/453R/454R, 45V/47M/393A/453R/454R, 45V/47M/408R/423C/453T/457G, 45V/47S/103M/393A/453R/454R/457G, 45V/47S/453R, 45V/103M/393A/453T/454R, 45V/453R/454R, 47M/103M/408R/453T/454R, 47M/393A/408R/453R/454R/457G, 47M/393A/453R/454R, 47S/408R/453R/457G, 53D/54V/244I/450T, 103M/393A/408R/453R/457G, 244I/450T, 408R/453T/454R, 446K/450T/464L, and 453R/457G. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 606 and one or more residue differences as compared to SEQ ID NO: 606, selected from T44G/I45V/C47M/G408R/D453R, T44G/I45V/C47M/D453H, T44G/I45V/C47M/D453R/T454R, T44G/C47M/E393A/D453R/S457G, T44S/I45V/Q103M/D453R/S457G, T44S/I45V/D453T/T454R/S457G, T44S/C47M/Q103M/E393A/G408R/D453R/T454R, T44S/C47M/Q103M/D453R/T454R/S457G, T44S/C47S/E393A/G408R/D453R/S457G, T44S/Q103M/G408R/D453R/T454R, I45V/C47M/E393A/D453R/T454R, I45V/C47M/G408R/W423C/D453T/S457G, I45V/C47S/Q103M/E393A/D453R/T454R/S457G, I45V/C47S/D453R, I45V/Q103M/E393A/D453T/T454R, I45V/D453R/T454R, C47M/Q103M/G408R/D453T/T454R, C47M/E393A/G408R/D453R/T454R/S457G, C47M/E393A/D453R/T454R, C47S/G408R/D453R/S457G, M53D/Q54V/V244I/E450T, Q103M/E393A/G408R/D453R/S457G, V244I/E450T, G408R/D453T/T454R, E446K/E450T/V464L, and D453R/S457G.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972,

selected from 38/45/53, 38/45/53/54/476, 38/45/53/244, 38/45/53/470/472, 38/45/54/244, 38/45/54/470/474/478, 38/45/472/476/478, 38/45/476, 38/53/54, 38/53/54/244, 38/53/54/244/472/478, 38/53/472, 38/54/244/470/474/476/478, 38/244/470/472/474/478, 38/478, 45/53/54/244/470/474/476/478, 45/53/54/470/474/476, 45/53/54/472/478, 45/54/244/474/478, 53, 53/54/470/476/478, 53/244/474/478, 54, 54/244, 54/244/472/476, 54/472, 244, and 474/478. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from 38F/45I/53D, 38F/45I/53D/54S/476S, 38F/45I/53L/244I, 38F/45I/53L/470N/472R, 38F/45I/54V/244T, 38F/45I/54V/470N/474S/478G, 38F/45I/472R/476S/478G, 38F/45I/476S, 38F/53D/54H/244I/472R/478G, 38F/53D/54S, 38F/53D/54S/244I, 38F/53D/54S/244I/472R/478G, 38F/53L/472R, 38F/54S/244I/470N/474S/476S/478G, 38F/244I/470R/472R/474S/478G, 38F/478G, 45I/53D/54H/244I/470N/474S/476S/478G, 45I/53D/54S/470R/474S/476S, 45I/53L/54S/472R/478G, 45I/54S/244T/474S/478G, 53D, 53D/244T/474S/478G, 53L/54S/470R/476S/478G, 54H, 54S/472R, 54V/244I/472R/476S, 54V/244T, 244T, and 474S/478G. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from C38F/V45I/M53D, C38F/V45I/M53D/Q54S/G476S, C38F/V45I/M53L/V244I, C38F/V45I/M53L/A470N/G472R, C38F/V45I/Q54V/V244T, C38F/V45I/Q54V/A470N/L474S/S478G, C38F/V45I/G472R/G476S/S478G, C38F/V45I/G476S, C38F/M53D/Q54H/V244I/G472R/S478G, C38F/M53D/Q54S, C38F/M53D/Q54S/V244I, C38F/M53D/Q54S/V244I/G472R/S478G, C38F/M53L/G472R, C38F/Q54S/V244I/A470N/L474S/G476S/S478G, C38F/V244I/A470R/G472R/L474S/S478G, C38F/S478G, V45I/M53D/Q54H/V244I/A470N/L474S/G476S/S478G, V45I/M53D/Q54S/A470R/L474S/G476S, V45I/M53L/Q54S/G472R/S478G, V45I/Q54S/V244T/L474S/S478G, M53D, M53D/V244T/L474S/S478G, M53L/Q54S/A470R/G476S/S478G, Q54H, Q54S/G472R, Q54V/V244I/G472R/G476S, Q54V/V244T, V244T, and L474S/S478G.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from 22/38/45/107/259/317/412, 22/38/45/259/317/412, 22/107, 22/107/317/412, 22/134, 22/134/220, 22/134/220/317, 22/134/317/412, 22/317, 38, 38/45/81/143/244, 38/53/243, 38/81, 38/81/244, 38/107/134/220, 38/132, 38/220/317, 45/81, 45/81/243, 45/81/243/343, 45/143/243/270/343, 45/244/270, 53/115/163/243/270/343, 53/143/243/270/343, 53/270, 64/134/220/317/412, 81/132/243, 107, 107/134/317, 115/244/270, 134/259, 143/243/270, 243, 243/270, 244, 244/270, 259, 270, and 317. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more

residue differences as compared to SEQ ID NO: 972, selected from 22T/38F/45I/107R/259A/317M/412H, 22T/38F/45I/259A/317M/412H, 22T/107R, 22T/107R/317M/412H, 22T/134G, 22T/134G/220R/317M, 22T/134G/317M/412H, 22T/134L/220R, 22T/317M, 38F, 38F/45I/81T/143R/244I, 38F/53D/243L, 38F/81T, 38F/81T/244I, 38F/107R/134G/220R, 38F/132G, 38F/220R/317M, 45I/81T, 45I/81T/243L, 45I/81T/243L/343R, 45I/143R/243L/270R/343R, 45I/244I/270R, 53D/115R/163L/243L/270R/343R, 53D/143R/243L/270R/343R, 53D/270R, 64R/134L/220R/317M/412H, 81T/132G/243L, 107R, 107R/134L/317M, 115R/244I/270R, 134G/259A, 143R/243L/270R, 243L, 243L/270R, 244I, 244I/270R, 259A, 270R, and 317M. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from P22T/C38F/V45I/D107R/S259A/R317M/G412H, P22T/C38F/V45I/S259A/R317M/G412H, P22T/D107R, P22T/D107R/R317M/G412H, P22T/T134G, P22T/T134G/P220R/R317M, P22T/T134G/R317M/G412H, P22T/T134L/P220R, P22T/R317M, C38F, C38F/V45I/V81T/D143R/V244I, C38F/M53D/V243L, C38F/V81T, C38F/V81T/V244I, C38F/D107R/T134G/P220R, C38F/F132G, C38F/P220R/R317M, V45I/V81T, V45I/V81T/V243L, V45I/V81T/V243L/E343R, V45I/D143R/V243L/K270R/E343R, V45I/V244I/K270R, M53D/K115R/A163L/V243L/K270R/E343R, M53D/D143R/V243L/K270R/E343R, M53D/K270R, P64R/T134L/P220R/R317M/G412H, V81T/F132G/V243L, D107R, D107R/T134L/R317M, K115R/V244I/K270R, T134G/S259A, D143R/V243L/K270R, V243L, V243L/K270R, V244I, V244I/K270R, S259A, K270R, and R317M.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from 22/38/45/107/259/317/412, 22/38/45/259/317/412, 22/38/64/134/220, 22/45/220/259, 22/64/107/259/317/412, 22/64/134/220, 22/107, 22/134, 22/134/317, 22/134/317/412, 22/220, 22/220/412/415, 22/259/317, 22/259/317/412, 22/317, 38, 38/45/81/143/244, 38/45/143/270, 38/53/243, 38/81, 38/81/143/243/270, 38/81/183/243/270, 38/81/243, 38/81/244, 38/107/134/220, 38/132, 38/220/317, 38/270/343, 45/81, 45/81/132/243/270, 45/81/243, 45/143/243/270/343, 45/244/270, 53/143/243/270/343, 53/244/270, 53/270, 64, 64/107/134, 64/107/317, 64/134, 64/134/220/317/412, 81/132, 81/132/243, 107, 107/134/317, 107/412, 115/244/270, 143/243/270, 220/259, 243, 243/270, 244/270, 255, 259, and 270. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from 22T/38F/45I/107R/259A/317M/412H, 22T/38F/45I/259A/317M/412H, 22T/38F/64R/134L/220R, 22T/45I/220R/259A, 22T/64R/107R/259A/317M/412H, 22T/64R/134L/220R, 22T/107R, 22T/134G, 22T/134G/317M/412H, 22T/134L/317M, 22T/134L/317M/412H, 22T/220R, 22T/220R/412H/415D, 22T/259A/317M, 22T/259A/317M/412H, 22T/317M, 38F, 38F/45I/81T/143R/244I, 38F/45I/143R/270R, 38F/53D/243L, 38F/81T, 38F/81T/143R/243L/270R, 38F/81T/183L/243L/270R, 38F/81T/243L,

- 38F/81T/244I, 38F/107R/134G/220R, 38F/132G, 38F/220R/317M, 38F/270R/343R, 45I/81T, 45I/81T/132G/243L/270R, 45I/81T/243L, 45I/143R/243L/270R/343R, 45I/244I/270R, 53D/143R/243L/270R/343R, 53D/244I/270R, 53D/270R, 64R, 64R/107R/134L, 64R/107R/317M, 64R/134L, 64R/134L/220R/317M/412H, 81T/132G, 81T/132G/243L, 107R, 107R/134L/317M, 107R/412H, 115R/244I/270R, 143R/243L/270R, 220R/259A, 243L, 243L/270R, 244I/270R, 255R, 259A, and 270R. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from
- P22T/C38F/V45I/D107R/S259A/R317M/G412H, P22T/C38F/V45I/S259A/R317M/G412H, P22T/C38F/P64R/T134L/P220R, P22T/V45I/P220R/S259A, P22T/P64R/D107R/S259A/R317M/G412H, P22T/P64R/T134L/P220R, P22T/D107R, P22T/T134G, P22T/T134G/R317M/G412H, P22T/T134L/R317M, P22T/T134L/R317M/G412H, P22T/P220R, P22T/P220R/G412H/G415D, P22T/S259A/R317M, P22T/S259A/R317M/G412H, P22T/R317M, C38F, C38F/V45I/V81T/D143R/V244I, C38F/V45I/D143R/K270R, C38F/M53D/V243L, C38F/V81T, C38F/V81T/D143R/V243L/K270R, C38F/V81T/V183L/V243L/K270R, C38F/V81T/V243L, C38F/V81T/V244I, C38F/D107R/T134G/P220R, C38F/F132G, C38F/P220R/R317M, C38F/K270R/E343R, V45I/V81T, V45I/V81T/F132G/V243L/K270R, V45I/V81T/V243L, V45I/D143R/V243L/K270R/E343R, V45I/V244I/K270R, M53D/D143R/V243L/K270R/E343R, M53D/V244I/K270R, M53D/K270R, P64R, P64R/D107R/T134L, P64R/D107R/R317M, P64R/T134L, P64R/T134L/P220R/R317M/G412H, V81T/F132G, V81T/F132G/V243L, D107R, D107R/T134L/R317M, D107R/G412H, K115R/V244I/K270R, D143R/V243L/K270R, P220R/S259A, V243L, V243L/K270R, V244I/K270R, A255R, S259A, and K270R.
- In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from 143, 143/175/272/273/415, 152/168/216/237, 152/168/423, 160, 160/201/272, 160/201/273/415, 160/226, 160/244/272/273/415, 160/415, 168, 168/187/216/255/423, 168/216/237, 168/255/423, 168/290/423, 187, 187/423, 199/201/226, 199/201/415, 201, 216, 216/423, 226/244/415, 226/415, 237, 237/290, 244, 255, 272, 290, 290/423, 415, and 423. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from 143R, 143R/175T/272W/273P/415S, 152S/168Q/216E/237S, 152S/168Q/423D, 160Q, 160Q/201G/272W, 160Q/201G/273P/415S, 160Q/226R, 160Q/244I/272W/273P/415S, 160Q/415S, 168Q, 168Q/187G/216E/255R/423D, 168Q/216E/237S, 168Q/255R/423D, 168Q/290R/423D, 187G, 187G/423D, 199W/201G/226R, 199W/201G/415S, 201G, 216E, 216E/423D, 226R/244I/415S, 226R/415S, 237S, 237S/290R, 244I, 255R, 272W, 290R, 290R/423D, 415S, and 423D. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%,

93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from D143R, D143R/D175T/K272W/D273P/G415S, E152S/F168Q/D216E/D237S, E152S/F168Q/W423D, G160Q, G160Q/N201G/K272W, G160Q/N201G/D273P/G415S, G160Q/E226R,

5 G160Q/V244I/K272W/D273P/G415S, G160Q/G415S, F168Q, F168Q/E187G/D216E/A255R/W423D, F168Q/D216E/D237S, F168Q/A255R/W423D, F168Q/K290R/W423D, E187G, E187G/W423D, S199W/N201G/E226R, S199W/N201G/G415S, N201G, D216E, D216E/W423D, E226R/V244I/G415S, E226R/G415S, D237S, D237S/K290R, V244I, A255R, K272W, K290R, K290R/W423D, G415S, and W423D.

10 In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from 143, 143/160/226/415, 143/175/272/273/415, 152/168/216/237, 152/168/423, 160, 160/201/272, 160/201/273/415, 160/226, 160/244/272/273/415, 160/415, 168, 168/187/216/255/423, 168/216/237, 168/216/237/255/343, 168/255/423, 168/290/423, 187, 199/201/226, 199/201/415, 201, 216, 216/423, 226/244/415, 226/415, 237, 237/290, 244, 255, 255/343, 272, 290, 290/423, 415, and 423.

In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from 143R, 143R/160T/226R/415S, 143R/175T/272W/273P/415S, 152S/168Q/216E/237S, 152S/168Q/423D, 160Q, 160Q/201G/272W, 160Q/201G/273P/415S, 160Q/226R, 160Q/244I/272W/273P/415S, 160Q/415S, 168Q, 168Q/187G/216E/255R/423D, 168Q/216E/237S, 168Q/216E/237S/255R/343R, 168Q/255R/423D, 168Q/290R/423D, 187G, 199W/201G/226R, 199W/201G/415S, 201G, 216E, 216E/423D, 226R/244I/415S, 226R/415S, 237S, 237S/290R, 244I, 255R, 255R/343R, 272W, 290R, 290R/423D, 415S, and 423D. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one

30 or more residue differences as compared to SEQ ID NO: 1082, selected from D143R, D143R/G160T/E226R/G415S, D143R/D175T/K272W/D273P/G415S, E152S/F168Q/D216E/D237S, E152S/F168Q/W423D, G160Q, G160Q/N201G/K272W, G160Q/N201G/D273P/G415S, G160Q/E226R, G160Q/V244I/K272W/D273P/G415S, G160Q/G415S, F168Q, F168Q/E187G/D216E/A255R/W423D, F168Q/D216E/D237S, F168Q/D216E/D237S/A255R/E343R, F168Q/A255R/W423D, F168Q/K290R/W423D, E187G, S199W/N201G/E226R, S199W/N201G/G415S, N201G, D216E, D216E/W423D, E226R/V244I/G415S, E226R/G415S, D237S, D237S/K290R, V244I, A255R, A255R/E343R, K272W, K290R, K290R/W423D, G415S, and W423D.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference

sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from 72, 90, 91, 92, 93, 95, 122, 129, 141, 142, 143, 144, 145, 150, 151, 151/155, 152, 154, 155, 157, 160, 161, 161/313, 164, 165, 168, 180, 181, 184, 193, 197, 204, 205, 205/472, 206, 299, and 320. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having

5 at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from 72M, 72Q, 90F, 91I, 91V, 92A, 92L, 92V, 93A, 93H, 93S, 95G, 95S, 122F, 122L, 122Q, 122V, 129F, 129I, 141Y, 142L, 143A, 143K, 144F, 144I, 144V, 144Y, 145F, 145L, 150D, 150E, 150G, 150R, 150S, 150T, 150V, 150W, 151A, 151Q, 151R,

10 151T/155T, 152D, 152H, 152L, 152Q, 152R, 152T, 152V, 152Y, 154E, 154K, 154L, 154P, 154Q, 154S, 154T, 154V, 154Y, 155E, 155G, 155H, 155Q, 155R, 155S, 155T, 157R, 160I, 160Q, 160V, 160W, 161C, 161C/313C, 161R, 161V, 164Q, 164R, 164S, 165G, 165S, 168E, 168K, 168L, 168N, 168R, 168V, 180C, 180L, 180T, 181L, 181V, 184A, 184G, 184H, 184S, 193H, 193L, 193M, 193V, 197H, 197R, 197Y, 204L, 204M, 204V, 204W, 204Y, 205I, 205L, 205L/472D, 206L, 299F, 299I, 320C, 320H, 320K, 320L, 320Q,

15 320R, 320S, and 320Y. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from R72M, R72Q, L90F, A91I, A91V, I92A, I92L, I92V, D93A, D93H, D93S, T95G, T95S, W122F, W122L, W122Q, W122V,

20 L129F, L129I, H141Y, Q142L, D143A, D143K, W144F, W144I, W144V, W144Y, W145F, W145L, H150D, H150E, H150G, H150R, H150S, H150T, H150V, H150W, P151A, P151Q, P151R, P151T/D155T, E152D, E152H, E152L, E152Q, E152R, E152T, E152V, E152Y, A154E, A154K, A154L, A154P, A154Q, A154S, A154T, A154V, A154Y, D155E, D155G, D155H, D155Q, D155R, D155S, D155T, K157R, G160I, G160Q, G160V, G160W, K161C, K161C/R313C, K161R, K161V, D164Q, D164R, D164S, A165G, A165S, F168E,

25 F168K, F168L, F168N, F168R, F168V, V180C, V180L, V180T, I181L, I181V, C184A, C184G, C184H, C184S, Y193H, Y193L, Y193M, Y193V, K197H, K197R, K197Y, F204L, F204M, F204V, F204W, F204Y, V205I, V205L, V205L/G472D, V206L, L299F, L299I, A320C, A320H, A320K, A320L, A320Q, A320R, A320S, and A320Y.

In some embodiments, the engineered transposase polypeptide with one or more improved

30 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from 90, 92, 122, 129, 141, 143, 144, 145, 150, 151, 151/155, 152, 154, 155, 155/265, 157, 160, 161, 164, 165, 168, 180, 181, 184, 193, 197, 204, 205, 205/472, 206, 299, and 320. In some

35 embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from 90F, 92A, 92L, 92P, 92V, 122F, 122L, 129F, 129I, 141Y, 143A, 143K, 144F, 144Y, 145F, 145L, 150D, 150E, 150G, 150R, 150S, 150V, 150W, 151A,

40 151K, 151Q, 151R, 151T/155T, 152D, 152H, 152L, 152Q, 152R, 152T, 152V, 152Y, 154E, 154K, 154L,

154Q, 154S, 154T, 154V, 154Y, 155E, 155G, 155H, 155Q, 155Q/265V, 155R, 155S, 155T, 157R, 160I, 160Q, 160V, 160W, 161C, 161R, 161V, 164Q, 164R, 164S, 165G, 165S, 168E, 168K, 168L, 168N, 168R, 168V, 180A, 180C, 180L, 180T, 181L, 181V, 184A, 184G, 184S, 193H, 193L, 193M, 193V, 197H, 197R, 204L, 204M, 204V, 204W, 204Y, 205I, 205L, 205L/472D, 206L, 299F, 299I, 320C, 320H, 320K, 320L, 320Q, 320R, 320S, and 320Y. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1082 and one or more residue differences as compared to SEQ ID NO: 1082, selected from L90F, I92A, I92L, I92P, I92V, W122F, W122L, L129F, L129I, H141Y, D143A, D143K, W144F, W144Y, W145F, W145L, H150D, H150E, H150G, H150R, H150S, H150V, H150W, P151A, P151K, P151Q, P151R, P151T/D155T, E152D, E152H, E152L, E152Q, E152R, E152T, E152V, E152Y, A154E, A154K, A154L, A154Q, A154S, A154T, A154V, A154Y, D155E, D155G, D155H, D155Q, D155Q/A265V, D155R, D155S, D155T, K157R, G160I, G160Q, G160V, G160W, K161C, K161R, K161V, D164Q, D164R, D164S, A165G, A165S, F168E, F168K, F168L, F168N, F168R, F168V, V180A, V180C, V180L, V180T, I181L, I181V, C184A, C184G, C184S, Y193H, Y193L, Y193M, Y193V, K197H, K197R, F204L, F204M, F204V, F204W, F204Y, V205I, V205L, V205L/G472D, V206L, L299F, L299I, A320C, A320H, A320K, A320L, A320Q, A320R, A320S, and A320Y.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1218 and one or more residue differences as compared to SEQ ID NO: 1218, selected from 129, 129/160, 129/160/164, 129/160/184, 129/164, 129/164/184, 129/205, 146/152/244, 150/155/226, 150/155/244, 152, 154, 154/160/184, 154/164, 154/205, 155, 160, 160/164, 160/164/184, 160/184, 164, 164/204, 164/205, 204, 205, 226, 226/244, and 244. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1218 and one or more residue differences as compared to SEQ ID NO: 1218, selected from 129I, 129I/160I, 129I/160I/164Q, 129I/160I/184A, 129I/164Q, 129I/164Q/184A, 129I/205I, 146L/152T/244I, 150D/155T/226R, 150D/155T/244I, 152T, 154K, 154K/160I/184A, 154K/164Q, 154K/205I, 155T, 160I, 160I/164Q, 160I/164Q/184A, 160I/184A, 164Q, 164Q/204L, 164Q/205I, 204L, 205I, 226R, 226R/244I, and 244I. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1218 and one or more residue differences as compared to SEQ ID NO: 1218, selected from L129I, L129I/G160I, L129I/G160I/D164Q, L129I/G160I/C184A, L129I/D164Q, L129I/D164Q/C184A, L129I/V205I, M146L/E152T/V244I, H150D/D155T/E226R, H150D/D155T/V244I, E152T, A154K, A154K/G160I/C184A, A154K/D164Q, A154K/V205I, D155T, G160I, G160I/D164Q, G160I/D164Q/C184A, G160I/C184A, D164Q, D164Q/F204L, D164Q/V205I, F204L, V205I, E226R, E226R/V244I, and V244I.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1218 and one or more residue differences as compared to SEQ ID NO: 1218,

5 selected from 129, 129/143, 129/154/164, 129/154/164/205, 129/154/205, 129/160, 129/160/164, 129/160/205, 129/164/205, 143, 143/150/155/226/244, 143/150/155/244, 143/150/226, 143/150/226/244/343, 143/150/226/343, 143/150/244, 143/150/343, 143/152, 143/152/226, 143/152/226/244, 143/152/226/343, 143/152/343, 143/155, 143/155/226/244, 143/155/226/244/343, 143/155/343, 143/226, 143/226/244/343, 143/226/343, 143/244, 143/244/343, 143/343, 146/152/244,

10 150/155/226, 150/155/244, 150/155/244/343, 150/226/244/343, 152, 152/174/244/343, 152/226, 152/226/343, 152/343, 154, 154/160/184, 154/164, 154/164/204, 154/164/204/205, 154/205, 155, 160, 160/164, 160/164/184, 164, 205, 226, 226/244, and 244. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to

15 reference sequence SEQ ID NO: 1218 and one or more residue differences as compared to SEQ ID NO: 1218, selected from 129I, 129I/143H, 129I/154K/164Q, 129I/154K/164Q/205I, 129I/154K/205I, 129I/160I, 129I/160I/164Q, 129I/160I/205I, 129I/164Q/205I, 143R, 143R/150D/155T/226R/244I, 143R/150D/155T/244I, 143R/150D/226R, 143R/150D/226R/244I/343R, 143R/150D/226R/343R, 143R/150D/244I, 143R/150D/343R, 143R/152T, 143R/152T/226R, 143R/152T/226R/244I,

20 143R/152T/226R/343R, 143R/152T/343R, 143R/155T, 143R/155T/226R/244I, 143R/155T/226R/244I/343R, 143R/155T/343R, 143R/226R, 143R/226R/244I/343R, 143R/226R/343R, 143R/244I, 143R/244I/343R, 143R/343R, 146L/152T/244I, 150D/155T/226R, 150D/155T/244I, 150D/155T/244I/343R, 150D/226R/244I/343R, 152T, 152T/174R/244I/343R, 152T/226R, 152T/226R/343R, 152T/343R, 154K, 154K/160I/184A, 154K/164Q, 154K/164Q/204L,

25 154K/164Q/204L/205I, 154K/205I, 155T, 160I, 160I/164Q, 160I/164Q/184A, 164Q, 205I, 226R, 226R/244I, and 244I. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1218 and one or more residue differences as compared to SEQ ID NO: 1218, selected from L129I,

30 L129I/D143H, L129I/A154K/D164Q, L129I/A154K/D164Q/V205I, L129I/A154K/V205I, L129I/G160I, L129I/G160I/D164Q, L129I/G160I/V205I, L129I/D164Q/V205I, D143R, D143R/H150D/D155T/E226R/V244I, D143R/H150D/D155T/V244I, D143R/H150D/E226R, D143R/H150D/E226R/V244I/E343R, D143R/H150D/E226R/E343R, D143R/H150D/V244I, D143R/H150D/E343R, D143R/E152T, D143R/E152T/E226R, D143R/E152T/E226R/V244I,

35 D143R/E152T/E226R/E343R, D143R/E152T/E343R, D143R/D155T, D143R/D155T/E226R/V244I, D143R/D155T/E226R/V244I/E343R, D143R/D155T/E343R, D143R/E226R, D143R/E226R/V244I/E343R, D143R/E226R/E343R, D143R/V244I, D143R/V244I/E343R, D143R/E343R, M146L/E152T/V244I, H150D/D155T/E226R, H150D/D155T/V244I, H150D/D155T/V244I/E343R, H150D/E226R/V244I/E343R, E152T, E152T/G174R/V244I/E343R, E152T/E226R, E152T/E226R/E343R, E152T/E343R, A154K,

A154K/G160I/C184A, A154K/D164Q, A154K/D164Q/F204L, A154K/D164Q/F204L/V205I, A154K/V205I, D155T, G160I, G160I/D164Q, G160I/D164Q/C184A, D164Q, V205I, E226R, E226R/V244I, and V244I.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 98/129/157/168, 129/157/168, 151, 151/152/255, 151/152/273, 151/255, 151/343, 157, 157/164, 157/164/168, 157/168, 164, 164/255, 168, 183, 255, 255/343, and 343. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 98I/129I/157R/168K, 129I/157R/168K, 151Q, 151Q/152T/255R, 151Q/152T/273Y, 151Q/255R, 151R/343R, 157R, 157R/164D, 157R/164R, 157R/164R/168K, 157R/168K, 164D, 164D/255R, 168K, 183I, 255R, 255R/343R, and 343R. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from L98I/L129I/K157R/Q168K, L129I/K157R/Q168K, P151Q, P151Q/E152T/A255R, P151Q/E152T/D273Y, P151Q/A255R, P151R/E343R, K157R, K157R/Q164D, K157R/Q164R, K157R/Q164R/Q168K, K157R/Q168K, Q164D, Q164D/A255R, Q168K, V183I, A255R, A255R/E343R, and E343R.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 129, 129/157, 129/157/168, 129/164, 129/164/168/183, 129/164/183, 143, 143/151, 143/151/152, 143/151/164, 143/164, 143/164/255, 143/164/343, 143/255, 143/343, 151, 151/152, 151/152/164/343, 151/152/343, 151/255, 151/255/343, 151/343, 152, 152/255, 157, 157/164/168, 157/168, 164, 164/168, 164/255, 164/447, 168, 183, 255, 255/343, 298/343, and 343. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 129I, 129I/157R, 129I/157R/168K, 129I/164R, 129I/164R/168K/183M, 129I/164R/183I, 143R, 143R/151Q, 143R/151R, 143R/151R/152T, 143R/151R/164D, 143R/164D, 143R/164D/255R, 143R/164D/343R, 143R/255R, 143R/343R, 151R, 151R/152T, 151R/152T/164D/343R, 151R/152T/343R, 151R/255R, 151R/255R/343R, 151R/343R, 152T, 152T/255R, 157R, 157R/164R/168K, 157R/168K, 164D, 164D/255R, 164R, 164R/168K, 164R/447T, 168K, 183I, 255R, 255R/343R, 298I/343R, and 343R. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference

sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from L129I, L129I/K157R, L129I/K157R/Q168K, L129I/Q164R, L129I/Q164R/Q168K/V183M, L129I/Q164R/V183I, D143R, D143R/P151Q, D143R/P151R, D143R/P151R/E152T, D143R/P151R/Q164D, D143R/Q164D, D143R/Q164D/A255R, D143R/Q164D/E343R, D143R/A255R, D143R/E343R, P151R, P151R/E152T, P151R/E152T/Q164D/E343R, P151R/E152T/E343R, P151R/A255R, P151R/A255R/E343R, P151R/E343R, E152T, E152T/A255R, K157R, K157R/Q164R/Q168K, K157R/Q168K, Q164D, Q164D/A255R, Q164R, Q164R/Q168K, Q164R/A447T, Q168K, V183I, A255R, A255R/E343R, V298I/E343R, and E343R.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 69, 69/164, 72, 73, 91, 92/164, 95, 96, 97, 98, 110, 112, 113, 121, 121/164, 122, 122/164, 124, 126, 127, 139, 140, 141, 144, 144/164, 145, 146, 147, 148, 150, 150/164, 151, 151/164, 153, 153/164, 155, 156, 157, 159, 160, 161, 163, 164, 164/172, 164/173, 164/182, 164/188, 164/191, 164/192, 164/195, 164/197, 164/199, 164/202, 164/205/249, 164/209, 165, 166, 167, 169, 170, 171, 172, 173, 177, 184, 187, 188, 190, 191, 192, 193, 194, 195, 197, 197/266, 198, 199, 200, 202, 203, 204, 205, 208, 209, 210, 321, 324, 331, and 349. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 69Q, 69V/164D, 72L, 72Q, 73Q, 73W, 91C, 91L, 91V, 92L/164D, 95G, 95S, 95V, 96G, 97T, 98V, 110R, 110T, 110V, 112A, 112H, 113V, 121A, 121L/164D, 121Q, 122C, 122F, 122L, 122V, 122V/164D, 124L, 124M, 126A, 126S, 127A, 127I, 139I, 140L, 140V, 141T, 144F, 144L, 144L/164D, 144V, 145F, 145G, 145K, 145L, 145Q, 145R, 145S, 145T, 146A, 146F, 146G, 146K, 146Q, 146R, 146V, 147W, 148A, 148E, 148F, 148G, 148R, 148V, 150A, 150A/164D, 150G, 150L, 150L/164D, 150R, 150R/164D, 150S/164D, 150T, 150V, 150W, 151A, 151C, 151E, 151G, 151I, 151K, 151L, 151L/164D, 151M, 151N, 151Q, 151S, 151T, 151V, 151W, 151Y, 151Y/164D, 153G, 153G/164D, 153N, 153P, 153R, 155F, 156C, 156G, 156L, 156Y, 157A, 157C, 157H, 157P, 157Q, 157R, 157S, 157V, 159A, 159G, 159P, 159Q, 159T, 160A, 160L, 160Q, 160S, 160V, 160Y, 161A, 161C, 161F, 161L, 161M, 161T, 163D, 163G, 163K, 163N, 163Q, 163W, 164D, 164D/172M, 164D/173M, 164D/182W, 164D/188G, 164D/191R, 164D/192E, 164D/195R, 164D/197A, 164D/199G, 164D/199L, 164D/202A, 164D/205L/249E, 164D/209R, 165H, 165T, 165V, 166M, 166R, 166V, 166W, 167L, 167R, 167V, 169I, 169S, 169T, 169Y, 170L, 171I, 171R, 171S, 172I, 172S, 172T, 172V, 173I, 177G, 177R, 177S, 177T, 177V, 177W, 177Y, 184E, 184G, 184Q, 184S, 184T, 187G, 188G, 190D, 190G, 190M, 190W, 191G, 191L, 191W, 192A, 192L, 193W, 194C, 194R, 194W, 195G, 195H, 195W, 197A, 197L/266V, 197M, 197S, 198A, 198G, 198I, 198K, 198L, 198N, 198R, 199G, 199L, 199W, 199Y, 200A, 200F, 200L, 200R, 200V, 200Y, 202F, 202I, 202L, 202S, 202V, 203Q, 204I, 204T, 204V, 205T, 208C, 209A, 209G, 209V, 210I, 210P, 210R, 210T, 210V, 321A, 321S, 324M, 331L, 349K, and 349L. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,

- 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from E69Q, E69V/Q164D, R72L, R72Q, K73Q, K73W, A91C, A91L, A91V, I92L/Q164D, T95G, T95S, T95V, T96G, S97T, L98V, K110R, K110T, K110V, G112A, G112H, S113V, W121A, W121L/Q164D, W121Q, W122C, W122F, W122L, W122V,
- 5 W122V/Q164D, H124L, H124M, T126A, T126S, L127A, L127I, L139I, I140L, I140V, H141T, W144F, W144L, W144L/Q164D, W144V, W145F, W145G, W145K, W145L, W145Q, W145R, W145S, W145T, M146A, M146F, M146G, M146K, M146Q, M146R, M146V, R147W, P148A, P148E, P148F, P148G, P148R, P148V, H150A, H150A/Q164D, H150G, H150L, H150L/Q164D, H150R, H150R/Q164D, H150S/Q164D, H150T, H150V, H150W, P151A, P151C, P151E, P151G, P151I, P151K, P151L,
- 10 P151L/Q164D, P151M, P151N, P151Q, P151S, P151T, P151V, P151W, P151Y, P151Y/Q164D, D153G, D153G/Q164D, D153N, D153P, D153R, D155F, E156C, E156G, E156L, E156Y, K157A, K157C, K157H, K157P, K157Q, K157R, K157S, K157V, S159A, S159G, S159P, S159Q, S159T, G160A, G160L, G160Q, G160S, G160V, G160Y, K161A, K161C, K161F, K161L, K161M, K161T, A163D, A163G, A163K, A163N, A163Q, A163W, Q164D, Q164D/R172M, Q164D/L173M, Q164D/S182W, Q164D/A188G, Q164D/Q191R,
- 15 Q164D/S192E, Q164D/Q195R, Q164D/K197A, Q164D/S199G, Q164D/S199L, Q164D/E202A, Q164D/V205L/K249E, Q164D/K209R, A165H, A165T, A165V, S166M, S166R, S166V, S166W, H167L, H167R, H167V, C169I, C169S, C169T, C169Y, R170L, Q171I, Q171R, Q171S, R172I, R172S, R172T, R172V, L173I, M177G, M177R, M177S, M177T, M177V, M177W, M177Y, C184E, C184G, C184Q, C184S, C184T, E187G, A188G, I190D, I190G, I190M, I190W, Q191G, Q191L, Q191W, S192A, S192L,
- 20 Y193W, L194C, L194R, L194W, Q195G, Q195H, Q195W, K197A, K197L/A266V, K197M, K197S, Q198A, Q198G, Q198I, Q198K, Q198L, Q198N, Q198R, S199G, S199L, S199W, S199Y, H200A, H200F, H200L, H200R, H200V, H200Y, E202F, E202I, E202L, E202S, E202V, R203Q, F204I, F204T, F204V, V205T, A208C, K209A, K209G, K209V, H210I, H210P, H210R, H210T, H210V, R321A, R321S, V324M, W331L, E349K, and E349L.
- 25 In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 69, 69/164, 72, 73, 91, 92/164, 95, 97, 98, 121/164, 126, 127, 139, 140, 142, 144, 144/164,
- 30 145, 146, 146/164, 148, 150, 150/164, 151, 153, 153/164, 155, 156, 157, 160, 163, 164/173, 164/188, 164/191, 164/192, 164/195, 164/199, 164/205/249, 164/209, 165, 166, 167, 169, 171, 172, 173, 177, 182, 187, 188, 190, 191, 192, 194, 195, 197/266, 198, 199, 200, 202, 203, 204, 205, 208, 209, 210, 324, 331, and 349. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%,
- 35 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 69Q, 69V, 69V/164D, 72L, 72Q, 73Q, 73W, 91C, 91V, 92L/164D, 95S, 97T, 98M, 98V, 121L/164D, 126A, 126S, 127I, 139I, 140L, 140V, 142C, 144F, 144L/164D, 144V, 144Y, 145F, 145K, 145L, 146A, 146F, 146K, 146L/164D, 146Q, 146R, 146V, 148R, 150A, 150A/164D, 150G, 150L, 150L/164D, 150R, 150R/164D, 150S/164D, 150T,
- 40 150V, 150W, 151C, 151E, 151K, 151M, 151N, 151Q, 151S, 151V, 151W, 151Y, 153G, 153G/164D,

153N, 153R, 155C, 155F, 155L, 155S, 155T, 156C, 156G, 156L, 156R, 156S, 156T, 156Y, 157A, 157C, 157P, 157Q, 157S, 157V, 160A, 160L, 160Q, 160S, 160V, 160Y, 163D, 163G, 163K, 163N, 163Q, 163W, 164D/173M, 164D/188G, 164D/191R, 164D/192E, 164D/195R, 164D/199G, 164D/199L, 164D/205L/249E, 164D/209R, 165G, 165T, 166M, 166V, 167L, 167R, 167V, 169I, 169S, 169T, 171I, 171R, 171S, 171V, 172I, 172T, 172V, 173I, 177G, 177R, 177S, 177T, 177V, 177W, 177Y, 182L, 187G, 188G, 190M, 191F, 191G, 192A, 192L, 194W, 195G, 195H, 195M, 195W, 195Y, 197L/266V, 198A, 198L, 198N, 198R, 199F, 199G, 199L, 199W, 199Y, 200A, 200F, 200L, 200R, 200S, 200V, 200Y, 202L, 203Q, 204I, 204T, 204V, 205T, 208S, 209V, 210T, 210V, 324M, 331L, and 349K. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from E69Q, E69V, E69V/Q164D, R72L, R72Q, K73Q, K73W, A91C, A91V, I92L/Q164D, T95S, S97T, L98M, L98V, W121L/Q164D, T126A, T126S, L127I, L139I, I140L, I140V, Q142C, W144F, W144L/Q164D, W144V, W144Y, W145F, W145K, W145L, M146A, M146F, M146K, M146L/Q164D, M146Q, M146R, M146V, P148R, H150A, H150A/Q164D, H150G, H150L, H150L/Q164D, H150R, H150R/Q164D, H150S/Q164D, H150T, H150V, H150W, P151C, P151E, P151K, P151M, P151N, P151Q, P151S, P151V, P151W, P151Y, D153G, D153G/Q164D, D153N, D153R, D155C, D155F, D155L, D155S, D155T, E156C, E156G, E156L, E156R, E156S, E156T, E156Y, K157A, K157C, K157P, K157Q, K157S, K157V, G160A, G160L, G160Q, G160S, G160V, G160Y, A163D, A163G, A163K, A163N, A163Q, A163W, Q164D/L173M, Q164D/A188G, Q164D/Q191R, Q164D/S192E, Q164D/Q195R, Q164D/S199G, Q164D/S199L, Q164D/V205L/K249E, Q164D/K209R, A165G, A165T, S166M, S166V, H167L, H167R, H167V, C169I, C169S, C169T, Q171I, Q171R, Q171S, Q171V, R172I, R172T, R172V, L173I, M177G, M177R, M177S, M177T, M177V, M177W, M177Y, S182L, E187G, A188G, I190M, Q191F, Q191G, S192A, S192L, L194W, Q195G, Q195H, Q195M, Q195W, Q195Y, K197L/A266V, Q198A, Q198L, Q198N, Q198R, S199F, S199G, S199L, S199W, S199Y, H200A, H200F, H200L, H200R, H200S, H200V, H200Y, E202L, R203Q, F204I, F204T, F204V, V205T, A208S, K209V, H210T, H210V, V324M, W331L, and E349K.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 73, 91, 93, 95, 97, 98, 110, 112, 122, 122/164, 124, 127, 139, 144, 145, 146, 146/164, 147, 148, 150, 150/164, 151, 151/164, 153, 153/164, 155, 156, 157, 159, 160, 161, 162, 163, 164, 164/182, 164/188, 164/191, 164/192, 164/195, 164/199, 164/205/249, 164/209, 165, 166, 167, 169, 171, 172, 177, 184, 187, 188, 190, 191, 192, 193, 195, 197/266, 198, 199, 200, 204, 205, 206, 207, 208, 209, 210, and 324. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 73Q, 73W, 91C, 91L, 93S, 95G, 95R, 95S, 95V, 97T, 98V, 110T, 110V, 112A, 112H, 122C, 122F, 122L, 122V, 122V/164D, 124A, 124L,

124M, 127I, 139I, 144F, 144Y, 145F, 145L, 145Q, 145R, 145S, 145T, 146A, 146F, 146G, 146K, 146L/164D, 146Q, 146R, 146V, 147W, 148E, 148F, 148G, 148R, 148V, 150A, 150G, 150G/164D, 150L, 150R/164D, 150S/164D, 150T, 150V, 150W, 151A, 151C, 151E, 151G, 151I, 151K, 151L, 151L/164D, 151M, 151N, 151Q, 151S, 151T, 151V, 151W, 151Y, 151Y/164D, 153G, 153G/164D, 153P, 153Q, 153R, 155L, 155S, 155W, 156G, 156S, 156Y, 157A, 157C, 157H, 157P, 157R, 157S, 157V, 159A, 159G, 159P, 159Q, 159T, 160A, 160L, 160M, 160S, 160V, 160Y, 161A, 161C, 161F, 161L, 161M, 161T, 162Y, 163D, 163G, 163W, 164D, 164D/182W, 164D/188G, 164D/191R, 164D/192E, 164D/195R, 164D/199G, 164D/199L, 164D/205L/249E, 164D/209R, 165G, 165T, 165V, 166M, 166V, 166W, 167L, 169I, 169T, 171R, 171S, 171V, 172I, 177W, 184Q, 184T, 187G, 188G, 190D, 190G, 190M, 190W, 191F, 191L, 191W, 192A, 193W, 195H, 195R, 195Y, 197L/266V, 198A, 198L, 198N, 198R, 198V, 199G, 199L, 199W, 199Y, 200R, 204L, 205T, 206L, 207K, 207Q, 208S, 208V, 209A, 209G, 209V, 210P, 210R, 210T, 210V, and 324M. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from K73Q, K73W, A91C, A91L, D93S, T95G, T95R, T95S, T95V, S97T, L98V, K110T, K110V, G112A, G112H, W122C, W122F, W122L, W122V, W122V/Q164D, H124A, H124L, H124M, L127I, L139I, W144F, W144Y, W145F, W145L, W145Q, W145R, W145S, W145T, M146A, M146F, M146G, M146K, M146L/Q164D, M146Q, M146R, M146V, R147W, P148E, P148F, P148G, P148R, P148V, H150A, H150G, H150G/Q164D, H150L, H150R/Q164D, H150S/Q164D, H150T, H150V, H150W, P151A, P151C, P151E, P151G, P151I, P151K, P151L, P151L/Q164D, P151M, P151N, P151Q, P151S, P151T, P151V, P151W, P151Y, P151Y/Q164D, D153G, D153G/Q164D, D153P, D153Q, D153R, D155L, D155S, D155W, E156G, E156S, E156Y, K157A, K157C, K157H, K157P, K157R, K157S, K157V, S159A, S159G, S159P, S159Q, S159T, G160A, G160L, G160M, G160S, G160V, G160Y, K161A, K161C, K161F, K161L, K161M, K161T, W162Y, A163D, A163G, A163W, Q164D, Q164D/S182W, Q164D/A188G, Q164D/Q191R, Q164D/S192E, Q164D/Q195R, Q164D/S199G, Q164D/S199L, Q164D/V205L/K249E, Q164D/K209R, A165G, A165T, A165V, S166M, S166V, S166W, H167L, C169I, C169T, Q171R, Q171S, Q171V, R172I, M177W, C184Q, C184T, E187G, A188G, I190D, I190G, I190M, I190W, Q191F, Q191L, Q191W, S192A, Y193W, Q195H, Q195R, Q195Y, K197L/A266V, Q198A, Q198L, Q198N, Q198R, Q198V, S199G, S199L, S199W, S199Y, H200R, F204L, V205T, V206L, R207K, R207Q, A208S, A208V, K209A, K209G, K209V, H210P, H210R, H210T, H210V, and V324M.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 69, 73, 91, 127, 140, 144, 145, 146, 146/164, 150, 150/164, 151, 153, 155, 156, 157, 160, 163, 164, 164/172, 164/188, 164/191, 164/192, 164/195, 164/199, 164/209, 165, 167, 169, 171, 172, 182, 183, 188, 191, 192, 193, 195, 196, 198, 199, 200, 204, 205, 208, 210, and 324. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more

sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from 69G, 69V, 73Q, 73W, 91G, 127I, 140V, 144F, 144Y, 145R, 146A, 146F, 146K, 146L/164D, 146Q, 146R, 150G, 150L, 150R, 150R/164D, 150S/164D, 150W, 151K, 151N, 153G, 153Q, 153R, 155C, 155F, 155L, 155M, 155P, 155S, 155T, 155W, 156C, 156G, 156R, 156S, 156T, 156W, 156Y, 157A, 157P, 157R, 157S, 157V, 160A, 160L, 160M, 160Q, 160R, 160S, 160V, 160Y, 163D, 163K, 163N, 163W, 164D, 164D/172M, 164D/188G, 164D/191R, 164D/192E, 164D/195R, 164D/199G, 164D/199L, 164D/209R, 165G, 167L, 167R, 169A, 169I, 169T, 171H, 171R, 171S, 171V, 172I, 182T, 183L, 188G, 191F, 191G, 191W, 192A, 192L, 192R, 193W, 195G, 195H, 195M, 195R, 195W, 195Y, 196C, 196G, 196L, 196W, 198A, 198L, 198N, 198R, 198V, 199L, 199W, 199Y, 200R, 200Y, 204L, 205T, 208S, 210V, and 324M. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 1580 and one or more residue differences as compared to SEQ ID NO: 1580, selected from E69G, E69V, K73Q, K73W, A91G, L127I, I140V, W144F, W144Y, W145R, M146A, M146F, M146K, M146L/Q164D, M146Q, M146R, H150G, H150L, H150R, H150R/Q164D, H150S/Q164D, H150W, P151K, P151N, D153G, D153Q, D153R, D155C, D155F, D155L, D155M, D155P, D155S, D155T, D155W, E156C, E156G, E156R, E156S, E156T, E156W, E156Y, K157A, K157P, K157R, K157S, K157V, G160A, G160L, G160M, G160Q, G160R, G160S, G160V, G160Y, A163D, A163K, A163N, A163W, Q164D, Q164D/R172M, Q164D/A188G, Q164D/Q191R, Q164D/S192E, Q164D/Q195R, Q164D/S199G, Q164D/S199L, Q164D/K209R, A165G, H167L, H167R, C169A, C169I, C169T, Q171H, Q171R, Q171S, Q171V, R172I, S182T, V183L, A188G, Q191F, Q191G, Q191W, S192A, S192L, S192R, Y193W, Q195G, Q195H, Q195M, Q195R, Q195W, Q195Y, D196C, D196G, D196L, D196W, Q198A, Q198L, Q198N, Q198R, Q198V, S199L, S199W, S199Y, H200R, H200Y, F204L, V205T, A208S, H210V, and V324M.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from 38/143/168/220/290/317/423, 38/152/154/164/168/220/290/317/423, 38/152/168/220/290/317/343/423, 38/152/168/220/290/317/423, 38/154/164/168/220/290/317/423, 38/168/220/290/317/423, and 38/220/317. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972, selected from 38F/143R/168Q/220R/290R/317M/423D, 38F/152T/154K/164R/168K/220R/290R/317M/423D, 38F/152T/168Q/220R/290R/317M/343R/423D, 38F/152T/168Q/220R/290R/317M/423D, 38F/154K/164Q/168Q/220R/290R/317M/423D, 38F/168Q/220R/290R/317M/423D, and 38F/220R/317M. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO:

972, selected from C38F/D143R/F168Q/P220R/K290R/R317M/W423D,
 C38F/E152T/A154K/D164R/F168K/P220R/K290R/R317M/W423D,
 C38F/E152T/F168Q/P220R/K290R/R317M/E343R/W423D,
 C38F/E152T/F168Q/P220R/K290R/R317M/W423D,

- 5 C38F/A154K/D164Q/F168Q/P220R/K290R/R317M/W423D, C38F/F168Q/P220R/K290R/R317M/W423D,
 and C38F/P220R/R317M.

- In some embodiments, the engineered transposase polypeptide with one or more improved
 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
 10 sequence SEQ ID NO: 2368 and one or more residue differences as compared to SEQ ID NO: 2368,
 selected from 73/143/169/191/193, 73/156, 73/156/192/435, 73/156/210, 73/163, 73/163/210, 73/192/193,
 127, 127/160/172/198, 127/172, 127/198, 146, 146/151/188, 146/324, 151/171/324, 151/324, 156/169,
 156/192, 157/171, 163, 163/169, 163/169/191/192/193, 165/171/324, 165/188/324, 165/324, 169,
 171/188/324, 171/324, 172, 188, 192, 192/193, 198, 210, and 324. In some embodiments, the engineered
 15 transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%,
 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to
 reference sequence SEQ ID NO: 2368 and one or more residue differences as compared to SEQ ID NO:
 2368, selected from 73Q/143K/169I/191F/193W, 73Q/156R, 73Q/156R/210V, 73Q/156Y/192A/435K,
 73Q/163D, 73Q/163D/210Q, 73Q/163N, 73Q/163N/210V, 73Q/192A/193W, 127I, 127I/160A/172I/198R,
 20 127I/172I, 127I/198R, 146L, 146L/151N/188G, 146L/324M, 151N/171V/324M, 151N/324M, 156Y/169I,
 156Y/192E, 157P/171V, 163D, 163D/169I/191F/192A/193W, 163N/169I, 165G/171R/324M,
 165G/188G/324M, 165G/324M, 169I, 171R/324M, 171V/188G/324M, 171V/324M, 172I, 188G, 192A,
 192E/193W, 198R, 210V, and 324M. In some embodiments, the engineered transposase polypeptide
 comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%,
 25 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ
 ID NO: 2368 and one or more residue differences as compared to SEQ ID NO: 2368, selected from
 K73Q/D143K/C169I/Q191F/Y193W, K73Q/E156R, K73Q/E156R/H210V, K73Q/E156Y/S192A/M435K,
 K73Q/A163D, K73Q/A163D/H210Q, K73Q/A163N, K73Q/A163N/H210V, K73Q/S192A/Y193W, L127I,
 L127I/G160A/R172I/Q198R, L127I/R172I, L127I/Q198R, M146L, M146L/P151N/A188G, M146L/V324M,
 30 P151N/Q171V/V324M, P151N/V324M, E156Y/C169I, E156Y/S192E, K157P/Q171V, A163D,
 A163D/C169I/Q191F/S192A/Y193W, A163N/C169I, A165G/Q171R/V324M, A165G/A188G/V324M,
 A165G/V324M, C169I, Q171R/V324M, Q171V/A188G/V324M, Q171V/V324M, R172I, A188G, S192A,
 S192E/Y193W, Q198R, H210V, and V324M.

- In some embodiments, the engineered transposase polypeptide with one or more improved
 35 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
 sequence SEQ ID NO: 2368 and one or more residue differences as compared to SEQ ID NO: 2368,
 selected from 73, 73/163/191/192/193, 127/160/172, 127/160/172/198, 144, 160, 160/172, 163/191/192,
 and 192. In some embodiments, the engineered transposase polypeptide comprises an amino acid
 40 sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%,

- 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2368 and one or more residue differences as compared to SEQ ID NO: 2368, selected from 73Q, 73Q/163N/191F/192A/193W, 127I/160A/172I, 127I/160A/172I/198R, 144Y, 160A, 160A/172I, 163N/191F/192A, and 192A. In some embodiments, the engineered transposase polypeptide comprises
- 5 an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2368 and one or more residue differences as compared to SEQ ID NO: 2368, selected from K73Q, K73Q/A163N/Q191F/S192A/Y193W, L127I/G160A/R172I, L127I/G160A/R172I/Q198R, W144Y, G160A, G160A/R172I, A163N/Q191F/S192A, and S192A.
- 10 In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2426 and one or more residue differences as compared to SEQ ID NO: 2426, selected from 8, 13, 17, 18, 20/146/157/167, 26, 34/38, 38/39/45, 67, 70, 73, 73/150/155/169,
- 15 73/150/171/192, 73/155, 73/155/169/199, 73/155/171, 73/192, 73/199, 78, 82, 84, 86, 86/384, 89, 90, 91, 95, 102, 110, 111, 113, 114, 115, 122/127, 126/127, 127/131/165, 127/133, 143, 143/146, 143/146/153, 143/146/157, 143/146/167, 143/146/167/188, 143/146/188, 143/153/167/188/246, 143/156/157/167/195, 143/156/188, 143/188, 143/188/195, 143/195, 144, 145, 145/151, 146, 146/152, 146/153/157, 146/153/157/167, 146/153/167, 146/157, 146/157/167, 146/167, 150, 150/155/169, 150/155/169/192,
- 20 150/155/192, 150/169, 150/192, 150/199, 151/152/154, 152/154, 152/154/155/160, 153/157/167, 155/171, 155/192, 156/270, 160/371, 167/188, 168/172, 169, 184, 190, 192/198, 198, 206, 208, 209, 210, 214/220, 215/220, 216/220, 217/220, 219/220, 220, 220/222, 220/225, 220/226, 231, 235, 242, 257, 259, 261, 262, 267, 268, 269, 270, 272, 274, 275, 276, 277, 278, 282, 283/290, 284/290, 287/290, 288/290, 289/290, 290, 290/291, 290/292, 290/293, 290/294, 304, 305, 305/373, 306, 307, 309, 317, 329, 333,
- 25 334, 339, 341, 344, 345, 346, 352, 353, 355, 358, 366, 368, 370, 372, 373, 375, 376, 378, 379, 380, 381, 382, 383, 384, 385, 386/393, 387/393, 389/393, 390/393, 391/393, 392/393, 393/394, 414, 415, 420/423, 422/423, 423, 423/425, 432, 440, 441, 443, 469, and 472. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to
- 30 reference sequence SEQ ID NO: 2426 and one or more residue differences as compared to SEQ ID NO: 2426, selected from 8D, 8V, 13W, 17V, 18D, 18L, 18P, 20D/146C/157P/167L, 26R, 34H/38C, 38C/39F/45I, 38C/39P/45I, 67G, 70G, 73H, 73Q/150N/155L/169T, 73Q/150N/171R/192A, 73Q/155L, 73Q/155L/169T/199L, 73Q/155L/171R, 73Q/192A, 73Q/199L, 73R, 78W, 82Q, 82R, 84T, 86F/384T, 86I, 89M, 90W, 91L, 91V, 95V, 102G, 110L, 111M, 113A, 113G, 113V, 114D, 114T, 115Q, 122A/127L,
- 35 126V/127L, 127L/131A/165T, 127L/133M, 143T, 143V, 143V/146C, 143V/146C/153G, 143V/146C/157P, 143V/146C/167L, 143V/146C/167L/188G, 143V/146C/188G, 143V/146K, 143V/153G/167L/188G/246R, 143V/156R/157P/167L/195H, 143V/156R/188G, 143V/188G, 143V/188G/195H, 143V/195H, 144F, 145L, 145V/151R, 146A/152A, 146C, 146C/153G/157P, 146C/153G/157P/167L, 146C/153G/167L, 146C/157P, 146C/157P/167L, 146C/167L, 146K, 146K/157P, 150N, 150N/155L/169T, 150N/155L/169T/192A,
- 40 150N/155L/192A, 150N/169T, 150N/192A, 150N/199L, 151E/152E/154A, 151R/152E/154A,

151T/152E/154A, 152E/154A, 152E/154A/155N/160G, 153G/157P/167L, 155L/171R, 155L/192A, 156G/270P, 160G/371W, 167L/188G, 168F/172V, 169T, 184A, 190G, 190S, 192L/198Q, 198D, 198G, 206F, 206W, 208V, 209G, 209M, 210G, 210L, 214G/220P, 214V/220P, 215F/220P, 216I/220P, 217D/220P, 217G/220P, 217L/220P, 217R/220P, 219G/220P, 220L, 220P/222C, 220P/222H, 220P/222R, 220P/225M, 220P/226R, 220P/226S, 231G, 235S, 242S, 257G, 259H, 261K, 262L, 267L, 268W, 269A, 269E, 269G, 270D, 270L, 270P, 272G, 272R, 272T, 274K, 274R, 275R, 275W, 276A, 276E, 276G, 277V, 278V, 282M, 283S/290K, 284A/290K, 284S/290K, 287P/290K, 288M/290K, 288P/290K, 288R/290K, 288V/290K, 289S/290K, 290C, 290K, 290K/291E, 290K/291G, 290K/291S, 290K/292A, 290K/293L, 290K/294K, 290K/294M, 290K/294R, 290K/294S, 290K/294W, 290Q, 290V, 304E, 304V, 305G/373C, 305R, 306G, 307S, 309V, 317R, 329A, 329R, 329T, 333L, 334L, 339G, 341L, 344G, 345T, 346T, 346V, 352L, 353V, 355C, 358C, 366S, 368W, 370S, 372G, 373T, 375Y, 376L, 378A, 378G, 378P, 378S, 379P, 380T, 380V, 381S, 382L, 383F, 383R, 384N, 384V, 385L, 386A/393E, 386S/393E, 386T/393E, 386V/393E, 387A/393E, 387R/393E, 389L/393E, 389M/393E, 389V/393E, 390E/393E, 390G/393E, 391L/393E, 392P/393E, 393E/394R, 414G, 415P, 415V, 420E/423W, 422G/423W, 422S/423W, 422V/423W, 423A, 423W/425A, 432T, 440G, 441Q, 441S, 441W, 443P, 469A, 472A, and 472R. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2426 and one or more residue differences as compared to SEQ ID NO: 2426, selected from Q8D, Q8V, H13W, A17V, A18D, A18L, A18P, G20D/M146C/K157P/H167L, K26R, D34H/F38C, F38C/S39F/V45I, F38C/S39P/V45I, S67G, V70G, K73H, K73Q/H150N/D155L/C169T, K73Q/H150N/Q171R/S192A, K73Q/D155L, K73Q/D155L/C169T/S199L, K73Q/D155L/Q171R, K73Q/S192A, K73Q/S199L, K73R, H78W, L82Q, L82R, Q84T, Y86F/A384T, Y86I, I89M, L90W, A91L, A91V, T95V, H102G, K110L, L111M, S113A, S113G, S113V, V114D, V114T, K115Q, W122A/I127L, T126V/I127L, I127L/G131A/A165T, I127L/T133M, D143T, D143V, D143V/M146C, D143V/M146C/D153G, D143V/M146C/K157P, D143V/M146C/H167L, D143V/M146C/H167L/A188G, D143V/M146C/A188G, D143V/M146K, D143V/D153G/H167L/A188G/S246R, D143V/E156R/K157P/H167L/Q195H, D143V/E156R/A188G, D143V/A188G, D143V/A188G/Q195H, D143V/Q195H, W144F, W145L, W145V/P151R, M146A/T152A, M146C, M146C/D153G/K157P, M146C/D153G/K157P/H167L, M146C/D153G/H167L, M146C/K157P, M146C/K157P/H167L, M146C/H167L, M146K, M146K/K157P, H150N, H150N/D155L/C169T, H150N/D155L/C169T/S192A, H150N/D155L/S192A, H150N/C169T, H150N/S192A, H150N/S199L, P151E/T152E/K154A, P151R/T152E/K154A, P151T/T152E/K154A, T152E/K154A, T152E/K154A/D155N/A160G, D153G/K157P/H167L, D155L/Q171R, D155L/S192A, E156G/K270P, A160G/P371W, H167L/A188G, K168F/I172V, C169T, C184A, I190G, I190S, S192L/R198Q, R198D, R198G, V206F, V206W, A208V, K209G, K209M, H210G, H210L, I214G/R220P, I214V/R220P, L215F/R220P, D216I/R220P, S217D/R220P, S217G/R220P, S217L/R220P, S217R/R220P, L219G/R220P, R220L, R220P/F222C, R220P/F222H, R220P/F222R, R220P/L225M, R220P/E226R, R220P/E226S, L231G, T235S, G242S, K257G, S259H, S261K, V262L, V267L, T268W, L269A, L269E, L269G, K270D, K270L, K270P, K272G, K272R, K272T, E274K, E274R, M275R, M275W, Q276A, Q276E, Q276G, P277V, I278V, Y282M, A283S/R290K, E284A/R290K, E284S/R290K, N287P/R290K,

S288M/R290K, S288P/R290K, S288R/R290K, S288V/R290K, P289S/R290K, R290C, R290K, R290K/D291E, R290K/D291G, R290K/D291S, R290K/M292A, R290K/E293L, R290K/P294K, R290K/P294M, R290K/P294R, R290K/P294S, R290K/P294W, R290Q, R290V, P304E, P304V, I305G/Y373C, I305R, E306G, T307S, S309V, M317R, K329A, K329R, K329T, T333L, G334L, R339G, R341L, E344G, P345T, R346T, R346V, V352L, S353V, L355C, I358C, R366S, A368W, T370S, P372G, Y373T, L375Y, R376L, K378A, K378G, K378P, K378S, G379P, L380T, L380V, V381S, D382L, A383F, A383R, A384N, A384V, N385L, E386A/A393E, E386S/A393E, E386T/A393E, E386V/A393E, L387A/A393E, L387R/A393E, S389L/A393E, S389M/A393E, S389V/A393E, Q390E/A393E, Q390G/A393E, S391L/A393E, C392P/A393E, A393E/T394R, N414G, G415P, G415V, S420E/D423W, K422G/D423W, K422S/D423W, K422V/D423W, D423A, D423W/Y425A, G432T, T440G, G441Q, G441S, G441W, A443P, L469A, G472A, and G472R.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2426 and one or more residue differences as compared to SEQ ID NO: 2426, selected from 8, 18, 25, 29, 30, 67, 73, 73/150/155/169, 73/150/155/171, 73/150/171/192, 73/171, 73/192, 73/199, 74, 82, 86, 86/384, 113, 115, 118, 127/128, 143, 150, 150/155/169, 150/155/169/192, 150/155/171, 150/155/171/199, 150/155/192, 150/155/199, 150/192, 150/199, 155/192, 168/171/172, 195/198, 197/198, 198/201, 198/203, 198/205, 235, 252, 256, 257, 262, 263, 267/343, 268, 269, 270, 272, 273, 276, 277, 288/290, 289/290, 290, 290/291, 290/292, 290/294, 298, 304, 305, 307, 333, 346, 355, 366, 370, 386/393, 389/393, 393/395, 440, and 472. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2426 and one or more residue differences as compared to SEQ ID NO: 2426, selected from 8D, 18L, 25G, 29F, 30R, 67G, 73H, 73Q/150N/155L/169T, 73Q/150N/155L/171R, 73Q/150N/171R/192A, 73Q/171R, 73Q/192A, 73Q/199L, 74R, 82F, 82Q, 86A, 86F/384T, 86I, 113E, 115A, 115Q, 118S, 127L/128I, 143H, 150N, 150N/155L/169T, 150N/155L/169T/192A, 150N/155L/171R, 150N/155L/171R/199L, 150N/155L/192A, 150N/155L/199L, 150N/192A, 150N/199L, 155L/192A, 168F/171P/172R, 195K/198Q, 195S/198Q, 197S/198Q, 198Q/201R, 198Q/203K, 198Q/205T, 235S, 252A, 256E, 256G, 257G, 262L, 263Q, 267P/343G, 268W, 269I, 270P, 272R, 273V, 276E, 277V, 288V/290K, 289S/290K, 290K/291E, 290K/292A, 290K/294S, 290Q, 290V, 298R, 304W, 305A, 305R, 307S, 333L, 346T, 355K, 366S, 370S, 386S/393E, 389M/393E, 393E/395R, 393E/395W, 440G, and 472R. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2426 and one or more residue differences as compared to SEQ ID NO: 2426, selected from Q8D, A18L, T25G, I29F, E30R, S67G, K73H, K73Q/H150N/D155L/C169T, K73Q/H150N/D155L/Q171R, K73Q/H150N/Q171R/S192A, K73Q/Q171R, K73Q/S192A, K73Q/S199L, A74R, L82F, L82Q, Y86A, Y86F/A384T, Y86I, S113E, K115A, K115Q, A118S, I127L/L128I, D143H, H150N, H150N/D155L/C169T, H150N/D155L/C169T/S192A, H150N/D155L/Q171R, H150N/D155L/Q171R/S199L,

H150N/D155L/S192A, H150N/D155L/S199L, H150N/S192A, H150N/S199L, D155L/S192A, K168F/Q171P/I172R, Q195K/R198Q, Q195S/R198Q, K197S/R198Q, R198Q/N201R, R198Q/R203K, R198Q/V205T, T235S, N252A, R256E, R256G, K257G, V262L, K263Q, V267P/E343G, T268W, L269I, K270P, K272R, D273V, Q276E, P277V, S288V/R290K, P289S/R290K, R290K/D291E, R290K/M292A, 5 R290K/P294S, R290Q, R290V, V298R, P304W, I305A, I305R, T307S, T333L, R346T, L355K, R366S, T370S, E386S/A393E, S389M/A393E, A393E/V395R, A393E/V395W, T440G, and G472R.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
10 sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540, selected from 66, 67, 70, 101, 102, 103, 104, 106, 107, 108, 113, 114, 115, 117, 118, 143, 152, 154, 235, 236/251, 237, 238, 242, 243, 246, 247, 249, 250, 251, 256/352, 257, 259, 285, 286, 287, 288, 289, 290, 292, 293, 294, 295, 296, 322, 323, 326, 329, 330, 333, 339, 340, 343, 344, 345, and 352. In some
15 embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540, selected from 66T, 67V, 70A, 101S, 101T, 102R, 102S, 103V, 104T, 106A, 107P, 107S, 107V, 108M, 113G, 113R, 114A, 114I, 114P, 114S, 115L, 115R, 117A, 118K, 143E, 143G, 143R, 143V, 152R, 154E, 154F, 154Q, 235R, 236V/251T, 237S, 238L, 242T, 243F, 20 243L, 246M, 247F, 249F, 249L, 249R, 249V, 249W, 250N, 251G, 251R, 251S, 256K/352A, 257L, 257R, 257V, 259H, 259P, 259Y, 285D, 286G, 286R, 287A, 287G, 287L, 287R, 287V, 288L, 288W, 289I, 289K, 289R, 289S, 290G, 290S, 292L, 292P, 292V, 293G, 294E, 294V, 295F, 295T, 295V, 296A, 296Q, 296R, 322G, 322Q, 322T, 323S, 323T, 326L, 326M, 329G, 329H, 330L, 330V, 333A, 333M, 333R, 333S, 339A, 339G, 339S, 340S, 343P, 343T, 344G, 344M, 345A, 345E, 345K, 345V, 352I, and 352L. In some
25 embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540, selected from V66T, S67V, V70A, K101S, K101T, H102R, H102S, Q103V, V104T, S106A, D107P, D107S, D107V, L108M, S113G, S113R, V114A, V114I, V114P, 30 V114S, K115L, K115R, K117A, A118K, D143E, D143G, D143R, D143V, T152R, K154E, K154F, K154Q, T235R, I236V/K251T, D237S, I238L, G242T, V243F, V243L, S246M, K247F, K249F, K249L, K249R, K249V, K249W, R250N, K251G, K251R, K251S, R256K/V352A, K257L, K257R, K257V, S259H, S259P, S259Y, E285D, T286G, T286R, N287A, N287G, N287L, N287R, N287V, S288L, S288W, P289I, P289K, P289R, P289S, R290G, R290S, M292L, M292P, M292V, E293G, P294E, P294V, L295F, L295T, L295V, 35 K296A, K296Q, K296R, W322G, W322Q, W322T, R323S, R323T, E326L, E326M, K329G, K329H, A330L, A330V, T333A, T333M, T333R, T333S, R339A, R339G, R339S, L340S, E343P, E343T, E344G, E344M, P345A, P345E, P345K, P345V, V352I, and V352L.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
40 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference

sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540, selected from 18, 18/113/270, 18/113/270/329/370, 18/113/270/370/384, 18/113/270/370/412, 18/113/329/384, 18/259/370/384, 18/270, 18/270/370, 18/270/370/384, 18/370, 18/370/384, 73, 73/115, 73/115/184/333, 73/115/184/341, 73/115/256, 73/184/256, 73/184/273/333, 73/184/333, 73/256, 73/256/375, 73/273, 73/273/333/375, 73/273/375, 73/333, 73/333/375, 73/375, 73/379, 77/233/238/376, 77/233/334/376, 77/262/376, 77/276, 77/334/376, 82, 82/127, 82/127/144/235/290, 82/127/144/266, 82/127/144/266/290, 82/127/144/378, 82/127/235, 82/127/290, 82/144/266/290/378, 82/144/266/378, 82/235/266, 82/235/266/378/383, 82/235/290, 82/235/378, 82/266, 82/266/378, 82/378, 113/270/370/384, 115/256/333/375, 115/256/375, 115/273/375, 115/375, 127/144/290, 127/149/150, 127/266/290/378, 127/378/383, 144/235/290, 144/266/290, 233/334, 233/376, 235, 235/290, 256/333, 262, 266/378, 270/370, 273, 276/352, 290, 290/378, 333, 358, 370, 375, and 378. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540, selected from 18D, 18D/113A/270L, 18D/113A/270L/329A/370S, 18D/259R/370S/384V, 18D/270L, 18D/270L/370S, 18D/270L/370S/384V, 18D/370S, 18L, 18L/113A/270L, 18L/113A/270L/370S/384V, 18L/113A/329A/384V, 18L/113E/270L/370S/412S, 18L/370S, 18L/370S/384V, 73H, 73H/115Q/184A/341G, 73H/184A/256G, 73H/273V/333L/375S, 73H/333L, 73R, 73R/115Q, 73R/115Q/184A/333L, 73R/115Q/256G, 73R/184A/273V/333L, 73R/184A/333L, 73R/256G, 73R/256G/375S, 73R/273V, 73R/273V/333L/375S, 73R/273V/375S, 73R/333L, 73R/333L/375S, 73R/375S, 73R/379V, 77V/233L/238V/376L, 77V/233L/334L/376L, 77V/262L/376L, 77V/276E, 77V/334L/376L, 82Q, 82Q/127L, 82Q/127L/144Y/235S/290V, 82Q/127L/144Y/266V, 82Q/127L/144Y/266V/290V, 82Q/127L/144Y/378S, 82Q/127L/235S, 82Q/127L/290Q, 82Q/144Y/266V/290Q/378S, 82Q/144Y/266V/378S, 82Q/235S/266V, 82Q/235S/266V/378S/383P, 82Q/235S/290V, 82Q/235S/378S, 82Q/266V, 82Q/266V/378S, 82Q/378S, 113A/270L/370S/384V, 115Q/256G/333L/375S, 115Q/256G/375S, 115Q/273V/375S, 115Q/375S, 127L/144Y/290V, 127L/149H/150Q, 127L/266V/290Q/378S, 127L/378S/383P, 144Y/235S/290Q, 144Y/266V/290Q, 233L/334L, 233L/376L, 235S, 235S/290Q, 256G/333L, 262L, 266V/378S, 270L/370S, 273V, 276E/352L, 290Q, 290V, 290V/378S, 333L, 358W, 370S, 375S, and 378S. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540, selected from A18D, A18D/S113A/K270L, A18D/S113A/K270L/K329A/T370S, A18D/S259R/T370S/A384V, A18D/K270L, A18D/K270L/T370S, A18D/K270L/T370S/A384V, A18D/T370S, A18L, A18L/S113A/K270L, A18L/S113A/K270L/T370S/A384V, A18L/S113A/K329A/A384V, A18L/S113E/K270L/T370S/G412S, A18L/T370S, A18L/T370S/A384V, Q73H, Q73H/K115Q/C184A/R341G, Q73H/C184A/R256G, Q73H/D273V/T333L/L375S, Q73H/T333L, Q73R, Q73R/K115Q, Q73R/K115Q/C184A/T333L, Q73R/K115Q/R256G, Q73R/C184A/D273V/T333L, Q73R/C184A/T333L, Q73R/R256G, Q73R/R256G/L375S, Q73R/D273V, Q73R/D273V/T333L/L375S, Q73R/D273V/L375S, Q73R/T333L, Q73R/T333L/L375S, Q73R/L375S, Q73R/G379V,

E77V/A233L/I238V/R376L, E77V/A233L/G334L/R376L, E77V/V262L/R376L, E77V/Q276E, E77V/G334L/R376L, L82Q, L82Q/I127L, L82Q/I127L/W144Y/T235S/R290V, L82Q/I127L/W144Y/A266V, L82Q/I127L/W144Y/A266V/R290V, L82Q/I127L/W144Y/K378S, L82Q/I127L/T235S, L82Q/I127L/R290Q, L82Q/W144Y/A266V/R290Q/K378S, L82Q/W144Y/A266V/K378S, L82Q/T235S/A266V, L82Q/T235S/A266V/K378S/A383P, L82Q/T235S/R290V, L82Q/T235S/K378S, L82Q/A266V, L82Q/A266V/K378S, L82Q/K378S, S113A/K270L/T370S/A384V, K115Q/R256G/T333L/L375S, K115Q/R256G/L375S, K115Q/D273V/L375S, K115Q/L375S, I127L/W144Y/R290V, I127L/D149H/H150Q, I127L/A266V/R290Q/K378S, I127L/K378S/A383P, W144Y/T235S/R290Q, W144Y/A266V/R290Q, A233L/G334L, A233L/R376L, T235S, T235S/R290Q, R256G/T333L, V262L, A266V/K378S, K270L/T370S, D273V, Q276E/V352L, R290Q, R290V, R290V/K378S, T333L, I358W, T370S, L375S, and K378S.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540, selected from 18, 18/113, 18/113/252, 18/113/252/336/384, 18/113/252/370, 18/113/270, 18/113/270/370/384, 18/259/370/384, 18/270, 18/270/370, 18/270/370/384, 18/370, 18/370/384, 73, 73/115/184/341, 73/115/341, 73/131/256/273/333/341/375, 73/184/256, 73/184/256/341, 73/184/256/375, 73/184/273/333, 73/184/333, 73/184/341/375, 73/256, 73/256/273, 73/256/273/375, 73/256/333/375, 73/256/375, 73/273, 73/273/333/341, 73/273/333/375, 73/273/375, 73/333, 73/333/341/375, 73/333/375, 73/341, 73/375, 73/379, 77, 77/118, 77/118/143/262/276/334/376, 77/118/143/334/352/376, 77/118/143/376, 77/118/233, 77/118/233/352/376, 77/118/262/334/376, 77/118/334, 77/143/262/276, 77/143/262/334, 77/143/376, 77/233/238/376, 77/233/262, 77/233/276/376, 77/233/334/376, 77/262/376, 77/276, 77/334/376, 77/352, 82, 82/127, 82/127/144/235/290, 82/127/144/266, 82/127/144/266/290, 82/127/144/378, 82/127/235, 82/127/235/266/358, 82/127/290, 82/127/290/358, 82/127/290/378, 82/144/266/290/378, 82/144/266/378, 82/150/235/266, 82/235/266, 82/235/266/290, 82/235/266/378/383, 82/235/290, 82/235/378, 82/266, 82/266/378, 82/290, 82/378, 113, 113/252, 113/252/270/370, 113/270/370/384, 113/336, 118/233/276/334, 118/334, 127/235/266, 127/266/290/378, 127/378/383, 143, 143/276, 143/276/334, 144/235/290, 144/266/290, 144/290/358, 233/334, 233/376, 235, 235/266/290/358/383, 235/358, 252/329/336/370, 252/384, 256/333, 256/375, 262, 262/276, 262/334, 266/290, 266/378, 270/370, 273, 276/352, 290, 290/378, 333, 334, 334/376, 336/370, 341, 358, 370, 375, and 378. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540, selected from 18D, 18D/113A/252A, 18D/113A/252A/336Q/384V, 18D/113A/270L, 18D/259R/370S/384V, 18D/270L, 18D/270L/370S, 18D/270L/370S/384V, 18D/370S, 18L, 18L/113A, 18L/113A/270L, 18L/113A/270L/370S/384V, 18L/113E, 18L/113E/252A/370S, 18L/370S, 18L/370S/384V, 73H, 73H/115Q/184A/341G, 73H/115Q/341G, 73H/184A/256G, 73H/256G/273V/375S, 73H/256G/333L/375S, 73H/273V/333L/341G, 73H/273V/333L/375S, 73H/333L, 73R, 73R/115Q/341G, 73R/131D/256G/273V/333L/341G/375S,

- 73R/184A/256G/341G, 73R/184A/256G/375S, 73R/184A/273V/333L, 73R/184A/333L,
 73R/184A/341G/375S, 73R/256G, 73R/256G/273V, 73R/256G/273V/375S, 73R/256G/375S, 73R/273V,
 73R/273V/333L/375S, 73R/273V/375S, 73R/333L, 73R/333L/341G/375S, 73R/333L/375S, 73R/341G,
 73R/375S, 73R/379V, 77V, 77V/118S, 77V/118S/143H/262L/276E/334L/376L,
 5 77V/118S/143H/334L/352L/376L, 77V/118S/143H/376L, 77V/118S/233L, 77V/118S/233L/352L/376L,
 77V/118S/262L/334L/376L, 77V/118S/334L, 77V/143H/262L/276E, 77V/143H/262L/334L,
 77V/143H/376L, 77V/233L/238V/376L, 77V/233L/262L, 77V/233L/276E/376L, 77V/233L/334L/376L,
 77V/262L/376L, 77V/276E, 77V/334L/376L, 77V/352L, 82Q, 82Q/127L, 82Q/127L/144Y/235S/290V,
 82Q/127L/144Y/266V, 82Q/127L/144Y/266V/290V, 82Q/127L/144Y/378S, 82Q/127L/235S,
 10 82Q/127L/235S/266V/358W, 82Q/127L/290Q, 82Q/127L/290Q/358W, 82Q/127L/290V/378S,
 82Q/144Y/266V/290Q/378S, 82Q/144Y/266V/378S, 82Q/150Q/235S/266V, 82Q/235S/266V,
 82Q/235S/266V/290Q, 82Q/235S/266V/378S/383P, 82Q/235S/290V, 82Q/235S/378S, 82Q/266V,
 82Q/266V/378S, 82Q/290V, 82Q/378S, 113A/270L/370S/384V, 113E, 113E/252A,
 113E/252A/270L/370S, 113E/336Q, 118S/233L/276E/334L, 118S/334L, 127L/235S/266V,
 15 127L/266V/290Q/378S, 127L/378S/383P, 143H, 143H/276E, 143H/276E/334L, 144Y/235S/290Q,
 144Y/266V/290Q, 144Y/290V/358W, 233L/334L, 233L/376L, 235S, 235S/266V/290Q/358W/383P,
 235S/358W, 252A/329A/336Q/370S, 252A/384V, 256G/333L, 256G/375S, 262L, 262L/276E, 262L/334L,
 266V/290Q, 266V/378S, 270L/370S, 273V, 276E/352L, 290Q, 290V/378S, 333L, 334L, 334L/376L,
 336Q/370S, 341G, 358W, 370S, 375S, and 378S. In some embodiments, the engineered transposase
 20 polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
 sequence SEQ ID NO: 2540 and one or more residue differences as compared to SEQ ID NO: 2540,
 selected from A18D, A18D/S113A/N252A, A18D/S113A/N252A/G336Q/A384V, A18D/S113A/K270L,
 A18D/S259R/T370S/A384V, A18D/K270L, A18D/K270L/T370S, A18D/K270L/T370S/A384V,
 25 A18D/T370S, A18L, A18L/S113A, A18L/S113A/K270L, A18L/S113A/K270L/T370S/A384V, A18L/S113E,
 A18L/S113E/N252A/T370S, A18L/T370S, A18L/T370S/A384V, Q73H, Q73H/K115Q/C184A/R341G,
 Q73H/K115Q/R341G, Q73H/C184A/R256G, Q73H/R256G/D273V/L375S, Q73H/R256G/T333L/L375S,
 Q73H/D273V/T333L/R341G, Q73H/D273V/T333L/L375S, Q73H/T333L, Q73R, Q73R/K115Q/R341G,
 Q73R/G131D/R256G/D273V/T333L/R341G/L375S, Q73R/C184A/R256G/R341G,
 30 Q73R/C184A/R256G/L375S, Q73R/C184A/D273V/T333L, Q73R/C184A/T333L,
 Q73R/C184A/R341G/L375S, Q73R/R256G, Q73R/R256G/D273V, Q73R/R256G/D273V/L375S,
 Q73R/R256G/L375S, Q73R/D273V, Q73R/D273V/T333L/L375S, Q73R/D273V/L375S, Q73R/T333L,
 Q73R/T333L/R341G/L375S, Q73R/T333L/L375S, Q73R/R341G, Q73R/L375S, Q73R/G379V, E77V,
 E77V/A118S, E77V/A118S/D143H/V262L/Q276E/G334L/R376L,
 35 E77V/A118S/D143H/G334L/V352L/R376L, E77V/A118S/D143H/R376L, E77V/A118S/A233L,
 E77V/A118S/A233L/V352L/R376L, E77V/A118S/V262L/G334L/R376L, E77V/A118S/G334L,
 E77V/D143H/V262L/Q276E, E77V/D143H/V262L/G334L, E77V/D143H/R376L,
 E77V/A233L/I238V/R376L, E77V/A233L/V262L, E77V/A233L/Q276E/R376L, E77V/A233L/G334L/R376L,
 E77V/V262L/R376L, E77V/Q276E, E77V/G334L/R376L, E77V/V352L, L82Q, L82Q/I127L,
 40 L82Q/I127L/W144Y/T235S/R290V, L82Q/I127L/W144Y/A266V, L82Q/I127L/W144Y/A266V/R290V,

- L82Q/I127L/W144Y/K378S, L82Q/I127L/T235S, L82Q/I127L/T235S/A266V/I358W, L82Q/I127L/R290Q, L82Q/I127L/R290Q/I358W, L82Q/I127L/R290V/K378S, L82Q/W144Y/A266V/R290Q/K378S, L82Q/W144Y/A266V/K378S, L82Q/H150Q/T235S/A266V, L82Q/T235S/A266V, L82Q/T235S/A266V/R290Q, L82Q/T235S/A266V/K378S/A383P, L82Q/T235S/R290V,
- 5 L82Q/T235S/K378S, L82Q/A266V, L82Q/A266V/K378S, L82Q/R290V, L82Q/K378S, S113A/K270L/T370S/A384V, S113E, S113E/N252A, S113E/N252A/K270L/T370S, S113E/G336Q, A118S/A233L/Q276E/G334L, A118S/G334L, I127L/T235S/A266V, I127L/A266V/R290Q/K378S, I127L/K378S/A383P, D143H, D143H/Q276E, D143H/Q276E/G334L, W144Y/T235S/R290Q, W144Y/A266V/R290Q, W144Y/R290V/I358W, A233L/G334L, A233L/R376L, T235S,
- 10 T235S/A266V/R290Q/I358W/A383P, T235S/I358W, N252A/K329A/G336Q/T370S, N252A/A384V, R256G/T333L, R256G/L375S, V262L, V262L/Q276E, V262L/G334L, A266V/R290Q, A266V/K378S, K270L/T370S, D273V, Q276E/V352L, R290Q, R290V/K378S, T333L, G334L, G334L/R376L, G336Q/T370S, R341G, I358W, T370S, L375S, and K378S.

- In some embodiments, the engineered transposase polypeptide with one or more improved
- 15 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3402 and one or more residue differences as compared to SEQ ID NO: 3402, selected from 18, 18/77/118/169/233/238/256/266/270, 18/77/150/256/376, 18/118,
- 18/118/143/169/238/256/266/333, 18/118/150/169/184/238/256/270/378, 18/118/169, 18/118/233/238,
- 20 18/143, 18/143/256, 18/150/169, 18/150/266, 18/169/184/290/333/376, 18/169/233/333, 18/169/392, 18/184/270, 18/233/256/334, 18/256/266/290/378, 77/113/118/256/266/290, 77/118/290, 77/169/270/290, 77/233, 77/233/256, 77/238, 113/118/169/184/256/376, 113/169, 113/233, 118, 118/143/238, 118/233/238/256/378, 118/233/238/270, 143/169/233/270, 143/233/266/290/376, 143/238/270/290/378, 143/256/334, 143/290, 169, 169/256/266, 183/270/333, 233/238, 233/238/376, 233/392, 238/256,
- 25 256/290/303, 270, 290, 376, and 378. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3402 and one or more residue differences as compared to SEQ ID NO: 3402, selected from 18L,
- 18L/77V/118S/169T/233L/238V/256R/266V/270P, 18L/77V/150N/256R/376L, 18L/118S,
- 30 18L/118S/143H/169T/238V/256R/266V/333T, 18L/118S/150N/169T/184A/238V/256R/270P/378S, 18L/118S/169T, 18L/118S/233L/238V, 18L/143H, 18L/143H/256R, 18L/150N/169T, 18L/150N/266V, 18L/169T/184A/290K/333T/376L, 18L/169T/233L/333T, 18L/169T/392P, 18L/184A/270P, 18L/233L/256R/334L, 18L/256R/266V/290K/378S, 77V/113E/118S/256R/266V/290K, 77V/118S/290K, 77V/169T/270P/290K, 77V/233L, 77V/233L/256R, 77V/238V, 113E/118S/169T/184A/256R/376L,
- 35 113E/169T, 113E/233L, 118S, 118S/143H/238V, 118S/233L/238V/256R/378S, 118S/233L/238V/270P, 143H/169T/233L/270P, 143H/233L/266V/290K/376L, 143H/238V/270P/290K/378S, 143H/256R/334L, 143H/290K, 169T, 169T/256R/266V, 183L/270P/333T, 233L/238V, 233L/238V/376L, 233L/392P, 238V/256R, 256R/290K/303G, 270P, 290K, 376L, and 378S. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%,
- 40 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to

reference sequence SEQ ID NO: 3402 and one or more residue differences as compared to SEQ ID NO:

3402, selected from A18L, A18L/E77V/A118S/C169T/A233L/I238V/G256R/A266V/K270P,

A18L/E77V/H150N/G256R/R376L, A18L/A118S,

A18L/A118S/D143H/C169T/I238V/G256R/A266V/L333T,

5 A18L/A118S/H150N/C169T/C184A/I238V/G256R/K270P/K378S, A18L/A118S/C169T,

A18L/A118S/A233L/I238V, A18L/D143H, A18L/D143H/G256R, A18L/H150N/C169T,

A18L/H150N/A266V, A18L/C169T/C184A/R290K/L333T/R376L, A18L/C169T/A233L/L333T,

A18L/C169T/C392P, A18L/C184A/K270P, A18L/A233L/G256R/G334L,

A18L/G256R/A266V/R290K/K378S, E77V/S113E/A118S/G256R/A266V/R290K, E77V/A118S/R290K,

10 E77V/C169T/K270P/R290K, E77V/A233L, E77V/A233L/G256R, E77V/I238V,

S113E/A118S/C169T/C184A/G256R/R376L, S113E/C169T, S113E/A233L, A118S, A118S/D143H/I238V,

A118S/A233L/I238V/G256R/K378S, A118S/A233L/I238V/K270P, D143H/C169T/A233L/K270P,

D143H/A233L/A266V/R290K/R376L, D143H/I238V/K270P/R290K/K378S, D143H/G256R/G334L,

D143H/R290K, C169T, C169T/G256R/A266V, V183L/K270P/L333T, A233L/I238V, A233L/I238V/R376L,

15 A233L/C392P, I238V/G256R, G256R/R290K/E303G, K270P, R290K, R376L, and K378S.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3402 and one or more residue differences as compared to SEQ ID NO: 3402,

20 selected from 18, 18/77/118/169/266/334/376, 18/77/150/256/376, 18/77/256/290/334, 18/113/238, 18/118, 18/118/143/169/238/256/266/333, 18/118/143/169/376, 18/118/169, 18/118/233/238, 18/118/266, 18/143, 18/143/169/233/238/358, 18/143/233/238/290/376, 18/143/256, 18/143/290, 18/143/333/334/378, 18/150/169, 18/150/266, 18/169/184/290/333/376, 18/169/233/333, 18/169/238/270/334/376, 18/169/392, 18/233/256/334, 18/233/266/290/376, 18/256/266/290/378, 18/266/270/378, 18/378,

25 77/113/121/143/233/334, 77/118/169/233/266/270, 77/118/238/376, 77/118/290, 77/169/270/290, 77/233, 77/238, 113/238/376, 118, 118/143/238, 118/143/392, 118/169/184/233/266/270/333, 118/233/238/256/378, 118/256/290/333/334/376/378, 118/256/334/376, 118/266/290, 143/169/233/270, 143/233/266/290/376, 143/238/270/290/378, 143/256/334, 143/290, 143/364, 150/233/333, 169, 169/256/266, 183/270/333, 233/238, 233/238/376, 233/392, 238/256, 238/266/270/378, 256/266/334,

30 266/270/376, 266/333, and 378. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3402 and one or more residue differences as compared to SEQ ID NO: 3402, selected from 18L, 18L/77V/118S/169T/266V/334L/376L, 18L/77V/150N/256R/376L, 18L/77V/256R/290K/334L,

35 18L/113E/238V, 18L/118S, 18L/118S/143H/169T/238V/256R/266V/333T, 18L/118S/143H/169T/376L, 18L/118S/169T, 18L/118S/233L/238V, 18L/118S/266V, 18L/143H, 18L/143H/169T/233L/238V/358W, 18L/143H/233L/238V/290K/376L, 18L/143H/256R, 18L/143H/290K, 18L/143H/333T/334L/378S, 18L/150N/169T, 18L/150N/266V, 18L/169T/184A/290K/333T/376L, 18L/169T/233L/333T, 18L/169T/238V/270P/334L/376L, 18L/169T/392P, 18L/233L/256R/334L, 18L/233L/266V/290K/376L,

40 18L/256R/266V/290K/378S, 18L/266V/270P/378S, 18L/378S, 77V/113E/121C/143H/233L/334L,

77V/118S/169T/233L/266V/270P, 77V/118S/238V/376L, 77V/118S/290K, 77V/169T/270P/290K, 77V/233L, 77V/238V, 113E/238V/376L, 118S, 118S/143H/238V, 118S/143H/392P, 118S/169T/184A/233L/266V/270P/333T, 118S/233L/238V/256R/378S, 118S/256R/290K/333T/334L/376L/378S, 118S/256R/334L/376L, 118S/266V/290K, 143H/169T/233L/270P, 143H/233L/266V/290K/376L, 143H/238V/270P/290K/378S, 143H/256R/334L, 143H/290K, 143H/364R, 150N/233L/333T, 169T, 169T/256R/266V, 183L/270P/333T, 233L/238V, 233L/238V/376L, 233L/392P, 238V/256R, 238V/266V/270P/378S, 256R/266V/334L, 266V/270P/376L, 266V/333T, and 378S. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3402 and one or more residue differences as compared to SEQ ID NO: 3402, selected from A18L, A18L/E77V/A118S/C169T/A266V/G334L/R376L, A18L/E77V/H150N/G256R/R376L, A18L/E77V/G256R/R290K/G334L, A18L/S113E/I238V, A18L/A118S, A18L/A118S/D143H/C169T/I238V/G256R/A266V/L333T, A18L/A118S/D143H/C169T/R376L, A18L/A118S/C169T, A18L/A118S/A233L/I238V, A18L/A118S/A266V, A18L/D143H, A18L/D143H/C169T/A233L/I238V/I358W, A18L/D143H/A233L/I238V/R290K/R376L, A18L/D143H/G256R, A18L/D143H/R290K, A18L/D143H/L333T/G334L/K378S, A18L/H150N/C169T, A18L/H150N/A266V, A18L/C169T/C184A/R290K/L333T/R376L, A18L/C169T/A233L/L333T, A18L/C169T/I238V/K270P/G334L/R376L, A18L/C169T/C392P, A18L/A233L/G256R/G334L, A18L/A233L/A266V/R290K/R376L, A18L/G256R/A266V/R290K/K378S, A18L/A266V/K270P/K378S, A18L/K378S, E77V/S113E/W121C/D143H/A233L/G334L, E77V/A118S/C169T/A233L/A266V/K270P, E77V/A118S/I238V/R376L, E77V/A118S/R290K, E77V/C169T/K270P/R290K, E77V/A233L, E77V/I238V, S113E/I238V/R376L, A118S, A118S/D143H/I238V, A118S/D143H/C392P, A118S/C169T/C184A/A233L/A266V/K270P/L333T, A118S/A233L/I238V/G256R/K378S, A118S/G256R/R290K/L333T/G334L/R376L/K378S, A118S/G256R/G334L/R376L, A118S/A266V/R290K, D143H/C169T/A233L/K270P, D143H/A233L/A266V/R290K/R376L, D143H/I238V/K270P/R290K/K378S, D143H/G256R/G334L, D143H/R290K, D143H/Q364R, H150N/A233L/L333T, C169T, C169T/G256R/A266V, V183L/K270P/L333T, A233L/I238V, A233L/I238V/R376L, A233L/C392P, I238V/G256R, I238V/A266V/K270P/K378S, G256R/A266V/G334L, A266V/K270P/R376L, A266V/L333T, and K378S.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3662 and one or more residue differences as compared to SEQ ID NO: 3662, selected from 77/156, 77/156/270, 77/184/244, 77/184/333, 77/244/270, 77/270, 143, 143/169/238/257/344, 143/169/255, 143/169/255/257, 143/169/343, 143/255, 143/255/257, 143/255/344, 143/257, 156, 156/270, 169/238, 169/238/255/257/344, 169/257/344, 169/344, 184, 184/244/324, 238, 238/255/344, 244/270, 257/344, and 270. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ

ID NO: 3662 and one or more residue differences as compared to SEQ ID NO: 3662, selected from 77E/156L, 77E/156L/270P, 77E/184A/244K, 77E/184A/333T, 77E/244K/270P, 77E/270P, 77E/270R, 143E, 143E/169T/238V/257R/344K, 143E/169T/255R, 143E/169T/255R/257R, 143E/255R, 143E/255R/257R, 143E/255R/344K, 143E/257R, 143H/169T/343R, 156L, 156L/270R, 169T/238V, 169T/238V/255R/257R/344M, 169T/257R/344M, 169T/344K, 169T/344M, 184A, 184A/244K/324M, 238V, 238V/255R/344M, 244K/270P, 257R/344K, and 270P. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3662 and one or more residue differences as compared to SEQ ID NO: 3662, selected from V77E/E156L, V77E/E156L/K270P, V77E/C184A/V244K, V77E/C184A/L333T, V77E/V244K/K270P, V77E/K270P, V77E/K270R, D143E, D143E/C169T/I238V/K257R/E344K, D143E/C169T/A255R, D143E/C169T/A255R/K257R, D143E/A255R, D143E/A255R/K257R, D143E/A255R/E344K, D143E/K257R, D143H/C169T/E343R, E156L, E156L/K270R, C169T/I238V, C169T/I238V/A255R/K257R/E344M, C169T/K257R/E344M, C169T/E344K, C169T/E344M, C184A, C184A/V244K/V324M, I238V, I238V/A255R/E344M, V244K/K270P, K257R/E344K, and K270P.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3662 and one or more residue differences as compared to SEQ ID NO: 3662, selected from 77, 77/156/184/244/324, 77/184/270, 143, 143/238, 143/238/255, 143/238/255/343, 143/238/255/392, 143/238/344, 143/255, 143/257/392, 143/344, 169, 169/238/255/344, 169/238/344, 169/255, 169/343, 169/392, 238/344, 255, 255/257, 255/257/392, 257, and 344. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3662 and one or more residue differences as compared to SEQ ID NO: 3662, selected from 77E, 77E/156L/184A/244K/324M, 77E/184A/270P, 143E/238V, 143E/238V/255R/343R, 143E/238V/255R/392P, 143E/238V/344M, 143E/257R/392P, 143E/344M, 143H, 143H/238V/255R, 143H/255R, 169T, 169T/238V/255R/344M, 169T/238V/344M, 169T/255R, 169T/343R, 169T/392P, 238V/344M, 255R, 255R/257R, 255R/257R/392P, 257R, and 344M.

In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3662 and one or more residue differences as compared to SEQ ID NO: 3662, selected from V77E, V77E/E156L/C184A/V244K/V324M, V77E/C184A/K270P, D143E/I238V, D143E/I238V/A255R/E343R, D143E/I238V/A255R/C392P, D143E/I238V/E344M, D143E/K257R/C392P, D143E/E344M, D143H, D143H/I238V/A255R, D143H/A255R, C169T, C169T/I238V/A255R/E344M, C169T/I238V/E344M, C169T/A255R, C169T/E343R, C169T/C392P, I238V/E344M, A255R, A255R/K257R, A255R/K257R/C392P, K257R, and E344M.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,

89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3788 and one or more residue differences as compared to SEQ ID NO: 3788, selected from 20, 22, 26, 42, 43, 44, 45, 50, 51, 52, 53, 54, 55, 60, 62, 64, 65, 95, 97, 99, 103, 104, 111, 112, 115, 115/143/169, 115/169/343, 115/343, 116, 116/343, 143, 143/169, 154/439, 169, 169/324, 169/324/343, 169/343, 186, 234, 236, 238, 238/246, 239, 239/244, 241, 242, 243, 244, 250, 251, 254, 255, 260, 283, 284, 295, 296, 298/326, 319, 322, 323, 324, 324/343, 326, 328, 329, 333, 334, 338, 340, 341, 343, 344, 347, 366, 434, 439, 442, 443, 444, 445, 446, and 447. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3788 and one or more residue differences as compared to SEQ ID NO: 3788, selected from 20R, 22F, 22L, 22N, 22R, 22V, 22W, 26R, 42G, 42P, 43L, 43P, 43V, 44G, 44P, 44V, 45A, 45E, 45G, 45L, 45M, 45Q, 45R, 45W, 50R, 51Q, 51R, 52K, 52P, 52R, 52T, 53A, 53H, 53R, 53S, 54R, 54S, 54T, 54V, 55G, 55M, 55S, 55T, 55V, 60L, 62G, 64A, 64D, 64E, 64G, 64R, 64S, 64T, 64V, 65A, 65M, 95G, 95R, 95V, 97A, 97C, 99G, 99R, 103D, 103F, 103G, 103R, 103S, 103T, 103V, 104G, 104T, 111V, 112R, 115E, 115E/169T/343R, 115G, 115L, 115P, 115Q, 115Q/143E/169T, 115Q/169T/343R, 115Q/343R, 115R, 115T, 115W, 116E/343R, 116S, 143E, 143E/169T, 154E/439S, 169T, 169T/324M, 169T/324M/343R, 169T/343R, 186A, 186P, 186Q, 186S, 234W, 236T, 236V, 236Y, 238L/246G, 238V, 239D, 239K/244F, 239L, 241R, 242M, 243E, 243G, 243L, 243Q, 243W, 244C, 244G, 244T, 250A, 250S, 251A, 251I, 251L, 251S, 254R, 254T, 255C, 255K, 255L, 255M, 255R, 260A, 260C, 260I, 260T, 283G, 283S, 283T, 284A, 284G, 284R, 295F, 295I, 295V, 296H, 298I/326A, 319E, 319K, 319R, 319W, 322G, 322S, 322T, 323H, 323T, 323V, 324M, 324M/343R, 326A, 326L, 326Q, 328Y, 329G, 329T, 329W, 333M, 333Q, 333R, 333W, 333Y, 334E, 334W, 338A, 340A, 341K, 343G, 343M, 343R, 343V, 344G, 344H, 344K, 344M, 344Q, 344S, 344T, 344V, 347R, 366S, 434Y, 439A, 439G, 439S, 439Y, 442G, 442H, 442Q, 443C, 443I, 443P, 443S, 444Q, 445G, 445L, 445M, 445T, 445V, 446F, 446G, 446R, 447L, 447N, and 447T. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 3788 and one or more residue differences as compared to SEQ ID NO: 3788, selected from G20R, P22F, P22L, P22N, P22R, P22V, P22W, K26R, S42G, S42P, I43L, I43P, I43V, T44G, T44P, T44V, V45A, V45E, V45G, V45L, V45M, V45Q, V45R, V45W, S50R, K51Q, K51R, A52K, A52P, A52R, A52T, M53A, M53H, M53R, M53S, Q54R, Q54S, Q54T, Q54V, E55G, E55M, E55S, E55T, E55V, F60L, R62G, P64A, P64D, P64E, P64G, P64R, P64S, P64T, P64V, N65A, N65M, T95G, T95R, T95V, S97A, S97C, S99G, S99R, Q103D, Q103F, Q103G, Q103R, Q103S, Q103T, Q103V, V104G, V104T, L111V, G112R, K115E, K115E/C169T/E343R, K115G, K115L, K115P, K115Q, K115Q/D143E/C169T, K115Q/C169T/E343R, K115Q/E343R, K115R, K115T, K115W, D116E/E343R, D116S, D143E, D143E/C169T, K154E/R439S, C169T, C169T/V324M, C169T/V324M/E343R, C169T/E343R, R186A, R186P, R186Q, R186S, Y234W, I236T, I236V, I236Y, I238L/S246G, I238V, P239D, P239K/V244F, P239L, K241R, G242M, V243E, V243G, V243L, V243Q, V243W, V244C, V244G, V244T, R250A, R250S, K251A, K251I, K251L, K251S, P254R, P254T, A255C, A255K, A255L, A255M, A255R, L260A, L260C, L260I, L260T, A283G, A283S, A283T, E284A, E284G, E284R, L295F, L295I, L295V, K296H, V298I/E326A, T319E, T319K, T319R,

T319W, W322G, W322S, W322T, R323H, R323T, R323V, V324M, V324M/E343R, E326A, E326L, E326Q, H328Y, K329G, K329T, K329W, L333M, L333Q, L333R, L333W, L333Y, G334E, G334W, E338A, L340A, R341K, E343G, E343M, E343R, E343V, E344G, E344H, E344K, E344M, E344Q, E344S, E344T, E344V, N347R, R366S, F434Y, R439A, R439G, R439S, R439Y, I442G, I442H, I442Q, A443C, A443I, A443P, A443S, S444Q, W445G, W445L, W445M, W445T, W445V, E446F, E446G, E446R, A447L, A447N, and A447T.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4150 and one or more residue differences as compared to SEQ ID NO: 4150, selected from 26, 26/236, 26/236/260/333, 26/236/295, 26/260, 26/260/295, 26/260/295/333, 26/295/319, 26/333, 64, 64/244, 64/251/260/347, 64/260/295/333, 64/260/347, 103/115/255, 103/115/343, 103/115/343/344, 103/143/255/284/343/344, 103/169/343/344, 103/255/284, 103/343/344, 115, 115/143/255, 115/143/255/343, 115/143/284, 115/143/343, 115/143/344, 115/169, 115/169/343, 115/255, 115/255/344, 115/284, 115/284/343/344, 115/284/344, 115/343, 115/343/344, 115/344, 143, 143/169/255/284/343/344, 143/169/343/344, 143/255, 143/255/284/343/344, 143/255/343, 143/255/343/344, 143/255/344, 143/343/344, 169, 169/255, 169/255/284, 169/255/343, 169/284, 169/343, 169/344, 236, 236/244/333, 236/260, 236/347, 244, 244/251, 244/251/319, 244/260, 244/295, 244/347, 251, 251/260/295/347, 251/333, 255, 255/284, 255/343, 255/343/344, 255/344, 259, 260, 260/295, 260/333, 284, 284/343/344, 284/344, 295, 295/333, 333, 343, 343/344, 344, and 347. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4150 and one or more residue differences as compared to SEQ ID NO: 4150, selected from 26R, 26R/236T, 26R/236T/260I/333Y, 26R/236T/295I, 26R/260I, 26R/260I/295I, 26R/260I/295I/333Y, 26R/295I/319K, 26R/333Y, 64A, 64A/244T, 64A/251A/260I/347R, 64A/260I/295I/333Y, 64A/260I/347R, 103G/115E/255K, 103G/115E/343M, 103G/115E/343M/344G, 103G/143E/255K/284G/343M/344G, 103G/169T/343M/344G, 103G/255K/284A, 103G/343M/344G, 115E, 115E/143E/255K, 115E/143E/344G, 115E/169T, 115E/255K, 115E/284A/344G, 115E/284G/343M/344G, 115E/343M, 115E/344G, 115T, 115T/143E/255K/343M, 115T/143E/284G, 115T/143E/343M, 115T/169T, 115T/169T/343M, 115T/255K, 115T/255K/344G, 115T/284G, 115T/343M/344G, 115T/344G, 143E, 143E/169T/255K/284G/343M/344G, 143E/169T/343M/344G, 143E/255K, 143E/255K/284A/343M/344G, 143E/255K/343M, 143E/255K/343M/344G, 143E/255K/344G, 143E/343M/344G, 169T, 169T/255K, 169T/255K/284G, 169T/255K/343M, 169T/284A, 169T/284G, 169T/343M, 169T/344G, 236T, 236T/244T/333Y, 236T/260I, 236T/347R, 244T, 244T/251A, 244T/251A/319K, 244T/260I, 244T/295I, 244T/347R, 251A, 251A/260I/295I/347R, 251A/333Y, 255K, 255K/284G, 255K/343M, 255K/343M/344G, 255K/344G, 259I, 260I, 260I/295I, 260I/333Y, 284A, 284A/343M/344G, 284G, 284G/344G, 295I, 295I/333Y, 333Y, 343M, 343M/344G, 344G, and 347R. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO:

- 4150 and one or more residue differences as compared to SEQ ID NO: 4150, selected from K26R, K26R/I236T, K26R/I236T/L260I/L333Y, K26R/I236T/L295I, K26R/L260I, K26R/L260I/L295I, K26R/L260I/L295I/L333Y, K26R/L295I/T319K, K26R/L333Y, P64A, P64A/V244T, P64A/K251A/L260I/N347R, P64A/L260I/L295I/L333Y, P64A/L260I/N347R, Q103G/K115E/A255K, Q103G/K115E/E343M, Q103G/K115E/E343M/E344G, Q103G/D143E/A255K/E284G/E343M/E344G, Q103G/C169T/E343M/E344G, Q103G/A255K/E284A, Q103G/E343M/E344G, K115E, K115E/D143E/A255K, K115E/D143E/E344G, K115E/C169T, K115E/A255K, K115E/E284A/E344G, K115E/E284G/E343M/E344G, K115E/E343M, K115E/E344G, K115T, K115T/D143E/A255K/E343M, K115T/D143E/E284G, K115T/D143E/E343M, K115T/C169T, K115T/C169T/E343M, K115T/A255K, K115T/A255K/E344G, K115T/E284G, K115T/E343M/E344G, K115T/E344G, D143E, D143E/C169T/A255K/E284G/E343M/E344G, D143E/C169T/E343M/E344G, D143E/A255K, D143E/A255K/E284A/E343M/E344G, D143E/A255K/E343M, D143E/A255K/E343M/E344G, D143E/A255K/E344G, D143E/E343M/E344G, C169T, C169T/A255K, C169T/A255K/E284G, C169T/A255K/E343M, C169T/E284A, C169T/E284G, C169T/E343M, C169T/E344G, I236T, I236T/V244T/L333Y, I236T/L260I, I236T/N347R, V244T, V244T/K251A, V244T/K251A/T319K, V244T/L260I, V244T/L295I, V244T/N347R, K251A, K251A/L260I/L295I/N347R, K251A/L333Y, A255K, A255K/E284G, A255K/E343M, A255K/E343M/E344G, A255K/E344G, S259I, L260I, L260I/L295I, L260I/L333Y, E284A, E284A/E343M/E344G, E284G, E284G/E344G, L295I, L295I/L333Y, L333Y, E343M, E343M/E344G, E344G, and N347R.
- In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4278 and one or more residue differences as compared to SEQ ID NO: 4278, selected from 115, 115/117, 115/117/118, 115/117/118/152, 115/117/118/169, 115/117/118/246, 115/117/152, 115/117/152/169, 115/117/152/169/246, 115/117/152/246, 115/117/169, 115/117/246, 115/118, 115/118/152, 115/118/152/169, 115/118/152/169/246, 115/118/169, 115/118/246, 115/152, 115/152/169, 115/152/169/246, 115/152/246, 115/152/246/374, 115/169, 115/169/246, 115/246, 117, 117/118, 117/118/152, 117/118/152/169, 117/118/152/246, 117/118/169, 117/118/246, 117/152, 117/152/169, 117/152/246/288, 117/169, 117/246, 118/246, 152, 152/169, 152/169/246, 152/246, 169, 169/246, and 246. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4278 and one or more residue differences as compared to SEQ ID NO: 4278, selected from 115E, 115E/117A, 115E/117A/118T, 115E/117A/118T/152R, 115E/117A/118T/169T, 115E/117A/118T/246D, 115E/117A/152R, 115E/117A/152R/169T, 115E/117A/152R/169T/246D, 115E/117A/152R/246D, 115E/117A/169T, 115E/117A/246D, 115E/118T, 115E/118T/152R, 115E/118T/152R/169T, 115E/118T/152R/169T/246D, 115E/118T/169T, 115E/118T/246D, 115E/152R, 115E/152R/169T, 115E/152R/169T/246D, 115E/152R/246D, 115E/152R/246D/374C, 115E/169T, 115E/169T/246D, 115E/246D, 117A, 117A/118T, 117A/118T/152R, 117A/118T/152R/169T, 117A/118T/152R/246D, 117A/118T/169T, 117A/118T/246D, 117A/152R, 117A/152R/169T, 117A/152R/246D/288G, 117A/169T,

117A/246D, 118T/246D, 152R, 152R/169T, 152R/169T/246D, 152R/246D, 169T, 169T/246D, and 246D.

In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4278 and one or more

- 5 residue differences as compared to SEQ ID NO: 4278, selected from K115E, K115E/K117A, K115E/K117A/S118T, K115E/K117A/S118T/T152R, K115E/K117A/S118T/C169T, K115E/K117A/S118T/S246D, K115E/K117A/T152R, K115E/K117A/T152R/C169T, K115E/K117A/T152R/C169T/S246D, K115E/K117A/T152R/S246D, K115E/K117A/C169T, K115E/K117A/S246D, K115E/S118T, K115E/S118T/T152R, K115E/S118T/T152R/C169T, K115E/S118T/T152R/C169T/S246D, K115E/S118T/C169T, K115E/S118T/S246D, K115E/T152R, K115E/T152R/C169T, K115E/T152R/C169T/S246D, K115E/T152R/S246D, K115E/T152R/S246D/Y374C, K115E/C169T, K115E/C169T/S246D, K115E/S246D, K117A, K117A/S118T, K117A/S118T/T152R, K117A/S118T/T152R/C169T, K117A/S118T/T152R/S246D, K117A/S118T/C169T, K117A/S118T/S246D, K117A/T152R, K117A/T152R/C169T, K117A/T152R/S246D/S288G, K117A/C169T, K117A/S246D, S118T/S246D, T152R, T152R/C169T, T152R/C169T/S246D, T152R/S246D, C169T, C169T/S246D, and S246D.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference

- 20 sequence SEQ ID NO: 4278 and one or more residue differences as compared to SEQ ID NO: 4278, selected from 73, 101, 114, 115, 118, 143, 145, 150, 151, 152, 153, 154, 156, 157, 160, 161, 163, 166, 167, 168, 169, 171, 172, 190, 191, 192, 195, 213, 214, 218, 219, 223, 225, 233, 233/343, 234, 235, 236, 237, 238, 239, 239/250, 243, 244, 249, 251, 254, 255, 257, 258, 259, 261, 262, 262/283, 277/292, 283, 284, 286, 289, 292, 295, and 296. In some embodiments, the engineered transposase polypeptide
- 25 comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4278 and one or more residue differences as compared to SEQ ID NO: 4278, selected from 73L, 73R, 101S, 114S, 115A, 115Q, 115R, 115S, 118R, 143E, 143L, 143Q, 143V, 145Y, 150I, 150L, 150M, 150T, 151A, 151D, 151F, 151H, 151R, 152L, 153G, 154P, 154S, 154Y, 156D, 156L, 156Q, 156W, 157L, 157P, 157R, 160R, 160T, 160V, 161R, 163K, 163L, 163Q, 163R, 166T, 167Q, 167R, 167T, 168A, 168N, 168V, 169L, 169V, 171I, 171L, 171M, 172L, 172R, 172V, 190F, 191H, 191L, 191V, 192E, 192G, 192R, 192S, 195L, 195M, 195T, 195V, 213R, 214L, 214V, 218P, 219R, 223A, 223D, 223Q, 225V, 233F, 233G, 233L, 233Q, 233R, 233T/343G, 233W, 234L, 234V, 235C, 235D, 235E, 235S, 236M, 236T, 237Q, 238L, 238V, 239A, 239E, 239N/250H, 239T, 243I, 243Q, 243V, 244A, 244F, 244L, 244S, 244T, 244W, 249G, 249I, 249L, 249R, 249S, 249T, 251H, 251Q, 254A, 254E, 255G, 255S, 255W, 257L, 257T, 258V, 259L, 259N, 261A, 261K, 261R, 261T, 261V, 262I/283V, 262L, 277S/292V, 283G, 283T, 284I, 284L, 284R, 284S, 286E, 286G, 286I, 286L, 286V, 289L, 292E, 292G, 295V, 296C, and 296R. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more
- 35 sequence identity to reference sequence SEQ ID NO: 4278 and one or more residue differences as
- 40

compared to SEQ ID NO: 4278, selected from Q73L, Q73R, K101S, V114S, K115A, K115Q, K115R, K115S, S118R, D143E, D143L, D143Q, D143V, W145Y, H150I, H150L, H150M, H150T, P151A, P151D, P151F, P151H, P151R, T152L, D153G, K154P, K154S, K154Y, E156D, E156L, E156Q, E156W, K157L, K157P, K157R, A160R, A160T, A160V, K161R, A163K, A163L, A163Q, A163R, S166T, H167Q, H167R, H167T, K168A, K168N, K168V, C169L, C169V, Q171I, Q171L, Q171M, I172L, I172R, I172V, I190F, Q191H, Q191L, Q191V, A192E, A192G, A192R, A192S, Q195L, Q195M, Q195T, Q195V, K213R, I214L, I214V, E218P, L219R, E223A, E223D, E223Q, L225V, A233F, A233G, A233L, A233Q, A233R, A233T/E343G, A233W, Y234L, Y234V, T235C, T235D, T235E, T235S, I236M, I236T, D237Q, I238L, I238V, P239A, P239E, P239N/R250H, P239T, G243I, G243Q, G243V, V244A, V244F, V244L, V244S, V244T, V244W, K249G, K249I, K249L, K249R, K249S, K249T, K251H, K251Q, P254A, P254E, A255G, A255S, A255W, K257L, K257T, A258V, S259L, S259N, S261A, S261K, S261R, S261T, S261V, V262I/A283V, V262L, P277S/M292V, A283G, A283T, E284I, E284L, E284R, E284S, T286E, T286G, T286I, T286L, T286V, P289L, M292E, M292G, I295V, K296C, and K296R.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4642 and one or more residue differences as compared to SEQ ID NO: 4642, selected from 93, 113, 113/114, 113/115, 113/143, 115/118, 115/177, 118/143/163/168/195/255/315, 118/143/255, 118/163/168/214/255/315, 118/168/214/315, 118/255, 124, 126, 129, 134, 143/154/255, 143/156/168/216/255, 143/163, 143/163/255, 143/168, 143/168/169/255/315, 143/255, 148, 156/214/315, 159, 163, 163/168, 163/168/169, 163/168/169/255/315, 163/168/169/315, 163/168/195/214/255/315, 163/214, 163/255, 163/255/315, 168/315, 186, 194, 197, 206, 207, 208, 210, 213, 255/315, 277, 293, 294, 315, 317, 320, 322, 328, 329, 330, 333, 352, 366, 392, and 445. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4642 and one or more residue differences as compared to SEQ ID NO: 4642, selected from 93A, 93F, 93G, 93H, 93M, 93S, 93T, 113C/115K, 113G/115K, 113S, 113S/114G, 113S/115E, 113S/143K, 113S/143R, 115K/118C, 115K/118G, 115R/177V, 118T/143Q/163V/168N/195E/255K/315L, 118T/143Q/255K, 118T/163V/168N/214L/255K/315L, 118T/168N/214L/315L, 118T/255K, 124T, 126V, 129M, 134R, 134V, 143Q/154L/255K, 143Q/156D/168N/216G/255K, 143Q/163V, 143Q/163V/255K, 143Q/168N, 143Q/168N/169L/255K/315L, 143Q/255K, 148L, 148Q, 148R, 156D/214L/315L, 159T, 163V, 163V/168M, 163V/168M/169L, 163V/168M/169T, 163V/168M/169V, 163V/168M/195E/214L/255K/315L, 163V/168N/169T/255K/315L, 163V/168N/169V/315L, 163V/214L, 163V/255K, 163V/255K/315L, 168N/315L, 186K, 194M, 197L, 206L, 206M, 207M, 208G, 210R, 213R, 255K/315L, 277D, 277G, 277S, 277W, 293G, 293P, 293R, 293S, 293T, 294G, 294R, 294V, 315L, 317A, 317C, 317G, 317L, 317Q, 317S, 320C, 320I, 320V, 320Y, 322F, 328F, 329L, 329S, 330S, 330T, 333G, 333Q, 333R, 352G, 352I, 366A, 366K, 392A, 445R, and 445Y. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 2 and one or more residue differences

- as compared to SEQ ID NO: 4642, selected from D93A, D93F, D93G, D93H, D93M, D93S, D93T, E113C/Q115K, E113G/Q115K, E113S, E113S/V114G, E113S/Q115E, E113S/D143K, E113S/D143R, Q115K/S118C, Q115K/S118G, Q115R/M177V, S118T/D143Q/A163V/K168N/Q195E/A255K/I315L, S118T/D143Q/A255K, S118T/A163V/K168N/I214L/A255K/I315L, S118T/K168N/I214L/I315L,
- 5 S118T/A255K, H124T, T126V, L129M, T134R, T134V, D143Q/K154L/A255K, D143Q/E156D/K168N/D216G/A255K, D143Q/A163V, D143Q/A163V/A255K, D143Q/K168N, D143Q/K168N/C169L/A255K/I315L, D143Q/A255K, P148L, P148Q, P148R, E156D/I214L/I315L, S159T, A163V, A163V/K168M, A163V/K168M/C169L, A163V/K168M/C169T, A163V/K168M/C169V, A163V/K168M/Q195E/I214L/A255K/I315L, A163V/K168N/C169T/A255K/I315L,
- 10 A163V/K168N/C169V/I315L, A163V/I214L, A163V/A255K, A163V/A255K/I315L, K168N/I315L, R186K, L194M, K197L, V206L, V206M, R207M, A208G, H210R, K213R, A255K/I315L, P277D, P277G, P277S, P277W, E293G, E293P, E293R, E293S, E293T, P294G, P294R, P294V, I315L, M317A, M317C, M317G, M317L, M317Q, M317S, A320C, A320I, A320V, A320Y, W322F, H328F, K329L, K329S, A330S, A330T, L333G, L333Q, L333R, V352G, V352I, R366A, R366K, C392A, W445R, and W445Y.
- 15 In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4642 and one or more residue differences as compared to SEQ ID NO: 4642, selected from 127, 128, 129, 140, 180, 182, 194, 277, 293, 294, 299, 300, 317, 319, 320, 324, 333, 356,
- 20 366, and 392. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4642 and one or more residue differences as compared to SEQ ID NO: 4642, selected from 127L, 128V, 129M, 140V, 180L, 182Q, 182T, 194M, 277A, 277G, 277H, 277S, 293G, 293K, 293P, 293S, 294G, 294L, 294Q, 299T,
- 25 300T, 317A, 317F, 317I, 317S, 319A, 320I, 320Y, 324I, 333D, 333T, 356G, 366K, 392H, and 392M. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4642 and one or more residue differences as compared to SEQ ID NO: 4642, selected from I127L, L128V, L129M, I140V, V180L,
- 30 S182Q, S182T, L194M, P277A, P277G, P277H, P277S, E293G, E293K, E293P, E293S, P294G, P294L, P294Q, L299T, L300T, M317A, M317F, M317I, M317S, T319A, A320I, A320Y, V324I, L333D, L333T, A356G, R366K, C392H, and C392M.
- In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
- 35 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4984 and one or more residue differences as compared to SEQ ID NO: 4984, selected from 113, 117, 118, 151, 191, 196, and 205. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
- 40 sequence SEQ ID NO: 4984 and one or more residue differences as compared to SEQ ID NO: 4984,

selected from 113A, 113C, 113K, 113R, 113S, 113T, 117R, 118A, 118K, 118R, 151A, 151C, 151E, 151F, 151G, 151H, 151I, 151K, 151L, 151M, 151N, 151Q, 151R, 151S, 151T, 151V, 151W, 151Y, 191G, 191K, 191L, 191M, 191R, 191S, 191V, 196A, 196F, 196G, 196H, 196I, 196K, 196L, 196M, 196R, 196S, 196V, 196W, 196Y, and 205T. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4984 and one or more residue differences as compared to SEQ ID NO: 4984, selected from E113A, E113C, E113K, E113R, E113S, E113T, K117R, S118A, S118K, S118R, P151A, P151C, P151E, P151F, P151G, P151H, P151I, P151K, P151L, P151M, P151N, P151Q, P151R, P151S, P151T, P151V, P151W, P151Y, Q191G, Q191K, Q191L, Q191M, Q191R, Q191S, Q191V, D196A, D196F, D196G, D196H, D196I, D196K, D196L, D196M, D196R, D196S, D196V, D196W, D196Y, and V205T.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4984 and one or more residue differences as compared to SEQ ID NO: 4984, selected from 114/115/143, 114/115/143/233, 114/115/233, 114/115/343, 114/143, 114/143/151/233/238/261, 114/143/213/261/343, 114/143/233, 114/143/233/238, 114/151/261/343, 114/213/238, 114/343, 115, 115/143, 115/143/151/154/213/261, 115/143/151/154/233/343, 115/143/151/233/261, 115/143/151/233/343, 115/143/154/213/233/238/261, 115/143/213, 115/143/233, 115/143/343, 115/151/154/233, 115/151/261, 115/213/233, 115/213/233/238, 115/213/261, 115/233/238, 115/238, 115/343, 115/348, 143, 143/151/154, 143/151/154/213, 143/154, 143/213/233, 143/213/233/238/261/343, 143/213/261/343, 143/233/261, 143/261, 151/154, 151/154/213/343, 151/154/343, 151/213, 151/213/233, 213/233/343, and 233/238/343. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4984 and one or more residue differences as compared to SEQ ID NO: 4984, selected from 114G/115E/143E/233T, 114G/115E/143V, 114G/115E/233T, 114G/115E/343G, 114G/143E/233Q, 114G/143L/233F/238V, 114G/143V, 114G/143V/151R/233T/238V/261R, 114G/143V/213R/261R/343G, 114G/151R/261R/343G, 114G/213R/238V, 114G/343G, 115E, 115E/143E, 115E/143E/213R, 115E/143L/151R/154S/213R/261R, 115E/143L/151R/154S/233Q/343G, 115E/143L/151R/233F/343G, 115E/143L/154S/213R/233Q/238V/261R, 115E/143L/233Q, 115E/143L/233R, 115E/143L/343G, 115E/143V/151R/233F/261R, 115E/151R/154S/233F, 115E/151R/261R, 115E/213R/233F/238V, 115E/213R/233R, 115E/213R/261R, 115E/233F/238V, 115E/238V, 115E/343G, 115E/348F, 143E, 143E/151R/154S, 143E/151R/154S/213R, 143E/213R/233F/238V/261R/343G, 143E/233F/261R, 143L/213R/233R, 143L/261R, 143V/154S, 143V/213R/233R, 143V/213R/261R/343G, 151R/154S, 151R/154S/213R/343G, 151R/154S/343G, 151R/213R, 151R/213R/233F, 213R/233R/343G, and 233T/238V/343G. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 4984 and one

- or more residue differences as compared to SEQ ID NO: 4984, selected from
V114G/Q115E/D143E/A233T, V114G/Q115E/D143V, V114G/Q115E/A233T, V114G/Q115E/E343G,
V114G/D143E/A233Q, V114G/D143L/A233F/I238V, V114G/D143V,
V114G/D143V/P151R/A233T/I238V/S261R, V114G/D143V/K213R/S261R/E343G,
5 V114G/P151R/S261R/E343G, V114G/K213R/I238V, V114G/E343G, Q115E, Q115E/D143E,
Q115E/D143E/K213R, Q115E/D143L/P151R/K154S/K213R/S261R,
Q115E/D143L/P151R/K154S/A233Q/E343G, Q115E/D143L/P151R/A233F/E343G,
Q115E/D143L/K154S/K213R/A233Q/I238V/S261R, Q115E/D143L/A233Q, Q115E/D143L/A233R,
Q115E/D143L/E343G, Q115E/D143V/P151R/A233F/S261R, Q115E/P151R/K154S/A233F,
10 Q115E/P151R/S261R, Q115E/K213R/A233F/I238V, Q115E/K213R/A233R, Q115E/K213R/S261R,
Q115E/A233F/I238V, Q115E/I238V, Q115E/E343G, Q115E/L348F, D143E, D143E/P151R/K154S,
D143E/P151R/K154S/K213R, D143E/K213R/A233F/I238V/S261R/E343G, D143E/A233F/S261R,
D143L/K213R/A233R, D143L/S261R, D143V/K154S, D143V/K213R/A233R,
D143V/K213R/S261R/E343G, P151R/K154S, P151R/K154S/K213R/E343G, P151R/K154S/E343G,
15 P151R/K213R, P151R/K213R/A233F, K213R/A233R/E343G, and A233T/I238V/E343G.

- In some embodiments, the engineered transposase polypeptide with one or more improved
properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
sequence SEQ ID NO: 5196 and one or more residue differences as compared to SEQ ID NO: 5196,
20 selected from 117, 117/118/143/168/169/392, 117/118/143/392, 117/118/168/169, 117/118/168/169/392,
117/143/168/169/170/392, 117/143/168/169/392, 117/168/169, 117/168/169/392, 117/168/392, 117/169,
117/392, 118/143, 118/143/168/169, 118/143/169, 118/143/392, 118/168, 118/168/392, 118/169/392,
118/392, 143/168/169/392, 143/169, 143/392, 168/169, 168/169/392, 168/392, 169/392, and 392. In some
embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at
25 least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,
99% or more sequence identity to reference sequence SEQ ID NO: 5196 and one or more residue
differences as compared to SEQ ID NO: 5196, selected from 117R, 117R/118R/143Q/168K/169T/392A,
117R/118R/143Q/392A, 117R/118R/168K/169M, 117R/118R/168K/169T/392A,
117R/118R/168K/169T/392P, 117R/143Q/168K/169T/170H/392A, 117R/143Q/168K/169T/392A,
30 117R/168K/169T, 117R/168K/169T/392P, 117R/168K/392A, 117R/169T, 117R/392P, 118R/143Q,
118R/143Q/168K/169T, 118R/143Q/169T, 118R/143Q/392A, 118R/143Q/392P, 118R/168K,
118R/168K/392P, 118R/169T/392A, 118R/392A, 118R/392P, 143Q/168K/169T/392P, 143Q/169T,
143Q/392P, 168K/169T, 168K/169T/392A, 168K/169T/392P, 168K/392A, 168K/392P, 169T/392S, 392A,
and 392P. In some embodiments, the engineered transposase polypeptide comprises an amino acid
35 sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%,
96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5196 and one or
more residue differences as compared to SEQ ID NO: 5196, selected from K117R,
K117R/S118R/D143Q/M168K/V169T/C392A, K117R/S118R/D143Q/C392A,
K117R/S118R/M168K/V169M, K117R/S118R/M168K/V169T/C392A,
40 K117R/S118R/M168K/V169T/C392P, K117R/D143Q/M168K/V169T/R170H/C392A,

K117R/D143Q/M168K/V169T/C392A, K117R/M168K/V169T, K117R/M168K/V169T/C392P, K117R/M168K/C392A, K117R/V169T, K117R/C392P, S118R/D143Q, S118R/D143Q/M168K/V169T, S118R/D143Q/V169T, S118R/D143Q/C392A, S118R/D143Q/C392P, S118R/M168K, S118R/M168K/C392P, S118R/V169T/C392A, S118R/C392A, S118R/C392P, D143Q/M168K/V169T/C392P, D143Q/V169T, D143Q/C392P, M168K/V169T, M168K/V169T/C392A, M168K/V169T/C392P, M168K/C392A, M168K/C392P, V169T/C392S, C392A, and C392P.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ

- 10 ID NO: 5286 and one or more residue differences as compared to SEQ ID NO: 5286, selected from 134, 134/182/194/294, 134/182/277/293, 134/182/293/320, 134/182/294, 134/194/277, 134/194/277/317, 134/194/293, 134/194/293/294, 134/194/293/317, 134/194/294, 134/194/317, 134/194/317/320, 134/317, 182, 182/194, 182/194/293/320, 182/293/294, 194, 194/277/293/317, 194/277/317/320, 194/277/320, 194/293/294/317, 194/293/320, 277, 293, 317, 317/320, and 320. In some embodiments, the engineered
- 15 transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5286 and one or more residue differences as compared to SEQ ID NO: 5286, selected from 134P, 134R/182T/277W/293G, 134R/182T/294Q, 134R/194M/277W, 134R/194M/277W/317S, 134R/194M/293P, 134R/194M/293P/294Q, 134R/194M/293P/317S,
- 20 134R/194M/294Q, 134R/194M/317S, 134R/194M/317S/320Y, 134R/317S, 134V, 134V/182T/194M/294Q, 134V/182T/293P/320Y, 134V/194M/277W, 134V/194M/293P, 182T, 182T/194M, 182T/194M/293G/320Y, 182T/293G/294Q, 194M, 194M/277W/293G/317S, 194M/277W/317S/320Y, 194M/277W/320Y, 194M/293P/294Q/317S, 194M/293P/320Y, 277W, 293P, 317S, 317S/320Y, and 320Y. In some
- embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at
- 25 least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5286 and one or more residue differences as compared to SEQ ID NO: 5286, selected from T134P, T134R/S182T/P277W/E293G, T134R/S182T/P294Q, T134R/L194M/P277W, T134R/L194M/P277W/M317S, T134R/L194M/E293P, T134R/L194M/E293P/P294Q, T134R/L194M/E293P/M317S, T134R/L194M/P294Q,
- 30 T134R/L194M/M317S, T134R/L194M/M317S/A320Y, T134R/M317S, T134V, T134V/S182T/L194M/P294Q, T134V/S182T/E293P/A320Y, T134V/L194M/P277W, T134V/L194M/E293P, S182T, S182T/L194M, S182T/L194M/E293G/A320Y, S182T/E293G/P294Q, L194M, L194M/P277W/E293G/M317S, L194M/P277W/M317S/A320Y, L194M/P277W/A320Y, L194M/E293P/P294Q/M317S, L194M/E293P/A320Y, P277W, E293P, M317S, M317S/A320Y, and
- 35 A320Y.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5412 and one or more residue differences as compared to SEQ ID NO: 5412,

40 selected from 12/134, 13/134, 13/134/210, 22/134/152, 22/134/235/344, 64/93/134/235, 64/134/344/347,

77/134, 79/134/210, 84/134, 86/134, 87/134, 89/134, 93/134/152/235/344, 93/134/235/344, 93/134/244/344, 93/134/344, 131, 134/152/344, 134/175, 134/210/382, 134/210/402, 134/210/456, 134/210/478, 134/344, 134/377, 134/378, 134/380, 134/381, 134/382, 134/384, 134/386, 134/387, 134/388, 134/390, 134/391, 134/393, 134/395, 134/398, 134/401, 134/402, 134/406, 134/407, 134/421, 5 134/456, 134/457, 134/459, 134/460, 134/464, 134/467, 134/468, 134/470, 134/471, 134/473, 134/473/474, 134/474, and 134/475. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5412 and one or more residue differences as compared to SEQ ID NO: 5412, selected from 10 12N/134R, 13C/134T/210R, 13L/134R, 13L/134T, 13R/134R, 13S/134R, 13T/134R, 13V/134R, 13V/134T/210R, 22R/134R/152P, 22R/134R/235S/344T, 64R/93T/134R/235S, 64R/134R/344Q/347R, 77Y/134R, 79V/134T/210R, 84V/134R, 86S/134R, 87G/134R, 87M/134R, 89C/134R, 89L/134R, 93T/134R/152P/235S/344T, 93T/134R/235S/344T, 93T/134R/244L/344T, 93T/134R/344T, 131I, 134R/152P/344Q, 134R/175C, 134R/210R/402I, 134R/344T, 134R/377I, 134R/378R, 134R/378V, 15 134R/380V, 134R/381D, 134R/382G, 134R/382K, 134R/382W, 134R/384F, 134R/386W, 134R/387G, 134R/387I, 134R/388F, 134R/390V, 134R/391H, 134R/391R, 134R/391T, 134R/393T, 134R/395I, 134R/395M, 134R/398S, 134R/401V, 134R/402L, 134R/402V, 134R/406C, 134R/407V, 134R/421V, 134R/456D, 134R/456T, 134R/457M, 134R/459L, 134R/460E, 134R/464A, 134R/467L, 134R/467R, 134R/467T, 134R/468S, 134R/470L, 134R/470P, 134R/471M, 134R/471S, 134R/473D/474I, 134R/473G, 20 134R/474T, 134R/475G, 134T/210R/382R, 134T/210R/456C, 134T/210R/478D, and 134T/380V. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5412 and one or more residue differences as compared to SEQ ID NO: 5412, selected from E12N/P134R, H13C/P134T/H210R, 25 H13L/P134R, H13L/P134T, H13R/P134R, H13S/P134R, H13T/P134R, H13V/P134R, H13V/P134T/H210R, P22R/P134R/T152P, P22R/P134R/T235S/E344T, P64R/D93T/P134R/T235S, P64R/P134R/E344Q/N347R, E77Y/P134R, T79V/P134T/H210R, Q84V/P134R, Y86S/P134R, P87G/P134R, P87M/P134R, I89C/P134R, I89L/P134R, D93T/P134R/T152P/T235S/E344T, D93T/P134R/T235S/E344T, D93T/P134R/V244L/E344T, D93T/P134R/E344T, G131I, 30 P134R/T152P/E344Q, P134R/D175C, P134R/H210R/K402I, P134R/E344T, P134R/K377I, P134R/K378R, P134R/K378V, P134R/L380V, P134R/V381D, P134R/D382G, P134R/D382K, P134R/D382W, P134R/A384F, P134R/E386W, P134R/L387G, P134R/L387I, P134R/E388F, P134R/Q390V, P134R/S391H, P134R/S391R, P134R/S391T, P134R/A393T, P134R/V395I, P134R/V395M, P134R/E398S, P134R/W401V, P134R/K402L, P134R/K402V, P134R/K406C, 35 P134R/M407V, P134R/L421V, P134R/Q456D, P134R/Q456T, P134R/S457M, P134R/V459L, P134R/V460E, P134R/V464A, P134R/E467L, P134R/E467R, P134R/E467T, P134R/M468S, P134R/A470L, P134R/A470P, P134R/D471M, P134R/D471S, P134R/E473D/L474I, P134R/E473G, P134R/L474T, P134R/L475G, P134T/H210R/D382R, P134T/H210R/Q456C, P134T/H210R/S478D, and P134T/L380V.

In some embodiments, the engineered transposase polypeptide with one or more improved properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5412 and one or more residue differences as compared to SEQ ID NO: 5412,

5 selected from 73/127/134/160/172/198/290, 73/127/134/160/198/333, 73/127/134/160/210/266, 73/127/134/172/192/266/333, 73/127/134/172/198, 73/127/134/210/290, 73/127/134/210/333, 73/127/134/290/333, 73/127/134/333, 73/134, 73/134/160/172/192/290, 73/134/160/172/198/333, 73/134/160/172/290, 73/134/160/266, 73/134/160/333, 73/134/172, 73/134/172/210/333, 73/134/172/266/333, 73/134/198/210, 73/134/198/210/266, 73/134/210/333, 73/134/333, 73/134/366,

10 127/134, 127/134/160/172/192/333, 127/134/160/172/210/266/290, 127/134/172/198/266/333, 127/134/192/210, 127/134/198/210/290, 127/134/210/266/290, 127/134/210/333, 134, 134/160/172/333, 134/160/198/266/290, 134/172/192/290/333, 134/192/198, 134/198, 134/198/266/333, 134/210, 134/210/290, and 134/333. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%,

15 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5412 and one or more residue differences as compared to SEQ ID NO: 5412, selected from 73K/127L/134R/160G/172R/198Q/290R, 73K/127L/134R/160G/198Q/333T, 73K/127L/134R/172R/192S/266A/333T, 73K/127L/134R/172R/198Q, 73K/127L/134R/333T, 73K/127L/134T/160G/210R/266A, 73K/127L/134T/210R/290R, 73K/127L/134T/210R/333T,

20 73K/127L/134T/290R/333T, 73K/134R, 73K/134R/160G/172R/290R, 73K/134R/160G/333T, 73K/134R/172R, 73K/134R/172R/210R/333T, 73K/134R/172R/266A/333T, 73K/134R/333T, 73K/134R/366H, 73K/134T/160G/172R/192S/290R, 73K/134T/160G/172R/198Q/333T, 73K/134T/160G/266A, 73K/134T/198Q/210R, 73K/134T/198Q/210R/266A, 73K/134T/210R/333T, 127L/134R, 127L/134R/160G/172R/192S/333T, 127L/134R/172R/198Q/266A/333T,

25 127L/134R/210R/266A/290R, 127L/134T, 127L/134T/160G/172R/210R/266A/290R, 127L/134T/192S/210R, 127L/134T/198Q/210R/290R, 127L/134T/210R/333T, 134R, 134R/160G/172R/333T, 134R/172R/192S/290R/333T, 134R/192S/198Q, 134R/198Q, 134R/198Q/266A/333T, 134R/333T, 134T/160G/198Q/266A/290R, 134T/210R, and 134T/210R/290R.

In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5412 and one or more residue differences as compared to SEQ ID NO: 5412, selected from

30 Q73K/I127L/P134R/A160G/I172R/R198Q/K290R, Q73K/I127L/P134R/A160G/R198Q/L333T, Q73K/I127L/P134R/I172R/A192S/V266A/L333T, Q73K/I127L/P134R/I172R/R198Q,

35 Q73K/I127L/P134R/L333T, Q73K/I127L/P134T/A160G/H210R/V266A, Q73K/I127L/P134T/H210R/K290R, Q73K/I127L/P134T/H210R/L333T, Q73K/I127L/P134T/K290R/L333T, Q73K/P134R, Q73K/P134R/A160G/I172R/K290R, Q73K/P134R/A160G/L333T, Q73K/P134R/I172R, Q73K/P134R/I172R/H210R/L333T, Q73K/P134R/I172R/V266A/L333T, Q73K/P134R/L333T, Q73K/P134R/R366H, Q73K/P134T/A160G/I172R/A192S/K290R,

40 Q73K/P134T/A160G/I172R/R198Q/L333T, Q73K/P134T/A160G/V266A, Q73K/P134T/R198Q/H210R,

Q73K/P134T/R198Q/H210R/V266A, Q73K/P134T/H210R/L333T, I127L/P134R,
 I127L/P134R/A160G/I172R/A192S/L333T, I127L/P134R/I172R/R198Q/V266A/L333T,
 I127L/P134R/H210R/V266A/K290R, I127L/P134T, I127L/P134T/A160G/I172R/H210R/V266A/K290R,
 I127L/P134T/A192S/H210R, I127L/P134T/R198Q/H210R/K290R, I127L/P134T/H210R/L333T, P134R,
 5 P134R/A160G/I172R/L333T, P134R/I172R/A192S/K290R/L333T, P134R/A192S/R198Q, P134R/R198Q,
 P134R/R198Q/V266A/L333T, P134R/L333T, P134T/A160G/R198Q/V266A/K290R, P134T/H210R, and
 P134T/H210R/K290R.

In some embodiments, the engineered transposase polypeptide with one or more improved
 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
 10 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
 sequence SEQ ID NO: 5422 and one or more residue differences as compared to SEQ ID NO: 5422,
 selected from 13, 13/178/381, 13/178/381/473, 13/178/457, 13/381, 13/381/457, 13/402/457, 13/457, 178,
 178/381/457, 178/381/457/473, 178/402/457, 178/473, 381, 381/393/457, 381/457/473, 381/473, and
 393/457. In some embodiments, the engineered transposase polypeptide comprises an amino acid
 15 sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%,
 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5422 and one or
 more residue differences as compared to SEQ ID NO: 5422, selected from 13L, 13L/178R/381D/473G,
 13L/178R/381W, 13L/178R/457M, 13L/381D/457M, 13L/381W, 13L/402L/457M, 13L/457M, 178R,
 178R/381D/457M/473G, 178R/381W/457M, 178R/402L/457M, 178R/473G, 381D/457M/473G, 381W,
 20 381W/393T/457M, 381W/473G, and 393T/457M. In some embodiments, the engineered transposase
 polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
 sequence SEQ ID NO: 5422 and one or more residue differences as compared to SEQ ID NO: 5422,
 selected from H13L, H13L/S178R/V381D/E473G, H13L/S178R/V381W, H13L/S178R/S457M,
 25 H13L/V381D/S457M, H13L/V381W, H13L/K402L/S457M, H13L/S457M, S178R,
 S178R/V381D/S457M/E473G, S178R/V381W/S457M, S178R/K402L/S457M, S178R/E473G,
 V381D/S457M/E473G, V381W, V381W/A393T/S457M, V381W/E473G, and A393T/S457M.

In some embodiments, the engineered transposase polypeptide with one or more improved
 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
 30 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
 sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO: 972,
 selected from
 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/194/198/220/243/260/266/27
 0/295/317/333/392/423,
 35 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/29
 5/317/333/392/423,
 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/29
 5/317/333/392/423/457,
 38/73/113/115/117/118/127/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/31
 40 7/333/392/423,

- 38/73/113/115/118/127/152/154/160/162/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423,
38/73/113/115/118/127/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/423,
5 38/73/113/115/118/127/152/154/160/164/168/172/184/189/192/198/220/243/260/266/270/295/317/333/423,
38/73/113/115/118/127/152/154/160/164/168/172/184/192/198/210/220/243/260/266/270/295/317/333/423,
38/73/113/115/118/127/152/154/160/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423,
38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423,
10 38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/243/266/270/317/333/423,
38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/266/270/317/333/423, and
38/113/127/152/154/160/164/168/172/198/220/290/317/423. In some embodiments, the engineered
transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%,
87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to
15 reference sequence SEQ ID NO: 972 and one or more residue differences as compared to SEQ ID NO:
972, selected from
38F/73Q/113E/115E/117R/118R/127I/134R/152T/154K/160A/163V/164R/168K/169T/172I/184A/192A/19
8R/220R/243G/260I/266V/270P/295I/317M/333L/392A/423D,
38F/73Q/113E/115E/117R/118R/127I/134R/152T/154K/160A/163V/164R/168K/169T/172I/184A/192A/19
20 8R/220R/243G/260I/266V/270P/295I/317M/333L/392A/423D/457M,
38F/73Q/113E/115E/117R/118R/127I/134V/152T/154K/160A/163V/164R/168K/169T/172I/184A/192A/194
M/198R/220R/243G/260I/266V/270P/295I/317M/333L/392A/423D,
38F/73Q/113E/115E/117R/118R/127I/152T/154K/160A/163V/164R/168K/169T/172I/184A/192A/198R/22
0R/243G/260I/266V/270P/295I/317M/333L/392A/423D,
25 38F/73Q/113E/115E/118S/127I/152T/154K/160A/163V/164R/168M/169V/172I/184A/192A/198R/220R/24
3G/260I/266V/270P/295I/317M/333L/423D,
38F/73Q/113E/115Q/118S/127I/152T/154K/160A/162L/164R/168K/172I/184A/192A/198R/220R/243G/26
0I/266V/270P/295I/317M/333L/423D,
38F/73Q/113E/115Q/118S/127I/152T/154K/160A/163V/164R/168M/169V/172I/184A/192A/198R/220R/24
30 3G/260I/266V/270P/295I/317M/333L/423D,
38F/73Q/113E/115Q/118S/127I/152T/154K/160A/164R/168K/172I/184A/189S/192A/198R/220R/243G/26
0I/266V/270P/295I/317M/333L/423D,
38F/73Q/113E/115Q/118S/127I/152T/154K/160A/164R/168K/172I/184A/192A/198R/210R/220R/243G/26
0I/266V/270P/295I/317M/333L/423D,
35 38F/73Q/113E/115Q/118S/127I/152T/154K/160A/164R/168K/172I/184A/192A/198R/220R/243G/260I/266
V/270P/295I/317M/333L/423D,
38F/73Q/113E/118S/127I/152T/154K/160A/164R/168K/172I/184A/192A/198R/220R/243G/260I/266V/270
P/295I/317M/333L/423D,
38F/73Q/113E/118S/127I/152T/154K/160A/164R/168K/172I/184A/192A/198R/220R/243G/266V/270P/31
40 7M/333L/423D,

38F/73Q/113E/118S/127I/152T/154K/160A/164R/168K/172I/184A/192A/198R/220R/266V/270P/317M/33
 3L/423D, and 38F/113E/127I/152T/154K/160A/164R/168K/172I/198R/220R/290R/317M/423D. In some
 embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at
 least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,
 5 99% or more sequence identity to reference sequence SEQ ID NO: 972 and one or more residue
 differences as compared to SEQ ID NO: 972, selected from
 C38F/K73Q/S113E/K115E/K117R/A118R/L127I/T134R/E152T/A154K/G160A/A163V/D164R/F168K/C16
 9T/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/C392A/W423D,
 C38F/K73Q/S113E/K115E/K117R/A118R/L127I/T134R/E152T/A154K/G160A/A163V/D164R/F168K/C16
 10 9T/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/C392A/W423D/
 S457M,
 C38F/K73Q/S113E/K115E/K117R/A118R/L127I/T134V/E152T/A154K/G160A/A163V/D164R/F168K/C16
 9T/R172I/C184A/S192A/L194M/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/C392A/
 W423D,
 15 C38F/K73Q/S113E/K115E/K117R/A118R/L127I/E152T/A154K/G160A/A163V/D164R/F168K/C169T/R17
 2I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/C392A/W423D,
 C38F/K73Q/S113E/K115E/A118S/L127I/E152T/A154K/G160A/A163V/D164R/F168M/C169V/R172I/C184
 A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D,
 C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/W162L/D164R/F168K/R172I/C184A/S192
 20 A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D,
 C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/A163V/D164R/F168M/C169V/R172I/C184
 A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D,
 C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/D189S/S192
 A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D,
 25 C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198
 R/H210R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D,
 C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198
 R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D,
 C38F/K73Q/S113E/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198R/P220
 30 R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D,
 C38F/K73Q/S113E/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198R/P220
 R/V243G/A266V/K270P/R317M/T333L/W423D,
 C38F/K73Q/S113E/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198R/P220
 R/A266V/K270P/R317M/T333L/W423D, and
 35 C38F/S113E/L127I/E152T/A154K/G160A/D164R/F168K/R172I/Q198R/P220R/K290R/R317M/W423D.

In some embodiments, the engineered transposase polypeptide with one or more improved
 properties comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%,
 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference
 sequence SEQ ID NO: 5704 and one or more residue differences as compared to SEQ ID NO: 5704,
 40 selected from 48, 116, 118, 120, 130, 137, 153, 157, 163, 175, 181, 187, 192, 213, 246, 263, 273, 296,

318, 334, 341, 394, 412, 424, 454, and 458. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5704 and one or more residue differences as compared to SEQ ID NO: 5704,
 5 selected from 48V, 116E, 118E, 120R, 130I, 137R, 153H, 157K, 163A, 175I, 181R, 187A, 192S, 213R, 246G, 263I, 273P, 296I, 318M, 334L, 341L, 394A, 412N, 424D, 454R, and 458M. In some embodiments, the engineered transposase polypeptide comprises an amino acid sequence having at least 60%, 70%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NO: 5704 and one or more residue differences as
 10 compared to SEQ ID NO: 5706, selected from I48V, S116E, Q118E, K120R, L130I, F137R, D153H, A157K, G163A, R175I, S181R, C187A, D192S, H213R, V246G, L263I, K273P, L296I, I318M, T334L, Q341L, E394A, K412N, W424D, E454R, and S458M.

As will be appreciated by the skilled artisan, in some embodiments, one or a combination of residue differences above that is selected can be kept constant (i.e., maintained) in the engineered
 15 transposase as a core feature, and additional residue differences at other residue positions incorporated into the sequence to generate additional engineered transposase polypeptides with improved properties. Accordingly, it is to be understood for any engineered transposase containing one or a subset of the residue differences above, the present invention contemplates other engineered transposases that comprise the one or subset of the residue differences, and additionally one or more residue differences at
 20 the other residue positions disclosed herein.

As noted above, the engineered transposase polypeptides are also capable of converting substrates (e.g., an adapter or donor polynucleotide and a target polynucleotide) to products (e.g., an annealed or ligated polynucleotide consisting of the adapter and the fragmented target polynucleotide). In some embodiments, the engineered transposase polypeptide is capable of converting the substrate
 25 compounds to the product compound with at least 1.2 fold, 1.5 fold, 2 fold, 3 fold, 4 fold, 5 fold, 10 fold, 20 fold, 30 fold, 40 fold, 50 fold, 60 fold, 70 fold, 80 fold, 90 fold, 100 fold, or more activity relative to the activity of the reference polypeptide of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

30 In some embodiments, the engineered transposase capable of converting the substrate compounds to the product compounds with at least 2-fold the activity relative to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, and comprises an amino acid sequence selected from the polypeptide sequences in SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

35 In some embodiments, the engineered transposase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, that increases soluble expression or isolated protein yield of the engineered transposase in a bacterial host cell, particularly in *E. coli*, as compared to a wild-type or

engineered reference transposase, and comprises an amino acid sequence selected from the polypeptide sequences in SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

In some embodiments, the engineered transposase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, that increases thermostability of the engineered transposase, as compared to a wild-type or engineered reference transposase, and comprises an amino acid sequence selected from the polypeptide sequences in SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

In some embodiments, the engineered transposase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, that increases the activity of the engineered transposase at high temperatures (by way of example and not limitation, 40 °C, 45 °C, 50 °C, 55 °C, 60 °C, or 65 °C), as compared to a wild-type or engineered reference transposase, and comprises an amino acid sequence selected from the polypeptide sequences in SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

In some embodiments, the engineered transposase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, that reduces the sequence insertion bias of the engineered transposase, as compared to a wild-type or engineered reference transposase, and comprises an amino acid sequence selected from the polypeptide sequences in SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

In some embodiments, the engineered transposase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, that reduces the inhibition, as compared to a wild-type or engineered reference transposase, and comprises an amino acid sequence selected from the polypeptide sequences in SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

In some embodiments, the engineered transposase has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, that increases specific activity of the engineered transposase on one or more an adapter or donor polynucleotide and target polynucleotide substrates as compared to a wild-type or engineered reference transposase, and comprises an amino acid sequence selected from the polypeptide sequences in SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

In some embodiments, the engineered transposase with one or more improved properties has an amino acid sequence comprising at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to a sequence selected from the polypeptide sequences of SEQ ID NOs: 2-2368, 2388-5694, and 5706-5756. In some embodiments, the engineered transposase with one or more improved properties has an amino acid sequence comprising at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to a sequence selected from

SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

In some embodiments, the engineered transposase with one or more improved properties has an amino acid sequence comprising a sequence selected from the polypeptide sequences of SEQ ID NOs:

5 4-2368, 2388-5694, and 5706-5756. In some embodiments, the engineered transposase with one or more improved properties has an amino acid sequence comprising a sequence selected from SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

10 In some embodiments, the engineered transposase, comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity to one of the polypeptide sequences of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, as provided in the Examples.

15 In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 4, 10, 396, 400, 608, 2000, 2002, 2014, 2058, 4000, 4012, 4560, 5000, and 5406.

20 In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 8, 40, 308, 900, 1058, 1208, 25 1688, 2246, 3460, 3490, 3982, 4864, 5026, and 5378.

In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino 30 acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 78, 146, 342, 746, 862, 894, 1006, 1668, 2444, 3680, 4782, 5234, 5436, and 5686.

In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 35 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 2, 972, 2368, 2980, 4144, 4642, 4934, 5196, 5286, 5300, 5500, 5656, 5692, and 5702.

In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences 40 selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%,

91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 14, 68, 80, 92, 100, 506, 1004, 2090, 3456, 4562, 4674, 4682, 4790, 5002, and 5604.

5 In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 58, 780, 1008, 1010, 1218, 1788, 2680, 2712, 2686, 3648, 4814, 5034, 5270, 5578, and 5688.

10 In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 4, 90, 98, 302, 1000, 1446, 15 1488, 1806, 2574, 3285, 4672, 4922, 5066, 5646, and 5694.

In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 14, 974, 1162, 1782, 2422, 20 2682, 4680, 4886, 4924, 5132, 5346, 5422, 5588, 5624, and 5680.

In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13) polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 25 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence identities to, comprising the amino acid sequences of, or consisting the amino acid sequences of SEQ ID NOs: 5706-5756. In some embodiments, exemplary engineered transposases useful in the methods described herein comprise one or more polypeptide sequences selected from polypeptide sequences comprising at least 85% (e.g., 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more) sequence 30 identities to the amino acid sequences of SEQ ID NOs: 5706-5756, wherein the polypeptide sequence has at least one mutation selected from the group consisting of I48V, S116E, Q118E, K120R, L130I, F137R, D153H, A157K, G163A, R175I, S181R, C187A, D192S, H213R, V246G, L263I, K273P, L296I, I318M, T334L, Q341L, E394A, K412N, W424D, E454R, and S458M. In addition to the residue positions specified above, any of the engineered transposase polypeptides disclosed herein can further 35 comprise other residue differences relative to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, at other residue positions (i.e., residue positions other than those included herein). Residue differences at these other residue positions can provide for additional variations in the amino acid sequence without adversely affecting the ability of the polypeptide to carry out the conversion 40 of substrate to product. Accordingly, in some embodiments, in addition to the amino acid residue

differences present in any one of the engineered transposase polypeptides selected from the polypeptide sequences in the range of SEQ ID NOs: 2-2368, 2388-5694, and 5706-5756, the sequence can further comprise 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-11, 1-12, 1-14, 1-15, 1-16, 1-18, 1-20, 1-22, 1-24, 1-26, 1-30, 1-35, 1-40, 1-45, 1-50, 1-100, or 1-150 residue differences at other amino acid residue positions as compared to the SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 100, or 150 residue positions. In some embodiments, the number of amino acid residue differences as compared to the reference sequence can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 residue positions. The residue differences at these other positions can be conservative changes or non-conservative changes. In some embodiments, the residue differences can comprise conservative substitutions and non-conservative substitutions as compared to the transposase polypeptide of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

In some embodiments, the present invention also provides engineered polypeptides that comprise a fragment of any of the engineered transposase polypeptides described herein that retains the functional activity and/or improved property of that engineered transposase. Accordingly, in some embodiments, the present invention provides a polypeptide fragment capable of converting substrate to product under suitable reaction conditions, wherein the fragment comprises at least about 90%, 95%, 96%, 97%, 98%, or 99% of a full-length or truncated amino acid sequence of an engineered transposase of the present invention, such as an exemplary transposase polypeptide selected from the polypeptide sequences in the range of SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756. In some embodiments, the engineered transposase can have an amino acid sequence comprising a deletion in any one of the transposase polypeptide sequences described herein, such as the exemplary engineered polypeptides of the polypeptide sequences in the range of SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756.

Thus, for each and every embodiment of the engineered transposase polypeptides of the invention, the amino acid sequence can comprise deletions of one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, up to 20% of the total number of amino acids, or up to 30% of the total number of amino acids of the transposase polypeptides, where the associated functional activity and/or improved properties of the engineered transposase described herein are maintained. In some embodiments, the deletions can comprise 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 amino acid residues. In some embodiments, the number of deletions can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, or 50 amino acid residues. In some embodiments, the deletions can comprise deletions of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 amino acid residues.

In some embodiments, the engineered transposase polypeptide herein can have an amino acid sequence comprising an insertion as compared to any one of the engineered transposase polypeptides described herein, such as the exemplary engineered polypeptides of the polypeptide sequences in the range of SEQ ID NOs: 2-2368, 2388-5694, and 5706-5756. Thus, for each and every embodiment of the transposase polypeptides of the invention, the insertions can comprise one or more amino acids, 2 or more amino acids, 3 or more amino acids, 4 or more amino acids, 5 or more amino acids, 6 or more amino acids, 8 or more amino acids, 10 or more amino acids, 15 or more amino acids, 20 or more amino acids, 30 or more amino acids, 40 or more amino acids, or 50 or more amino acids, where the associated functional activity and/or improved properties of the engineered transposase described herein is maintained. The insertions can be to amino or carboxy terminus, or internal portions of the transposase polypeptide.

In some embodiments, the engineered transposase described herein can have an amino acid sequence comprising a sequence selected from the polypeptide sequences in the range of SEQ ID NOs: 4-2368, 2388-5694, and 5706-5756, and optionally one or several (e.g., up to 3, 4, 5, or up to 10) amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, 1-50, 1-75, 1-100, or 1-150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally around 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

In the above embodiments, the suitable reaction conditions for the engineered polypeptides are provided in Tables 5.1, 6.1, 7.1, 8.1, 9.1, 10.1, 10.4, 11.1, 11.3, 12.1, 12.4, 13.1, 14.1, 14.4, 14.7, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3, and as described in the Examples herein.

In some embodiments, the polypeptides of the present invention are fusion polypeptides in which the engineered polypeptides are fused to other polypeptides, such as, by way of example and not limitation, antibody tags (e.g., myc epitope), purification sequences (e.g., His tags for binding to metals), and cell localization signals (e.g., secretion signals). Thus, the engineered polypeptides described herein can be used with or without fusions to other polypeptides.

It is to be understood that the polypeptides described herein are not restricted to the genetically encoded amino acids. In addition to the genetically encoded amino acids, the polypeptides described herein may be comprised, either in whole or in part, of naturally occurring and/or synthetic non-encoded amino acids. Certain commonly encountered non-encoded amino acids of which the polypeptides described herein may be comprised include, but are not limited to: the D-stereoisomers of the genetically-encoded amino acids; 2,3-diaminopropionic acid (Dpr); α -aminoisobutyric acid (Aib); ϵ -aminohexanoic acid (Aha); δ -aminovaleric acid (Ava); N-methylglycine or sarcosine (MeGly or Sar); ornithine (Orn); citrulline (Cit); t-butylalanine (Bua); t-butylglycine (Bug); N-methylisoleucine (Melle); phenylglycine (Phg); cyclohexylalanine (Cha); norleucine (Nle); naphthylalanine (Nal); 2-chlorophenylalanine (Ocf); 3-chlorophenylalanine (Mcf); 4-chlorophenylalanine (Pcf); 2-fluorophenylalanine (Off); 3-fluorophenylalanine

(Mff); 4-fluorophenylalanine (Pff); 2-bromophenylalanine (Obf); 3-bromophenylalanine (Mbf); 4-bromophenylalanine (Pbf); 2-methylphenylalanine (Omf); 3-methylphenylalanine (Mmf); 4-methylphenylalanine (Pmf); 2-nitrophenylalanine (Onf); 3-nitrophenylalanine (Mnf); 4-nitrophenylalanine (Pnf); 2-cyanophenylalanine (Ocf); 3-cyanophenylalanine (Mcf); 4-cyanophenylalanine (Pcf); 2-trifluoromethylphenylalanine (Otf); 3-trifluoromethylphenylalanine (Mtf); 4-trifluoromethylphenylalanine (Ptf); 4-aminophenylalanine (Paf); 4-iodophenylalanine (Pif); 4-aminomethylphenylalanine (Pamf); 2,4-dichlorophenylalanine (Opef); 3,4-dichlorophenylalanine (Mpcf); 2,4-difluorophenylalanine (Opff); 3,4-difluorophenylalanine (Mpff); pyrid-2-ylalanine (2pAla); pyrid-3-ylalanine (3pAla); pyrid-4-ylalanine (4pAla); naphth-1-ylalanine (1nAla); naphth-2-ylalanine (2nAla); thiazolylalanine (taAla); benzothienylalanine (bAla); thienylalanine (tAla); furylalanine (fAla); homophenylalanine (hPhe); homotyrosine (hTyr); homotryptophan (hTrp); pentafluorophenylalanine (5ff); styrylalanine (sAla); authrylalanine (aAla); 3,3-diphenylalanine (Dfa); 3-amino-5-phenylpentanoic acid (Afp); penicillamine (Pen); 1,2,3,4-tetrahydroisoquinoline-3-carboxylic acid (Tic); β -2-thienylalanine (Thi); methionine sulfoxide (Mso); N(w)-nitroarginine (nArg); homolysine (hLys); phosphonomethylphenylalanine (pmPhe); phosphoserine (pSer); phosphothreonine (pThr); homoaspartic acid (hAsp); homoglutamic acid (hGlu); 1-aminocyclopent-(2 or 3)-ene-4 carboxylic acid; pipecolic acid (PA), azetidine-3-carboxylic acid (ACA); 1-aminocyclopentane-3-carboxylic acid; allylglycine (aGly); propargylglycine (pgGly); homoalanine (hAla); norvaline (nVal); homoleucine (hLeu), homovaline (hVal); homoisoleucine (hIle); homoarginine (hArg); N-acetyl lysine (AcLys); 2,4-diaminobutyric acid (Dbu); 2,3-diaminobutyric acid (Dab); N-methylvaline (MeVal); homocysteine (hCys); homoserine (hSer); hydroxyproline (Hyp) and homoproline (hPro). Additional non-encoded amino acids of which the polypeptides described herein may be comprised will be apparent to those of skill in the art (See e.g., the various amino acids provided in Fasman, CRC Practical Handbook of Biochemistry and Molecular Biology, CRC Press, Boca Raton, FL, pp. 3-70 [1989], and the references cited therein, all of which are incorporated by reference). These amino acids may be in either the L- or D-configuration.

Those of skill in the art will recognize that amino acids or residues bearing side chain protecting groups may also comprise the polypeptides described herein. Non-limiting examples of such protected amino acids, which in this case belong to the aromatic category, include (protecting groups listed in parentheses), but are not limited to: Arg(tos), Cys(methylbenzyl), Cys (nitropyridinesulfonyl), Glu(δ -benzylester), Gln(xanthyl), Asn(N- δ -xanthyl), His(bom), His(benzyl), His(tos), Lys(fmoc), Lys(tos), Ser(O-benzyl), Thr (O-benzyl) and Tyr(O-benzyl).

Non-encoding amino acids that are conformationally constrained of which the polypeptides described herein may be composed include, but are not limited to, N-methyl amino acids (L-configuration); 1-aminocyclopent-(2 or 3)-ene-4-carboxylic acid; pipecolic acid; azetidine-3-carboxylic acid; homoproline (hPro); and 1-aminocyclopentane-3-carboxylic acid.

In some embodiments, the engineered polypeptides can be in various forms, for example, such as an isolated preparation, as a substantially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, and/or as cell extracts and/or lysates of such cells. The enzymes can be lyophilized, spray-dried, precipitated or be in the form of a crude paste, as further discussed below.

In some embodiments, the engineered polypeptides can be in the form of a biocatalytic composition. In some embodiments, the biocatalytic composition comprises (a) a means for conversion of an adapter or donor polynucleotide and a target polynucleotide to an annealed or ligated polynucleotide consisting of the adapter and the optionally fragmented target polynucleotide by contact with a
5 transposase and (b) a suitable cofactor. The suitable cofactor may be magnesium, manganese, or any other suitable cofactor.

In some embodiments, the polypeptides described herein are provided in the form of kits. The enzymes in the kits may be present individually or as a plurality of enzymes. The kits can further include reagents for carrying out the enzymatic reactions, substrates for assessing the activity of enzymes, as
10 well as reagents for detecting the products. The kits can also include reagent dispensers and instructions for use of the kits.

In some embodiments, the kits of the present invention include arrays comprising a plurality of different transposase polypeptides at different addressable position, wherein the different polypeptides are different variants of a reference sequence each having at least one different improved enzyme property.

In some embodiments, a plurality of polypeptides immobilized on solid supports are configured on an array at various locations, addressable for robotic delivery of reagents, or by detection methods and/or instruments. The array can be used to test a variety of substrate compounds for conversion by the polypeptides. Such arrays comprising a plurality of engineered polypeptides and methods of their use are known in the art (See e.g., WO2009/008908A2).

Polynucleotides Encoding Engineered Transposases, Expression Vectors and Host Cells

In another aspect, the present invention provides polynucleotides encoding the engineered transposase polypeptides described herein. The polynucleotides may be operatively linked to one or more heterologous regulatory sequences that control gene expression to create a recombinant polynucleotide
25 capable of expressing the polypeptide. Expression constructs containing a heterologous polynucleotide encoding the engineered transposase are introduced into appropriate host cells to express the corresponding transposase polypeptide.

As will be apparent to the skilled artisan, availability of a protein sequence and the knowledge of the codons corresponding to the various amino acids provide a description of all the polynucleotides
30 capable of encoding the subject polypeptides. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons, allows an extremely large number of nucleic acids to be made, all of which encode the improved transposase enzymes. Thus, having knowledge of a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the sequence of one or more codons in a way which does not
35 change the amino acid sequence of the protein. In this regard, the present invention specifically contemplates each and every possible variation of polynucleotides that could be made encoding the polypeptides described herein by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide described herein, including the amino acid sequences presented in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5,
40 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3,

18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3, and disclosed in the sequence listing incorporated by reference herein as the polypeptide sequences in the range of SEQ ID NOs: 2-2368, 2388-5694, and 5706-5756.

In various embodiments, the codons are preferably selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria are used to express the gene in bacteria; preferred codons used in yeast are used for expression in yeast; and preferred codons used in mammals are used for expression in mammalian cells. In some embodiments, all codons need not be replaced to optimize the codon usage of the transposase since the natural sequence will comprise preferred codons and because use of preferred codons may not be required for all amino acid residues. Consequently, codon optimized polynucleotides encoding the transposase enzymes may contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full-length coding region.

In some embodiments, the polynucleotide comprises a codon optimized nucleotide sequence encoding the transposase polypeptide amino acid sequence, as represented by SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences encoding the polypeptide sequences in the range of SEQ ID NOs: 2-2368, 2388-5694, and 5706-5756. In some embodiments, the polynucleotide has a nucleic acid sequence comprising at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequences in the polynucleotide sequences in the range of SEQ ID NOs: 1-2367, 2387-5693, and 5705-5755. In some embodiments, the codon optimized sequences of the polynucleotide sequences in the range of SEQ ID NOs: 1-2367, 2387-5693, and 5705-5755, enhance expression of the encoded transposase, providing preparations of enzyme capable of converting substrate to product.

In some embodiments, the polynucleotide sequence comprises at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 1, 3, 155, 301, 337, 605, 971, 1081, 1217, 1579, 1597, 1599, 1603, 1719, 2367, 2425, 2539, 3401, 3661, 3787, 4149, 4277, 4641, 4983, 5195, 5285, 5411, 5421, and/or 5703, and/or a functional fragment thereof, wherein said polynucleotide sequence encodes an engineered polypeptide comprising at least one substitution at one or more amino acid positions.

In some embodiments, the polynucleotide sequence encodes at least one engineered transposase comprising a sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

In some embodiments, the polynucleotide sequence comprises SEQ ID NOs: 1, 3, 155, 301, 337, 605, 971, 1081, 1217, 1579, 1597, 1599, 1603, 1719, 2367, 2425, 2539, 3401, 3661, 3787, 4149, 4277, 4641, 4983, 5195, 5285, 5411, 5421, and/or 5703.

In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference sequence selected from the polynucleotide sequences in SEQ ID NOs: 1-2367, 2387-5693, and 5705-5755, or a complement thereof, and encode a transposase.

In some embodiments, as described above, the polynucleotide encodes an engineered
5 transposase polypeptide with improved properties as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, wherein the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to a reference sequence selected from SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972,
10 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, and one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, wherein the sequence is selected from the polypeptide sequences in the range of SEQ ID NOs: 2-2368, 2388-5694, and 5706-
15 5756. In some embodiments, the reference amino acid sequence is selected from the polypeptide sequences in the range of SEQ ID NOs: 2-2368, 2388-5694, and 5704. In some embodiments, the reference amino acid sequence is SEQ ID NO: 2, while in some other embodiments, the reference sequence is SEQ ID NO: 4, while in some other embodiments, the reference sequence is SEQ ID NO: 156. In some embodiments, the reference amino acid sequence is SEQ ID NO: 302, while in some other
20 embodiments, the reference sequence is SEQ ID NO: 338, while in some other embodiments, the reference sequence is SEQ ID NO: 606. In some embodiments, the reference amino acid sequence is SEQ ID NO: 972, while in some other embodiments, the reference sequence is SEQ ID NO: 1082, while in some other embodiments, the reference sequence is SEQ ID NO: 1218. In some embodiments, the reference amino acid sequence is SEQ ID NO: 1580, while in some other embodiments, the reference
25 sequence is SEQ ID NO: 1598, while in some other embodiments, the reference sequence is SEQ ID NO: 1600. In some embodiments, the reference amino acid sequence is SEQ ID NO: 1604, while in some other embodiments, the reference sequence is SEQ ID NO: 1720, while in some other embodiments, the reference sequence is SEQ ID NO: 2368. In some embodiments, the reference amino acid sequence is SEQ ID NO: 2426, while in some other embodiments, the reference sequence is SEQ ID NO: 2540, while
30 in some other embodiments, the reference sequence is SEQ ID NO: 3402. In some embodiments, the reference amino acid sequence is SEQ ID NO: 3662, while in some other embodiments, the reference sequence is SEQ ID NO: 3788, while in some other embodiments, the reference sequence is SEQ ID NO: 4150. In some embodiments, the reference amino acid sequence is SEQ ID NO: 4278, while in some other embodiments, the reference sequence is SEQ ID NO: 4642, while in some other embodiments, the
35 reference sequence is SEQ ID NO: 4984. In some embodiments, the reference amino acid sequence is SEQ ID NO: 5196, while in some other embodiments, the reference sequence is SEQ ID NO: 5286, while in some other embodiments, the reference sequence is SEQ ID NO: 5412. In some embodiments, the reference amino acid sequence is SEQ ID NO: 5422. In some embodiments, the reference amino acid sequence is SEQ ID NO: 5500. In some embodiments, the reference amino acid sequence is SEQ ID NO:
40 5704.

In some embodiments, the polynucleotide encodes a transposase polypeptide capable of converting one or more substrates to product with improved properties as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, wherein the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to reference sequence SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

In some embodiments, the polynucleotide encoding the engineered transposase comprises a polynucleotide sequence selected from the polynucleotide sequences in the range of SEQ ID NOs: 1-2367, 2387-5693, and 5705-5755.

In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from the polynucleotide sequences in the range of SEQ ID NOs: 1-2367, 2387-5693, and 5705-5755 or a complement thereof, and encode a transposase polypeptide with one or more of the improved properties described herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a transposase comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, that has an amino acid sequence comprising one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, as described above and in the Examples, below.

In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transposase polypeptide with improved properties comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered transposase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 1-2367, 2387-5693, and 5705-5755.

In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes an engineered transposase polypeptide with improved properties comprising an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704. In some embodiments, the polynucleotides encode the polypeptides described herein but have at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

or 99% or more sequence identity at the nucleotide level to a reference polynucleotide encoding the engineered transposase. In some embodiments, the reference polynucleotide sequence is selected from SEQ ID NOs: 1-2367, 2387-5693, and 5705-5755.

In some embodiments, an isolated polynucleotide encoding any of the engineered transposase polypeptides provided herein is manipulated in a variety of ways to provide for expression of the polypeptide. In some embodiments, the polynucleotides encoding the polypeptides are provided as expression vectors where one or more control sequences is present to regulate the expression of the polynucleotides and/or polypeptides. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

In some embodiments, the control sequences include among other sequences, promoters, leader sequences, polyadenylation sequences, propeptide sequences, signal peptide sequences, and transcription terminators. As known in the art, suitable promoters can be selected based on the host cells used. For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present application, include, but are not limited to the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (*dagA*), *Bacillus subtilis* levansucrase gene (*sacB*), *Bacillus licheniformis* alpha-amylase gene (*amyL*), *Bacillus stearothermophilus* maltogenic amylase gene (*amyM*), *Bacillus amyloliquefaciens* alpha-amylase gene (*amyQ*), *Bacillus licheniformis* penicillinase gene (*penP*), *Bacillus subtilis* *xylA* and *xylB* genes, and prokaryotic beta-lactamase gene (See e.g., Villa-Kamaroff et al., Proc. Natl Acad. Sci. USA 75: 3727-3731 [1978]), as well as the *tac* promoter (See e.g., DeBoer et al., Proc. Natl Acad. Sci. USA 80: 21-25 [1983]). Exemplary promoters for filamentous fungal host cells, include promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (*glaA*), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (See e.g., WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof. Exemplary yeast cell promoters can be from the genes can be from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are known in the art (See e.g., Romanos et al., Yeast 8:423-488 [1992]).

In some embodiments, the control sequence is a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice finds use in the present invention. For example, exemplary transcription terminators for filamentous fungal host cells can be obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease. Exemplary

terminators for yeast host cells can be obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are known in the art (See e.g., Romanos et al., *supra*).

5 In some embodiments, the control sequence is a suitable leader sequence, a non-translated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used. Exemplary leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose
10 phosphate isomerase. Suitable leaders for yeast host cells include, but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP). The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid
15 sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells include, but are not limited to those from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-
20 like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are also known in the art (See e.g., Guo and Sherman, *Mol. Cell. Bio.*, 15:5983-5990 [1995]).

In some embodiments, the control sequence is a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may
25 inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. Any signal peptide coding region that directs the expressed polypeptide into the secretory pathway of a host cell of choice finds use for expression of the engineered transposase polypeptides provided herein.

30 Effective signal peptide coding regions for bacterial host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NCIB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. Further signal peptides are known in the art (See e.g., Simonen and Palva, *Microbiol. Rev.*, 57:109-137 [1993]).

35 Effective signal peptide coding regions for filamentous fungal host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase. Useful signal peptides for yeast host cells include, but are not limited to those from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces*
40 *cerevisiae* invertase.

In some embodiments, the control sequence is a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is referred to as a "proenzyme," "propeptide," or "zymogen," in some cases). A propeptide can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propeptide. The propeptide coding region includes, but is not limited to the genes for *Bacillus subtilis* alkaline protease (*aprE*), *Bacillus subtilis* neutral protease (*nprT*), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (See e.g., WO 95/33836). Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

In some embodiments, regulatory sequences are also utilized. These sequences facilitate the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include, but are not limited to the *lac*, *tac*, and *trp* operator systems. In yeast host cells, suitable regulatory systems include, but are not limited to the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include, but are not limited to the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

The present invention also provides recombinant expression vectors comprising a polynucleotide encoding an engineered transposase polypeptide, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. In some embodiments, the various nucleic acid and control sequences described above are combined together to produce a recombinant expression vector which includes one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the variant transposase polypeptide at such sites. Alternatively, the polynucleotide sequence(s) of the present invention are expressed by inserting the polynucleotide sequence or a nucleic acid construct comprising the polynucleotide sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

The recombinant expression vector may be any vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and can result in the expression of the variant transposase polynucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extra-chromosomal element, a minichromosome, or an artificial chromosome). The vector may contain any means for assuring self-replication. In some alternative embodiments, the vector may be one which, when introduced into the host cell, is integrated into the

genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

In some embodiments, the expression vector preferably contains one or more selectable markers, which permit easy selection of transformed cells. A "selectable marker" is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophy, and the like. Examples of bacterial selectable markers include, but are not limited to the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferases), *bar* (phosphinothricin acetyltransferase), *hph* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenylyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof. In another aspect, the present invention provides a host cell comprising a polynucleotide encoding at least one engineered transposase polypeptide of the present invention, the polynucleotide being operatively linked to one or more control sequences for expression of the engineered transposase enzyme(s) in the host cell. Host cells for use in expressing the polypeptides encoded by the expression vectors of the present invention are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, *Vibrio fluvialis*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* and *Pichia pastoris* [ATCC Accession No. 201178]); insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, BHK, 293, and Bowes melanoma cells; and plant cells. Exemplary host cells are *Escherichia coli* strains (e.g., W3110 (Δ fhuA) and BL21).

In some embodiments, the host cell strain comprises a knockout of one or more genes, in particular phosphatase genes. In some embodiments, the host cell comprises a knockout or single gene deletion of *E. coli* genes *aphA*, *surE*, *phoA*, and/or *cpdB*, as described below in the Examples. In some embodiments, the host cell comprising a knockout of one or more phosphatase genes has increased production of the product and/or decreased de-phosphorylation of the product or substrate.

Accordingly, in another aspect, the present invention provides methods for producing the engineered transposase polypeptides, where the methods comprise culturing a host cell capable of expressing a polynucleotide encoding the engineered transposase polypeptide under conditions suitable for expression of the polypeptide. In some embodiments, the methods further comprise the steps of isolating and/or purifying the transposase polypeptides, as described herein.

Appropriate culture media and growth conditions for the above-described host cells are well known in the art. Polynucleotides for expression of the transposase polypeptides may be introduced into cells by various methods known in the art. Techniques include, among others, electroporation, biolistic particle bombardment, liposome mediated transfection, calcium chloride transfection, and protoplast fusion.

The engineered transposases with the properties disclosed herein can be obtained by subjecting the polynucleotide encoding the naturally occurring or engineered transposase polypeptide to

mutagenesis and/or directed evolution methods known in the art, and as described herein. An exemplary directed evolution technique is mutagenesis and/or DNA shuffling (See e.g., Stemmer, Proc. Natl. Acad. Sci. USA 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and U.S. Pat. 6,537,746). Other directed evolution procedures that can be used include, among others, staggered extension process (StEP), in vitro recombination (See e.g., Zhao et al., Nat. Biotechnol., 16:258-261 [1998]), mutagenic PCR (See e.g., Caldwell et al., PCR Methods Appl., 3:S136-S140 [1994]), and cassette mutagenesis (See e.g., Black et al., Proc. Natl. Acad. Sci. USA 93:3525-3529 [1996]).

For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed. Mutagenesis and directed evolution methods are well known in the art (See e.g., US Patent Nos. 5,605,793, 5,811,238, 5,830,721, 5,834,252, 5,837,458, 5,928,905, 6,096,548, 6,117,679, 6,132,970, 6,165,793, 6,180,406, 6,251,674, 6,265,201, 6,277,638, 6,287,861, 6,287,862, 6,291,242, 6,297,053, 6,303,344, 6,309,883, 6,319,713, 6,319,714, 6,323,030, 6,326,204, 6,335,160, 6,335,198, 6,344,356, 6,352,859, 6,355,484, 6,358,740, 6,358,742, 6,365,377, 6,365,408, 6,368,861, 6,372,497, 6,337,186, 6,376,246, 6,379,964, 6,387,702, 6,391,552, 6,391,640, 6,395,547, 6,406,855, 6,406,910, 6,413,745, 6,413,774, 6,420,175, 6,423,542, 6,426,224, 6,436,675, 6,444,468, 6,455,253, 6,479,652, 6,482,647, 6,483,011, 6,484,105, 6,489,146, 6,500,617, 6,500,639, 6,506,602, 6,506,603, 6,518,065, 6,519,065, 6,521,453, 6,528,311, 6,537,746, 6,573,098, 6,576,467, 6,579,678, 6,586,182, 6,602,986, 6,605,430, 6,613,514, 6,653,072, 6,686,515, 6,703,240, 6,716,631, 6,825,001, 6,902,922, 6,917,882, 6,946,296, 6,961,664, 6,995,017, 7,024,312, 7,058,515, 7,105,297, 7,148,054, 7,220,566, 7,288,375, 7,384,387, 7,421,347, 7,430,477, 7,462,469, 7,534,564, 7,620,500, 7,620,502, 7,629,170, 7,702,464, 7,747,391, 7,747,393, 7,751,986, 7,776,598, 7,783,428, 7,795,030, 7,853,410, 7,868,138, 7,783,428, 7,873,477, 7,873,499, 7,904,249, 7,957,912, 7,981,614, 8,014,961, 8,029,988, 8,048,674, 8,058,001, 8,076,138, 8,108,150, 8,170,806, 8,224,580, 8,377,681, 8,383,346, 8,457,903, 8,504,498, 8,589,085, 8,762,066, 8,768,871, 9,593,326, and all related US, as well as PCT and non-US counterparts; Ling et al., Anal. Biochem., 254(2):157-78 [1997]; Dale et al., Meth. Mol. Biol., 57:369-74 [1996]; Smith, Ann. Rev. Genet., 19:423-462 [1985]; Botstein et al., Science, 229:1193-1201 [1985]; Carter, Biochem. J., 237:1-7 [1986]; Kramer et al., Cell, 38:879-887 [1984]; Wells et al., Gene, 34:315-323 [1985]; Minshull et al., Curr. Op. Chem. Biol., 3:284-290 [1999]; Christians et al., Nat. Biotechnol., 17:259-264 [1999]; Cramer et al., Nature, 391:288-291 [1998]; Cramer et al., Nat. Biotechnol., 15:436-438 [1997]; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 94:4504-4509 [1997]; Cramer et al., Nat. Biotechnol., 14:315-319 [1996]; Stemmer, Nature, 370:389-391 [1994]; Stemmer, Proc. Nat. Acad. Sci. USA, 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; and WO 2009/152336, all of which are incorporated herein by reference).

In some embodiments, the enzyme clones obtained following mutagenesis treatment are screened by subjecting the enzymes to a defined temperature (or other assay conditions, such as testing the enzyme's activity over a broad range of substrates) and measuring the amount of enzyme activity remaining after heat treatments or other assay conditions. Clones containing a polynucleotide encoding a transposase polypeptide are then sequenced to identify the nucleotide sequence changes (if any), and

used to express the enzyme in a host cell. Measuring enzyme activity from the expression libraries can be performed using any suitable method known in the art (e.g., standard biochemistry techniques, such as HPLC analysis).

In some embodiments, the clones obtained following mutagenesis treatment can be screened for engineered transposases having one or more desired improved enzyme properties (e.g., reduced sequence bias). Measuring enzyme activity from the expression libraries can be performed using the standard biochemistry techniques, such as HPLC analysis, LC-MS analysis, RapidFire-MS analysis, and/or capillary electrophoresis analysis.

When the sequence of the engineered polypeptide is known, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides encoding portions of the transposase can be prepared by chemical synthesis as known in the art (e.g., the classical phosphoramidite method of Beaucage et al., Tet. Lett. 22:1859-69 [1981], or the method described by Matthes et al., EMBO J. 3:801-05 [1984]) as typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors. In addition, essentially any nucleic acid can be obtained from any of a variety of commercial sources. In some embodiments, additional variations can be created by synthesizing oligonucleotides containing deletions, insertions, and/or substitutions, and combining the oligonucleotides in various permutations to create engineered transposases with improved properties.

Accordingly, in some embodiments, a method for preparing the engineered transposase polypeptide comprises: (a) synthesizing a polynucleotide encoding a polypeptide comprising an amino acid sequence having at least about 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% or more sequence identity to an amino acid sequence selected from the polypeptide sequences of SEQ ID NOs: 2-2368 and 5704-5756, and having one or more residue differences as compared to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704; and (b) expressing the transposase polypeptide encoded by the polynucleotide.

In some embodiments of the method, the polynucleotide encodes an engineered transposase that has optionally one or several (e.g., up to 3, 4, 5, or up to 10) amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, 1-50, 1-75, 1-100, or 1-150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally around 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, 50, 55, 60, 65, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

In some embodiments, any of the engineered transposase enzymes expressed in a host cell can be recovered from the cells and/or the culture medium using any one or more of the well-known techniques for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography. Suitable solutions for lysing and the high efficiency extraction of proteins from bacteria, such as *E. coli*, are commercially available (e.g., CellLytic B™, Sigma-Aldrich, St. Louis MO).

Chromatographic techniques for isolation of the transposase polypeptide include, among others, reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme will depend, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., and will be apparent to those having skill in the art.

In some embodiments, affinity techniques may be used to isolate the improved transposase enzymes. For affinity chromatography purification, any antibody which specifically binds the transposase polypeptide may be used. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., may be immunized by injection with a transposase polypeptide, or a fragment thereof. The transposase polypeptide or fragment may be attached to a suitable carrier, such as BSA, by means of a side chain functional group or linkers attached to a side chain functional group. In some embodiments, the affinity purification can use a specific ligand bound by the transposase or dye affinity column (See e.g., EP0641862; Stellwagen, "Dye Affinity Chromatography," In Current Protocols in Protein Science, Unit 9.2-9.2.16 [2001]).

In some embodiments, a polynucleotide that encodes an engineered transposase (e.g., any one of the engineered transposases described herein, such as in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5, 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3) is modified to replace the segment of nucleotides that encode the C-terminal 6-histidine tag with a different tag (e.g., a different affinity tag). In some embodiments, the polynucleotide that encodes an engineered transposase (e.g., any one of the engineered transposases described herein, such as in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5, 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3) is modified to encode an engineered transposase with one or more additional tags (e.g., one or more affinity tags), such as the C-terminal 6-histidine tag and one or more additional tags. Tags such as affinity tags (e.g., affinity tags suitable for protein purification) are known in the art and include, for instance, a polyhistidine tag (e.g., five or more histidine residues), a polyarginine tag (e.g., five or more arginine residues), and a FLAG tag.

In some embodiments, a polynucleotide that encodes an engineered transposase (e.g., any one of the engineered transposases described herein, such as in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5, 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3) is modified to replace the segment of nucleotides

that encodes the C-terminal GGSG (SEQ ID NO: 5777) peptide spacer with a different peptide spacer. Suitable peptide spacers are known in the art and comprise flexible amino acid residues such as glycine, alanine, and serine.

In some embodiments, a peptide spacer includes motifs, such as multiple or repeating motifs, of
 5 GA, GS, GG, GGA, GGS, GGG, GGGA (SEQ ID NO: 5778), GGGS (SEQ ID NO: 5779), GGGG (SEQ ID NO: 5780), GGGGA (SEQ ID NO: 5781), GGGGS (SEQ ID NO: 5782), GGGGG (SEQ ID NO: 5783), GGAG (SEQ ID NO: 5784), AGGG (SEQ ID NO: 5785), or SGGG (SEQ ID NO: 5786). In some embodiments, a peptide spacer includes 2 to 12 amino acids including motifs of GA or GS, e.g., GA, GS, GAGA (SEQ ID NO: 5787), GSGS (SEQ ID NO: 5788), GAGAGA (SEQ ID NO: 5789), GSGSGS (SEQ ID NO: 5790), GAGAGAGA (SEQ ID NO: 5791), GSGSGSGS (SEQ ID NO: 5792), GAGAGAGAGA (SEQ ID NO: 5793), GSGSGSGSGS (SEQ ID NO: 5794), GAGAGAGAGAGA (SEQ ID NO: 5795), and GSGSGSGSGSGS (SEQ ID NO: 5796).

In some embodiments, a peptide spacer includes 3 to 12 amino acids including motifs of GGA or GGS, e.g., GGA, GGS, GGAGGA (SEQ ID NO: 5797), GSGGGS (SEQ ID NO: 5798), GGAGGAGGA (SEQ ID NO: 5799), GSGGSGGS (SEQ ID NO: 5780), GGAGGAGGAGGA (SEQ ID NO: 5801), and GSGGSGGSGGS (SEQ ID NO: 5802).

In yet some embodiments, a peptide spacer includes 4 to 12 amino acids including motifs of GGAG (SEQ ID NO: 5803), GGAGGGAG (SEQ ID NO: 5804), GSGGGSG (SEQ ID NO: 5805), GGAGGGAGGGAG (SEQ ID NO: 5806), and GSGGGSGGGSG (SEQ ID NO: 5807). In some
 20 embodiments, a spacer includes motifs of GGGGA (SEQ ID NO: 5808), or GGGGS (SEQ ID NO: 5809), e.g., GGGGAGGGGAGGGGA (SEQ ID NO: 5810) and GGGGSGGGGSGGGGS (SEQ ID NO: 5811).

In some embodiments, a peptide spacer is cleavable. In some embodiments, a peptide spacer contains a protease binding site and can be cleaved by a protease via site-specific enzyme activity.

In some embodiments, an engineered transposase (e.g., an engineered transposase of the
 25 invention) is modified to include one or more tags and/or linkers on either the N- or C- termini. In some embodiments, a tag (e.g., a polyhistidine tag such as a 6-histidine tag) and/or a portion of a peptide spacer of an engineered transposase (e.g., an engineered transposase of the invention) is removed following or during purification. In some embodiments, the tag (e.g., an affinity tag) and/or a portion of the peptide spacer is removed, e.g., via proteolytic cleavage.

30 In some embodiments, a polynucleotide that encodes an engineered transposase (e.g., any one of the engineered transposases described herein, such as in Tables 5.2, 6.2, 7.2, 8.2, 9.2, 10.2, 10.3, 10.5, 11.2, 11.4, 11.5, 12.2, 12.3, 12.5, 12.6, 13.2, 13.3, 13.4, 14.2, 14.3, 14.5, 14.6, 14.8, 14.9, 15.1, 16.2, 16.3, 17.2, 17.3, 18.2, 19.2, 19.3, 20.2, 20.3, 21.2, 21.3, 22.2, 23.2, 24.2, 25.2, 26.2, 27.2, 28.2, 29.2, 30.2, 31.2, 32.2, 33.2, 34.2, 35.1, 36.1, and 36.3) is modified to remove the segment of nucleotides
 35 that encode the C-terminal 6-histidine tag and/or the C-terminal GGSG (SEQ ID NO: 5777) peptide spacer.

Methods of Using the Engineered Transposase Enzymes

In some embodiments, the transposase enzymes described herein find use in processes for
 40 conversion of one or more suitable substrates to a product.

In some embodiments, the engineered transposase polypeptides disclosed herein can be used in a process for the conversion of an adapter or donor polynucleotide and a target polynucleotide to an annealed or ligated polynucleotide consisting of the adapter and the optionally fragmented target polynucleotide. In some embodiments, this process results in a pool of fragmented polynucleotides with
5 ligated adapters, useful as a next generation sequencing library.

In the embodiments provided herein and illustrated in the Examples, various ranges of suitable reaction conditions that can be used in the processes, include but are not limited to, substrate loading, co-substrate loading, pH, temperature, buffer, solvent system, cofactor, polypeptide loading, and reaction time. Further suitable reaction conditions for carrying out the process for biocatalytic conversion of
10 substrate compounds to product compounds using an engineered transposase described herein can be readily optimized in view of the guidance provided herein by routine experimentation that includes, but is not limited to, contacting the engineered transposase polypeptide and one or more substrate compounds under experimental reaction conditions of concentration, pH, temperature, and solvent conditions, and detecting the product compound.

The substrate compound(s) in the reaction mixtures can be varied, taking into consideration, for example, the desired amount of product compound, the effect of each substrate concentration on enzyme activity, stability of enzyme under reaction conditions, and the percent conversion of each substrate to product. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each substrate of at least about 0.1 μM to 1 μM , 1 μM to 2 μM , 2 μM to 3 μM , 3 μM to 5 μM , 5 μM to
20 10 μM , or 10 μM or greater. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each substrate of at least about 0.5 to about 25 g/L, 1 to about 25 g/L, 5 to about 25 g/L, about 10 to about 25 g/L, or 20 to about 25 g/L. In some embodiments, the suitable reaction conditions comprise a substrate compound loading for each substrate of at least about 0.5 g/L, at least about 1 g/L, at least about 5 g/L, at least about 10 g/L, at least about 15 g/L, at least about 20 g/L, or at
25 least about 30 g/L, or even greater.

In carrying out the transposase-mediated synthesis processes described herein, the engineered polypeptide may be added to the reaction mixture in the form of a purified enzyme, partially purified enzyme, whole cells transformed with gene(s) encoding the enzyme, as cell extracts and/or lysates of such cells, and/or as an enzyme immobilized on a solid support. Whole cells transformed with gene(s)
30 encoding the engineered transposase enzyme or cell extracts, lysates thereof, and isolated enzymes may be employed in a variety of different forms, including solid (e.g., lyophilized, spray-dried, and the like) or semisolid (e.g., a crude paste). The cell extracts or cell lysates may be partially purified by precipitation (ammonium sulfate, polyethyleneimine, heat treatment or the like, followed by a desalting procedure prior to lyophilization (e.g., ultrafiltration, dialysis, etc.). Any of the enzyme preparations (including whole cell
35 preparations) may be stabilized by crosslinking using known crosslinking agents, such as, for example, glutaraldehyde or immobilization to a solid phase (e.g., Eupergit C, and the like).

The gene(s) encoding the engineered transposase polypeptides can be transformed into host cell separately or together into the same host cell. For example, in some embodiments one set of host cells can be transformed with gene(s) encoding one engineered transposase polypeptide, and another set can
40 be transformed with gene(s) encoding another transposase. Both sets of transformed cells can be utilized

together in the reaction mixture in the form of whole cells, or in the form of lysates or extracts derived therefrom. In other embodiments, a host cell can be transformed with gene(s) encoding multiple engineered transposase polypeptides. In some embodiments the engineered polypeptides can be expressed in the form of secreted polypeptides, and the culture medium containing the secreted

5 polypeptides can be used for the transposase reaction.

In some embodiments, the improved activity of the engineered transposase polypeptides disclosed herein provides for processes wherein higher percentage conversion can be achieved with lower concentrations of the engineered polypeptide. In some embodiments of the process, the suitable reaction conditions comprise an engineered polypeptide amount of about 1% (w/w), 2% (w/w), 5% (w/w),

10 10% (w/w), 20% (w/w), 30% (w/w), 40% (w/w), 50% (w/w), 75% (w/w), 100% (w/w) or more of substrate compound loading.

In some embodiments, the engineered polypeptide is present at a molar ratio of engineered polypeptide to substrate of about 50 to 1, 25 to 1, 10 to 1, 5 to 1, 1 to 1, 1 to 5, 1 to 10, 1 to 25 or 1 to 50. In some embodiments, the engineered polypeptide is present at a molar ratio of engineered polypeptide to

15 substrate from a range of about 50 to 1 to a range of about 1 to 50.

In some embodiments, the engineered polypeptide is present at about 0.01 g/L to about 50 g/L; about 0.01 to about 0.1 g/L; about 0.05 g/L to about 50 g/L; about 0.1 g/L to about 40 g/L; about 1 g/L to about 40 g/L; about 2 g/L to about 40 g/L; about 5 g/L to about 40 g/L; about 5 g/L to about 30 g/L; about 0.1 g/L to about 10 g/L; about 0.5 g/L to about 10 g/L; about 1 g/L to about 10 g/L; about 0.1 g/L to about

20 5 g/L; about 0.5 g/L to about 5 g/L; or about 0.1 g/L to about 2 g/L. In some embodiments, the transposase polypeptide is present at about 0.01 g/L, 0.05 g/L, 0.1 g/L, 0.2 g/L, 0.5 g/L, 1, 2 g/L, 5 g/L, 10 g/L, 15 g/L, 20 g/L, 25 g/L, 30 g/L, 35 g/L, 40 g/L, or 50 g/L.

In some embodiments, the suitable reaction conditions comprise a divalent metal cofactor. In some embodiments, the divalent metal cofactor is magnesium. In some embodiments, magnesium is present in the reaction mixture as magnesium (II) chloride. In some embodiments, the magnesium (II) chloride is present at concentrations of about 1 to 100 mM; about 1 to 50 mM; about 1 to 20 mM; or about

25 1 to 10 mM. In some embodiments, the magnesium (II) chloride is present at concentrations of about 3 mM; about 5 mM; about 7 mM, or about 10 mM.

During the course of the reaction, the pH of the reaction mixture may change. The pH of the reaction mixture may be maintained at a desired pH or within a desired pH range. This may be done by the addition of an acid or a base, before and/or during the course of the reaction. Alternatively, the pH may be controlled by using a buffer. Accordingly, in some embodiments, the reaction condition comprises a buffer. Suitable buffers to maintain desired pH ranges are known in the art and include, by way of example and not limitation, borate, phosphate, 2-(N-morpholino)ethanesulfonic acid (MES), 3-(N-

35 morpholino)propanesulfonic acid (MOPS), acetate, triethanolamine, and 2-amino-2-hydroxymethyl-propane-1,3-diol (Tris), and the like. In some embodiments, the reaction conditions comprise water as a suitable solvent with no buffer present.

In the embodiments of the process, the reaction conditions comprise a suitable pH. The desired pH or desired pH range can be maintained by use of an acid or base, an appropriate buffer, or a

40 combination of buffering and acid or base addition. The pH of the reaction mixture can be controlled

before and/or during the course of the reaction. In some embodiments, the suitable reaction conditions comprise a solution pH from about 4 to about 10, pH from about 5 to about 10, pH from about 5 to about 9, pH from about 6 to about 9, pH from about 6 to about 8. In some embodiments, the reaction conditions comprise a solution pH of about 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, or 10.

5 In the embodiments of the processes herein, a suitable temperature is used for the reaction conditions, for example, taking into consideration the increase in reaction rate at higher temperatures, and the activity of the enzyme during the reaction time period. Accordingly, in some embodiments, the suitable reaction conditions comprise a temperature of about 10 °C to about 95 °C, about 10 °C to about 75 °C, about 15 °C to about 95 °C, about 20 °C to about 95 °C, about 20 °C to about 65 °C, about 25 °C to about 70 °C, or about 50 °C to about 70 °C. In some embodiments, the suitable reaction conditions comprise a
10 temperature of about 10 °C, 15 °C, 20 °C, 25 °C, 30 °C, 35 °C, 40 °C, 45 °C, 50 °C, 55 °C, 60 °C, 65 °C, 70 °C, 75 °C, 80 °C, 85 °C, 90 °C or 95 °C. In some embodiments, the temperature during the enzymatic reaction can be maintained at a specific temperature throughout the course of the reaction. In some embodiments, the temperature during the enzymatic reaction can be adjusted over a temperature profile
15 during the course of the reaction.

In some embodiments, the processes of the invention are carried out in a solvent. Suitable solvents include water, aqueous buffer solutions, organic solvents, polymeric solvents, and/or co-solvent systems, which generally comprise aqueous solvents, organic solvents and/or polymeric solvents. The aqueous solvent (water or aqueous co-solvent system) may be pH-buffered or unbuffered. In some
20 embodiments, the processes using the engineered transposase polypeptides can be carried out in an aqueous co-solvent system comprising an organic solvent (e.g., ethanol, isopropanol (IPA), dimethyl sulfoxide (DMSO), dimethylformamide (DMF) ethyl acetate, butyl acetate, 1-octanol, heptane, octane, methyl t butyl ether (MTBE), toluene, and the like), ionic or polar solvents (e.g., 1-ethyl 4 methylimidazolium tetrafluoroborate, 1-butyl-3-methylimidazolium tetrafluoroborate, 1-butyl 3
25 methylimidazolium hexafluorophosphate, glycerol, polyethylene glycols, and the like). In some embodiments, the co-solvent can be a polar solvent, such as a polyol, dimethylsulfoxide (DMSO), or lower alcohol. The non-aqueous co-solvent component of an aqueous co-solvent system may be miscible with the aqueous component, providing a single liquid phase, or may be partly miscible or immiscible with the aqueous component, providing two liquid phases. Exemplary aqueous co-solvent systems can comprise
30 water and one or more co-solvents selected from an organic solvent, polar solvent, and polyol solvent. In general, the co-solvent component of an aqueous co-solvent system is chosen such that it does not adversely inactivate the transposase enzyme under the reaction conditions. Appropriate co-solvent systems can be readily identified by measuring the enzymatic activity of the specified engineered transposase enzyme with a defined substrate of interest in the candidate solvent system, utilizing an
35 enzyme activity assay, such as those described herein.

In some embodiments of the process, the suitable reaction conditions comprise an aqueous co-solvent, where the co-solvent comprises DMSO at about 1% to about 50% (v/v), about 1 to about 40% (v/v), about 2% to about 40% (v/v), about 5% to about 30% (v/v), about 10% to about 30% (v/v), or about 10% to about 20% (v/v). In some embodiments of the process, the suitable reaction conditions can
40 comprise an aqueous co-solvent comprising ethanol at about 1% (v/v), about 5% (v/v), about 10% (v/v),

about 15% (v/v), about 20% (v/v), about 25% (v/v), about 30% (v/v), about 35% (v/v), about 40% (v/v), about 45% (v/v), or about 50% (v/v).

In some embodiments, the reaction conditions comprise a surfactant for stabilizing or enhancing the reaction. Surfactants can comprise non-ionic, cationic, anionic and/or amphiphilic surfactants.

5 Exemplary surfactants, include by way of example and not limitation, nonyl phenoxypolyethoxyethanol (NP40), TRITON™ X-100 polyethylene glycol *tert*-octylphenyl ether, polyoxyethylene-stearylamine, cetyltrimethylammonium bromide, sodium oleylamidosulfate, polyoxyethylene-sorbitanmonostearate, hexadecyldimethylamine, etc. Any surfactant that may stabilize or enhance the reaction may be employed. The concentration of the surfactant to be employed in the reaction may be generally from 0.1
10 to 50 mg/mL, particularly from 1 to 20 mg/mL.

In some embodiments, the reaction conditions include an antifoam agent, which aids in reducing or preventing formation of foam in the reaction solution, such as when the reaction solutions are mixed or sparged. Anti-foam agents include non-polar oils (e.g., minerals, silicones, etc.), polar oils (e.g., fatty acids, alkyl amines, alkyl amides, alkyl sulfates, etc.), and hydrophobic (e.g., treated silica, polypropylene,
15 etc.), some of which also function as surfactants. Exemplary anti-foam agents include Y-30® (Dow Corning), poly-glycol copolymers, oxy/ethoxylated alcohols, and polydimethylsiloxanes. In some embodiments, the anti-foam can be present at about 0.001% (v/v) to about 5% (v/v), about 0.01% (v/v) to about 5% (v/v), about 0.1% (v/v) to about 5% (v/v), or about 0.1% (v/v) to about 2% (v/v). In some
20 embodiments, the anti-foam agent can be present at about 0.001% (v/v), about 0.01% (v/v), about 0.1% (v/v), about 0.5% (v/v), about 1% (v/v), about 2% (v/v), about 3% (v/v), about 4% (v/v), or about 5% (v/v) or more as desirable to promote the reaction.

The quantities of reactants used in the transposase reaction will generally vary depending on the quantities of product desired, and concomitantly the amount of substrates employed. Those having ordinary skill in the art will readily understand how to vary these quantities to tailor them to the desired
25 level of productivity and scale of production.

In some embodiments, the order of addition of reactants is not critical. The reactants may be added together at the same time to a solvent (e.g., monophasic solvent, biphasic aqueous co-solvent system, and the like), or alternatively, some of the reactants may be added separately, and some together at different time points. For example, the cofactor, co-substrate and substrate may be added first to the
30 solvent.

The solid reactants (e.g., enzyme, salts, etc.) may be provided to the reaction in a variety of different forms, including powder (e.g., lyophilized, spray dried, and the like), solution, emulsion, suspension, and the like. The reactants can be readily lyophilized or spray dried using methods and equipment that are known to those having ordinary skill in the art. For example, the protein solution can
35 be frozen at -80°C in small aliquots, then added to a pre-chilled lyophilization chamber, followed by the application of a vacuum.

For improved mixing efficiency when an aqueous co-solvent system is used, the transposase, and co-substrate may be added and mixed into the aqueous phase first. The substrate may be added and mixed in, followed by the organic phase or the substrate may be dissolved in the organic phase and mixed

in. Alternatively, the substrate may be premixed in the organic phase, prior to addition to the aqueous phase.

The processes of the present invention are generally allowed to proceed until further conversion of substrate to product does not change significantly with reaction time (e.g., less than 10% of substrate being converted, or less than 5% of substrate being converted). In some embodiments, the reaction is

allowed to proceed until there is complete or near complete conversion of substrate to product. Transformation of substrate to product can be monitored using known methods by detecting substrate and/or product, with or without derivatization. Suitable analytical methods include gas chromatography, HPLC, MS, and the like. In some embodiments, after suitable conversion to product, the reactants are separated from the product and additional reactants are added.

Any of the processes disclosed herein using the engineered polypeptides for the preparation of products can be carried out under a range of suitable reaction conditions, including but not limited to ranges of substrates, temperature, pH, solvent system, substrate loading, polypeptide loading, cofactor loading, and reaction time.

In some embodiments, the enzyme loading is between 1-30% w/w. In some embodiments, additional reaction components or additional techniques are carried out to supplement the reaction conditions. These can include taking measures to stabilize or prevent inactivation of the enzyme, reduce product inhibition, shift reaction equilibrium to formation of the desired product.

In some embodiments, the engineered transposase polypeptides can be provided on a solid support, such as a membrane, resin, solid carrier, or other solid phase material. A solid support can be composed of organic polymers such as polystyrene, polyethylene, polypropylene, polyfluoroethylene, polyethyleneoxy, and polyacrylamide, as well as co-polymers and grafts thereof. A solid support can also be inorganic, such as glass, silica, controlled pore glass (CPG), reverse phase silica or metal, such as gold or platinum. The configuration of a solid support can be in the form of beads, spheres, particles, granules, a gel, a membrane or a surface. Surfaces can be planar, substantially planar, or non-planar. Solid supports can be porous or non-porous, and can have swelling or non-swelling characteristics. A solid support can be configured in the form of a well, depression, or other container, vessel, feature, or location.

In some embodiments, the engineered transposase polypeptides of the present invention can be immobilized on a solid support such that they retain their improved activity, and/or other improved properties relative to the reference polypeptide of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704. In such embodiments, the immobilized polypeptides can facilitate the biocatalytic conversion of the substrate compounds or other suitable substrates to the product and after the reaction is complete are easily retained (e.g., by retaining beads on which polypeptide is immobilized) and then reused or recycled in subsequent reactions. Such immobilized enzyme processes allow for further efficiency and cost reduction.

Methods of enzyme immobilization are well-known in the art. The engineered polypeptides can be bound non-covalently or covalently. Various methods for conjugation and immobilization of enzymes to solid supports (e.g., resins, membranes, beads, glass, etc.) are well known in the art (See e.g., Yi et al.,

Proc. Biochem., 42(5): 895-898 [2007]; Martin et al., Appl. Microbiol. Biotechnol., 76(4): 843-851 [2007]; Koszelewski et al., J. Mol. Cat. B: Enzymatic, 63: 39-44 [2010]; Truppo et al., Org. Proc. Res. Dev., published online: dx.doi.org/10.1021/op200157c; Hermanson, Bioconjugate Techniques, 2nd ed., Academic Press, Cambridge, MA [2008]; Mateo et al., Biotechnol. Prog., 18(3):629-34 [2002]; and

5 "Bioconjugation Protocols: Strategies and Methods," In Methods in Molecular Biology, Niemeyer (ed.), Humana Press, New York, NY [2004]; the disclosures of each which are incorporated by reference herein). Solid supports useful for immobilizing the engineered transposase of the present invention include but are not limited to beads or resins comprising polymethacrylate with epoxide functional groups, polymethacrylate with amino epoxide functional groups, styrene/DVB copolymer or polymethacrylate with

10 octadecyl functional groups. Exemplary solid supports useful for immobilizing the engineered transposase polypeptides of the present invention include, but are not limited to, EnginZyme (including, EziG-1, EziG-1, and EziG-3), chitosan beads, Eupergit C, and SEPABEADs (Mitsubishi) (including EC-EP, EC-HFA/S, EXA252, EXE119 and EXE120).

Methods are provided for using engineered transposases, i.e., to reduce the complexity and time

15 required for generating a mixture of altered nucleic acids or a plurality of nucleic acid fragments, i.e., for generating nucleic acid libraries from nucleic acid samples for nucleic acid sequencing, while simultaneously improving performance of the nucleic acid libraries. In some instances, the methods may be used for the generation of nucleic acid fragments (e.g., a sequence library) amenable to high throughput sequencing technologies. The engineered transposases are used, e.g., to insert excised

20 transposable DNA elements for simultaneous fragmentation and tagging (e.g., barcoding) of a sample during sequence library preparation via the methods described herein.

In some instances, the methods described herein include contacting a target nucleic acid (e.g., nucleic acids of a nucleic acid sample) with a transposase (e.g., an engineered transposase) under conditions and for a time sufficient for the transposase to carry out a transposition event. In some

25 instances, the transposition event occurs in reaction conditions additionally including a terminal deoxynucleotide transferase, dNTP, and/or buffer components suitable for the addition of deoxynucleotides to the 3' terminus of each of the plurality of nucleic acid fragments. In some instances, the transposition event occurs in reaction conditions additionally including a DNA ligase and/or buffer components suitable for a ligation reaction.

30 In some instances, each altered nucleic acid of the mixture of altered nucleic acids is between 30 and 30000 (e.g., 30-30000, 30-25000, 30-20000, 30-15000, 30-10000, 30-7500, 30-5000, 30-4000, 30-3000, 30-2500, 30-2000, 30-1500, 30-1000, 30-900, 30-800, 30-700, 30-600, 30-500, 30-400, 30-300, 30-250, 30-200, 30-150, 30-100, 30-50, 100-500, 100-1000, 100-1500, 100-2500, 100-5000, 100-10000, 100-20000, 100-30000, 500-1000, 500-2500, 500-5000, 500-10000, 500-20000, 500-30000, 1000-2500,

35 1000-5000, 1000-10000, 1000-20000, 1000-30000, 5000-10000, 5000-20000, 5000-30000, 10000-20000, 20000-30000, 2500-5000, 7500-10000, or 15000-25000) nucleotides in length. In some instances, each altered nucleic acid of the mixture of altered nucleic acids further includes a sample tag and/or a unique molecular identifier (UMI). In some instances, each altered nucleic acid of the mixture of altered nucleic acids includes an identifiable sequence tag (IST). In some instances, the IST is between 6 and 30 (e.g., 6,

7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30) nucleotides in length.

In some instances, each nucleic acid fragment of the plurality of nucleic acid fragments is between 30 and 30000 (e.g., 30-30000, 30-25000, 30-20000, 30-15000, 30-10000, 30-7500, 30-5000, 30-4000, 30-3000, 30-2500, 30-2000, 30-1500, 30-1000, 30-900, 30-800, 30-700, 30-600, 30-500, 30-400, 30-300, 30-250, 30-200, 30-150, 30-100, 30-50, 100-500, 100-1000, 100-1500, 100-2500, 100-5000, 100-10000, 100-20000, 100-30000, 500-1000, 500-2500, 500-5000, 500-10000, 500-20000, 500-30000, 1000-2500, 1000-5000, 1000-10000, 1000-20000, 1000-30000, 5000-10000, 5000-20000, 5000-30000, 10000-20000, 20000-30000, 2500-5000, 7500-10000, or 15000-25000) nucleotides in length. In some instances, each nucleic acid fragment of the subset of nucleic acid fragments further includes a sample tag and/or a unique molecular identifier (UMI). In some instances, each nucleic acid fragment of the subset of nucleic acid fragments includes an identifiable sequence tag (IST). In some instances, the IST is between 6 and 30 (e.g., 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30) nucleotides in length.

Suitable samples and applications

Methods described herein (e.g., using engineered transposases) may be suitable for any samples and applications wherein sequencing technologies (e.g., next-generation sequencing (NGS)) may be performed. A sample may derive from a single cell or a plurality of single cells. A sample may derive from a whole population of cells or from a larger cell population. A sample may derive from one cell type or a plurality of cell types. A sample may derive from bacteria, fungi, viruses, animal tissue, or plant tissue. A sample may be isolated or partially isolated. One or more samples may be used for applications including but not limited to whole genome sequencing, whole exome sequencing, transcriptome analysis, epigenome sequencing, sequencing of target regions, deep sequencing, among other clinical and research applications.

A sample may be one or more nucleic acids of interest that may be used to generate a sequence library suitable for NGS or similar sequencing methods. A nucleic acid may be DNA (e.g., genomic, mitochondrial DNA, or a combination thereof), RNA (e.g., mRNA, rRNA, tRNA, microRNA, non-coding RNA, or a combination thereof), or a combination of both DNA and RNA. In some instances, an RNA sample is converted to complementary DNA (cDNA) using enrichment and reverse transcription methodologies known in the art. In some instances, an RNA sample can be transformed into an RNA/DNA duplex by reverse-transcription using a suitable reverse transcriptase, including, e.g., a Moloney murine leukemia virus (M-MLV) reverse transcriptase, a human immunodeficiency virus (HIV) reverse transcriptase, and an avian sarcoma-leukosis virus (ASLV) reverse transcriptase. Avian Sarcoma-Leukosis Virus (ASLV) reverse transcriptase includes, but is not limited to, Rous Sarcoma Virus (RSV) reverse transcriptase, Avian Myeloblastosis Virus (AMV) reverse transcriptase, Avian Erythroblastosis Virus (AEV) Helper Virus MCAV reverse transcriptase, Avian Myelocytomatosis Virus MC29 Helper Virus MCAV reverse transcriptase, Avian Reticuloendotheliosis Virus (REV-T) Helper Virus REV-A reverse transcriptase, Avian Sarcoma Virus UR2 Helper Virus UR2AV reverse transcriptase, Avian Sarcoma Virus Y73 Helper Virus YAV reverse transcriptase, Rous Associated Virus (RAV) reverse transcriptase, and

Myeloblastosis Associated Virus (MAV) reverse transcriptase. In some instances, a nucleic acid may include genomic DNA or cDNAs from a single cell. In some instances, a nucleic acid may include nucleic acids from a plurality of haplotypes. In some instances, a nucleic acid may include crosslinked via histones or chromatin from single or multiple cells. In some instances, a nucleic acid may include nucleic acid that has been condensed or optionally treated with one or more condensing agents.

Sequence Library Construction

Provided are methods using transposases with improved activity in cleaving or fragmenting polynucleotides and ligating or tagging polynucleotides (e.g., the engineered transposases described herein) that overcome the limitations of existing NGS library preparation methods. Improved transposases with increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias compared to another wild-type or engineered transposase find utility in library preparation for NGS, but are not limited to these uses. Improved transposases (e.g., the engineered transposases described herein) can find use in any application involving cleaving or fragmenting double-stranded polynucleotides and/or ligating adapters or tagging polynucleotides, including applications where a donor polynucleotide is transferred into a target polynucleotide.

The present methods employ novel transposases that have improved activity in the fragmentation or cleavage of double stranded DNA and the ligation of adapters or tagging of the DNA fragments. The transposases employed in the methods have increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias as compared to another wild-type transposase or engineered transposase. The engineered polypeptides employed in the methods are variants of SEQ ID NO: 2 or SEQ ID NO: 5704, a chimeric protein based on an IS4 transposase from *Parashewanella curva* or a modified *Escherichia coli* Tn5 transposase, respectively, which recognize the 19 bp Tn5 mosaic end (ME) transposon donor recognition sequence. These engineered transposases are capable of improved preparation of NGS libraries from polynucleotide samples.

In nature, transposases facilitate the transposition of fragments of DNA (transposons) from a donor strand to a host strand at sites flanked by short inverted-repeat sequences or end sequences (ESs). The transposition mechanism is initiated by the binding of two transposases to donor DNA ESs to form the dimeric synaptic complex, which catalyzes cleavage of the donor DNA. The synaptic complex (i.e., transposome; dimeric transposase/DNA complex) then binds to the target DNA to transfer the transposon DNA to the target DNA.

Transposase-mediated NGS library preparation is similar to naturally-occurring transposition, except that the donor DNA or polynucleotide consists of two, separate adapter sequences rather than one continuous loop of donor DNA or polynucleotide. These adapter sequences (e.g., i5 and i7 adapter oligonucleotides) are bound to transposases in a loading step that replaces the initial donor DNA or polynucleotide binding and cleavage step. Following adapter loading, the dimerized synaptic complex

(i.e., transposome) proceeds to bind the target DNA or polynucleotide sample. Cleavage of the target polynucleotide and ligation of the adapter sequences results in fragmentation of the target DNA or polynucleotide and ligation of sequencing adapters to the target fragments.

The resulting DNA or polynucleotide fragments with ligated adapters form the next generation sequencing library. An ideal library preparation method would generate random fragments in the polynucleotide sequence, without bias toward the sequence composition of the target site. Compared to traditional library preparation methods, transposase-mediated fragmentation has lower sequence insertion bias. However, certain transposases are known to have sequence insertion bias. As an example, Tn5 has a slight G/C sequence insertion bias. In some instances, the engineered transposases employed in the methods have reduced sequence insertion bias or increased sequence promiscuity as compared to a wild-type or reference transposase.

The methods described herein can be used for the generation of nucleic acid libraries having amplicons using adapter oligonucleotides and barcoded amplifier oligonucleotides. In some instances, the methods include generating the nucleic acid libraries using (a) a DNA polymerase; (b) a first synaptic complex including a first transposase (e.g., an engineered transposase described herein) and a first adapter oligonucleotide, and a second synaptic complex including a second transposase (e.g., an engineered transposase described herein) and a second adapter oligonucleotide; and (c) magnesium ions (e.g., Mg^{2+}). In some instances, each adapter oligonucleotide includes an adapter priming region. The method may further include using (i) a first amplifier oligonucleotide including a first universal primer region, a first amplifier barcode region (e.g., a pool tag, e.g., an i5 or an i7 tag), and a first amplifier priming region; and (ii) a second amplifier oligonucleotide including a second universal primer region, a second amplifier barcode region (e.g., a pool tag, e.g., an i5 or an i7 tag), and a second amplifier priming region, wherein the first adapter priming region of the first adapter oligonucleotide is homologous to the first amplifier priming region of the first amplifier oligonucleotide and the second adapter priming region of the second adapter oligonucleotide is homologous to the second amplifier priming region of the second amplifier oligonucleotide. In some instances, each adapter oligonucleotide includes a universal primer region and an adapter barcode region (e.g., an identifiable sequence tag (IST) or a sample tag).

Transposition

The nucleic acid samples of a sequencing library (e.g., a NGS library) may be generated in a single reaction vessel through a method that includes a transposition reaction (e.g., using engineered transposases described herein), a polymerization reaction, and a PCR reaction for a tagmentation-based sequence library preparation. The method includes first combining the DNA polymerase, the first synaptic complex and the second synaptic complex, magnesium ions (e.g., Mg^{2+}), the first and second amplifier oligonucleotides, and the nucleic acid sample in a first reaction vessel. The engineered transposases of the synaptic complexes as described herein will fragment the nucleic acid sample into nucleic acid fragments and add the adapter oligonucleotide sequences to the 5' ends of the two strands of a nucleic acid duplex containing a nucleic acid fragment and its complement sequence in a transposition reaction to generate intermediate nucleic acids. In some instances, the transposition reaction occurs at a transposition reaction temperature between 25-65 °C (e.g., between 35-65 °C, between 40-65 °C,

between 45-65 °C, between 50-65 °C, between 55-65 °C, between 60-65 °C, between 25-60 °C, between 25-55 °C, between 25-50 °C, between 25-45 °C, between 25-40 °C, between 25-35 °C, between 25-30 °C, between 40-50 °C, or between 53-57 °C; e.g., at about 25 °C, at about 30 °C, about 35 °C, about 40 °C, about 45 °C, about 50 °C, about 53 °C, about 54 °C, about 55 °C, about 56 °C, about 57 °C, about 60

5 °C, or about 65 °C). In some instances, a transposition reaction temperature is selected following consideration of the rate of reaction and transposase activity at a given temperature or range of temperatures. In some instances, a specific temperature is maintained over the course of the enzymatic reaction to maximize enzymatic activity or reaction efficiency. In other instances, the temperature increases or decreases over a range of temperatures to maximize enzymatic activity or reaction

10 efficiency.

In some instances, the transposition reaction occurs for a first reaction duration between 1 and 30 minutes (e.g., between 1 and 25 minutes, between 1 and 20 minutes, between 1 and 15 minutes, between 1 and 10 minutes, between 1 and 5 minutes, between 5 and 10 minutes, between 5 and 20 minutes, between 5 and 30 minutes, between 10 and 30 minutes, between 15 and 30 minutes, between

15 20 and 30 minutes, between 25 and 30 minutes, between 10 and 20 minutes, between 15 and 25 minutes; e.g., about 1 minute, about 5 minutes, about 10 minutes, about 13 minutes, about 14 minutes, about 15 minutes, about 16 minutes, about 17 minutes, about 20 minutes, about 25 minutes, or about 30 minutes). In some instances, the duration of the transposition reaction is determined by the temperature or range of temperatures at which the transposition reaction occurs. In some instances, the duration of the

20 transposition reaction is determined by the expected efficiency or stability of a transposase.

In some instances, the nucleic acid sample for fragmentation is present in the transposition reaction at variable amounts. The sample may be present at low concentrations at about 1 to 10 ng/μL (e.g., about 10 ng/μL, about 9 ng/μL, about 8 ng/μL, about 7 ng/μL, about 6 ng/μL, about 5 ng/μL, about 4 ng/μL, about 3 ng/μL, about 2 ng/μL, about 1 ng/μL, or less than 1 ng/μL). The sample may be present at

25 medium concentrations at about 10 ng/μL to 100 ng/μL (e.g., about 100 ng/μL, about 90 ng/μL, about 80 ng/μL, about 70 ng/μL, about 60 ng/μL, about 50 ng/μL, about 40 ng/μL, about 30 ng/μL, about 20 ng/μL, or about 10 ng/μL). Further, the sample may be present at high concentrations at about 100 ng/μL or higher (e.g., about 100 ng/μL, about 150 ng/μL, about 200 ng/μL, about 250 ng/μL, about 300 ng/μL, or higher). For each of the sample tagging reaction concentration ranges, a relatively fixed and limited

30 sample-tagging condition is used such that the number of tags (e.g., sample tags or identifiable sequence tags (ISTs)) per prescribed unit mass of nucleic acid is limited. The resulting tagmented samples thereby have a similar number of tagmentation occurrences but still a varying amount of nucleic acid molecules. The plurality of tagged DNA samples is pooled to create a sample pool in which a multiplicity of potentially differently tagged DNA samples produced by transposition processes are combined into one

35 heterogeneous mixture. The important feature of this pool is that while it may include different masses of nucleic acid molecules from each sample in rough proportion to the mass of nucleic acid molecules in each sample, the number of sample tagging instances from each sample present in the pool is fixed and normalized.

In some instances, one or more incorporated tags may consist of the same known oligonucleotide

40 sequence, or a different oligonucleotide sequence. One or more tags may consist of oligonucleotide

sequences of the same lengths or different lengths, wherein said lengths range from 4 to 10 base pairs (e.g., 4 base pairs, 5 base pairs, 6 base pairs, 7 base pairs, 8 base pairs, 9 base pairs, or 10 base pairs), 8 to 15 base pairs (e.g., 8 base pairs, 9 base pairs, 10 base pairs, 11 base pairs, 12 base pairs, 13 base pairs, 14 base pairs, or 15 base pairs), or 10 to 20 base pairs (e.g., 10 base pairs, 11 base pairs, 12 base pairs, 13 base pairs, 14 base pairs, 15 base pairs, 16 base pairs, 17 base pairs, 18 base pairs, 19 base pairs, or 20 base pairs).

Polymerization

After the transposition reaction, a polymerization reaction by a polymerase (e.g., a DNA polymerase) can extend the 3' ends of the intermediate nucleic acids to generate full duplexes. In some instances, the polymerization reaction occurs at a polymerization reaction temperature between 55-95 °C (e.g., between 60-95 °C, between 65-95 °C, between 70-95 °C, between 75-95 °C, between 80-95 °C, between 85-95 °C, between 90-95 °C, between 55-90 °C, between 55-85 °C, between 55-80 °C, between 55-75 °C, between 55-70 °C, between 55-65 °C, between 55-60 °C, between 65-85 °C, between 70-80 °C, between 73-77 °C; e.g., at about 55 °C, at about 60 °C, at about 65 °C, at about 70 °C, at about 73 °C, at about 74 °C, at about 75 °C, at about 76 °C, at about 77 °C, at about 80 °C, at about 85 °C, at about 90 °C, at about 95 °C). In some instances, the temperature of the polymerization reaction is dependent on the enzymatic activity and efficiency of the DNA polymerase. In some instances, the transposases dissociate as the first reaction vessel is heated to the polymerization temperature. In some instances, the DNA polymerase may be a thermostable DNA polymerase. In some instances, the DNA polymerase may be a high-fidelity DNA polymerase. In some instances, the DNA polymerase may be a hot-start DNA polymerase. In some instances, the hot-start DNA polymerase may contain an antibody or aptamer bound to the DNA polymerase, that is released when the hot-start DNA polymerase heated to and incubated at a suitable temperature, which may or may not be higher than the optimal polymerization temperature. In some instances, the DNA polymerase is selected based on the GC content of the sample or desired coverage following amplification.

In some instances, the polymerization reaction occurs for a second reaction duration between 1 and 60 minutes (e.g., between 1 and 55 minutes, between 1 and 50 minutes, between 1 and 45 minutes, between 1 and 40 minutes, between 1 and 35 minutes, between 1 and 30 minutes, between 1 and 25 minutes, between 1 and 20 minutes, between 1 and 15 minutes, between 1 and 10 minutes, between 10 and 60 minutes, between 15 and 60 minutes, between 20 and 60 minutes, between 25 and 60 minutes, between 30 and 60 minutes, between 35 and 60 minutes, between 40 and 60 minutes, between 45 and 60 minutes, between 50 and 60 minutes, between 55 and 60 minutes, between 15 and 35 minutes, between 30 and 50 minutes, between 10 and 20 minutes, between 20 and 40 minutes, or between 13 and 17 minutes; e.g., about 1 minute, about 5 minutes, about 10 minutes, about 13 minutes, about 14 minutes, about 15 minutes, about 16 minutes, about 17 minutes, about 20 minutes, about 25 minutes, about 30 minutes, about 35 minutes, about 40 minutes, about 45 minutes, about 50 minutes, about 55 minutes, or about 60 minutes). In some instances, the duration of the polymerization reaction is dependent on the temperature or range of temperatures at which the polymerization reaction occurs.

40

Amplification

In some instances, nucleic acid samples (e.g., nucleic acid fragment) may be amplified through polymerase chain reaction (PCR), multiple displacement amplification (MDA), ligase chain reaction (LCR), loop mediated isothermal amplification (LAMP), rolling circle amplification (RCA), or strand displacement amplification (SDA).

For example, a PCR reaction generates the amplicons of the sequence library by using the amplifier oligonucleotides as primers. In some instances, the amplicons are generated in a PCR reaction with a first universal primer and a second universal primer, wherein the first universal primer and the second universal primer bind to respective complementary universal primer regions in the sequencing oligonucleotides. In some instances, the amplicons may be further amplified in a PCR reaction with a first universal primer and a second universal primer, wherein the first universal primer and the second universal primer bind to respective complementary universal primer regions in the amplicons.

In some instances, the amplicons include (a) a nucleic acid sequence including a first universal primer region, a first amplifier barcode region (e.g., a pool tag, e.g., an i5 or an i7 tag), a first amplifier priming region, a homologous sequence of a first nucleic acid fragment, a complement sequence of the second amplifier priming region, a complement sequence of the second amplifier barcode region (e.g., a pool tag, e.g., an i5 or an i7 tag), and a complement sequence of the second universal primer region; and (b) the complement sequence thereof. In some instances, a nucleic acid duplex including the first nucleic acid fragment and its complement is generated from the nucleic acid sample by transposase activity. In some instances, the intermediate nucleic acid is the template for the PCR reaction. In some instances, the PCR reaction produces amplicons from the intermediate nucleic acids using the amplifier oligonucleotides. In some instances, the PCR reaction includes 1-35 cycles (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, or 35 cycles). In some instances, the PCR reaction includes more than 35 cycles. In some instances, the PCR reaction may be monitored by tracking the cycle threshold (Ct) via fluorescence detection using well-established methods in the art. In some instances, the resulting sequence library following transposition, tagmentation, and amplification may be evaluated for quality (e.g., size distribution, purity, concentration, among other quality considerations). In some instances, the sequence library may be assessed for the size distribution of the resulting fragments by gel electrophoresis compared to a reference sample (e.g., a DNA ladder).

Reaction Considerations

In the instances provided herein, various ranges of suitable reaction conditions that can be used in the processes, include but are not limited to, substrate loading, co-substrate loading, pH, temperature, buffer, solvent system, cofactor, polypeptide loading, and reaction time. Further suitable reaction conditions for carrying out the process for biocatalytic conversion of nucleic acid samples to a sequencing library and its intermediates using an engineered transposase described herein can be readily optimized in view of the guidance provided herein by routine experimentation that includes, but is not limited to, contacting the transposase and one or more substrate compounds (e.g., nucleic acid samples; e.g., DNA)

under experimental reaction conditions of concentration, pH, temperature, and solvent conditions, and detecting the product compound.

In some instances, one or more enzymes (e.g., a reverse transcriptase, a transposase (e.g., an engineered transposase described herein), or a DNA polymerase) are inactivated by the application of heat or an inhibitor following a chemical reaction (e.g., transposition, polymerization, or amplification) in the tagmentation-based sequence library preparation. In some instances, one or more enzymes are inactivated by changes in buffer or solvent condition (e.g., changes in pH or changes in salt concentration). In some instances, the nucleic acid sample is purified or partially purified following a chemical reaction. In some instances, the nucleic acid sample is concentrated or diluted following a chemical reaction.

In some instances, one or more enzymes (e.g., a reverse transcriptase, a transposase (e.g., an engineered transposase described herein), or a DNA polymerase) are immobilized to improve reaction efficiency, enzyme activity, or sample recovery. In some instances, an enzyme is immobilized on a solid support, a surface, a resin, a membrane, a bead, a nanoparticle, a biological cofactor (e.g., a protein, an antibody or antibody fragment, a peptide, an oligonucleotide, an aptamer), or a combination thereof. In some instances, an enzyme is immobilized through an irreversible or a reversible reaction. In some instances, an enzyme is immobilized by cross-linking, adsorption, entrapment, encapsulation, click chemistry reactions, ionic bonding, covalent bonding, hydrophobic interactions, among other non-limiting methods of immobilization.

In some instances, the nucleic acid sample, the first adapter oligonucleotide, the second adapter oligonucleotide, the first amplifier oligonucleotide, the second amplifier oligonucleotide, the intermediate nucleic acids, and/or the amplicons may include DNA. In some instances, the nucleic acid sample may include double-stranded DNA.

In some instances, further processing steps are used for library preparation, including gap filling, normalization, pooling of tagged fragment samples, and/or ligation of additional adapters to create dual-tagged libraries. A variety of parameters may be manipulated to create sequencing libraries with various characteristics. For example, varying the concentration of transposase and adapters can change the fragment length. Copy number and sample coverage may also be influenced or controlled using methods known to those of skill in the art.

Methods of Sequencing

Methods described herein can be used for determining the nucleic acid sequences of the sequencing oligonucleotides or amplicons through nucleic acid sequencing (e.g., next-generation sequencing (NGS)) or other methods known in the art. In some instances, sequencing can be performed by various systems that are currently available, e.g., a sequencing system by Pacific Biosciences (PACBIO®), ILLUMINA®, Oxford NANOPORE®, Genapsys®, or ThermoFisher (ION TORRENT®). Alternatively or in addition, sequencing may be performed using nucleic acid amplification, sequencing by ligation (e.g., SOLiD or polony-based sequencing), sequencing by synthesis (e.g., Illumina dye sequencing, single-molecule real-time (SMRT) sequencing, or pyrosequencing), nanopore sequencing, polymerase chain reaction (PCR) (e.g., digital PCR, quantitative PCR, or real time PCR), or isothermal

amplification. In some instances, the sequencing oligonucleotides and amplicons described herein can be uniquely identified based on the nucleic acid sequences of the nucleic acid fragment and the nucleic acid sequences of the adapter barcode regions (e.g., an identifiable sequence tag (IST) or a sample tag) or amplifier barcode regions (e.g., pool tags, e.g., i5 or i7 tags) of the sequencing oligonucleotides or amplicons, respectively. The invention further includes data generated by nucleic acid sequencing, as well as methods for generating and analyzing such sequence data, and reaction mixtures used in and formed by such methods.

Methods of Evaluating Transposition Activity

Methods described herein can be used for assaying transposition activity of a transposase described herein (e.g., an engineered transposase described herein. In some instances, the methods include performing a transposition reaction on a nucleic acid (e.g., DNA, e.g., double-stranded DNA (dsDNA)), i.e., to insert a nucleic acid sequence of interest (e.g., DNA, e.g., dsDNA, e.g., dsDNA adapters) into a target nucleic acid sequence (e.g., DNA, e.g., dsDNA). In some instances, synaptic complexes (i.e., transposomes) are prepared by incubating transposases (e.g., engineered transposases described herein) with nucleic acid adapters (e.g., DNA adapter oligonucleotides). In some instances, 1 - 50 μ M (e.g., 1-40, 1-30, 1-20, 1-10, 1-5, 5-50, 5-40, 5-30, 5-20, 5-10, 10-50, 10-40, 10-30, 10-20, 20-50, 20-40, or 20-30 μ M; e.g., about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 45, or 50 μ M) of the adapter may be incubated with engineered transposases for about 60 minutes to load the transposases and generate synaptic complexes. In some instances, a target substrate (e.g., a target dsDNA) is added and incubated with the synaptic complexes until a transposition event occurs. In some instances, the mixture of the synaptic complexes and the target substrate is incubated (e.g., heated in a thermocycler). In some instances, the mixture is incubated at about 55 °C for 5-15 minutes (e.g., about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 minutes). In some instances, the target substrate is any of the nucleic acid samples described herein, including DNA or cDNA, and may be prokaryotic, eukaryotic, or artificial in origin.

In some instances, the activity of two different transposases (e.g., any transposases described herein or known to one of skill in the art, e.g., any naturally occurring transposases or any engineered transposases described herein), i.e., a first transposase and a second transposase, may be compared by comparing the amount of transposition reactions that occur when using synaptic complexes comprising the first transposase versus synaptic complexes comprising the second transposase. Preferably, the other reaction conditions are maintained similar or identical for the comparison (e.g., target nucleic acid concentration, reaction temperature, reaction duration, adapter concentration, transposase concentration, buffer conditions). In some instances, an activity fold improvements over positive control (FIOP) can be calculated as the percent product of transposition reactions produced by the first transposase (e.g., an engineered transposase) compared with the percent product of transposition reactions produced by the second transposase (e.g., a reference transposase), i.e., as an "FIOP Percent Product."

Methods of Evaluating Transposition Performance

Methods described herein can be used for assaying transposition performance of a transposase employed in the methods (e.g., an engineered transposase described herein. In some instances, the methods include generating an NGS library from a nucleic acid sample (e.g., DNA, e.g., double-stranded DNA (dsDNA)), i.e., by performing a tagmentation reaction (e.g., to fragment the nucleic acid sample and insert adapter nucleic acids DNA, e.g., dsDNA, e.g., dsDNA adapters) and amplifying the nucleic acid fragments resulting from the tagmentation reaction. In some instances, synaptic complexes (i.e., transposomes) are prepared by incubating transposases (e.g., engineered transposases described herein) with nucleic acid adapters (e.g., DNA adapter oligonucleotides, e.g., i5 and i7 oligonucleotides) to generate loaded i5 and i7 transposases. In some instances, a nucleic acid sample is mixed with the synaptic complexes (e.g., loaded i5 and i7 transposases) and incubated until a transposition reaction occurs. In some instances, the mixture is incubated at about 55 °C for 5-15 minutes (e.g., about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 minutes). In some instances, the transposition reaction is performed using a method described herein. In some instances, the product of the transposition reaction is amplified, e.g., using a method of nucleic acid amplification known in the art (e.g., a PCR reaction, e.g., using primers suitable for the i5 and i7 adapter sequences). In some instances, the nucleic acid sample is any of the nucleic acid samples described herein, including DNA or cDNA, and may be prokaryotic, eukaryotic, or artificial in origin. In some instances, the NGS library is performed using methods described herein, e.g., in the section above titled "Sequence Library Construction."

In some instances, the performance of two different transposases (e.g., any transposases described herein or known to one of skill in the art, e.g., any naturally occurring transposases or any engineered transposases described herein), i.e., a first transposase and a second transposase, may be compared by comparing the insertion bias (e.g., sequence insertion bias, e.g., G/C bias) of transposition reactions that occur when using synaptic complexes comprising the first transposase versus synaptic complexes comprising the second transposase. Preferably, the other reaction conditions are maintained similar or identical for the comparison (e.g., target nucleic acid concentration, reaction temperature, reaction duration, adapter concentration, transposase concentration, buffer conditions). In some instances, an activity fold improvements over positive control (FIOP) can be calculated as the percent product of transposition reactions produced by the first transposase (e.g., an engineered transposase) compared with the percent product of transposition reactions produced by the second transposase (e.g., a reference transposase), i.e., as an "FIOP Percent Product." In some instances, a FIOP can be calculated as the relative calculated enzyme activity of the first transposase (e.g., an engineered transposase) compared to the second transposase (e.g., a reference transposase), such as by comparing fragmenting activity of a nucleic acid sequence (e.g., a double-stranded nucleic acid sequence) and/or transposition of a sequence.

In some instances, transposase promiscuity or target bias of a first transposase (e.g., an engineered transposase) relative to a second transposase (e.g., a reference transposase) can be calculated as the ratio of the product peak area produced by the first transposase (e.g., the product of a first pair of annealed oligos vs. product of a second pair of annealed oligos (e.g., pairs of annealed oligonucleotides comprising a polynucleotide sequence selected from SEQ ID NOs: 2370-2387 and 5696-

5700)) divided by the ratio of the product peak area observed by the reaction with the second transposase (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate), e.g., as a “Promiscuity FIOP” or “FIOP Target Product Ratio.”

In some instances, transposon insertion frequency is biased by the sequence of the insertion sites over the first 15 nucleotides of NGS reads putatively comprising the region of the target bound by the transposome construct. In some instances, bias plots can be generated from NGS data to represent this bias as a function of the distance from the insertion site. In some instances, overall relative bias for each transposase (e.g., an engineered transposase) can be calculated as the root-mean-squared deviation (RMSD) of the observed vs. expected representation of all four nucleotides at each position as shown in Scheme 1. Additionally, the mean RMSD over, for example, the first 15 nucleotides (RMSD₁₅) can be used to summarize sequence insertion bias with a single number.

Scheme 1.

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}}$$

In some instances, relative target bias for each transposase (e.g., engineered transposase) can be calculated as the mean RMSD₁₅ of a second transposase (e.g., a reference transposase) divided by the mean RMSD₁₅ of a first transposase (e.g., an engineered transposase). In this manner, engineered transposases with target bias mean RMSD₁₅ less than a reference transposase will have a ratio greater than 1.

In further embodiments, any of the above described processes for the conversion of one or more substrate compounds to product compound can further comprise one or more steps selected from: extraction; isolation; purification; and crystallization of product compound. As is known to those skilled in the art, acidic compounds may exist in various salt forms that can be used interchangeably in the methods described herein. All such forms are specifically envisaged for use in the methods described herein. Methods, techniques, and protocols for extracting, isolating, purifying, and/or crystallizing the product from biocatalytic reaction mixtures produced by the above disclosed processes are known to the ordinary artisan and/or accessed through routine experimentation. Additionally, illustrative methods are provided in the Examples below.

Various features and embodiments of the invention are illustrated in the following representative examples, which are intended to be illustrative, and not limiting.

EXAMPLES

The following Examples, including experiments and results achieved, are provided for illustrative purposes only and are not to be construed as limiting the present invention. Indeed, there are various suitable sources for many of the reagents and equipment described below. It is not intended that the present invention be limited to any particular source for any reagent or equipment item.

In the experimental disclosure below, the following abbreviations apply: M (molar); mM (millimolar), uM and μM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg

(micrograms); L and l (liter); ml and mL (milliliter); cm (centimeters); mm (millimeters); μm and μm (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); psi and PSI (pounds per square inch); °C (degrees Celsius); RT and rt (room temperature); CV (coefficient of variability); CAM and cam (chloramphenicol); PMBS (polymyxin B sulfate);

5 IPTG (isopropyl β -D-l-thiogalactopyranoside); LB (lysogeny broth); TB (terrific broth); SFP (shake flask powder); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); nt (nucleotide; polynucleotide); aa (amino acid; polypeptide); *E. coli* W3110 (commonly used laboratory *E. coli* strain, available from the Coli Genetic Stock Center [CGSC], New Haven, CT); HTP (high throughput); HPLC (high pressure liquid chromatography); HPLC-UV (HPLC-Ultraviolet Visible Detector); ¹H NMR (proton nuclear

10 magnetic resonance spectroscopy); FIOPC (fold improvements over positive control); Sigma and Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO); Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Microfluidics (Microfluidics, Westwood, MA); Life Technologies (Life Technologies, a part of Fisher Scientific, Waltham, MA); Amresco (Amresco, LLC, Solon, OH); Carbosynth (Carbosynth, Ltd., Berkshire, UK); Varian (Varian Medical Systems, Palo Alto, CA); Agilent (Agilent Technologies, Inc., Santa Clara, CA);

15 Infors (Infors USA Inc., Annapolis Junction, MD); and Thermotron (Thermotron, Inc., Holland, MI).

EXAMPLE 1

Transposase Gene Acquisition and Construction of Expression Vectors

The wild-type (WT) polypeptide described by UniProt accession number A0A3L8PTF2 from species

20 *Parashewanella curva* is a predicted IS4 family transposase. A synthetic gene (SEQ ID NO: 1) encoding a C-terminal 6-histidine tagged version of the transposase was designed *in silico* to recognize the 19 bp Tn5 mosaic end (ME) transposon donor recognition sequence with codon optimization for *E. coli* expression, was synthesized, and was subcloned into the *E. coli* expression vector pCK100900i (See e.g., US Pat. No. 7,629,157 and US Pat. Appln. Publn. 2016/0244787, both of which are hereby incorporated by reference).

25 This plasmid construct was transformed into an *E. coli* strain derived from BL-21. Directed evolution techniques were used to generate libraries of gene variants from these plasmids (See e.g., US Pat. No. 8,383,346 and WO 2010/144103, both of which are hereby incorporated by reference).

EXAMPLE 2

Transposase Expression and Lysate Processing for High Throughput (HTP) Screening

High Throughput (HTP) Growth of Transposase Enzyme and Variants

Transformed *E. coli* cells were selected by plating onto LB agar plates containing 1% glucose and 30 $\mu\text{g/mL}$ chloramphenicol. After overnight incubation at 37°C, colonies were placed into the wells of 96-well shallow flat bottom NUNC™ (Thermo-Scientific) plates filled with 190 μL /well LB medium supplemented

35 with 1% glucose and 30 $\mu\text{g/mL}$ chloramphenicol. The cultures were allowed to grow overnight for 18-20 hours in a shaker (200 rpm, 30°C, and 85% relative humidity; Kuhner). Overnight growth samples (20 μL) were transferred into Costar 96-well deep plates filled with 380 μL of Terrific Broth supplemented with 30 $\mu\text{g/mL}$ chloramphenicol. The plates were incubated for 120 minutes in a shaker (250 rpm, 30°C, and 85% relative humidity; Kuhner) until the OD₆₀₀ reached between 0.4-0.8. The cells were then induced with 40 μL

40 of 10 mM IPTG in sterile water and incubated overnight for 18-20 hours in a shaker (250 rpm, 20-30°C, and

85% relative humidity; Kuhner). The cells were pelleted (4,000 rpm for 20 min), the supernatants were discarded, and the cells were frozen at -80 °C prior to analysis.

For lysis, 300 µL lysis buffer containing 1 g/L lysozyme were added to the cell pellet in each well. The cells were resuspended by vigorous shaking at room temperature for 30-90 minutes on a bench top shaker. A 150-µL aliquot of the re-suspended cells was transferred to a 96-well format 200 µL BioRad PCR plate and heated at the lysis temperature from 30-60°C for 30 mins as described in the examples below. Following heat-treatment, the cell debris was pelleted by centrifugation (4,000 rpm, 4°C, 10 mins) and clear supernatants were then used in biocatalytic reactions to determine their activity levels.

EXAMPLE 3

Shake Flask Expression and Purification of Transposase

Shake Flask Expression

Selected HTP cultures grown as described above were plated onto LB agar plates with 1% glucose and 30 µg/mL chloramphenicol and grown overnight at 37°C. A single colony from each culture was transferred to 5 mL of LB broth with 1% glucose and 30 µg/mL chloramphenicol. The cultures were grown for 20 h at 30 °C, 250 rpm, and subcultured at a dilution of approximately 1:50 into 250 mL of Terrific Broth with 30 µg/mL of chloramphenicol. The cultures were incubated for approximately 3-4 hours at 30 °C, 250 rpm, to an OD₆₀₀ of about 0.6-0.8, and then induced with the addition of IPTG at a final concentration of 1 mM. The induced cultures were incubated for 18-20 h at 20°C-30°C, 250 rpm. Following this incubation period, the cultures were centrifuged at 4,000 rpm for 10 min. The culture supernatant was discarded, and the pellets were resuspended in 20 mL of 50 mM Tris, 500 mM NaCl, 20 mM Imidazole, pH 8.0. This cell suspension was chilled in an ice bath and lysed using a Microfluidizer cell disruptor (Microfluidics M-110L). The crude lysate was pelleted by centrifugation (10,000 rpm for 60 mins at 4 °C), and the supernatant was then filtered through a 0.2 µm PES membrane to further clarify the lysate.

Purification of Transposase from Shake Flask Lysates

Transposase lysates were purified using an AKTA Start purification system and a 5mL HisTrap FF column (GE Healthcare) using the AC Step HiF setting (the run parameters are provided below). The shake flask purification wash buffer comprised of 50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazole, pH 8.0. The shake flask purification elution buffer comprised of 50 mM Tris-HCl, 500 mM NaCl, 250 mM imidazole, pH 8.0.

Table 3.1: Purification Parameters

Parameter	Volume
Column volume	5 mL
Sample application flow rate	8 mL/min
Sample wash flow rate	12 mL/min
Sample elution flow rate	5 mL/min
Pressure limit	0.3 MPa

Sample volume	40-50 mL
Equilibration volume	5 column volumes (CV) = 25mL
Wash Unbound volume	20 CV = 100 mL
Elution	Isocratic (step)
Elution volume	3.4 CV = 17 mL
Fraction volume	0.75-1.5 mL
RE-equilibration volume	5 CV = 25 mL

Elution fractions containing protein were identified by UV absorption (A280) and pooled, then dialyzed overnight in dialysis buffer (50 mM Tris, 150 mM NaCl, 0.1 mM EDTA, 50% glycerol, pH 8.0) in a 3.5K Slide-A-Lyzer™ dialysis cassette (Thermo Fisher) for buffer exchange. Transposase concentrations in the preparations were measured by absorption at 280 nm.

EXAMPLE 4

Capillary Electrophoresis (CE) Analysis of Oligonucleotides

Sample preparation for reaction analysis using CE:

For analysis of the assay reaction samples, capillary electrophoresis was performed using an ABI 3500xL or SeqStudio 24 Flex Genetic Analyzers (ThermoFisher). 2 µL of the final diluted assay samples were further diluted 10x in a new 96-well MicroAmp Optical PCR plate or 384-well MicroAmp Optical PCR plate with 18 µL of Hi-Di™ Formamide (ThermoFisher) containing an appropriate size standard (LIZ or Alexa633).

CE analysis was done using POP6 polymer and a capillary length of 50 cm. Oven temperature was set at 50°C. The prerun, run and injection voltages were 18, 19.5 and 5 kVolts respectively. The prerun, run and injection times were 180, 850 and 5 secs. FAM-labeled oligo substrates (forward/reverse [odd/even] annealed oligos summarized in **Table 4.1**) and products were identified by their sizes relative to the sizing ladder.

Table 4.1: Target oligos used for transposase capillary electrophoresis assays	
SEQ ID NO:	Sequence
2369	TTTTTTTGGCGGTTTTAACTGTGATT/36-FAM/
2370	AATCACAGTTTAAACCGCCAAAAA*A
2371	TTTTTTGGGTGGTTTTAACTCTGCTTTT/36-FAM/
2372	AAAAGCAGAGTTTAAACCCACCCAAAAA*A
2373	TTTTTTGGGCCGTTTTAAACGGTGCTTTTTT/36-FAM/
2374	AAAAAAGCACCGTTTAAACGGCCAAAAA*A
2375	TTTTTTGGGCGTTTTTAAATGTGCTTTTTTT/36-FAM/
2376	AAAAAAAAGCACATTTTAAACGCCAAAAA*A
2377	TTTTTTGGGCGGGTTTAATCTGTGCTTTTTTTTTT/36-FAM/
2378	AAAAAAAAAAGCACAGATTAAACCCGCCAAAAA*A
2379	TTTTTTGGGCGGTTTTAACTGTGCTTTTTTTTTTT/36-FAM/

2380	AAAAAAAAAAAAAGCACAGTTTAAAACCGCCCAAAAAA*A
2381	TTTTTTGGGCGGTTGTCAACTGTGCTTTTTTTTTTTTTTT/36-FAM/
2382	AAAAAAAAAAAAAGCACAGTTGACAACCGCCCAAAAAA*A
2383	TTTTTTGGGCGGTTTGAAACTGTGCTTTTTTTTTTTTTTT/36-FAM/
2384	AAAAAAAAAAAAAGCACAGTTTCAAACCGCCCAAAAAA*A
2385	TTTTTTGGGCGATTTTAAATTGTGCTTTTTTTTTTTTTTT/36-FAM/
2386	AAAAAAAAAAAAAGCACAAATTTAAATCGCCCAAAAAA*A
“*” is a phosphorothioate bond between these base pairs “/36-FAM/” is a 3'-6-carboxyfluorescein modification	

EXAMPLE 5

Improvements Over SEQ ID NO: 2 in the Transposition of Double-Stranded DNA (dsDNA) Adapters into dsDNA Target Sequences

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 2 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2. Activity of the transposase variants was

10 determined using a capillary electrophoresis (CE) assay as shown in **Figure 2**.

Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μM seqWell UR1 adapter supplemented with 7 mM MgCl₂ at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final

15 concentration of 0.03-0.3 μM. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 5.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 5.1**, and analyzed on CE as

20 described in Example 4.

Table 5.1 Reaction Conditions
Transposase expression: 20°C in <i>E. coli</i> BL-21, Lysis buffer: 50 mM Tris-HCL, 150mM NaCl, 0.1 mM EDTA, 50% glycerol, pH 7.5, 1.0 g/L lysozyme; Lysis conditions: 90 minute resuspension at room temperature followed by incubation at 31°C for 30 minutes; Adapter concentration: 14.3 μM; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, at 0.3 μM; Final dilution before CE analysis: 40x

Activity relative to SEQ ID NO: 2 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID

NO: 2 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 5.2**.

Table 5.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	FIOP Percent Product (Relative to SEQ ID NO: 2)¹
3/4	Q340L	+++
5/6	D291Y/Q340L	+++
7/8	G359A	+++
9/10	M177I/G359A	+++
11/12	R346D	+++
13/14	I358V	+++
15/16	S10A	+++
17/18	I317W	+++
19/20	I448L	+++
21/22	Q191Y	+++
23/24	H167R	+++
25/26	S217G	++
27/28	L194F	++
29/30	I305V	++
31/32	P371L	++
33/34	D291G	++
35/36	K289P	++
37/38	T44P	++
39/40	V352L	++
41/42	S309E	++
43/44	L362I	++
45/46	E55K	++
47/48	F396I	++
49/50	Q54T	++
51/52	D107G	++
53/54	S39P	++
55/56	S437R	++
57/58	L82R	++
59/60	S192E	+
61/62	A447V	+
63/64	A443P	+
65/66	V281I	+
67/68	R62D/R361H	+
69/70	I358M	+
71/72	Y193L	+
73/74	G415P	+
75/76	L428I	+
77/78	E310D	+
79/80	D164E	+

Table 5.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2)	FIOF Percent Product (Relative to SEQ ID NO: 2) ¹
81/82	L128A	+
83/84	I89V	+
85/86	S353A	+
87/88	N414D/M435L	+
89/90	G75H	+
91/92	R24L	+
93/94	M35L	+
95/96	I137L	+
97/98	K51E	+
99/100	H13E	+
101/102	I92L	+
103/104	F357I	+
105/106	A337I	+
107/108	S46A	+
109/110	N201G	+
111/112	R72L	+
113/114	W331L	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2, and defined as follows: "+" 1.09 to 2.20; "++" > 2.20; and "+++" > 4.07.		

EXAMPLE 6

Improvements Over SEQ ID NO: 4 in the Transposition of Double-Stranded DNA (dsDNA) Adapters into dsDNA Target Sequences

HTP Screening for Improved Transposase Variants

SEQ ID NO: 4 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 6.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 6.1**, and analyzed on CE as described in Example 4.

Table 6.1 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 50% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 90-minute resuspension at room temperature followed by incubation at 31°C for 30 minutes; Adapter concentration: 17 µM; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, at 0.3 µM; Final dilution before CE analysis: 70x

- Activity relative to SEQ ID NO: 4 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID NO: 4 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 6.2**.

Table 6.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4)	FIOP Percent Product (Relative to SEQ ID NO: 4) ¹
115/116	A118S	+++
117/118	K259S	+++
119/120	N18A	+++
121/122	R411K	+++
123/124	S10A/R346D/G359A	+++
125/126	K289P	+++
127/128	S182A	+++
129/130	A368S	+++
131/132	T286I	+++
133/134	T287N	+++
135/136	R346D	+++
137/138	H313R	+++
139/140	V262L	+++
141/142	S10A	+++
143/144	S10A/R346D	+++
145/146	S309A	+++
147/148	K16S	+++
149/150	E226K	+++
151/152	C146M	++
153/154	I317W	++
155/156	S10A/I317W/G359A	++
157/158	V298L	++
159/160	H13V	++
161/162	S427A	++
163/164	T6A	++
165/166	S10A/G359A	++

Table 6.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4)	FIOF Percent Product (Relative to SEQ ID NO: 4) ¹
167/168	P68A	++
169/170	E12S	++
171/172	S74A	++
173/174	S10A/I358V	++
175/176	C355L	++
177/178	T126V	++
179/180	V70A	++
181/182	C73K	++
183/184	V264S	++
185/186	E30N/A154V	++
187/188	T19L	++
189/190	I317W/R346D	++
191/192	F4H	++
193/194	S10A/I317W/I358V	++
195/196	I305V	++
197/198	K263R	++
199/200	Y282L	++
201/202	S288P	++
203/204	E30N	++
205/206	S106E	+
207/208	S10A/I317W	+
209/210	V280A	+
211/212	Q8D	+
213/214	T31V	+
215/216	D291G	+
217/218	E293T	+
219/220	K135R	+
221/222	M35L	+
223/224	S246K	+
225/226	R37K	+
227/228	E15S	+
229/230	E310Q	+
231/232	I448L	+
233/234	A163L	+
235/236	K247R	+
237/238	D34Q	+
239/240	D93E	+
241/242	M292E	+
243/244	D237S	+
245/246	E12S/Y373C	+
247/248	I358V	+
249/250	I360V	+

Table 6.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4)	FOP Percent Product (Relative to SEQ ID NO: 4) ¹
251/252	A208S	+
253/254	G359A	+
255/256	T134F	+
257/258	T454A	+
259/260	K26A	+
261/262	C38Y	+
263/264	D143E	+
265/266	I317W/G359A	+
267/268	N5R	+
269/270	A154P/D453E	+
271/272	D453E	+
273/274	I369F	+
275/276	A320H	+
277/278	K422Q	+
279/280	A351M	+
281/282	P410G	+
283/284	Q426M	+
285/286	R207C/T286I	+
287/288	E446G	+
289/290	A356S	+
291/292	L428I	+
293/294	I317W/W445L	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4, and defined as follows: "+" 1.13 to 174; "++" > 174; and "+++" > 250.		

EXAMPLE 7

**Improvements Over SEQ ID NO: 156 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters Into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 156 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
15 concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 7.1**.

Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 7.1**, and analyzed on CE as described in Example 4.

Table 7.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 50% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 90-minute resuspension at room temperature followed by incubation at 31°C for 30 minutes; Adapter concentration: 17 µM; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, at 0.3 µM; Final dilution before CE analysis: 70x

- 5 Activity relative to SEQ ID NO: 156 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380 produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380 observed by the reaction with SEQ ID NO: 156 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in Table 7.2.

10

Table 7.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 156)	FIOP Percent Product (Relative to SEQ ID NO: 156)¹
295/296	S74A/K259S	+++
297/298	K259S	+++
299/300	N18A/S74A/K259S	+++
301/302	N18A/C73K/S74A/K259S	++
303/304	N18A/S74A/K259S/T286I/K289P	++
305/306	N18A/A118S/K259S	++
307/308	K289P	++
309/310	N18A/K259S/T286I/K289P/R411K	++
311/312	S74A/K259S/K289P	++
313/314	N18A/K289P/R411K	+
315/316	N18A/S74A/A118S/K259S/T286I/K289P/R411K	+
317/318	C73K/S74A/K259S/K289P	+
319/320	N18A/S74A/K259S/T286I/R411K	+
321/322	N18A/C73K/T286I/K289P/R411K	+
323/324	C73K/K259S/T286I	+
325/326	K259S/T286I/K289P	+
327/328	S74A/T286I/K289P/R411K	+
329/330	N18A/K259S/K289P/R411K	+
331/332	C73K/S74A/K259S/T286I/K289P	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 156, and defined as follows: "+" 1.53 to 1.87; "++" > 1.87; and "+++" > 2.22.		

EXAMPLE 8**Improvements Over SEQ ID NO: 302 in the Transposition of Double-Stranded DNA (dsDNA)****Adapters into dsDNA Target Sequences**

HTP Screening for Improved Transposase Variants

5 SEQ ID NO: 302 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 8.1**.

15 Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 8.1**, and analyzed on CE as described in Example 4.

Table 8.1 Reaction Conditions

Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 50% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 41°C for 30 minutes; Adapter concentration: 17 μ M; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, at 0.3 μ M; Final dilution before CE analysis: 70x

20 Activity relative to SEQ ID NO: 302 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID NO: 302 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 8.2**.

Table 8.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 302)	FIOP Percent Product (Relative to SEQ ID NO: 302) ¹
333/334	K16S/C146M/E226K/T287N/K289P/C355L	+++
335/336	K251L	+++
337/338	C146M/N150H/T287N/K289P/E310Q/C355L	+++
339/340	K16S/T19L/C355L/S427A	+++
341/342	K16S/R37K/C146M/T286I/T287N/K289P/E310Q/H313R/C355L	+++

Table 8.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 302)	FIOF Percent Product (Relative to SEQ ID NO: 302) ¹
343/344	C146M/T286I/T287N/K289P/C355L/S427A	+++
345/346	R37K/A74S/E310Q/C355L/S427A	+++
347/348	T287N/C355L/S427A	+++
349/350	C146M/E226K/T286I/C355L	+++
351/352	K135R/C355L	+++
353/354	S182A/C355L	+++
355/356	C146M/S182A/C355L	+++
357/358	A74S/E310Q/H313R/C355L	+++
359/360	A74S/C355L/S427A	+++
361/362	T19L/R37K/C146M/T286I/T287N/E310Q/C355L	+++
363/364	Q54V	+++
365/366	C355L	+++
367/368	R37K/V70A/E226K/E310Q/H313R/C355L	+++
369/370	C146M/S427A	+++
371/372	C146M/K289P	+++
373/374	T19L/K135R/T286I/T287N/K289P/C355L	+++
375/376	A74S/E310Q/C355L	+++
377/378	V70A/A74S/C355L/A368S	+++
379/380	C38F	+++
381/382	Q54S	+++
383/384	T19L/S182A/C355L	++
385/386	K408R	++
387/388	D453R	++
389/390	T454R	++
391/392	V244T	++
393/394	K135R/E226K/T286I/K289P/C355L/A368S	++
395/396	E393A	++
397/398	A470R	++
399/400	T44S	++
401/402	M53D	++
403/404	D453T	++
405/406	N150H/E226K/E310Q/H313R	++
407/408	R37K/K135R/C146M/S182A/E226K/T286I/T287N/ C355L/A368S	++
409/410	S478G	++
411/412	V70A/K135R/C146M/N150H/S182A/T286I/K289P/ H313R/S427A	++
413/414	E450L	++
415/416	R37K/C146M/T287N/K289P/S427A	++
417/418	L474S	++
419/420	I45V	++
421/422	A470N	++
423/424	V70A/A74S/E226K/T286I/T287N/K289P/A368S/S4 27A	++

Table 8.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 302)	FOP Percent Product (Relative to SEQ ID NO: 302) ¹
425/426	K249F	++
427/428	D453H	++
429/430	T44G	++
431/432	E450M	++
433/434	C47S	++
435/436	C146M	++
437/438	R37K/C146M/E226K/K289P/S427A	++
439/440	S182A/E226K/C355L	++
441/442	L469F	++
443/444	M407L	++
445/446	T44V	++
447/448	G476S	++
449/450	I45M	++
451/452	Q103R	++
453/454	G476T	++
455/456	R439G	++
457/458	D397S	++
459/460	S457A	++
461/462	K117R	+
463/464	K466R	+
465/466	K249I	+
467/468	E446Q	+
469/470	Q426W	+
471/472	A470M	+
473/474	Q426H	+
475/476	S478C	+
477/478	V244I	+
479/480	T454L	+
481/482	K249T	+
483/484	I43V	+
485/486	E446K	+
487/488	W317R/R411N	+
489/490	K247L	+
491/492	L474R	+
493/494	S478V	+
495/496	K249V	+
497/498	Q103M	+
499/500	S50D	+
501/502	G476L	+
503/504	E398V	+
505/506	D471S	+
507/508	V464L	+

Table 8.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 302)	FIOF Percent Product (Relative to SEQ ID NO: 302) ¹
509/510	W413Y	+
511/512	E55M	+
513/514	V460T	+
515/516	E450T	+
517/518	G472R	+
519/520	S478L	+
521/522	L300F	+
523/524	R411D	+
525/526	A470V	+
527/528	Q426V	+
529/530	V464M	+
531/532	Q54H	+
533/534	Q426A	+
535/536	S444G	+
537/538	L475P	+
539/540	K41R	+
541/542	D453G	+
543/544	K249Y	+
545/546	I43L	+
547/548	Q48D	+
549/550	Q426S	+
551/552	M405V	+
553/554	Q426M	+
555/556	L431M	+
557/558	K409G	+
559/560	M53L	+
561/562	S42G	+
563/564	K249L	+
565/566	C47M	+
567/568	M405F	+
569/570	W401V	+
571/572	E467G	+
573/574	D399V	+
575/576	D471A	+
577/578	K249S	+
579/580	G476D	+
581/582	A429G	+
583/584	C38N	+
585/586	V459M	+
587/588	K101S	+
589/590	A470Y	+

Table 8.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 302)	FIOF Percent Product (Relative to SEQ ID NO: 302) ¹
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 302, and defined as follows: "+" 1.01 to 1.49; "++" > 1.49; and "+++" > 1.93.		

EXAMPLE 9

**Improvements Over SEQ ID NO: 338 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 338 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
15 concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 9.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 9.1**, and analyzed on CE as described in Example 4.

Table 9.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 17 μ M; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, at 0.3 μ M; Final dilution before CE analysis: 70x

20 Activity relative to SEQ ID NO: 338 (Activity FIOF) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID NO: 338 (where the percent product may be set as the average of replicates or else the highest single
25 sample as appropriate). The results are shown in **Table 9.2**.

Table 9.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 338)	FOP Percent Product (Relative to SEQ ID NO: 338) ¹
591/592	H150N/K408G/R411N	+++
593/594	R37K/W317R/M405V/K408G	+++
595/596	H150N/H313R/M405V/R411N	+++
597/598	R37K/W317R/M405V/Q426M	++
599/600	K16S/H313R	++
601/602	K16S	++
603/604	H150N/K408G	++
605/606	R37K/H313R/W317R/K408G/R411N	++
607/608	R37K/M405V/K409D/S427M	+
609/610	H150N/M405V/K406S	+
611/612	E226K/M405V	+
613/614	W317R/V460L	+
615/616	R37K/H150N/K408G	+
617/618	K16S/K409D/R411N/S427M	+
619/620	H150N	+
621/622	K408G/Q426M	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 338, and defined as follows: "+" 3.25 to 4.01; "++" > 4.01; and "+++" > 4.99.		

EXAMPLE 10

Improvements Over SEQ ID NO: 606 in the Transposition of Double-Stranded DNA (dsDNA) Adapters into dsDNA Target Sequences

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 606 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- 10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
- 15 concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 10.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 10.1**, and analyzed on CE as described in Example 4.

Table 10.1 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 40-50% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 60-minute resuspension at room temperature followed by incubation at 41°C for 30 minutes; Adapter concentration: 11 µM; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 µM; Final dilution before CE analysis: 18x

- Activity relative to SEQ ID NO: 606 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID
- 5 NO: 606 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 10.2**.

Table 10.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOP Percent Product (Relative to SEQ ID NO: 606) ¹
623/624	S199W	+++
625/626	V183L	+++
627/628	D143R	+++
629/630	E223R	+++
631/632	P220R	+++
633/634	E343R	+++
635/636	F168Q	+++
637/638	A163L	+++
639/640	I140L	+++
641/642	A255R	+++
643/644	E326A	+++
645/646	I92T	+++
647/648	A163T	+++
649/650	E343G	+++
651/652	S106Q	+++
653/654	E293V	+++
655/656	G160T	+++
657/658	D107R	+++
659/660	E12M	+++
661/662	E223V	+++
663/664	I92L	+++
665/666	I140V	+++
667/668	D107G	++
669/670	G160Q	++
671/672	P22T	++
673/674	E152S	++

Table 10.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOF Percent Product (Relative to SEQ ID NO: 606) ¹
675/676	Q390R	++
677/678	D107H	++
679/680	D143V	++
681/682	A320S	++
683/684	F132G	++
685/686	G415S	++
687/688	L380R	++
689/690	T136L	++
691/692	D273P	++
693/694	T79Y	++
695/696	S259A	++
697/698	R179G	++
699/700	T370F	++
701/702	P22S	++
703/704	E326L	++
705/706	S199T	++
707/708	T134G	++
709/710	G412A	++
711/712	G412H	++
713/714	H167R	++
715/716	D143G	++
717/718	I140C	++
719/720	R346N	++
721/722	Y282G	++
723/724	D34G	++
725/726	S113K	++
727/728	G412R	++
729/730	F132R	++
731/732	G412I	++
733/734	D273T	+
735/736	L215V	+
737/738	G408R	+
739/740	S199R	+
741/742	Q195G	+
743/744	A80T	+
745/746	D175T	+
747/748	A320R	+
749/750	A266P	+
751/752	D382A	+
753/754	H13I	+
755/756	V243M	+
757/758	T319R	+

Table 10.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOF Percent Product (Relative to SEQ ID NO: 606) ¹
759/760	E226L	+
761/762	K290T	+
763/764	M292W	+
765/766	K290R	+
767/768	Y374L	+
769/770	T134E	+
771/772	D216E	+
773/774	E218G	+
775/776	S389G	+
777/778	E202M	+
779/780	S309G	+
781/782	A18Q	+
783/784	S261E	+
785/786	K290A	+
787/788	L82Q	+
789/790	D273K	+
791/792	V81T	+
793/794	A233R	+
795/796	T31M/V262A	+
797/798	L215R	+
799/800	V243L	+
801/802	K270R	+
803/804	T235V	+
805/806	N414R	+
807/808	I442K	+
809/810	G408A	+
811/812	N201G	+
813/814	E226Q	+
815/816	L365I	+
817/818	D382R	+
819/820	A424V	+
821/822	E386D	+
823/824	D237S	+
825/826	N201Q	+
827/828	R346G	+
829/830	K115R	+
831/832	T19L	+
833/834	F60L	+
835/836	I236T	+
837/838	D237G	+
839/840	K26R	+
841/842	T268A	+

Table 10.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOF Percent Product (Relative to SEQ ID NO: 606) ¹
843/844	V262L	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 606, and defined as follows: "+" 1.01 to 1.31; "++" > 1.31; and "+++" > 1.82.		

5 Additionally, target bias relative to SEQ ID NO: 606 (Promiscuity FIOF) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 606 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 10.3**.

Table 10.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOF Target Product Ratio (Relative to SEQ ID NO: 606) ¹
623/624	S199W	+++
627/628	D143R	+++
845/846	S113E	+++
631/632	P220R	+++
625/626	V183L	+++
629/630	E223R	+++
847/848	P254F	+++
849/850	Q195S	+++
851/852	H13N	+++
811/812	N201G	+++
853/854	P277K	+++
673/674	E152S	+++
757/758	T319R	+++
855/856	E69A	+++
633/634	E343R	+++
857/858	E344V	+++
685/686	G415S	+++
655/656	G160T	+++
649/650	E343G	+++
651/652	S106Q	+++
635/636	F168Q	+++
671/672	P22T	+++
859/860	E88L	+++
861/862	Y224A	+++
637/638	A163L	+++
647/648	A163T	+++

Table 10.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOP Target Product Ratio (Relative to SEQ ID NO: 606) ¹
653/654	E293V	+++
659/660	E12M	+++
863/864	S227Y	+++
695/696	S259A	++
865/866	E386Q	++
641/642	A255R	++
661/662	E223V	++
693/694	T79Y	++
731/732	G412I	++
867/868	H150R	++
869/870	E30H	++
665/666	I140V	++
669/670	G160Q	++
697/698	R179G	++
705/706	S199T	++
681/682	A320S	++
687/688	L380R	++
719/720	R346N	++
871/872	S2A	++
873/874	A320L	++
875/876	D175V	++
877/878	Y224S	++
639/640	I140L	++
675/676	Q390R	++
683/684	F132G	++
745/746	D175T	++
879/880	V418S	++
881/882	E15R	++
821/822	E386D	++
657/658	D107R	++
763/764	M292W	++
883/884	T134L	++
885/886	K135T	++
887/888	D291S	++
889/890	H13Y	++
715/716	D143G	++
727/728	G412R	++
769/770	T134E	++
779/780	S309G	++
803/804	T235V	++
891/892	L231V	++
893/894	Y373R	++

Table 10.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOP Target Product Ratio (Relative to SEQ ID NO: 606) ¹
691/692	D273P	++
817/818	D382R	++
895/896	S217F	++
897/898	L215P	++
899/900	L355G	++
677/678	D107H	+
699/700	T370F	+
711/712	G412H	+
733/734	D273T	+
741/742	Q195G	+
783/784	S261E	+
901/902	K115S	+
667/668	D107G	+
689/690	T136L	+
701/702	P22S	+
707/708	T134G	+
717/718	I140C	+
903/904	P304R	+
905/906	K272L	+
907/908	Q84R	+
909/910	T235A	+
723/724	D34G	+
729/730	F132R	+
789/790	D273K	+
833/834	F60L	+
911/912	V114P	+
913/914	L128F	+
915/916	P289T	+
917/918	L380G	+
747/748	A320R	+
709/710	G412A	+
721/722	Y282G	+
795/796	T31M/V262A	+
735/736	L215V	+
793/794	A233R	+
835/836	I236T	+
919/920	Q198S	+
827/828	R346G	+
831/832	T19L	+
921/922	S217G	+
923/924	K16S	+
759/760	E226L	+

Table 10.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOP Target Product Ratio (Relative to SEQ ID NO: 606) ¹
825/826	N201Q	+
679/680	D143V	+
739/740	S199R	+
755/756	V243M	+
753/754	H13I	+
773/774	E218G	+
775/776	S389G	+
781/782	A18Q	+
799/800	V243L	+
815/816	L365I	+
805/806	N414R	+
839/840	K26R	+
925/926	L219S	+
927/928	P87T	+
929/930	N287L	+
931/932	V267P	+
933/934	N287R	+
935/936	M468T	+
937/938	A320N	+
939/940	E12L	+
941/942	K51R	+
943/944	K270A	+
945/946	T307L	+
713/714	H167R	+
749/750	A266P	+
761/762	K290T	+
801/802	K270R	+
819/820	A424V	+
841/842	T268A	+
947/948	W423D	+
787/788	L82Q	+
809/810	G408A	+
949/950	E69L	+
951/952	M146K	+
953/954	K272P	+
955/956	N279G	+
957/958	M275L	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 606, and defined as follows: "+" 1.00 to 1.23; "++" > 1.23; and "+++" > 1.54.		

Additional reactions were run as described above but with modified conditions. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 10.4**.

Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 10.4** and analyzed on CE as described in Example 4.

Table 10.4 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 50% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 48°C for 30 minutes; Adapter concentration: 11 µM; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, at 0.03 µM; Final dilution before CE analysis: 17x

- 5 Activity relative to SEQ ID NO: 606 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID NO: 606 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 10.5**.

10

Table 10.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOP Percent Product (Relative to SEQ ID NO: 606)¹
959/960	C47M/E393A/D453R/T454R	+++
961/962	T44G/I45V/C47M/D453R/T454R	+++
963/964	I45V/C47S/Q103M/E393A/D453R/T454R/S457G	+++
965/966	I45V/C47S/D453R	+++
967/968	T44S/I45V/Q103M/D453R/S457G	+++
969/970	M53D/Q54V/V244I/E450T	++
971/972	I45V/C47M/E393A/D453R/T454R	++
973/974	I45V/D453R/T454R	++
975/976	T44S/I45V/D453T/T454R/S457G	++
977/978	I45V/Q103M/E393A/D453T/T454R	++
979/980	T44S/C47M/Q103M/D453R/T454R/S457G	++
981/982	T44G/I45V/C47M/D453H	++
983/984	T44G/I45V/C47M/G408R/D453R	++
985/986	C47M/Q103M/G408R/D453T/T454R	+
987/988	T44S/C47S/E393A/G408R/D453R/S457G	+
989/990	G408R/D453T/T454R	+
991/992	E446K/E450T/V464L	+
993/994	D453R/S457G	+
995/996	T44G/C47M/E393A/D453R/S457G	+
997/998	T44S/C47M/Q103M/E393A/G408R/D453R/T454R	+
999/1000	T44S/Q103M/G408R/D453R/T454R	+
1001/1002	C47S/G408R/D453R/S457G	+

Table 10.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 606)	FIOP Percent Product (Relative to SEQ ID NO: 606) ¹
1003/1004	Q103M/E393A/G408R/D453R/S457G	+
1005/1006	C47M/E393A/G408R/D453R/T454R/S457G	+
1007/1008	I45V/C47M/G408R/W423C/D453T/S457G	+
1009/1010	V244I/E450T	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 606, and defined as follows: "+" 3.58 to 5.14; "++" > 5.14; and "+++" > 9.45.		

EXAMPLE 11

**Improvements Over SEQ ID NO: 972 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 972 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 11.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 11.1**, and analyzed on CE as described in Example 4.

Table 11.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 15% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 50°C for 30 minutes; Adapter concentration: 11 μ M; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 μ M; Final dilution before CE analysis: 18x

20 Activity relative to SEQ ID NO: 972 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID NO: 972 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 11.2**.

Table 11.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP Percent Product (Relative to SEQ ID NO: 972) ¹
1011/1012	C38F/V45I/Q54V/V244T	+++
1013/1014	M53L/Q54S/A470R/G476S/S478G	+++
1015/1016	L474S/S478G	+++
1017/1018	Q54S/G472R	+++
1019/1020	C38F/V45I/M53D	+++
1021/1022	Q54V/V244I/G472R/G476S	++
1023/1024	C38F/S478G	++
1025/1026	C38F/V244I/A470R/G472R/L474S/S478G	++
1027/1028	C38F/V45I/G472R/G476S/S478G	++
1029/1030	C38F/V45I/Q54V/A470N/L474S/S478G	++
1031/1032	M53D/V244T/L474S/S478G	++
1033/1034	C38F/V45I/M53L/V244I	++
1035/1036	V244T	++
1037/1038	C38F/M53D/Q54S	++
1039/1040	V45I/M53L/Q54S/G472R/S478G	+
1041/1042	M53D	+
1043/1044	C38F/M53D/Q54S/V244I/G472R/S478G	+
1045/1046	C38F/M53D/Q54S/V244I	+
1047/1048	C38F/V45I/M53D/Q54S/G476S	+
1049/1050	Q54H	+
1051/1052	C38F/M53L/G472R	+
1053/1054	C38F/V45I/M53L/A470N/G472R	+
1055/1056	Q54V/V244T	+
1057/1058	V45I/Q54S/V244T/L474S/S478G	+
1059/1060	V45I/M53D/Q54H/V244I/A470N/L474S/G476S/S478G	+
1061/1062	C38F/V45I/G476S	+
1063/1064	V45I/M53D/Q54S/A470R/L474S/G476S	+
1065/1066	C38F/Q54S/V244I/A470N/L474S/G476S/S478G	+
1067/1068	C38F/M53D/Q54H/V244I/G472R/S478G	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 972, and defined as follows: "+" 1.24 to 1.41; "++" > 1.41; and "+++" > 1.76.		

Additional reactions were run as described above but with modified conditions. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 11.3**.

- 5 Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 11.3** and analyzed on CE as described in Example 4.

Table 11.3 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 30-40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 48°C for 30 minutes; Adapter concentration: 11 µM; Target substrate: annealed oligos SEQ ID NO: 2369/SEQ ID NO: 2370, SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 µM; Final dilution before CE analysis: 18x

- Activity relative to SEQ ID NO: 972 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID
- 5 NO: 972 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 11.4**.

Table 11.4

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP Percent Product (Relative to SEQ ID NO: 972) ¹
1069/1070	C38F	+++
1071/1072	V243L	+++
1073/1074	M53D/D143R/V243L/K270R/E343R	+++
1075/1076	K115R/V244I/K270R	+++
1077/1078	D143R/V243L/K270R	+++
1079/1080	C38F/F132G	+++
1081/1082	C38F/P220R/R317M	+++
1083/1084	P22T/T134G/P220R/R317M	++
1085/1086	C38F/V45I/V81T/D143R/V244I	++
1087/1088	C38F/V81T/V244I	++
1089/1090	V81T/F132G/V243L	++
1091/1092	M53D/K115R/A163L/V243L/K270R/E343R	++
1093/1094	V243L/K270R	++
1095/1096	V45I/D143R/V243L/K270R/E343R	++
1097/1098	V45I/V244I/K270R	++
1099/1100	C38F/V81T	++
1101/1102	P22T/T134G	++
1103/1104	V244I/K270R	++
1105/1106	S259A	++
1107/1108	C38F/M53D/V243L	+
1109/1110	V45I/V81T/V243L/E343R	+
1111/1112	V45I/V81T	+
1113/1114	P22T/T134G/R317M/G412H	+

Table 11.4		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP Percent Product (Relative to SEQ ID NO: 972) ¹
1115/1116	T134G/S259A	+
1117/1118	P64R/T134L/P220R/R317M/G412H	+
1119/1120	P22T/T134L/P220R	+
1121/1122	P22T/C38F/V45I/D107R/S259A/R317M/G412H	+
1123/1124	M53D/K270R	+
1125/1126	P22T/D107R/R317M/G412H	+
1127/1128	K270R	+
1129/1130	C38F/D107R/T134G/P220R	+
1131/1132	D107R/T134L/R317M	+
1133/1134	P22T/C38F/V45I/S259A/R317M/G412H	+
1135/1136	P22T/D107R	+
1137/1138	D107R	+
1139/1140	P22T/R317M	+
1141/1142	V244I	+
1143/1144	V45I/V81T/V243L	+
1145/1146	R317M	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 972, and defined as follows: "+" 3.48 to 5.22; "++" > 5.22; and "+++" > 9.80.		

- 5 Additionally, target bias relative to SEQ ID NO: 972 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 972 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 11.5**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 11.3**; results (unreported) were similar to those reported below.

Table 11.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP Target Product Ratio (Relative to SEQ ID NO: 972) ¹
1095/1096	V45I/D143R/V243L/K270R/E343R	+++
1073/1074	M53D/D143R/V243L/K270R/E343R	+++
1077/1078	D143R/V243L/K270R	+++
1081/1082	C38F/P220R/R317M	+++
1071/1072	V243L	+++
1069/1070	C38F	+++
1085/1086	C38F/V45I/V81T/D143R/V244I	+++
1147/1148	C38F/V81T/D143R/V243L/K270R	+++
1097/1098	V45I/V244I/K270R	+++
1149/1150	M53D/V244I/K270R	+++

Table 11.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP Target Product Ratio (Relative to SEQ ID NO: 972) ¹
1151/1152	C38F/V45I/D143R/K270R	+++
1143/1144	V45I/V81T/V243L	++
1139/1140	P22T/R317M	++
1105/1106	S259A	++
1117/1118	P64R/T134L/P220R/R317M/G412H	++
1153/1154	K270R	++
1121/1122	P22T/C38F/V45I/D107R/S259A/R317M/G412H	++
1131/1132	D107R/T134L/R317M	++
1093/1094	V243L/K270R	++
1155/1156	C38F/V81T/V183L/V243L/K270R	++
1157/1158	V244I/K270R	++
1159/1160	V81T/F132G	++
1161/1162	P22T/T134L/R317M/G412H	++
1129/1130	C38F/D107R/T134G/P220R	++
1163/1164	C38F/K270R/E343R	++
1101/1102	P22T/T134G	++
1079/1080	C38F/F132G	++
1111/1112	V45I/V81T	++
1165/1166	V45I/V81T/F132G/V243L/K270R	+
1133/1134	P22T/C38F/V45I/S259A/R317M/G412H	+
1167/1168	P64R/T134L	+
1107/1108	C38F/M53D/V243L	+
1169/1170	P22T/C38F/P64R/T134L/P220R	+
1171/1172	C38F/V81T/V243L	+
1113/1114	P22T/T134G/R317M/G412H	+
1173/1174	P220R/S259A	+
1135/1136	P22T/D107R	+
1175/1176	P22T/V45I/P220R/S259A	+
1075/1076	K115R/V244I/K270R	+
1177/1178	P22T/P64R/T134L/P220R	+
1089/1090	V81T/F132G/V243L	+
1123/1124	M53D/K270R	+
1179/1180	P22T/T134L/R317M	+
1137/1138	D107R	+
1181/1182	P22T/S259A/R317M	+
1183/1184	P64R/D107R/T134L	+
1185/1186	P22T/P64R/D107R/S259A/R317M/G412H	+
1187/1188	D107R/G412H	+
1189/1190	P64R/D107R/R317M	+
1191/1192	P22T/S259A/R317M/G412H	+
1193/1194	P22T/P220R	+
1087/1088	C38F/V81T/V244I	+

Table 11.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP Target Product Ratio (Relative to SEQ ID NO: 972) ¹
1099/1100	C38F/V81T	+
1195/1196	P22T/P220R/G412H/G415D	+
1197/1198	P64R	+
1199/1200	A255R	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 972, and defined as follows: "+" 1.00 to 1.37; "++" > 1.29; and "+++" > 1.55.		

EXAMPLE 12

Improvements Over SEQ ID NO: 1082 in the Transposition of Double-Stranded DNA (dsDNA) Adapters into dsDNA Target Sequences

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 1082 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
15 concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 12.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 12.1**, and analyzed on CE as described in Example 4.

Table 12.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 48°C for 30 minutes; Adapter concentration: 11 μ M; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 μ M; Final dilution before CE analysis: 18x

20 Activity relative to SEQ ID NO: 1082 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID NO: 1082 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 12.2**.

Table 12.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOP Percent Product (Relative to SEQ ID NO: 1082) ¹
1201/1202	E187G	+++
1203/1204	F168Q/E187G/D216E/A255R/W423D	+++
1205/1206	S199W/N201G/G415S	+++
1207/1208	S199W/N201G/E226R	+++
1209/1210	G160Q/N201G/K272W	+++
1211/1212	F168Q/A255R/W423D	+++
1213/1214	E187G/W423D	++
1215/1216	D237S	++
1217/1218	F168Q/K290R/W423D	++
1219/1220	K272W	++
1221/1222	K290R/W423D	++
1223/1224	F168Q	++
1225/1226	F168Q/D216E/D237S	++
1227/1228	D216E	++
1229/1230	D237S/K290R	++
1231/1232	G160Q	++
1233/1234	V244I	+
1235/1236	E152S/F168Q/D216E/D237S	+
1237/1238	N201G	+
1239/1240	G415S	+
1241/1242	E226R/G415S	+
1243/1244	A255R	+
1245/1246	G160Q/E226R	+
1247/1248	G160Q/G415S	+
1249/1250	E152S/F168Q/W423D	+
1251/1252	W423D	+
1253/1254	K290R	+
1255/1256	G160Q/N201G/D273P/G415S	+
1257/1258	D143R	+
1259/1260	G160Q/V244I/K272W/D273P/G415S	+
1261/1262	D216E/W423D	+
1263/1264	E226R/V244I/G415S	+
1265/1266	D143R/D175T/K272W/D273P/G415S	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1082, and defined as follows: "+" 1.01 to 2.22; "++" > 2.22; and "+++" > 2.39.		

Additionally, target bias relative to SEQ ID NO: 1082 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 1082 (where the control

5

ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 12.3**.

Table 12.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOP Target Product Ratio Relative to SEQ ID NO: 1082¹
1205/1206	S199W/N201G/G415S	+++
1249/1250	E152S/F168Q/W423D	+++
1207/1208	S199W/N201G/E226R	+++
1231/1232	G160Q	+++
1253/1254	K290R	+++
1221/1222	K290R/W423D	+++
1211/1212	F168Q/A255R/W423D	++
1217/1218	F168Q/K290R/W423D	++
1267/1268	F168Q	++
1209/1210	G160Q/N201G/K272W	++
1201/1202	E187G	++
1227/1228	D216E	++
1235/1236	E152S/F168Q/D216E/D237S	++
1269/1270	D237S	++
1229/1230	D237S/K290R	++
1243/1244	A255R	++
1271/1272	D143R/G160T/E226R/G415S	++
1219/1220	K272W	+
1233/1234	V244I	+
1225/1226	F168Q/D216E/D237S	+
1237/1238	N201G	+
1251/1252	W423D	+
1257/1258	D143R	+
1247/1248	G160Q/G415S	+
1239/1240	G415S	+
1241/1242	E226R/G415S	+
1245/1246	G160Q/E226R	+
1265/1266	D143R/D175T/K272W/D273P/G415S	+
1255/1256	G160Q/N201G/D273P/G415S	+
1203/1204	F168Q/E187G/D216E/A255R/W423D	+
1273/1274	F168Q/D216E/D237S/A255R/E343R	+
1259/1260	G160Q/V244I/K272W/D273P/G415S	+
1261/1262	D216E/W423D	+
1263/1264	E226R/V244I/G415S	+
1275/1276	A255R/E343R	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1082, and defined as follows: "+" 1.10 to 2.75; "++" > 2.75; and "+++" > 4.18.		

Additional reactions were run as described above but with modified conditions. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 12.4**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 12.4** and analyzed on CE as described in Example 4.

5

Table 12.4 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 45°C for 30 minutes; Adapter concentration: 11 µM; Target substrate: annealed oligos SEQ ID NO: 2369/SEQ ID NO: 2370, SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 µM; Final dilution before CE analysis: 18x

Activity relative to SEQ ID NO: 1082 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID NO: 1082 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 12.5**.

10

Table 12.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOP Percent Product (Relative to SEQ ID NO: 1082)¹
1277/1278	D164Q	+++
1279/1280	W122F	+++
1281/1282	E152D	+++
1283/1284	D155T	+++
1285/1286	A154S	+++
1287/1288	V205L	+++
1289/1290	W145F	+++
1291/1292	H150D	+++
1293/1294	C184S	+++
1295/1296	E152Q	+++
1297/1298	D155Q	+++
1299/1300	A320Y	+++
1301/1302	A320Q	+++
1303/1304	A154K	+++
1305/1306	K161R	+++
1307/1308	D93A	+++

Table 12.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOP Percent Product (Relative to SEQ ID NO: 1082) ¹
1309/1310	D143A	+++
1311/1312	K161V	+++
1313/1314	I92A	+++
1315/1316	K161C	+++
1317/1318	C184A	+++
1319/1320	F204L	+++
1321/1322	K157R	+++
1323/1324	E152V	+++
1325/1326	H150W	++
1327/1328	D155G	++
1329/1330	Y193H	++
1331/1332	A320H	++
1333/1334	A154Y	++
1335/1336	L299I	++
1337/1338	A154Q	++
1339/1340	H141Y	++
1341/1342	A320R	++
1343/1344	D164S	++
1345/1346	F204Y	++
1347/1348	I92V	++
1349/1350	P151Q	++
1351/1352	K197R	++
1353/1354	F168V	++
1355/1356	P151A	++
1357/1358	H150G	++
1359/1360	V205I	++
1361/1362	G160W	++
1363/1364	L129I	++
1365/1366	H150E	++
1367/1368	A154T	++
1369/1370	Y193V	++
1371/1372	D164R	++
1373/1374	I181V	++
1375/1376	L90F	++
1377/1378	Y193L	++
1379/1380	A154P	++
1381/1382	D155E	++
1383/1384	Y193M	++
1385/1386	I181L	++
1387/1388	A91I	++
1389/1390	A165S	++
1391/1392	V206L	++

Table 12.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOP Percent Product (Relative to SEQ ID NO: 1082) ¹
1393/1394	C184G	++
1395/1396	A91V	++
1397/1398	A320C	+
1399/1400	C184H	+
1401/1402	F168L	+
1403/1404	T95G	+
1405/1406	D143K	+
1407/1408	H150S	+
1409/1410	I92L	+
1411/1412	F168N	+
1413/1414	W122L	+
1415/1416	F168E	+
1417/1418	D155R	+
1419/1420	E152Y	+
1421/1422	V205L/G472D	+
1423/1424	P151T/D155T	+
1425/1426	A154V	+
1427/1428	E152L	+
1429/1430	G160I	+
1431/1432	E152H	+
1433/1434	E152R	+
1435/1436	E152T	+
1437/1438	A154L	+
1439/1440	W145L	+
1441/1442	F204M	+
1443/1444	D93S	+
1445/1446	A320L	+
1447/1448	A320K	+
1449/1450	W122Q	+
1451/1452	K197H	+
1453/1454	A154E	+
1455/1456	F204V	+
1457/1458	V180C	+
1459/1460	A320S	+
1461/1462	W144F	+
1463/1464	W122V	+
1465/1466	L299F	+
1467/1468	W144Y	+
1469/1470	P151R	+
1471/1472	L129F	+
1473/1474	T95S	+
1475/1476	D155H	+

Table 12.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOP Percent Product (Relative to SEQ ID NO: 1082) ¹
1477/1478	G160Q	+
1479/1480	K161C/R313C	+
1481/1482	F204W	+
1483/1484	G160V	+
1485/1486	V180L	+
1487/1488	V180T	+
1489/1490	H150V	+
1491/1492	F168R	+
1493/1494	W144V	+
1495/1496	F168K	+
1497/1498	A165G	+
1499/1500	D155S	+
1501/1502	H150R	+
1503/1504	K197Y	+
1505/1506	R72Q	+
1507/1508	D93H	+
1509/1510	H150T	+
1511/1512	Q142L	+
1513/1514	R72M	+
1515/1516	W144I	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1082, and defined as follows: "+" 1.01 to 4.48; "++" > 4.48; and "+++" > 7.47.		

- 5 Additionally, target bias relative to SEQ ID NO: 1082 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 1082 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 12.6**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 12.3**; results (unreported) were similar to those reported below.

Table 12.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1082) ¹
1281/1282	E152D	+++
1291/1292	H150D	+++
1283/1284	D155T	+++
1405/1406	D143K	+++
1277/1278	D164Q	+++
1435/1436	E152T	+++

Table 12.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOF Target Product Ratio (Relative to SEQ ID NO: 1082) ¹
1393/1394	C184G	+++
1309/1310	D143A	+++
1287/1288	V205L	+++
1303/1304	A154K	+++
1429/1430	G160I	+++
1343/1344	D164S	+++
1319/1320	F204L	+++
1363/1364	L129I	+++
1329/1330	Y193H	+++
1411/1412	F168N	+++
1321/1322	K157R	+++
1279/1280	W122F	+++
1341/1342	A320R	+++
1313/1314	I92A	+++
1297/1298	D155Q	++
1353/1354	F168V	++
1373/1374	I181V	++
1299/1300	A320Y	++
1437/1438	A154L	++
1355/1356	P151A	++
1371/1372	D164R	++
1359/1360	V205I	++
1453/1454	A154E	++
1391/1392	V206L	++
1295/1296	E152Q	++
1517/1518	D155Q/A265V	++
1325/1326	H150W	++
1285/1286	A154S	++
1349/1350	P151Q	++
1409/1410	I92L	++
1345/1346	F204Y	++
1289/1290	W145F	++
1331/1332	A320H	++
1431/1432	E152H	++
1335/1336	L299I	++
1459/1460	A320S	++
1365/1366	H150E	++
1323/1324	E152V	++
1519/1520	I92P	++
1419/1420	E152Y	++
1339/1340	H141Y	++
1347/1348	I92V	++

Table 12.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOF Target Product Ratio (Relative to SEQ ID NO: 1082) ¹
1417/1418	D155R	++
1375/1376	L90F	++
1301/1302	A320Q	++
1351/1352	K197R	++
1425/1426	A154V	+
1465/1466	L299F	+
1367/1368	A154T	+
1317/1318	C184A	+
1337/1338	A154Q	+
1389/1390	A165S	+
1381/1382	D155E	+
1423/1424	P151T/D155T	+
1439/1440	W145L	+
1445/1446	A320L	+
1407/1408	H150S	+
1333/1334	A154Y	+
1385/1386	I181L	+
1433/1434	E152R	+
1415/1416	F168E	+
1497/1498	A165G	+
1469/1470	P151R	+
1369/1370	Y193V	+
1427/1428	E152L	+
1495/1496	F168K	+
1361/1362	G160W	+
1489/1490	H150V	+
1305/1306	K161R	+
1413/1414	W122L	+
1397/1398	A320C	+
1327/1328	D155G	+
1521/1522	P151K	+
1447/1448	A320K	+
1481/1482	F204W	+
1357/1358	H150G	+
1293/1294	C184S	+
1485/1486	V180L	+
1477/1478	G160Q	+
1467/1468	W144Y	+
1471/1472	L129F	+
1451/1452	K197H	+
1401/1402	F168L	+
1441/1442	F204M	+

Table 12.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1082)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1082) ¹
1483/1484	G160V	+
1377/1378	Y193L	+
1455/1456	F204V	+
1383/1384	Y193M	+
1421/1422	V205L/G472D	+
1457/1458	V180C	+
1523/1524	V180A	+
1461/1462	W144F	+
1491/1492	F168R	+
1499/1500	D155S	+
1501/1502	H150R	+
1311/1312	K161V	+
1475/1476	D155H	+
1315/1316	K161C	+
1487/1488	V180T	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1082, and defined as follows: "+" 1.00 to 2.45; "++" > 2.45; and "+++" > 3.48.		

EXAMPLE 13

**Improvements Over SEQ ID NO: 1218 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 1218 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 13.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 13.1**, and analyzed on CE as described in Example 4.

Table 13.1 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30minute resuspension at room temperature followed by incubation at 51°C for 30 minutes; Adapter concentration: 11 µM; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 µM; Final dilution before CE analysis: 18x

- Activity relative to SEQ ID NO: 1218 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID
- 5 NO: 1218 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 13.2**.

Table 13.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1218)	FIOP Percent Product (Relative to SEQ ID NO: 1218) ¹
1525/1526	V244I	+++
1527/1528	E226R/V244I	+++
1529/1530	D155T	+++
1531/1532	H150D/D155T/V244I	+++
1533/1534	D164Q	+++
1535/1536	G160I	+++
1537/1538	A154K	++
1539/1540	E152T	++
1541/1542	M146L/E152T/V244I	++
1543/1544	E226R	++
1545/1546	L129I/D164Q/C184A	++
1547/1548	A154K/V205I	++
1549/1550	G160I/D164Q/C184A	++
1551/1552	G160I/D164Q	++
1553/1554	L129I/G160I	+
1555/1556	D164Q/V205I	+
1557/1558	L129I	+
1559/1560	L129I/G160I/C184A	+
1561/1562	L129I/V205I	+
1563/1564	H150D/D155T/E226R	+
1565/1566	F204L	+
1567/1568	L129I/G160I/D164Q	+
1569/1570	A154K/G160I/C184A	+
1571/1572	D164Q/F204L	+
1573/1574	G160I/C184A	+
1575/1576	L129I/D164Q	+

Table 13.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1218)	FIOF Percent Product (Relative to SEQ ID NO: 1218) ¹
1577/1578	V205I	+
1579/1580	A154K/D164Q	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1218, and defined as follows: "+" 1.00 to 1.13; "++" > 1.13; and "+++" > 1.22.		

- 5 Additionally, target bias relative to SEQ ID NO: 1218 (Promiscuity FIOF) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 1218 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 13.3**.

Table 13.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1218)	FIOF Target Product Ratio (Relative to SEQ ID NO: 1218) ¹
1581/1582	D143R/E152T/E343R	+++
1583/1584	D143R/E152T/E226R	+++
1585/1586	D143R/E152T/E226R/V244I	+++
1587/1588	D143R/E152T	+++
1589/1590	D143R/E152T/E226R/E343R	+++
1591/1592	D143R/E226R/E343R	+++
1541/1542	M146L/E152T/V244I	+++
1593/1594	E152T/E226R/E343R	+++
1595/1596	D143R/D155T/E226R/V244I	+++
1597/1598	E152T	+++
1599/1600	E152T/E343R	+++
1601/1602	D143R/H150D/V244I	++
1603/1604	D143R	++
1605/1606	D143R/D155T/E226R/V244I/E343R	++
1607/1608	D143R/V244I/E343R	++
1609/1610	D143R/E226R/V244I/E343R	++
1611/1612	E152T/E226R	++
1613/1614	D143R/D155T/E343R	++
1615/1616	D143R/E226R	++
1617/1618	D143R/E343R	++
1619/1620	D143R/V244I	++
1621/1622	E152T/G174R/V244I/E343R	++
1579/1580	A154K/D164Q	++
1623/1624	D143R/D155T	++
1625/1626	D143R/H150D/E226R	++

Table 13.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1218)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1218) ¹
1627/1628	D143R/H150D/D155T/E226R/V244I	++
1629/1630	D143R/H150D/E226R/E343R	++
1631/1632	A154K	++
1633/1634	D164Q	+
1635/1636	D143R/H150D/D155T/V244I	+
1535/1536	G160I	+
1637/1638	L129I/A154K/D164Q	+
1639/1640	D143R/H150D/E226R/V244I/E343R	+
1641/1642	V244I	+
1531/1532	H150D/D155T/V244I	+
1567/1568	L129I/G160I/D164Q	+
1643/1644	A154K/D164Q/F204L	+
1543/1544	E226R	+
1551/1552	G160I/D164Q	+
1527/1528	E226R/V244I	+
1645/1646	L129I/A154K/D164Q/V205I	+
1557/1558	L129I	+
1647/1648	L129I/G160I	+
1547/1548	A154K/V205I	+
1529/1530	D155T	+
1569/1570	A154K/G160I/C184A	+
1649/1650	D143R/H150D/E343R	+
1651/1652	L129I/D143H	+
1653/1654	H150D/D155T/V244I/E343R	+
1563/1564	H150D/D155T/E226R	+
1655/1656	L129I/A154K/V205I	+
1657/1658	L129I/G160I/V205I	+
1659/1660	A154K/D164Q/F204L/V205I	+
1549/1550	G160I/D164Q/C184A	+
1661/1662	V205I	+
1663/1664	H150D/E226R/V244I/E343R	+
1665/1666	L129I/D164Q/V205I	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1218, and defined as follows: "+" 1.00 to 1.88; "++" > 1.88; and "+++" > 2.43.		

EXAMPLE 14**Improvements Over SEQ ID NO: 1580 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences****HTP Screening for Improved Transposase Variants**

5 SEQ ID NO: 1580 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 14.1**.

15 Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 14.1**, and analyzed on CE as described in Example 4.

Table 14.1 Reaction Conditions

Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 51°C for 30 minutes; Adapter concentration: 11 μ M; Target substrate: annealed oligos SEQ ID NO: 2369/SEQ ID NO: 2370, SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 μ M; Final dilution before CE analysis: 18x
--

20 Activity relative to SEQ ID NO: 1580 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID NO: 1580 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 14.2**.

Table 14.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
1667/1668	K157R	+++
1669/1670	P151Q/A255R	+++
1671/1672	K157R/Q164D	+++

Table 14.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
1673/1674	P151Q	++
1675/1676	K157R/Q168K	++
1677/1678	P151Q/E152T/D273Y	++
1679/1680	Q164D/A255R	++
1681/1682	Q164D	++
1683/1684	V183I	++
1685/1686	K157R/Q164R/Q168K	+
1687/1688	Q168K	+
1689/1690	L129I/K157R/Q168K	+
1691/1692	A255R/E343R	+
1693/1694	E343R	+
1695/1696	P151Q/E152T/A255R	+
1697/1698	L98I/L129I/K157R/Q168K	+
1699/1700	P151R/E343R	+
1701/1702	A255R	+
1703/1704	K157R/Q164R	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1580, and defined as follows: "+" 1.07 to 1.49; "++" > 1.49; and "+++" > 2.21.		

- Additionally, target bias relative to SEQ ID NO: 1580 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 1580 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 14.3**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 14.1**; results (unreported) were similar to those reported below.

Table 14.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
1705/1706	D143R/A255R	+++
1707/1708	D143R/Q164D/E343R	+++
1709/1710	L129I	+++
1711/1712	D143R/Q164D/A255R	+++
1713/1714	D143R/P151R/E152T	+++
1715/1716	Q168K	+++
1717/1718	E152T/A255R	+++
1719/1720	Q164R/Q168K	++
1721/1722	P151R/E152T/E343R	++
1723/1724	L129I/Q164R	++
1725/1726	E152T	++

Table 14.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOF Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
1727/1728	L129I/Q164R/Q168K/V183M	++
1729/1730	Q164R	++
1693/1694	E343R	++
1731/1732	D143R/P151R	++
1733/1734	L129I/Q164R/V183I	++
1735/1736	D143R	++
1737/1738	P151R/E343R	++
1739/1740	V298I/E343R	++
1701/1702	A255R	+
1681/1682	Q164D	+
1741/1742	Q164R/A447T	+
1743/1744	P151R	+
1745/1746	P151R/A255R/E343R	+
1747/1748	P151R/E152T	+
1683/1684	V183I	+
1749/1750	D143R/P151R/Q164D	+
1685/1686	K157R/Q164R/Q168K	+
1751/1752	A255R/E343R	+
1753/1754	D143R/E343R	+
1755/1756	K157R	+
1675/1676	K157R/Q168K	+
1757/1758	P151R/E152T/Q164D/E343R	+
1759/1760	L129I/K157R	+
1761/1762	D143R/Q164D	+
1763/1764	P151R/A255R	+
1765/1766	L129I/K157R/Q168K	+
1679/1680	Q164D/A255R	+
1767/1768	D143R/P151Q	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1580, and defined as follows: "+" 1.02 to 1.53; "++" > 1.53; and "+++" > 2.46.		

Additional reactions were run as described above but with modified conditions. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 14.4**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 14.4** and analyzed on CE as described in Example 4.

Table 14.4 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 41°C for 30 minutes; Adapter concentration: 11 μM; Target substrate: annealed oligos SEQ ID NO: 2369/SEQ ID NO: 2370, SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 μM; Final dilution before CE analysis: 18x

- Activity relative to SEQ ID NO: 1580 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID
- 5 NO: 1580 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 14.5**.

Table 14.5

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
1769/1770	W122V/Q164D	+++
1771/1772	K157V	+++
1773/1774	K161C	+++
1775/1776	P151W	+++
1777/1778	W145L	+++
1779/1780	C169I	+++
1781/1782	H150T	+++
1783/1784	E202I	+++
1785/1786	M177S	+++
1787/1788	K73W	+++
1789/1790	R203Q	+++
1791/1792	H150W	+++
1793/1794	A188G	+++
1795/1796	H200V	+++
1797/1798	Q198R	+++
1799/1800	A165H	+++
1801/1802	L98V	+++
1803/1804	P151T	+++
1805/1806	S192A	+++
1807/1808	W121L/Q164D	+++
1809/1810	S192L	+++
1811/1812	S199G	+++
1813/1814	S199Y	+++

Table 14.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
1815/1816	M146F	+++
1817/1818	H150A	+++
1819/1820	E202L	+++
1821/1822	M177V	+++
1823/1824	Q195G	+++
1825/1826	S199W	+++
1827/1828	A163G	+++
1829/1830	E349L	+++
1831/1832	K161M	+++
1833/1834	I140L	+++
1835/1836	R172V	+++
1837/1838	G160L	+++
1839/1840	E202S	+++
1841/1842	M146R	+++
1843/1844	A208C	+++
1845/1846	E156Y	+++
1847/1848	W122V	+++
1849/1850	A163W	+++
1851/1852	K110T	+++
1853/1854	K161T	+++
1855/1856	C169S	+++
1857/1858	T95V	+++
1859/1860	H200A	+++
1861/1862	K157Q	+++
1863/1864	L139I	+++
1865/1866	M146K	+++
1867/1868	I140V	+++
1869/1870	K157H	++
1871/1872	Q191G	++
1873/1874	W144V	++
1875/1876	G160Q	++
1877/1878	P151Q	++
1879/1880	P151E	++
1881/1882	A165V	++
1883/1884	I190G	++
1885/1886	Q164D/Q191R	++
1887/1888	S159Q	++
1889/1890	E349K	++
1891/1892	P151L/Q164D	++
1893/1894	P151S	++
1895/1896	L127A	++
1897/1898	P151V	++

Table 14.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
1899/1900	R321S	++
1901/1902	S159A	++
1903/1904	E202V	++
1905/1906	H150A/Q164D	++
1907/1908	K110V	++
1909/1910	P148G	++
1911/1912	P151Y	++
1913/1914	W122L	++
1915/1916	H210R	++
1917/1918	D153G/Q164D	++
1919/1920	K110R	++
1921/1922	P151Y/Q164D	++
1923/1924	Q195H	++
1925/1926	K161L	++
1927/1928	P151M	++
1929/1930	W122C	++
1931/1932	A91V	++
1933/1934	P151G	++
1935/1936	H210I	++
1937/1938	K161A	++
1939/1940	L194R	++
1941/1942	W145S	++
1943/1944	D153N	++
1945/1946	G112H	++
1947/1948	K157C	++
1949/1950	P151L	++
1951/1952	I190D	++
1953/1954	P151I	++
1955/1956	Q164D/S192E	++
1957/1958	W144L/Q164D	++
1959/1960	I190W	++
1961/1962	K157A	++
1963/1964	K209V	++
1965/1966	Q195W	++
1967/1968	Q198I	++
1969/1970	S166W	++
1971/1972	G160V	++
1973/1974	H150S/Q164D	++
1975/1976	R147W	++
1977/1978	S199L	++
1979/1980	G112A	++
1981/1982	M146V	++

Table 14.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
1983/1984	P151C	++
1985/1986	Q164D/S182W	++
1987/1988	T126A	++
1989/1990	A165T	++
1991/1992	E156L	++
1993/1994	Q164D/S199G	++
1995/1996	S159T	++
1997/1998	W121Q	++
1999/2000	W145Q	++
2001/2002	E156G	++
2003/2004	H124L	++
2005/2006	M177R	++
2007/2008	Q198G	++
2009/2010	W145G	++
2011/2012	H141T	++
2013/2014	H150G	++
2015/2016	H210P	++
2017/2018	Q164D/L173M	+
2019/2020	P148E	+
2021/2022	Q164D/S199L	+
2023/2024	W121A	+
2025/2026	P151K	+
2027/2028	S166V	+
2029/2030	A91C	+
2031/2032	M177W	+
2033/2034	W145T	+
2035/2036	W331L	+
2037/2038	D155F	+
2039/2040	P148R	+
2041/2042	P151A	+
2043/2044	S159G	+
2045/2046	E69V/Q164D	+
2047/2048	H210T	+
2049/2050	S166M	+
2051/2052	C169Y	+
2053/2054	P151N	+
2055/2056	Q164D/V205L/K249E	+
2057/2058	F204V	+
2059/2060	H167R	+
2061/2062	H210V	+
2063/2064	Q164D/A188G	+
2065/2066	A163D	+

Table 14.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
2067/2068	L194W	+
2069/2070	Q164D	+
2071/2072	C184Q	+
2073/2074	E187G	+
2075/2076	Q164D/Q195R	+
2077/2078	W145F	+
2079/2080	E69Q	+
2081/2082	P148F	+
2083/2084	Q164D/E202A	+
2085/2086	M146A	+
2087/2088	Q164D/R172M	+
2089/2090	Q164D/K209R	+
2091/2092	S159P	+
2093/2094	L194C	+
2095/2096	Q191W	+
2097/2098	T95S	+
2099/2100	H124M	+
2101/2102	H150L/Q164D	+
2103/2104	K157R	+
2105/2106	C184G	+
2107/2108	F204T	+
2109/2110	H150V	+
2111/2112	H200L	+
2113/2114	P148V	+
2115/2116	H200F	+
2117/2118	Q198N	+
2119/2120	H167V	+
2121/2122	K157S	+
2123/2124	E156C	+
2125/2126	G160S	+
2127/2128	H150L	+
2129/2130	H150R	+
2131/2132	R170L	+
2133/2134	R321A	+
2135/2136	W145K	+
2137/2138	Y193W	+
2139/2140	C184T	+
2141/2142	I92L/Q164D	+
2143/2144	K197S	+
2145/2146	K209G	+
2147/2148	Q198A	+
2149/2150	R172S	+

Table 14.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
2151/2152	A163Q	+
2153/2154	K197L/A266V	+
2155/2156	W144L	+
2157/2158	A163K	+
2159/2160	A163N	+
2161/2162	C184S	+
2163/2164	I190M	+
2165/2166	K209A	+
2167/2168	D153R	+
2169/2170	F204I	+
2171/2172	K161F	+
2173/2174	M146G	+
2175/2176	G160Y	+
2177/2178	L173I	+
2179/2180	M177T	+
2181/2182	R172T	+
2183/2184	D153G	+
2185/2186	K157P	+
2187/2188	K197M	+
2189/2190	L127I	+
2191/2192	R172I	+
2193/2194	W144F	+
2195/2196	M146Q	+
2197/2198	M177Y	+
2199/2200	T95G	+
2201/2202	V324M	+
2203/2204	C184E	+
2205/2206	K197A	+
2207/2208	Q171R	+
2209/2210	R72L	+
2211/2212	E202F	+
2213/2214	Q171I	+
2215/2216	R72Q	+
2217/2218	S166R	+
2219/2220	S97T	+
2221/2222	T126S	+
2223/2224	W122F	+
2225/2226	H200R	+
2227/2228	P148A	+
2229/2230	D153P	+
2231/2232	M177G	+
2233/2234	A91L	+

Table 14.5		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
2235/2236	H167L	+
2237/2238	H200Y	+
2239/2240	Q164D/K197A	+
2241/2242	Q191L	+
2243/2244	Q198K	+
2245/2246	S113V	+
2247/2248	K73Q	+
2249/2250	V205T	+
2251/2252	H150R/Q164D	+
2253/2254	Q171S	+
2255/2256	Q198L	+
2257/2258	C169T	+
2259/2260	W145R	+
2261/2262	G160A	+
2263/2264	T96G	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1580 and defined as follows: "+" 1.00 to 1.37; "++" > 1.37; and "+++" > 1.58.		

- 5 Additionally, target bias relative to SEQ ID NO: 1580 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 1580 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 14.6**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 14.1**; results (unreported) were similar to those reported below.

Table 14.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
1787/1788	K73W	+++
1779/1780	C169I	+++
1845/1846	E156Y	+++
1797/1798	Q198R	+++
1789/1790	R203Q	+++
1785/1786	M177S	+++
2035/2036	W331L	+++
1807/1808	W121L/Q164D	+++
1885/1886	Q164D/Q191R	+++
2265/2266	M146L/Q164D	+++
1791/1792	H150W	+++

Table 14.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
1771/1772	K157V	+++
1795/1796	H200V	+++
1805/1806	S192A	+++
1827/1828	A163G	+++
1801/1802	L98V	+++
2001/2002	E156G	+++
1993/1994	Q164D/S199G	+++
1811/1812	S199G	+++
1813/1814	S199Y	+++
1793/1794	A188G	+++
1825/1826	S199W	+++
2101/2102	H150L/Q164D	+++
1981/1982	M146V	+++
2185/2186	K157P	+++
1823/1824	Q195G	+++
2089/2090	Q164D/K209R	+++
2059/2060	H167R	+++
2063/2064	Q164D/A188G	+++
1781/1782	H150T	+++
1971/1972	G160V	+++
1809/1810	S192L	+++
2031/2032	M177W	+++
2119/2120	H167V	+++
1973/1974	H150S/Q164D	++
1821/1822	M177V	++
1961/1962	K157A	++
1815/1816	M146F	++
2257/2258	C169T	++
1955/1956	Q164D/S192E	++
2235/2236	H167L	++
2077/2078	W145F	++
2057/2058	F204V	++
1855/1856	C169S	++
1841/1842	M146R	++
2267/2268	Q171V	++
2075/2076	Q164D/Q195R	++
2207/2208	Q171R	++
1865/1866	M146K	++
2269/2270	E156S	++
2037/2038	D155F	++
2191/2192	R172I	++
2157/2158	A163K	++

Table 14.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
2147/2148	Q198A	++
2261/2262	G160A	++
2109/2110	H150V	++
2045/2046	E69V/Q164D	++
2271/2272	D155T	++
1905/1906	H150A/Q164D	++
1923/1924	Q195H	++
2013/2014	H150G	++
2247/2248	K73Q	++
2273/2274	D155S	++
2097/2098	T95S	++
2195/2196	M146Q	++
2115/2116	H200F	++
1911/1912	P151Y	++
2255/2256	Q198L	++
2029/2030	A91C	++
2111/2112	H200L	++
1873/1874	W144V	++
2179/2180	M177T	++
1927/1928	P151M	++
1871/1872	Q191G	++
2065/2066	A163D	++
2167/2168	D153R	++
2159/2160	A163N	++
2061/2062	H210V	++
1833/1834	I140L	++
1777/1778	W145L	++
2275/2276	E156T	++
2017/2018	Q164D/L173M	++
2117/2118	Q198N	++
2067/2068	L194W	++
2005/2006	M177R	+
1837/1838	G160L	+
2121/2122	K157S	++
1965/1966	Q195W	+
2079/2080	E69Q	+
2277/2278	L98M	+
2021/2022	Q164D/S199L	+
1867/1868	I140V	+
1879/1880	P151E	+
2085/2086	M146A	+
2025/2026	P151K	+

Table 14.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
2279/2280	E156R	+
2053/2054	P151N	+
2281/2282	E69V	+
2193/2194	W144F	+
2283/2284	Q195Y	+
1957/1958	W144L/Q164D	+
2209/2210	R72L	+
2123/2124	E156C	+
1875/1876	G160Q	+
2251/2252	H150R/Q164D	+
2183/2184	D153G	+
2221/2222	T126S	+
2285/2286	D155C	+
2049/2050	S166M	+
1917/1918	D153G/Q164D	+
2253/2254	Q171S	+
2197/2198	M177Y	+
2039/2040	P148R	+
2189/2190	L127I	+
1877/1878	P151Q	+
1849/1850	A163W	+
1889/1890	E349K	+
2177/2178	L173I	+
1987/1988	T126A	+
1819/1820	E202L	+
2107/2108	F204T	+
1835/1836	R172V	+
2125/2126	G160S	+
2047/2048	H210T	+
2287/2288	Q191F	+
1893/1894	P151S	+
2129/2130	H150R	+
1977/1978	S199L	+
2289/2290	Q195M	+
1859/1860	H200A	+
1775/1776	P151W	+
2291/2292	D155L	+
2027/2028	S166V	+
2215/2216	R72Q	+
2201/2202	V324M	+
2293/2294	A208S	+
2295/2296	S199F	+

Table 14.6		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
2151/2152	A163Q	+
1863/1864	L139I	+
2219/2220	S97T	+
2297/2298	Q142C	+
2073/2074	E187G	+
2055/2056	Q164D/V205L/K249E	+
1943/1944	D153N	+
2141/2142	I92L/Q164D	+
1989/1990	A165T	+
2299/2300	S182L	+
2175/2176	G160Y	+
2225/2226	H200R	+
1861/1862	K157Q	+
2169/2170	F204I	+
1817/1818	H150A	+
2301/2302	W144Y	+
1947/1948	K157C	+
2249/2250	V205T	+
2153/2154	K197L/A266V	+
2135/2136	W145K	+
2303/2304	A165G	+
1931/1932	A91V	+
2181/2182	R172T	+
1991/1992	E156L	+
2231/2232	M177G	+
2163/2164	I190M	+
2237/2238	H200Y	+
1897/1898	P151V	+
2305/2306	H200S	+
1983/1984	P151C	+
2127/2128	H150L	+
1963/1964	K209V	+
2213/2214	Q171I	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1580 and defined as follows: "+" 1.00 to 1.21; "++" > 1.21; and "+++" > 1.39.		

Further additional reactions were run as described above but with modified conditions. The exact concentration of UR1 adapter, target substrate identity and concentration are shown in **Table 14.7**.

Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted

5 with water to a final dilution as described in **Table 14.7** and analyzed on CE as described in Example 4.

Table 14.7 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30-minute resuspension at room temperature followed by incubation at 51°C for 30 minutes; Adapter concentration: 11 µM; Target substrate: annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.03 µM; Final dilution before CE analysis: 18x

- Activity relative to SEQ ID NO: 1580 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380, observed by the reaction with SEQ ID
- 5 NO: 1580 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 14.8**.

Table 14.8

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
1857/1858	T95V	+++
1831/1832	K161M	+++
1847/1848	W122V	+++
1769/1770	W122V/Q164D	+++
1937/1938	K161A	+++
2171/2172	K161F	+++
1773/1774	K161C	+++
1925/1926	K161L	+++
2201/2202	V324M	+++
1945/1946	G112H	+++
1929/1930	W122C	+++
1853/1854	K161T	+++
2073/2074	E187G	+++
1911/1912	P151Y	+++
1955/1956	Q164D/S192E	+++
2053/2054	P151N	+++
2307/2308	H150G/Q164D	+++
1779/1780	C169I	+++
1961/1962	K157A	+++
1921/1922	P151Y/Q164D	+++
1901/1902	S159A	+++
1983/1984	P151C	+++
1975/1976	R147W	+++
2061/2062	H210V	+++
2047/2048	H210T	+++
2081/2082	P148F	+++

Table 14.8		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
2199/2200	T95G	+++
2063/2064	Q164D/A188G	+++
2029/2030	A91C	+++
2265/2266	M146L/Q164D	+++
1771/1772	K157V	+++
2013/2014	H150G	+++
1881/1882	A165V	+++
1787/1788	K73W	+++
2267/2268	Q171V	+++
1951/1952	I190D	+++
1949/1950	P151L	++
1879/1880	P151E	++
1891/1892	P151L/Q164D	++
2121/2122	K157S	++
2185/2186	K157P	++
2065/2066	A163D	++
1985/1986	Q164D/S182W	++
1877/1878	P151Q	++
2195/2196	M146Q	++
1995/1996	S159T	++
2309/2310	Q198V	++
1917/1918	D153G/Q164D	++
1907/1908	K110V	++
1979/1980	G112A	++
1969/1970	S166W	++
2165/2166	K209A	++
2043/2044	S159G	++
1909/1910	P148G	++
1947/1948	K157C	++
1897/1898	P151V	++
1941/1942	W145S	++
2039/2040	P148R	++
1999/2000	W145Q	++
1793/1794	A188G	++
1993/1994	Q164D/S199G	++
2247/2248	K73Q	++
2021/2022	Q164D/S199L	++
2099/2100	H124M	++
2095/2096	Q191W	++
1797/1798	Q198R	++
2147/2148	Q198A	++
2041/2042	P151A	++

Table 14.8		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
1887/1888	S159Q	++
1893/1894	P151S	++
2027/2028	S166V	++
2097/2098	T95S	++
1981/1982	M146V	++
2137/2138	Y193W	++
2167/2168	D153R	++
2145/2146	K209G	++
1849/1850	A163W	++
2089/2090	Q164D/K209R	++
1781/1782	H150T	++
2091/2092	S159P	++
2183/2184	D153G	++
1927/1928	P151M	++
1977/1978	S199L	++
2303/2304	A165G	++
2173/2174	M146G	++
2241/2242	Q191L	++
1791/1792	H150W	++
2055/2056	Q164D/V205L/K249E	++
2071/2072	C184Q	++
2235/2236	H167L	+
2117/2118	Q198N	+
2033/2034	W145T	+
1973/1974	H150S/Q164D	+
2103/2104	K157R	+
1775/1776	P151W	+
2191/2192	R172I	+
2175/2176	G160Y	+
1883/1884	I190G	+
1805/1806	S192A	+
1777/1778	W145L	+
2311/2312	D93S	+
1815/1816	M146F	+
1933/1934	P151G	+
2251/2252	H150R/Q164D	+
1803/1804	P151T	+
2313/2314	W162Y	+
2069/2070	Q164D	+
1825/1826	S199W	+
2189/2190	L127I	+
2109/2110	H150V	+

Table 14.8		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
2259/2260	W145R	+
2207/2208	Q171R	+
2257/2258	C169T	+
1963/1964	K209V	+
2113/2114	P148V	+
1913/1914	W122L	+
1953/1954	P151I	+
2015/2016	H210P	+
2223/2224	W122F	+
2287/2288	Q191F	+
2315/2316	Q195R	+
1989/1990	A165T	+
2019/2020	P148E	+
2139/2140	C184T	+
2317/2318	R207K	+
2319/2320	V206L	+
2127/2128	H150L	+
2261/2262	G160A	+
2075/2076	Q164D/Q195R	+
2321/2322	D155W	+
2253/2254	Q171S	+
2233/2234	A91L	+
2193/2194	W144F	+
2293/2294	A208S	+
1813/1814	S199Y	+
2323/2324	A208V	+
1869/1870	K157H	+
1837/1838	G160L	+
1841/1842	M146R	+
2229/2230	D153P	+
1971/1972	G160V	+
2219/2220	S97T	+
1851/1852	K110T	+
1817/1818	H150A	+
2283/2284	Q195Y	+
2225/2226	H200R	+
2325/2326	T95R	+
2003/2004	H124L	+
2327/2328	G160M	+
1845/1846	E156Y	+
2329/2330	H124A	+
1863/1864	L139I	+

Table 14.8		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOP Percent Product (Relative to SEQ ID NO: 1580) ¹
2269/2270	E156S	+
2025/2026	P151K	+
1885/1886	Q164D/Q191R	+
2331/2332	D153Q	+
2077/2078	W145F	+
2291/2292	D155L	+
2031/2032	M177W	+
2049/2050	S166M	+
1923/1924	Q195H	+
2001/2002	E156G	+
2333/2334	F204L	+
2249/2250	V205T	+
2335/2336	R207Q	+
2085/2086	M146A	+
1811/1812	S199G	+
2301/2302	W144Y	+
1915/1916	H210R	+
1865/1866	M146K	+
2255/2256	Q198L	+
2125/2126	G160S	+
1801/1802	L98V	+
1959/1960	I190W	+
1827/1828	A163G	+
2153/2154	K197L/A266V	+
2163/2164	I190M	+
2273/2274	D155S	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1580 and defined as follows: "+" 1.00 to 1.81; "++" > 1.81; and "+++" > 2.39.		

- 5 Additionally, target bias relative to SEQ ID NO: 1580 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 1580 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 14.9**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 14.1**; results (unreported) were similar to those reported below.

Table 14.9		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOF Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
2265/2266	M146L/Q164D	+++
2247/2248	K73Q	+++
1955/1956	Q164D/S192E	+++
2235/2236	H167L	+++
1787/1788	K73W	+++
2063/2064	Q164D/A188G	+++
2147/2148	Q198A	+++
2279/2280	E156R	+++
2185/2186	K157P	+++
2261/2262	G160A	+++
1779/1780	C169I	+++
1793/1794	A188G	+++
1797/1798	Q198R	+++
2257/2258	C169T	+++
2267/2268	Q171V	+++
2201/2202	V324M	+++
2309/2310	Q198V	+++
2337/2338	D155M	+++
2339/2340	D196W	+++
1977/1978	S199L	+++
1805/1806	S192A	+++
1845/1846	E156Y	+++
2291/2292	D155L	+++
2341/2342	E156W	++
2315/2316	Q195R	++
2175/2176	G160Y	++
2287/2288	Q191F	++
2089/2090	Q164D/K209R	++
1971/1972	G160V	++
2271/2272	D155T	++
2269/2270	E156S	++
2343/2344	G160R	++
2207/2208	Q171R	++
2255/2256	Q198L	++
2273/2274	D155S	++
2303/2304	A165G	++
2283/2284	Q195Y	++
2275/2276	E156T	++
2001/2002	E156G	++
2125/2126	G160S	++
2117/2118	Q198N	++
2189/2190	L127I	++

Table 14.9		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	FIOF Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
2333/2334	F204L	++
2195/2196	M146Q	++
2183/2184	D153G	++
2191/2192	R172I	++
2251/2252	H150R/Q164D	++
1923/1924	Q195H	++
2059/2060	H167R	++
1791/1792	H150W	++
1885/1886	Q164D/Q191R	++
2053/2054	P151N	++
1825/1826	S199W	++
2327/2328	G160M	++
1809/1810	S192L	++
2061/2062	H210V	++
1961/1962	K157A	++
1865/1866	M146K	+
2021/2022	Q164D/S199L	+
2293/2294	A208S	+
1813/1814	S199Y	+
1841/1842	M146R	+
2037/2038	D155F	+
2121/2122	K157S	+
2065/2066	A163D	+
2253/2254	Q171S	+
2345/2346	D155P	+
2159/2160	A163N	+
2237/2238	H200Y	+
1837/1838	G160L	+
1875/1876	G160Q	+
1823/1824	Q195G	+
2123/2124	E156C	+
2285/2286	D155C	+
2157/2158	A163K	+
2347/2348	S192R	+
2095/2096	Q191W	+
1849/1850	A163W	+
1815/1816	M146F	+
2321/2322	D155W	+
2349/2350	A91G	+
2351/2352	D196C	+
2353/2354	D196G	+
2137/2138	Y193W	+

Table 14.9		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 1580)	Flop Target Product Ratio (Relative to SEQ ID NO: 1580) ¹
1973/1974	H150S/Q164D	+
2013/2014	H150G	+
2167/2168	D153R	+
2301/2302	W144Y	+
2193/2194	W144F	+
2025/2026	P151K	+
2331/2332	D153Q	+
2129/2130	H150R	+
1965/1966	Q195W	+
1871/1872	Q191G	+
2355/2356	C169A	+
2357/2358	E69G	+
1771/1772	K157V	+
2359/2360	V183L	+
2069/2070	Q164D	+
2075/2076	Q164D/Q195R	+
2281/2282	E69V	+
2249/2250	V205T	+
2289/2290	Q195M	+
2361/2362	Q171H	+
1867/1868	I140V	+
2363/2364	S182T	+
1993/1994	Q164D/S199G	+
2103/2104	K157R	+
2259/2260	W145R	+
2127/2128	H150L	+
2225/2226	H200R	+
2085/2086	M146A	+
2087/2088	Q164D/R172M	+
2365/2366	D196L	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 1580, and defined as follows: "+" 1.00 to 1.17; "++" > 1.17; and "+++" > 1.41.		

EXAMPLE 15**Relative Performance Measurements of Shake-Flask Purified Transposase Variants over SEQ ID NO: 972 in the Preparation of Libraries for Next Generation Sequencing Applications**

5 SEQ ID NO: 1720 was selected as the parent transposase enzyme. Variants of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptide encoded by the variants of SEQ ID NO: 1720, including SEQ ID NO: 2368, were produced in shake flask and purified as described in Example 3. Additionally, transposase variants of SEQ ID NO: 972, 1082, 1218, 1580, 1598, 1600, and

1604 were produced in shake flask and purified as described in Example 3. Transposase concentrations of select variants were measured by absorption at 280 nm and performance of the normalized enzymes was measured in the preparation of libraries for next generation sequencing as follows.

Reactions were prepared using seqWell reagents and were evaluated on an Illumina sequencer. To load the transposase enzymes the shake purified enzyme variants were diluted to 2.5uM in seqWell MTSB and combined in 1:1 ratio with the seqWell i5 tagging reagents (premixed 2 barcode blend) and incubated at room temperature for 1 hour. The same set of samples at concentration of 2.5uM were combined with seqWell i7 oligos in 1:1 ratio in a separate set of reactions and incubated at room temperature for one hour. The loaded i5 and i7 transposases were then combined in equal volumes and diluted to 15nM prior to library preparation. To prepare the library for sequencing 10ng of E.coli genomic DNA was mixed in equal volumes (5uL) with i5/i7 transposase mix and coding buffer. The reaction was incubated at 55°C for 15 minutes and terminated with 7.5uL of seqWell X-solution at 68°C for 10 minutes. The reaction was then purified with (1x) MAGwise beads and eluted in 25uL of 10mM Tris pH 8.0. This was followed by 12 cycle amplification and amplification purification with MAGwise beads. The samples were evaluated using Agilent Bioanalyzer and the Illumina MiSeq assay.

When bound to appropriate adapter sequences, active transposase variants are capable of binding target sequences, cleaving the target strand, and ligating the adapter oligonucleotides to the target fragment as shown in **Figure 1**.

Transposon insertion frequency is biased by the sequence of the insertion sites over the first 15 nucleotides of NGS reads putatively comprising the region of the target bound by the transposome construct. Bias plots generated from NGS data represent this bias as a function of the distance from the insertion site. Overall relative bias for each variant can be calculated as the root-mean-squared deviation (RMSD) of the observed vs. expected representation of all four nucleotides at each position as shown in Scheme 1, shown herein. Additionally, the mean RMSD over, for example, the first 15 nucleotides (RMSD₁₅) can be used to summarize insertion bias with a single number.

Relative target bias for each variant was calculated as the mean RMSD₁₅ of SEQ ID NO: 972 divided by the mean RMSD₁₅ of the transposase variant. Variants with target bias mean RMSD₁₅ less than SEQ ID NO: 972 will have a ratio greater than 1. The results are shown in **Table 15.1**.

Table 15.1		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP RMSD15 of SEQ ID NO:972 Relative to Variant¹
2367/2368	C38F/E152T/A154K/D164R/F168K/P220R/K290R/R317M/W423D	+++
1603/1604	C38F/D143R/F168Q/P220R/K290R/R317M/W423D	++
1597/1598	C38F/E152T/F168Q/P220R/K290R/R317M/W423D	++
1599/1600	C38F/E152T/F168Q/P220R/K290R/R317M/E343R/W423D	+
1579/1580	C38F/A154K/D164Q/F168Q/P220R/K290R/R317M/W423D	+

Table 15.1		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP RMSD15 of SEQ ID NO:972 Relative to Variant ¹
1217/1218	C38F/F168Q/P220R/K290R/R317M/W423D	+
1081/1082	C38F/P220R/R317M	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 972 and defined as follows: "+" 1.11 to 1.31; "++" > 1.31; and "+++" > 1.35.		

EXAMPLE 16

**Improvements Over SEQ ID NO: 2368 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 2368 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 16.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 16.1**, and analyzed on CE as described in Example 4.

Table 16.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 51°C for 30 minutes; Adapter concentration: 11 μ M; Target substrates: annealed oligos SEQ ID NO: 2375/SEQ ID NO: 2376 and SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.035 μ M; Final dilution before CE analysis: 18x

20 Activity relative to SEQ ID NO: 2368 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2375/SEQ ID NO: 2376 and SEQ ID NO: 2385/SEQ ID NO: 2386, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2375/SEQ ID NO: 2376 and SEQ ID NO: 2385/SEQ ID NO: 2386, observed by the reaction with SEQ ID NO: 2368 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate).
25 The results are shown in **Table 16.2**.

Table 16.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2368)	FIOP Percent Product (Relative to SEQ ID NO: 2368) ¹
2387/2388	V324M	+++
2389/2390	Q171V/V324M	+++
2391/2392	S192E/Y193W	+++
2393/2394	C169I	+++
2395/2396	P151N/V324M	+++
2397/2398	A163N/C169I	+++
2399/2400	M146L/V324M	+++
2401/2402	Q171V/A188G/V324M	++
2403/2404	A165G/V324M	++
2405/2406	S192A	++
2407/2408	A163D	++
2409/2410	E156Y/S192E	++
2411/2412	K73Q/D143K/C169I/Q191F/Y193W	+
2413/2414	P151N/Q171V/V324M	+
2415/2416	M146L	+
2417/2418	K73Q/E156R	+
2419/2420	K73Q/S192A/Y193W	+
2421/2422	K73Q/E156R/H210V	+
2423/2424	L127I/Q198R	+
2425/2426	L127I/G160A/R172I/Q198R	+
2427/2428	A163D/C169I/Q191F/S192A/Y193W	+
2429/2430	K157P/Q171V	+
2431/2432	A165G/Q171R/V324M	+
2433/2434	Q171R/V324M	+
2435/2436	L127I/R172I	+
2437/2438	H210V	+
2439/2440	Q198R	+
2441/2442	K73Q/E156Y/S192A/M435K	+
2443/2444	R172I	+
2445/2446	K73Q/A163N	+
2447/2448	K73Q/A163N/H210V	+
2449/2450	A165G/A188G/V324M	+
2451/2452	L127I	+
2453/2454	M146L/P151N/A188G	+
2455/2456	K73Q/A163D/H210Q	+

Table 16.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2368)	FIOP Percent Product (Relative to SEQ ID NO: 2368) ¹
2457/2458	A188G	+
2459/2460	K73Q/A163D	+
2461/2462	E156Y/C169I	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2368, and defined as follows: "+" 1.12 to 2.03; "++" 2.03 to 3.00; and "+++" > 3.00.		

- 5 Additionally, target bias relative to SEQ ID NO: 2368 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2375/SEQ ID NO: 2376 vs. product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 2368 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 16.3**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 16.1**; results (unreported) were similar to those reported below.

Table 16.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2368)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2368) ¹
2405/2406	S192A	+
2425/2426	L127I/G160A/R172I/Q198R	++
2463/2464	G160A/R172I	+++
2465/2466	G160A	+++
2467/2468	L127I/G160A/R172I	++
2469/2470	A163N/Q191F/S192A	+
2471/2472	W144Y	+
2473/2474	K73Q/A163N/Q191F/S192A/Y193W	+
2475/2476	K73Q	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2368 and defined as follows: "+" 1.04 to 1.32; "++" 1.32 to 1.54; and "+++" > 1.54.		

EXAMPLE 17**Improvements Over SEQ ID NO: 2426 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences****HTP Screening for Improved Transposase Variants**

5 SEQ ID NO: 2426 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 17.1**.

15 Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 17.1**, and analyzed on CE as described in Example 4.

Table 17.1 Reaction Conditions

Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 11 μ M; Target substrates: annealed oligos SEQ ID NO: 2375/SEQ ID NO: 2376 and SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.035 μ M; Final dilution before CE analysis: 18x
--

20 Activity relative to SEQ ID NO: 2426 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2375/SEQ ID NO: 2376 and SEQ ID NO: 2385/SEQ ID NO: 2386 produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2375/SEQ ID NO: 2376 and SEQ ID NO: 2385/SEQ ID NO: 2386, observed by the reaction with SEQ ID NO: 2426 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 17.2**.

25

Table 17.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426)¹
2477/2478	M146C/K157P	+++
2479/2480	M146K/K157P	+++
2481/2482	M146C/D153G/K157P	+++
2483/2484	T152E/K154A	+++

Table 17.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426) ¹
2485/2486	D143V/M146C/K157P	+++
2487/2488	D153G/K157P/H167L	+++
2489/2490	M146C/D153G/K157P/H167L	+++
2491/2492	H150N/C169T	+++
2493/2494	W145L	+++
2495/2496	D143V/M146C/D153G	+++
2497/2498	D143V/M146C/H167L/A188G	+++
2499/2500	M146C	+++
2501/2502	K270P	+++
2503/2504	K329A	+++
2505/2506	D143V/D153G/H167L/A188G/S246R	+++
2507/2508	R220P/E226S	+++
2509/2510	P151T/T152E/K154A	+++
2511/2512	L269E	+++
2513/2514	T333L	+++
2515/2516	H150N	+++
2517/2518	P151R/T152E/K154A	+++
2519/2520	T126V/I127L	+++
2521/2522	I190G	+++
2523/2524	T152E/K154A/D155N/A160G	+++
2525/2526	L380V	+++
2527/2528	T235S	+++
2529/2530	K73H	+++
2531/2532	R220P/F222R	+++
2533/2534	D143V/M146C	+++
2535/2536	G472R	+++
2537/2538	H210L	+++
2539/2540	K73Q/S192A	++
2541/2542	K209G	++
2543/2544	T307S	++
2545/2546	S288M/R290K	++
2547/2548	H150N/D155L/C169T/S192A	++
2549/2550	I127L/G131A/A165T	++
2551/2552	R290K/M292A	++
2553/2554	D143V/E156R/A188G	++

Table 17.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426) ¹
2555/2556	R220P/E226R	++
2557/2558	E156G/K270P	++
2559/2560	A18L	++
2561/2562	D143V/M146K	++
2563/2564	S217D/R220P	++
2565/2566	H150N/S199L	++
2567/2568	D155L/S192A	++
2569/2570	K73Q/D155L/C169T/S199L	++
2571/2572	S389L/A393E	++
2573/2574	Y86F/A384T	++
2575/2576	R346T	++
2577/2578	A443P	++
2579/2580	D143V/M146C/H167L	++
2581/2582	L375Y	++
2583/2584	L469A	++
2585/2586	H150N/D155L/S192A	++
2587/2588	S288P/R290K	++
2589/2590	S113V	++
2591/2592	V114D	++
2593/2594	K73Q/D155L	++
2595/2596	D143V/E156R/K157P/H167L/Q195H	++
2597/2598	G242S	++
2599/2600	D143V	++
2601/2602	H167L/A188G	++
2603/2604	K110L	++
2605/2606	D143V/A188G/Q195H	++
2607/2608	K270L	++
2609/2610	D143T	++
2611/2612	S389V/A393E	++
2613/2614	K272R	++
2615/2616	I214V/R220P	++
2617/2618	E306G	++
2619/2620	D143V/A188G	++
2621/2622	M146C/H167L	++
2623/2624	P277V	++

Table 17.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426) ¹
2625/2626	K272T	++
2627/2628	L90W	++
2629/2630	A91L	++
2631/2632	L387A/A393E	++
2633/2634	K378P	++
2635/2636	H150N/D155L/C169T	++
2637/2638	V352L	++
2639/2640	M146A/T152A	++
2641/2642	H150N/S192A	++
2643/2644	K378S	++
2645/2646	M146C/K157P/H167L	++
2647/2648	R290K/D291E	++
2649/2650	N385L	++
2651/2652	A383R	++
2653/2654	S113G	++
2655/2656	D143V/Q195H	++
2657/2658	E386S/A393E	++
2659/2660	I278V	++
2661/2662	S353V	++
2663/2664	R290K/P294S	++
2665/2666	R290K	++
2667/2668	S309V	++
2669/2670	A368W	++
2671/2672	G334L	+
2673/2674	G379P	+
2675/2676	K73R	+
2677/2678	Q276E	+
2679/2680	K209M	+
2681/2682	A91V	+
2683/2684	G441Q	+
2685/2686	G441W	+
2687/2688	Q390G/A393E	+
2689/2690	K73Q/H150N/Q171R/S192A	+
2691/2692	C392P/A393E	+
2693/2694	S113A	+

Table 17.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426) ¹
2695/2696	E386T/A393E	+
2697/2698	T95V	+
2699/2700	H13W	+
2701/2702	K73Q/H150N/D155L/C169T	+
2703/2704	H102G	+
2705/2706	R339G	+
2707/2708	S67G	+
2709/2710	K73Q/S199L	+
2711/2712	G20D/M146C/K157P/H167L	+
2713/2714	K270D	+
2715/2716	L380T	+
2717/2718	R290K/P294W	+
2719/2720	P304E	+
2721/2722	W122A/I127L	+
2723/2724	D143V/M146C/A188G	+
2725/2726	M146C/D153G/H167L	+
2727/2728	L269G	+
2729/2730	Q276G	+
2731/2732	K422G/D423W	+
2733/2734	R220P/F222C	+
2735/2736	S192L/R198Q	+
2737/2738	Q390E/A393E	+
2739/2740	Q8D	+
2741/2742	C169T	+
2743/2744	K272G	+
2745/2746	T268W	+
2747/2748	D382L	+
2749/2750	L82Q	+
2751/2752	M146K	+
2753/2754	S288V/R290K	+
2755/2756	L111M	+
2757/2758	K329R	+
2759/2760	L355C	+
2761/2762	R220P/L225M	+
2763/2764	R198G	+

Table 17.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426) ¹
2765/2766	M275R	+
2767/2768	R290V	+
2769/2770	Y373T	+
2771/2772	I89M	+
2773/2774	C184A	+
2775/2776	A160G/P371W	+
2777/2778	S217R/R220P	+
2779/2780	R341L	+
2781/2782	E274K	+
2783/2784	S259H	+
2785/2786	P151E/T152E/K154A	+
2787/2788	R290K/P294M	+
2789/2790	V262L	+
2791/2792	R290Q	+
2793/2794	T440G	+
2795/2796	D216I/R220P	+
2797/2798	P345T	+
2799/2800	M275W	+
2801/2802	A383F	+
2803/2804	A384N	+
2805/2806	N414G	+
2807/2808	R220L	+
2809/2810	D155L/Q171R	+
2811/2812	L82R	+
2813/2814	W144F	+
2815/2816	R290K/P294K	+
2817/2818	R198D	+
2819/2820	L219G/R220P	+
2821/2822	L215F/R220P	+
2823/2824	I214G/R220P	+
2825/2826	K168F/I172V	+
2827/2828	V206W	+
2829/2830	S217G/R220P	+
2831/2832	R220P/F222H	+
2833/2834	S389M/A393E	+

Table 17.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426) ¹
2835/2836	N287P/R290K	+
2837/2838	I305R	+
2839/2840	V114T	+
2841/2842	A18P	+
2843/2844	V267L	+
2845/2846	K422V/D423W	+
2847/2848	Q8V	+
2849/2850	D34H/F38C	+
2851/2852	V70G	+
2853/2854	S391L/A393E	+
2855/2856	V206F	+
2857/2858	G441S	+
2859/2860	S217L/R220P	+
2861/2862	I358C	+
2863/2864	T370S	+
2865/2866	R290K/P294R	+
2867/2868	A208V	+
2869/2870	S288R/R290K	+
2871/2872	G432T	+
2873/2874	K257G	+
2875/2876	K422S/D423W	+
2877/2878	E284S/R290K	+
2879/2880	K378G	+
2881/2882	E386A/A393E	+
2883/2884	L231G	+
2885/2886	I127L/T133M	+
2887/2888	Y86I	+
2889/2890	A393E/T394R	+
2891/2892	R366S	+
2893/2894	G472A	+
2895/2896	D423W/Y425A	+
2897/2898	S261K	+
2899/2900	K115Q	+
2901/2902	R290C	+
2903/2904	K73Q/D155L/Q171R	+

Table 17.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426) ¹
2905/2906	A18D	+
2907/2908	D423A	+
2909/2910	H78W	+
2911/2912	P372G	+
2913/2914	I190S	+
2915/2916	P289S/R290K	+
2917/2918	P304V	+
2919/2920	F38C/S39P/V45I	+
2921/2922	K329T	+
2923/2924	F38C/S39F/V45I	+
2925/2926	Q276A	+
2927/2928	Y282M	+
2929/2930	E386V/A393E	+
2931/2932	W145V/P151R	+
2933/2934	Q84T	+
2935/2936	R290K/D291G	+
2937/2938	G415V	+
2939/2940	E274R	+
2941/2942	R376L	+
2943/2944	R290K/E293L	+
2945/2946	K26R	+
2947/2948	L269A	+
2949/2950	H210G	+
2951/2952	E284A/R290K	+
2953/2954	G415P	+
2955/2956	V381S	+
2957/2958	S420E/D423W	+
2959/2960	R346V	+
2961/2962	K378A	+
2963/2964	I305G/Y373C	+
2965/2966	E344G	+
2967/2968	M317R	+
2969/2970	L387R/A393E	+
2971/2972	A384V	+
2973/2974	A283S/R290K	+

Table 17.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Percent Product (Relative to SEQ ID NO: 2426) ¹
2975/2976	A17V	+
2977/2978	R290K/D291S	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2426, and defined as follows: "+" 1.01 to 1.49; "++" 1.49 to 1.90; and "+++" > 1.90.		

5 Additionally, target bias relative to SEQ ID NO: 2426 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2375/SEQ ID NO: 2376 vs. product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 2426 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 17.3**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 17.1**; results (unreported) were similar to those reported below.

Table 17.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2426) ¹
2501/2502	K270P	++
2513/2514	T333L	++
2515/2516	H150N	+
2527/2528	T235S	+
2529/2530	K73H	++
2535/2536	G472R	+
2539/2540	K73Q/S192A	++
2543/2544	T307S	+
2547/2548	H150N/D155L/C169T/S192A	++
2551/2552	R290K/M292A	+
2559/2560	A18L	++
2565/2566	H150N/S199L	+
2567/2568	D155L/S192A	+
2573/2574	Y86F/A384T	+
2575/2576	R346T	++
2585/2586	H150N/D155L/S192A	++
2613/2614	K272R	+
2623/2624	P277V	+
2635/2636	H150N/D155L/C169T	+

Table 17.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2426) ¹
2641/2642	H150N/S192A	++
2647/2648	R290K/D291E	+
2657/2658	E386S/A393E	+
2663/2664	R290K/P294S	+
2677/2678	Q276E	+
2689/2690	K73Q/H150N/Q171R/S192A	++
2701/2702	K73Q/H150N/D155L/C169T	+
2707/2708	S67G	+
2709/2710	K73Q/S199L	++
2739/2740	Q8D	+
2745/2746	T268W	+
2749/2750	L82Q	++
2753/2754	S288V/R290K	+
2767/2768	R290V	+
2789/2790	V262L	+
2791/2792	R290Q	+
2793/2794	T440G	+
2833/2834	S389M/A393E	+
2837/2838	I305R	+
2863/2864	T370S	+
2873/2874	K257G	+
2887/2888	Y86I	+
2891/2892	R366S	+
2899/2900	K115Q	+++
2915/2916	P289S/R290K	+
2979/2980	S113E	+++
2981/2982	A74R	+++
2983/2984	T25G	+++
2985/2986	V298R	++
2987/2988	K197S/R198Q	++
2989/2990	P304W	++
2991/2992	K73Q/H150N/D155L/Q171R	++
2993/2994	H150N/D155L/S199L	++
2995/2996	R256G	++
2997/2998	A118S	++

Table 17.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2426)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2426) ¹
2999/3000	I29F	++
3001/3002	V267P/E343G	+
3003/3004	H150N/D155L/Q171R/S199L	+
3005/3006	N252A	+
3007/3008	R198Q/V205T	+
3009/3010	K115A	+
3011/3012	D143H	+
3013/3014	L269I	+
3015/3016	L82F	+
3017/3018	A393E/V395W	+
3019/3020	D273V	+
3021/3022	Q195K/R198Q	+
3023/3024	R198Q/N201R	+
3025/3026	E30R	+
3027/3028	K73Q/Q171R	+
3029/3030	Y86A	+
3031/3032	K263Q	+
3033/3034	L355K	+
3035/3036	I127L/L128I	+
3037/3038	A393E/V395R	+
3039/3040	H150N/D155L/Q171R	+
3041/3042	I305A	+
3043/3044	K168F/Q171P/I172R	+
3045/3046	R198Q/R203K	+
3047/3048	Q195S/R198Q	+
3049/3050	R256E	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2426, and defined as follows: "+" 1.03 to 1.16; "++" 1.16 to 1.39; and "+++" > 1.39.		

EXAMPLE 18

**Improvements Over SEQ ID NO: 2540 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 2540 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation

mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 18.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 18.1**, and analyzed on CE as described in Example 4.

Table 18.1 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 5 μ M; Target substrate: annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, at 0.035 μ M; Final dilution before CE analysis: 12x

- Activity relative to SEQ ID NO: 2540 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, observed by the reaction with SEQ ID NO: 2540 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 18.2**.

Table 18.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Percent Product (Relative to SEQ ID NO: 2540) ¹
3051/3052	D143E	+++
3053/3054	S259H	+++
3055/3056	V114S	+++
3057/3058	V243F	+++
3059/3060	K154Q	+++
3061/3062	E343T	+++
3063/3064	T286G	+++
3065/3066	K154F	+++
3067/3068	K249F	+++
3069/3070	L295F	+++
3071/3072	S106A	+++
3073/3074	P289R	+++

Table 18.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Percent Product (Relative to SEQ ID NO: 2540) ¹
3075/3076	W322G	+++
3077/3078	P294V	+++
3079/3080	P345V	+++
3081/3082	P345A	+++
3083/3084	K101T	+++
3085/3086	D143V	++
3087/3088	K249V	++
3089/3090	V114I	++
3091/3092	R290G	++
3093/3094	K249L	++
3095/3096	G242T	++
3097/3098	K251R	++
3099/3100	S113G	++
3101/3102	K329G	++
3103/3104	D107S	++
3105/3106	S246M	++
3107/3108	V114A	++
3109/3110	L295T	++
3111/3112	S288L	++
3113/3114	W322T	++
3115/3116	N287G	++
3117/3118	P289S	++
3119/3120	K249W	++
3121/3122	S259Y	++
3123/3124	I236V/K251T	++
3125/3126	T333S	++
3127/3128	E344G	++
3129/3130	E285D	++
3131/3132	S259P	++
3133/3134	L108M	++
3135/3136	E343P	++
3137/3138	M292V	++
3139/3140	E344M	++
3141/3142	P289I	++
3143/3144	P345K	++

Table 18.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Percent Product (Relative to SEQ ID NO: 2540) ¹
3145/3146	V104T	++
3147/3148	K251S	++
3149/3150	P345E	++
3151/3152	K257L	++
3153/3154	K249R	++
3155/3156	K257R	++
3157/3158	S67V	+
3159/3160	V114P	+
3161/3162	K247F	+
3163/3164	A330L	+
3165/3166	T152R	+
3167/3168	N287A	+
3169/3170	W322Q	+
3171/3172	T333M	+
3173/3174	K115L	+
3175/3176	R339G	+
3177/3178	R290S	+
3179/3180	T235R	+
3181/3182	V352I	+
3183/3184	D143R	+
3185/3186	R339A	+
3187/3188	S288W	+
3189/3190	K101S	+
3191/3192	Q103V	+
3193/3194	K329H	+
3195/3196	S113R	+
3197/3198	R339S	+
3199/3200	V70A	+
3201/3202	D143G	+
3203/3204	K115R	+
3205/3206	E293G	+
3207/3208	I238L	+
3209/3210	A330V	+
3211/3212	K296A	+
3213/3214	P294E	+

Table 18.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Percent Product (Relative to SEQ ID NO: 2540) ¹
3215/3216	T333A	+
3217/3218	V243L	+
3219/3220	K117A	+
3221/3222	K296R	+
3223/3224	R256K/V352A	+
3225/3226	D237S	+
3227/3228	D107V	+
3229/3230	P289K	+
3231/3232	V66T	+
3233/3234	L295V	+
3235/3236	T333R	+
3237/3238	H102R	+
3239/3240	V352L	+
3241/3242	R250N	+
3243/3244	E326L	+
3245/3246	L340S	+
3247/3248	M292P	+
3249/3250	M292L	+
3251/3252	E326M	+
3253/3254	K257V	+
3255/3256	R323S	+
3257/3258	N287V	+
3259/3260	A118K	+
3261/3262	H102S	+
3263/3264	K251G	+
3265/3266	K296Q	+
3267/3268	D107P	+
3269/3270	T286R	+
3271/3272	K154E	+
3273/3274	R323T	+
3275/3276	N287L	+
3277/3278	N287R	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2540, and defined as follows: "+" 1.03 to 1.48; "++" 1.48 to 1.80; and "+++" > 1.80.		

EXAMPLE 19

Improvements Over SEQ ID NO: 2540 in the Transposition of Double Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences

HTP Screening for Improved Transposase Variants

- 5 SEQ ID NO: 2540 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.
- 10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 11 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 19.1**.
- 15 Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 19.1**, and analyzed on CE as described in Example 4.

Table 19.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 11 μ M; Target substrates: SEQ ID NO: 2369/SEQ ID NO: 2370, SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384 and SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.035 μ M; Final dilution before CE analysis: 18x

- 20 Activity relative to SEQ ID NO: 2540 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2369/SEQ ID NO: 2370, SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384, and SEQ ID NO: 2385/SEQ ID NO: 2386, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2369/SEQ ID NO: 2370, SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384, and SEQ ID NO: 2385/SEQ ID NO: 2386, observed by the reaction with SEQ ID NO: 2540 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 19.2**.

30

Table 19.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Percent Product (Relative to SEQ ID NO: 2540) ¹
3279/3280	A18L/S113A/K270L	+++
3281/3282	A18L/S113A/K329A/A384V	+++
3283/3284	A18D/S113A/K270L	+++
3285/3286	A18D/K270L	+++
3287/3288	A18L	+++
3289/3290	A18D/S113A/K270L/K329A/T370S	+++
3291/3292	A18D/K270L/T370S	+++
3293/3294	T333L	+++
3295/3296	A18D/T370S	+++
3297/3298	K270L/T370S	+++
3299/3300	A18D	+++
3301/3302	Q73R/T333L	+++
3303/3304	Q73R/G379V	+++
3305/3306	S113A/K270L/T370S/A384V	+++
3307/3308	Q73H/T333L	+++
3309/3310	Q73R/T333L/L375S	+++
3311/3312	Q73R/L375S	++
3313/3314	A18D/K270L/T370S/A384V	++
3315/3316	Q73H/D273V/T333L/L375S	++
3317/3318	Q73R/C184A/T333L	++
3319/3320	A18L/T370S	++
3321/3322	A18L/S113A/K270L/T370S/A384V	++
3323/3324	Q73R	++
3325/3326	Q73R/C184A/D273V/T333L	++
3327/3328	A18L/T370S/A384V	++
3329/3330	Q73R/K115Q/C184A/T333L	++
3331/3332	Q73H	++
3333/3334	L375S	++
3335/3336	R290V/K378S	+
3337/3338	W144Y/T235S/R290Q	+
3339/3340	I127L/K378S/A383P	+
3341/3342	A18D/S259R/T370S/A384V	+
3343/3344	Q73R/D273V/T333L/L375S	+
3345/3346	K115Q/D273V/L375S	+
3347/3348	K378S	+

Table 19.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Percent Product (Relative to SEQ ID NO: 2540) ¹
3349/3350	Q73R/D273V	+
3351/3352	W144Y/A266V/R290Q	+
3353/3354	A266V/K378S	+
3355/3356	T370S	+
3357/3358	K115Q/L375S	+
3359/3360	L82Q/I127L/W144Y/A266V/R290V	+
3361/3362	L82Q	+
3363/3364	L82Q/I127L/W144Y/K378S	+
3365/3366	E77V/A233L/G334L/R376L	+
3367/3368	L82Q/A266V	+
3369/3370	T235S	+
3371/3372	L82Q/W144Y/A266V/K378S	+
3373/3374	L82Q/W144Y/A266V/R290Q/K378S	+
3375/3376	R290Q	+
3377/3378	L82Q/K378S	+
3379/3380	V262L	+
3381/3382	Q73R/K115Q	+
3383/3384	A233L/R376L	+
3385/3386	L82Q/I127L	+
3387/3388	A233L/G334L	+
3389/3390	L82Q/T235S/A266V/K378S/A383P	+
3391/3392	L82Q/I127L/W144Y/T235S/R290V	+
3393/3394	L82Q/A266V/K378S	+
3395/3396	L82Q/T235S/K378S	+
3397/3398	I127L/A266V/R290Q/K378S	+
3399/3400	K115Q/R256G/T333L/L375S	+
3401/3402	R256G/T333L	+
3403/3404	D273V	+
3405/3406	Q73H/C184A/R256G	+
3407/3408	L82Q/T235S/R290V	+
3409/3410	Q73R/K115Q/R256G	+
3411/3412	Q73R/D273V/L375S	+
3413/3414	L82Q/I127L/T235S	+
3415/3416	Q73R/R256G	+
3417/3418	E77V/V262L/R376L	+

Table 19.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Percent Product (Relative to SEQ ID NO: 2540) ¹
3419/3420	L82Q/I127L/W144Y/A266V	+
3421/3422	I127L/W144Y/R290V	+
3423/3424	K115Q/R256G/L375S	+
3425/3426	A18L/S113E/K270L/T370S/G412S	+
3427/3428	I127L/D149H/H150Q	+
3429/3430	L82Q/I127L/R290Q	+
3431/3432	E77V/Q276E	+
3433/3434	L82Q/T235S/A266V	+
3435/3436	E77V/A233L/I238V/R376L	+
3437/3438	Q73H/K115Q/C184A/R341G	+
3439/3440	R290V	+
3441/3442	T235S/R290Q	+
3443/3444	E77V/G334L/R376L	+
3445/3446	Q73R/R256G/L375S	+
3447/3448	Q276E/V352L	+
3449/3450	I358W	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2540, and defined as follows: "+" 1.08 to 1.62; "++" 1.62 to 2.12; and "+++" > 2.12.		

5 Additionally, target bias relative to SEQ ID NO: 2540 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 2540 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 19.3**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 19.1**; results (unreported) were similar to those reported below.

Table 19.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2540) ¹
3279/3280	A18L/S113A/K270L	+++
3283/3284	A18D/S113A/K270L	+
3285/3286	A18D/K270L	+
3287/3288	A18L	+++
3291/3292	A18D/K270L/T370S	+++

Table 19.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2540) ¹
3293/3294	T333L	++
3295/3296	A18D/T370S	++
3297/3298	K270L/T370S	+++
3299/3300	A18D	+++
3301/3302	Q73R/T333L	++
3303/3304	Q73R/G379V	++
3305/3306	S113A/K270L/T370S/A384V	++
3307/3308	Q73H/T333L	+
3309/3310	Q73R/T333L/L375S	++
3311/3312	Q73R/L375S	++
3313/3314	A18D/K270L/T370S/A384V	+
3315/3316	Q73H/D273V/T333L/L375S	+
3317/3318	Q73R/C184A/T333L	+
3319/3320	A18L/T370S	+++
3321/3322	A18L/S113A/K270L/T370S/A384V	+++
3323/3324	Q73R	+++
3325/3326	Q73R/C184A/D273V/T333L	+
3327/3328	A18L/T370S/A384V	++
3331/3332	Q73H	+++
3333/3334	L375S	+++
3335/3336	R290V/K378S	+++
3337/3338	W144Y/T235S/R290Q	+
3339/3340	I127L/K378S/A383P	+
3341/3342	A18D/S259R/T370S/A384V	+
3343/3344	Q73R/D273V/T333L/L375S	++
3347/3348	K378S	++
3349/3350	Q73R/D273V	+++
3351/3352	W144Y/A266V/R290Q	+
3353/3354	A266V/K378S	+++
3355/3356	T370S	+
3359/3360	L82Q/I127L/W144Y/A266V/R290V	+
3361/3362	L82Q	+
3363/3364	L82Q/I127L/W144Y/K378S	+
3365/3366	E77V/A233L/G334L/R376L	+++
3367/3368	L82Q/A266V	++

Table 19.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2540) ¹
3369/3370	T235S	+
3371/3372	L82Q/W144Y/A266V/K378S	+
3373/3374	L82Q/W144Y/A266V/R290Q/K378S	++
3375/3376	R290Q	++
3377/3378	L82Q/K378S	+
3379/3380	V262L	+
3383/3384	A233L/R376L	+
3385/3386	L82Q/I127L	+
3387/3388	A233L/G334L	+++
3389/3390	L82Q/T235S/A266V/K378S/A383P	++
3391/3392	L82Q/I127L/W144Y/T235S/R290V	+
3393/3394	L82Q/A266V/K378S	++
3395/3396	L82Q/T235S/K378S	+++
3397/3398	I127L/A266V/R290Q/K378S	+
3401/3402	R256G/T333L	+
3403/3404	D273V	+
3405/3406	Q73H/C184A/R256G	++
3407/3408	L82Q/T235S/R290V	++
3411/3412	Q73R/D273V/L375S	+
3413/3414	L82Q/I127L/T235S	++
3415/3416	Q73R/R256G	++
3417/3418	E77V/V262L/R376L	+
3419/3420	L82Q/I127L/W144Y/A266V	+
3429/3430	L82Q/I127L/R290Q	++
3431/3432	E77V/Q276E	++
3433/3434	L82Q/T235S/A266V	+
3435/3436	E77V/A233L/I238V/R376L	+++
3437/3438	Q73H/K115Q/C184A/R341G	+
3443/3444	E77V/G334L/R376L	++
3445/3446	Q73R/R256G/L375S	+
3447/3448	Q276E/V352L	++
3449/3450	I358W	++
3451/3452	A18L/S113E/N252A/T370S	+++
3453/3454	S113E/N252A	+++
3455/3456	A18L/S113A	+++

Table 19.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2540) ¹
3457/3458	A18L/S113E	+++
3459/3460	A266V/R290Q	+++
3461/3462	E77V/A118S/D143H/G334L/V352L/R376L	+++
3463/3464	L82Q/H150Q/T235S/A266V	++
3465/3466	E77V/D143H/R376L	++
3467/3468	Q73R/R341G	++
3469/3470	E77V	++
3471/3472	L82Q/R290V	++
3473/3474	D143H/Q276E	++
3475/3476	A118S/G334L	+
3477/3478	E77V/A118S/A233L	+
3479/3480	E77V/A118S/D143H/R376L	+
3481/3482	E77V/A118S/D143H/V262L/Q276E/G334L/R376L	+
3483/3484	E77V/A233L/V262L	+
3485/3486	I127L/T235S/A266V	+
3487/3488	N252A/K329A/G336Q/T370S	+
3489/3490	Q73H/K115Q/R341G	+
3491/3492	L82Q/I127L/R290V/K378S	+
3493/3494	E77V/A118S	+
3495/3496	L82Q/I127L/T235S/A266V/I358W	+
3497/3498	Q73R/C184A/R256G/L375S	+
3499/3500	A118S/A233L/Q276E/G334L	+
3501/3502	E77V/A233L/Q276E/R376L	+
3503/3504	E77V/D143H/V262L/Q276E	+
3505/3506	G334L/R376L	+
3507/3508	N252A/A384V	+
3509/3510	Q73R/C184A/R341G/L375S	+
3511/3512	Q73R/C184A/R256G/R341G	+
3513/3514	R256G/L375S	+
3515/3516	R341G	+
3517/3518	V262L/G334L	+
3519/3520	G334L	+
3521/3522	L82Q/I127L/R290Q/I358W	+
3523/3524	Q73R/R256G/D273V/L375S	+

Table 19.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 2540)	FIOP Target Product Ratio (Relative to SEQ ID NO: 2540) ¹
3525/3526	A18D/S113A/N252A/G336Q/A384V	+
3527/3528	E77V/V352L	+
3529/3530	T235S/I358W	+
3531/3532	S113E	+
3533/3534	D143H	+
3535/3536	E77V/A118S/V262L/G334L/R376L	+
3537/3538	T235S/A266V/R290Q/I358W/A383P	+
3539/3540	D143H/Q276E/G334L	+
3541/3542	Q73H/R256G/T333L/L375S	+
3543/3544	Q73R/T333L/R341G/L375S	+
3545/3546	V262L/Q276E	+
3547/3548	A18D/S113A/N252A	+
3549/3550	E77V/A118S/A233L/V352L/R376L	+
3551/3552	G336Q/T370S	+
3553/3554	L82Q/T235S/A266V/R290Q	+
3555/3556	Q73H/D273V/T333L/R341G	+
3557/3558	Q73H/R256G/D273V/L375S	+
3559/3560	Q73R/R256G/D273V	+
3561/3562	E77V/A118S/G334L	+
3563/3564	E77V/D143H/V262L/G334L	+
3565/3566	Q73R/G131D/R256G/D273V/T333L/R341G/L375S	+
3567/3568	Q73R/K115Q/R341G	+
3569/3570	S113E/N252A/K270L/T370S	+
3571/3572	W144Y/R290V/I358W	+
3573/3574	S113E/G336Q	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 2540, and defined as follows: "+" 1.05 to 1.23; "++" 1.23 to 1.38; and "+++" > 1.38.		

EXAMPLE 20**Improvements Over SEQ ID NO: 3402 in the Transposition of Double-Stranded DNA (dsDNA)**

5

Adapters into dsDNA Target Sequences

HTP Screening for Improved Transposase Variants

SEQ ID NO: 3402 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation

mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 20.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 20.1**, and analyzed on CE as described in Example 4.

Table 20.1 Reaction Conditions

Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 5 μ M; Target substrate: annealed oligos SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384 and SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.035 μ M; Final dilution before CE analysis: 18x
--

- Activity relative to SEQ ID NO: 3402 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384 and SEQ ID NO: 2385/SEQ ID NO: 2386 produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2371/SEQ ID NO: 2372, SEQ ID NO: 2373/SEQ ID NO: 2374, SEQ ID NO: 2375/SEQ ID NO: 2376, SEQ ID NO: 2377/SEQ ID NO: 2378, SEQ ID NO: 2379/SEQ ID NO: 2380, SEQ ID NO: 2381/SEQ ID NO: 2382, SEQ ID NO: 2383/SEQ ID NO: 2384 and SEQ ID NO: 2385/SEQ ID NO: 2386, observed by the reaction with SEQ ID NO: 3402 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 20.2**.

Table 20.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3402)	FIOP Percent Product (Relative to SEQ ID NO: 3402) ¹
3575/3576	A18L/G256R/A266V/R290K/K378S	+++
3577/3578	A18L/D143H/G256R	+++
3579/3580	C169T/G256R/A266V	+++
3581/3582	C169T	+++

Table 20.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3402)	FIOP Percent Product (Relative to SEQ ID NO: 3402) ¹
3583/3584	A18L/H150N/C169T	+++
3585/3586	A118S/A233L/I238V/G256R/K378S	+++
3587/3588	A18L	++
3589/3590	I238V/G256R	++
3591/3592	A18L/C169T/C392P	++
3593/3594	D143H/C169T/A233L/K270P	++
3595/3596	E77V/A233L	++
3597/3598	A18L/E77V/H150N/G256R/R376L	++
3599/3600	K270P	++
3601/3602	A18L/A118S/H150N/C169T/C184A/I238V/G256R/K270P/K378S	++
3603/3604	A18L/C184A/K270P	++
3605/3606	D143H/R290K	+
3607/3608	D143H/G256R/G334L	+
3609/3610	A118S	+
3611/3612	A18L/C169T/A233L/L333T	+
3613/3614	A18L/D143H	+
3615/3616	A18L/A118S/C169T	+
3617/3618	E77V/I238V	+
3619/3620	R376L	+
3621/3622	A18L/A118S/D143H/C169T/I238V/G256R/A266V/L333T	+
3623/3624	E77V/C169T/K270P/R290K	+
3625/3626	A18L/A233L/G256R/G334L	+
3627/3628	A18L/A118S	+
3629/3630	A233L/C392P	+
3631/3632	A18L/A118S/A233L/I238V	+
3633/3634	A233L/I238V/R376L	+
3635/3636	A18L/H150N/A266V	+
3637/3638	E77V/A233L/G256R	+
3639/3640	A118S/A233L/I238V/K270P	+
3641/3642	A233L/I238V	+
3643/3644	G256R/R290K/E303G	+
3645/3646	S113E/A118S/C169T/C184A/G256R/R376L	+
3647/3648	S113E/C169T	+

Table 20.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3402)	FIOP Percent Product (Relative to SEQ ID NO: 3402) ¹
3649/3650	A118S/D143H/I238V	+
3651/3652	D143H/A233L/A266V/R290K/R376L	+
3653/3654	D143H/I238V/K270P/R290K/K378S	+
3655/3656	A18L/C169T/C184A/R290K/L333T/R376L	+
3657/3658	A18L/E77V/A118S/C169T/A233L/I238V/G256 R/A266V/K270P	+
3659/3660	E77V/A118S/R290K	+
3661/3662	E77V/S113E/A118S/G256R/A266V/R290K	+
3663/3664	V183L/K270P/L333T	+
3665/3666	K378S	+
3667/3668	R290K	+
3669/3670	S113E/A233L	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 3402, and defined as follows: "+" 1.05 to 2.05; "++" 2.05 to 3.17; and "+++" > 3.17.		

- 5 Additionally, target bias relative to SEQ ID NO: 3402 (Promiscuity FIOP) was calculated as the ratio of the product peak area produced by the variant (product of the annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386 vs. product of the annealed oligos SEQ ID NO: 2379/SEQ ID NO: 2380) divided by the ratio of the product peak area observed by the reaction with SEQ ID NO: 3402 (where the control ratio may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 20.3**. Additional target bias data were collected and calculated for each target substrate/ annealed oligo pair in **Table 20.1**; results (unreported) were similar to those reported below.

Table 20.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3402)	FIOP Target Product Ratio (Relative to SEQ ID NO: 3402) ¹
3575/3576	A18L/G256R/A266V/R290K/K378S	+++
3577/3578	A18L/D143H/G256R	+++
3579/3580	C169T/G256R/A266V	++
3581/3582	C169T	++
3583/3584	A18L/H150N/C169T	++
3585/3586	A118S/A233L/I238V/G256R/K378S	+++
3587/3588	A18L	+
3589/3590	I238V/G256R	+++
3591/3592	A18L/C169T/C392P	+

Table 20.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3402)	FIOP Target Product Ratio (Relative to SEQ ID NO: 3402) ¹
3593/3594	D143H/C169T/A233L/K270P	++
3595/3596	E77V/A233L	+
3597/3598	A18L/E77V/H150N/G256R/R376L	+
3605/3606	D143H/R290K	++
3607/3608	D143H/G256R/G334L	+++
3609/3610	A118S	+
3611/3612	A18L/C169T/A233L/L333T	+
3613/3614	A18L/D143H	+
3615/3616	A18L/A118S/C169T	++
3617/3618	E77V/I238V	+++
3621/3622	A18L/A118S/D143H/C169T/I238V/G256R/A266V/L333T	+
3623/3624	E77V/C169T/K270P/R290K	+
3625/3626	A18L/A233L/G256R/G334L	+++
3627/3628	A18L/A118S	+
3629/3630	A233L/C392P	+
3631/3632	A18L/A118S/A233L/I238V	+
3633/3634	A233L/I238V/R376L	+
3635/3636	A18L/H150N/A266V	+
3641/3642	A233L/I238V	++
3649/3650	A118S/D143H/I238V	++
3651/3652	D143H/A233L/A266V/R290K/R376L	++
3653/3654	D143H/I238V/K270P/R290K/K378S	+
3655/3656	A18L/C169T/C184A/R290K/L333T/R376L	+
3659/3660	E77V/A118S/R290K	+
3663/3664	V183L/K270P/L333T	+
3665/3666	K378S	+
3671/3672	A18L/D143H/C169T/A233L/I238V/I358W	+++
3673/3674	A18L/E77V/A118S/C169T/A266V/G334L/R376L	+++
3675/3676	G256R/A266V/G334L	+++
3677/3678	A18L/C169T/I238V/K270P/G334L/R376L	+++
3679/3680	A118S/G256R/G334L/R376L	++
3681/3682	A18L/A118S/D143H/C169T/R376L	++
3683/3684	A18L/A233L/A266V/R290K/R376L	++

Table 20.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3402)	FIOP Target Product Ratio (Relative to SEQ ID NO: 3402) ¹
3685/3686	E77V/S113E/W121C/D143H/A233L/G334L	++
3687/3688	A18L/K378S	+
3689/3690	A266V/L333T	+
3691/3692	A118S/G256R/R290K/L333T/G334L/R376L/K378S	+
3693/3694	A18L/D143H/R290K	+
3695/3696	A18L/E77V/G256R/R290K/G334L	+
3697/3698	A266V/K270P/R376L	+
3699/3700	H150N/A233L/L333T	+
3701/3702	S113E/I238V/R376L	+
3703/3704	A18L/A118S/A266V	+
3705/3706	D143H/Q364R	+
3707/3708	E77V/A118S/C169T/A233L/A266V/K270P	+
3709/3710	A118S/C169T/C184A/A233L/A266V/K270P/L333T	+
3711/3712	A18L/S113E/I238V	+
3713/3714	A18L/A266V/K270P/K378S	+
3715/3716	A18L/D143H/A233L/I238V/R290K/R376L	+
3717/3718	E77V/A118S/I238V/R376L	+
3719/3720	A118S/A266V/R290K	+
3721/3722	A118S/D143H/C392P	+
3723/3724	I238V/A266V/K270P/K378S	+
3725/3726	A18L/D143H/L333T/G334L/K378S	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 3402, and defined as follows: "+" 1.06 to 1.28; "++" 1.28 to 1.46; and "+++" > 1.46.		

EXAMPLE 21**Improvements Over SEQ ID NO: 3662 in the Transposition of Double-Stranded DNA (dsDNA)****Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 3662 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at

room-temperature for 60 mins. After 60 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 21.1**.

- 5 Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 21.1**, and analyzed on CE as described in Example 4.

Table 21.1 Reaction Conditions	
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 5 μ M; Target substrate: annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698 or SEQ ID NO: 5699/SEQ ID NO: 5700, at 0.035 μ M; Final dilution before CE analysis: 12x	

- 10 Activity relative to SEQ ID NO: 3662 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, observed by the reaction with SEQ ID NO: 3662 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 21.2**.

Table 21.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3662)	FIOP Percent Product (Relative to SEQ ID NO: 3662)¹
3727/3728	D143E/A255R/E344K	+++
3729/3730	C169T/E344K	+++
3731/3732	D143E/A255R	+++
3733/3734	D143E/K257R	++
3735/3736	D143E/C169T/I238V/K257R/E344K	++
3737/3738	C184A	++
3739/3740	K257R/E344K	++
3741/3742	D143E	++
3743/3744	V77E/K270P	++
3745/3746	D143E/C169T/A255R/K257R	++
3747/3748	V77E/E156L/K270P	++
3749/3750	D143E/C169T/A255R	++
3751/3752	K270P	++
3753/3754	C184A/V244K/V324M	++
3755/3756	V77E/K270R	+

Table 21.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3662)	FIOP Percent Product (Relative to SEQ ID NO: 3662) ¹
3757/3758	I238V/A255R/E344M	+
3759/3760	D143H/C169T/E343R	+
3761/3762	C169T/E344M	+
3763/3764	V244K/K270P	+
3765/3766	C169T/I238V	+
3767/3768	V77E/V244K/K270P	+
3769/3770	D143E/A255R/K257R	+
3771/3772	E156L	+
3773/3774	V77E/E156L	+
3775/3776	I238V	+
3777/3778	C169T/K257R/E344M	+
3779/3780	V77E/C184A/V244K	+
3781/3782	C169T/I238V/A255R/K257R/E344M	+
3783/3784	V77E/C184A/L333T	+
3785/3786	E156L/K270R	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 3662, and defined as follows: "+" 1.05 to 1.42; "++" 1.42 to 1.77; and "+++" > 1.77.		

- Activity relative to SEQ ID NO: 3662 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5699/SEQ ID NO: 5700, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5699/SEQ ID NO: 5700, observed by the reaction with SEQ ID NO: 3662 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 21.3**.

Table 21.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3662)	FIOP Percent Product (Relative to SEQ ID NO: 3662) ¹
3787/3788	V77E/C184A/K270P	+++
3789/3790	C169T/E343R	+++
3791/3792	V77E	+++
3793/3794	D143E/E344M	++
3795/3796	C169T/A255R	++
3797/3798	A255R/K257R	++
3799/3800	C169T	++
3801/3802	C169T/I238V/A255R/E344M	+

Table 21.3		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3662)	FIOP Percent Product (Relative to SEQ ID NO: 3662) ¹
3803/3804	D143E/I238V/E344M	+
3805/3806	D143E/I238V	+
3807/3808	D143E/I238V/A255R/E343R	+
3809/3810	K257R	+
3811/3812	I238V/E344M	+
3813/3814	A255R	+
3815/3816	D143H	+
3817/3818	C169T/I238V/E344M	+
3819/3820	D143E/K257R/C392P	+
3821/3822	D143E/I238V/A255R/C392P	+
3823/3824	E344M	+
3825/3826	D143H/I238V/A255R	+
3827/3828	D143H/A255R	+
3829/3830	C169T/C392P	+
3831/3832	A255R/K257R/C392P	+
3833/3834	V77E/E156L/C184A/V244K/V324M	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 3662, and defined as follows: "+" 1.01 to 1.41; "++" 1.41 to 1.80; and "+++" > 1.80.		

EXAMPLE 22

Improvements Over SEQ ID NO: 3788 in the Transposition of Double-Stranded DNA (dsDNA)

Adapters into dsDNA Target Sequences

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 3788 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
 15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 22.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 22.1**, and analyzed on CE as described in Example 4.

Table 22.1 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 48°C for 30 minutes; Adapter concentration: 5 µM; Target substrate: annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, at 0.035 µM; Final dilution before CE analysis: 12x

- Activity relative to SEQ ID NO: 3788 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698 produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698 observed by the reaction with SEQ ID NO: 3788 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 22.2**.

Table 22.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3788)	FIOP Percent Product (Relative to SEQ ID NO: 3788) ¹
3835/3836	K115T	+++
3837/3838	E343M	+++
3839/3840	W445M	+++
3841/3842	E284A	+++
3843/3844	N347R	+++
3845/3846	W322T	+++
3847/3848	A443P	+++
3849/3850	P64A	+++
3851/3852	K115R	+++
3853/3854	E343R	+++
3855/3856	L260C	+++
3857/3858	K26R	+++
3859/3860	D116S	+++
3861/3862	S99G	+++
3863/3864	V243W	+++
3865/3866	Q103G	+++
3867/3868	E284G	+++
3869/3870	P239L	+++
3871/3872	R323V	+++
3873/3874	S97A	+++
3875/3876	E55V	+++
3877/3878	E344T	+++

Table 22.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3788)	FIOP Percent Product (Relative to SEQ ID NO: 3788) ¹
3879/3880	W322S	+++
3881/3882	W445L	+++
3883/3884	K51R	+++
3885/3886	T95G	+++
3887/3888	A255R	+++
3889/3890	V243L	+++
3891/3892	E55T	+++
3893/3894	R439G	+++
3895/3896	A447L	+++
3897/3898	E344Q	++
3899/3900	S97C	++
3901/3902	V45R	++
3903/3904	K154E/R439S	++
3905/3906	R186S	++
3907/3908	I442H	++
3909/3910	V45W	++
3911/3912	Q103R	++
3913/3914	Q54V	++
3915/3916	G334W	++
3917/3918	R439A	++
3919/3920	P22F	++
3921/3922	T44P	++
3923/3924	A443S	++
3925/3926	L333Y	++
3927/3928	K251I	++
3929/3930	A52P	++
3931/3932	E344S	++
3933/3934	V324M	++
3935/3936	E344G	++
3937/3938	R186Q	++
3939/3940	L260I	++
3941/3942	W445T	++
3943/3944	T44G	++
3945/3946	W445V	++
3947/3948	E343V	++

Table 22.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3788)	FIOP Percent Product (Relative to SEQ ID NO: 3788) ¹
3949/3950	T319K	++
3951/3952	C169T	++
3953/3954	L340A	++
3955/3956	E284R	++
3957/3958	E326Q	++
3959/3960	V244C	++
3961/3962	E344H	++
3963/3964	A255M	++
3965/3966	P64G	++
3967/3968	L295I	++
3969/3970	Q103T	++
3971/3972	V45Q	++
3973/3974	M53A	++
3975/3976	P22W	++
3977/3978	V45E	++
3979/3980	D143E	++
3981/3982	T44V	++
3983/3984	I236T	++
3985/3986	V104G	++
3987/3988	A255K	++
3989/3990	C169T/V324M/E343R	++
3991/3992	A447N	++
3993/3994	V104T	++
3995/3996	D143E/C169T	++
3997/3998	S42G	++
3999/4000	E446R	++
4001/4002	L333Q	++
4003/4004	Q103S	++
4005/4006	K251A	++
4007/4008	K329T	++
4009/4010	V324M/E343R	++
4011/4012	V243E	+
4013/4014	K251L	+
4015/4016	L295F	+
4017/4018	V45A	+

Table 22.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3788)	FIOP Percent Product (Relative to SEQ ID NO: 3788) ¹
4019/4020	R323T	+
4021/4022	E326A	+
4023/4024	P254T	+
4025/4026	V45G	+
4027/4028	V45M	+
4029/4030	D116E/E343R	+
4031/4032	A447T	+
4033/4034	I43L	+
4035/4036	N65M	+
4037/4038	V244T	+
4039/4040	L111V	+
4041/4042	W445G	+
4043/4044	P64D	+
4045/4046	S50R	+
4047/4048	A52K	+
4049/4050	E344V	+
4051/4052	P64E	+
4053/4054	V243Q	+
4055/4056	P64V	+
4057/4058	A255C	+
4059/4060	E344M	+
4061/4062	L295V	+
4063/4064	R186A	+
4065/4066	A283G	+
4067/4068	F60L	+
4069/4070	P22V	+
4071/4072	K115P	+
4073/4074	K115L	+
4075/4076	Q54R	+
4077/4078	F434Y	+
4079/4080	T95R	+
4081/4082	A283T	+
4083/4084	E55M	+
4085/4086	G334E	+
4087/4088	L333M	+

Table 22.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3788)	FIOP Percent Product (Relative to SEQ ID NO: 3788) ¹
4089/4090	G112R	+
4091/4092	I442G	+
4093/4094	H328Y	+
4095/4096	K251S	+
4097/4098	P64R	+
4099/4100	Q103F	+
4101/4102	C169T/E343R	+
4103/4104	A52R	+
4105/4106	V45L	+
4107/4108	S42P	+
4109/4110	L333R	+
4111/4112	P239D	+
4113/4114	E446G	+
4115/4116	K115G	+
4117/4118	M53S	+
4119/4120	A255L	+
4121/4122	R250A	+
4123/4124	A443C	+
4125/4126	Q103V	+
4127/4128	Q54S	+
4129/4130	Y234W	+
4131/4132	L260A	+
4133/4134	Q54T	+
4135/4136	R439S	+
4137/4138	M53H	+
4139/4140	P22R	+
4141/4142	R323H	+
4143/4144	K115E	+
4145/4146	P22N	+
4147/4148	W322G	+
4149/4150	V243G	+
4151/4152	E344K	+
4153/4154	E343G	+
4155/4156	I238V	+
4157/4158	P64S	+

Table 22.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3788)	FIOP Percent Product (Relative to SEQ ID NO: 3788) ¹
4159/4160	M53R	+
4161/4162	A443I	+
4163/4164	L260T	+
4165/4166	P239K/V244F	+
4167/4168	E326L	+
4169/4170	I238L/S246G	+
4171/4172	R250S	+
4173/4174	I442Q	+
4175/4176	R366S	+
4177/4178	S99R	+
4179/4180	A283S	+
4181/4182	Q103D	+
4183/4184	P254R	+
4185/4186	T319R	+
4187/4188	R439Y	+
4189/4190	I43V	+
4191/4192	E55G	+
4193/4194	P22L	+
4195/4196	E55S	+
4197/4198	K241R	+
4199/4200	K329W	+
4201/4202	K115E/C169T/E343R	+
4203/4204	L333W	+
4205/4206	R186P	+
4207/4208	K115Q/C169T/E343R	+
4209/4210	I43P	+
4211/4212	C169T/V324M	+
4213/4214	T319E	+
4215/4216	K329G	+
4217/4218	K115Q/E343R	+
4219/4220	K115Q	+
4221/4222	K115Q/D143E/C169T	+
4223/4224	I236V	+
4225/4226	N65A	+
4227/4228	A52T	+

Table 22.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 3788)	FIOP Percent Product (Relative to SEQ ID NO: 3788) ¹
4229/4230	R341K	+
4231/4232	G20R	+
4233/4234	V298I/E326A	+
4235/4236	E446F	+
4237/4238	E338A	+
4239/4240	K51Q	+
4241/4242	T319W	+
4243/4244	V244G	+
4245/4246	R62G	+
4247/4248	K115W	+
4249/4250	S444Q	+
4251/4252	G242M	+
4253/4254	T95V	+
4255/4256	K296H	+
4257/4258	I236Y	+
4259/4260	P64T	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 3788, and defined as follows: "+" 1.01 to 2.00; "++" 2.00 to 2.88; and "+++" > 2.88.		

EXAMPLE 23

**Improvements Over SEQ ID NO: 4150 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 4150 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- 10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
- 15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 23.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 23.1**, and analyzed on CE as described in Example 4.

Table 23.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 30% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 2 µM; Target substrate: annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386, at 0.035 µM; Final dilution before CE analysis: 12x

- Activity relative to SEQ ID NO: 4150 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 2385/SEQ ID NO: 2386, observed by the reaction with SEQ ID NO: 4150 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 23.2**.
- 5

Table 23.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4150)	FIOP Percent Product (Relative to SEQ ID NO: 4150)¹
4261/4262	L295I/L333Y	+++
4263/4264	Q103G/C169T/E343M/E344G	+++
4265/4266	E343M	+++
4267/4268	D143E	+++
4269/4270	D143E/C169T/E343M/E344G	+++
4271/4272	A255K/E343M	+++
4273/4274	P64A/L260I/L295I/L333Y	+++
4275/4276	Q103G/E343M/E344G	+++
4277/4278	L260I/L295I	+++
4279/4280	E343M/E344G	+++
4281/4282	C169T/E343M	+++
4283/4284	L333Y	+++
4285/4286	E284G	+++
4287/4288	S259I	+++
4289/4290	D143E/A255K/E343M	+++
4291/4292	K115T/D143E/E343M	+++
4293/4294	C169T/E284G	+++
4295/4296	E284A	++
4297/4298	Q103G/D143E/A255K/E284G/E343M/E344G	++
4299/4300	C169T	++
4301/4302	C169T/E344G	++
4303/4304	C169T/E284A	++

Table 23.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4150)	FIOP Percent Product (Relative to SEQ ID NO: 4150) ¹
4305/4306	V244T/N347R	++
4307/4308	Q103G/K115E/E343M/E344G	++
4309/4310	A255K/E284G	++
4311/4312	E344G	++
4313/4314	A255K	++
4315/4316	E284A/E343M/E344G	++
4317/4318	D143E/A255K	++
4319/4320	E284G/E344G	++
4321/4322	D143E/A255K/E343M/E344G	++
4323/4324	D143E/A255K/E284A/E343M/E344G	++
4325/4326	K115T/C169T	++
4327/4328	L260I	++
4329/4330	K26R	++
4331/4332	V244T/L295I	++
4333/4334	K26R/L295I/T319K	++
4335/4336	L295I	+
4337/4338	A255K/E344G	+
4339/4340	K115T/D143E/A255K/E343M	+
4341/4342	K115T/E284G	+
4343/4344	K115T/C169T/E343M	+
4345/4346	I236T/N347R	+
4347/4348	K115T/E344G	+
4349/4350	C169T/A255K	+
4351/4352	D143E/C169T/A255K/E284G/E343M/E344G	+
4353/4354	D143E/A255K/E344G	+
4355/4356	K115T	+
4357/4358	K115T/E343M/E344G	+
4359/4360	K115E/E343M	+
4361/4362	V244T/L260I	+
4363/4364	I236T	+
4365/4366	V244T	+
4367/4368	K115E/C169T	+
4369/4370	P64A	+
4371/4372	V244T/K251A/T319K	+
4373/4374	K115T/A255K/E344G	+

Table 23.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4150)	FIOP Percent Product (Relative to SEQ ID NO: 4150) ¹
4375/4376	P64A/K251A/L260I/N347R	+
4377/4378	L260I/L333Y	+
4379/4380	D143E/E343M/E344G	+
4381/4382	K115E	+
4383/4384	K26R/L333Y	+
4385/4386	K251A	+
4387/4388	K115E/A255K	+
4389/4390	C169T/A255K/E284G	+
4391/4392	K251A/L260I/L295I/N347R	+
4393/4394	K115E/E284A/E344G	+
4395/4396	K26R/L260I/L295I	+
4397/4398	C169T/A255K/E343M	+
4399/4400	K115E/E284G/E343M/E344G	+
4401/4402	P64A/L260I/N347R	+
4403/4404	Q103G/A255K/E284A	+
4405/4406	N347R	+
4407/4408	P64A/V244T	+
4409/4410	K26R/I236T	+
4411/4412	K26R/I236T/L295I	+
4413/4414	I236T/V244T/L333Y	+
4415/4416	V244T/K251A	+
4417/4418	Q103G/K115E/E343M	+
4419/4420	K26R/L260I	+
4421/4422	K26R/L260I/L295I/L333Y	+
4423/4424	K115E/D143E/A255K	+
4425/4426	I236T/L260I	+
4427/4428	A255K/E343M/E344G	+
4429/4430	K115T/A255K	+
4431/4432	K115E/D143E/E344G	+
4433/4434	K115T/D143E/E284G	+
4435/4436	K115E/E344G	+
4437/4438	K26R/I236T/L260I/L333Y	+
4439/4440	Q103G/K115E/A255K	+
4441/4442	K251A/L333Y	+

Table 23.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4150)	FIOP Percent Product (Relative to SEQ ID NO: 4150) ¹
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4150, and defined as follows: "+" 1.03 to 1.67; "++" 1.67 to 2.21; and "+++" > 2.21.		

EXAMPLE 24

**Improvements Over SEQ ID NO: 4278 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 4278 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 24.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 24.1**, and analyzed on CE as described in Example 4.

Table 24.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 5 μ M; Target substrate: annealed oligos SEQ ID NO: 5695/SEQ ID NO: 5696, at 0.035 μ M; Final dilution before CE analysis: 12x

20 Activity relative to SEQ ID NO: 4278 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5695/SEQ ID NO: 5696, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5695/SEQ ID NO: 5696, observed by the reaction with SEQ ID NO: 4278 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 24.2**.

25

Table 24.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4278)	FIOP Percent Product (Relative to SEQ ID NO: 4278) ¹
4443/4444	K117A	+++
4445/4446	T152R/S246D	+++
4447/4448	C169T/S246D	+++
4449/4450	T152R/C169T	+++
4451/4452	K117A/T152R/C169T	+++
4453/4454	K117A/T152R	+++
4455/4456	S246D	+++
4457/4458	K115E/T152R/C169T	++
4459/4460	K115E/S118T/T152R/C169T/S246D	++
4461/4462	K117A/C169T	++
4463/4464	T152R/C169T/S246D	++
4465/4466	K115E/T152R/C169T/S246D	++
4467/4468	K115E/K117A	++
4469/4470	K115E/K117A/T152R/S246D	++
4471/4472	T152R	++
4473/4474	K117A/S118T	++
4475/4476	K115E/C169T/S246D	++
4477/4478	K117A/S246D	++
4479/4480	C169T	++
4481/4482	K115E/S118T	+
4483/4484	K115E	+
4485/4486	K115E/T152R/S246D	+
4487/4488	K115E/S118T/S246D	+
4489/4490	K117A/S118T/T152R	+
4491/4492	K115E/S118T/T152R/C169T	+
4493/4494	K117A/S118T/C169T	+
4495/4496	K115E/T152R/S246D/Y374C	+
4497/4498	K115E/K117A/T152R/C169T/S246D	+
4499/4500	K115E/K117A/T152R	+
4501/4502	K115E/K117A/S118T	+
4503/4504	K115E/K117A/S118T/S246D	+
4505/4506	K117A/S118T/T152R/S246D	+
4507/4508	K115E/K117A/C169T	+
4509/4510	K117A/S118T/S246D	+
4511/4512	K115E/T152R	+

Table 24.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4278)	FIOP Percent Product (Relative to SEQ ID NO: 4278) ¹
4513/4514	S118T/S246D	+
4515/4516	K115E/S246D	+
4517/4518	K117A/T152R/S246D/S288G	+
4519/4520	K115E/S118T/T152R	+
4521/4522	K115E/K117A/T152R/C169T	+
4523/4524	K115E/K117A/S118T/C169T	+
4525/4526	K117A/S118T/T152R/C169T	+
4527/4528	K115E/K117A/S246D	+
4529/4530	K115E/K117A/S118T/T152R	+
4531/4532	K115E/C169T	+
4533/4534	K115E/S118T/C169T	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4278, and defined as follows: "+" 1.10 to 3.60; "++" 3.60 to 5.61; and "+++" > 5.61.		

EXAMPLE 25**Improvements Over SEQ ID NO: 4278 in the Transposition of Double-Stranded DNA (dsDNA)****Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 4278 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- 10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
- 15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 25.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 25.1**, and analyzed on CE as described in Example 4.

Table 25.1-Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 2 µM; Target substrate: annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, at 0.035 µM; Final dilution before CE analysis: 12x

- Activity relative to SEQ ID NO: 4278 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, observed by the reaction with SEQ ID NO: 4278 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 25.2**.
- 5

Table 25.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4278)	FIOP Percent Product (Relative to SEQ ID NO: 4278)¹
4535/4536	I238V	+++
4537/4538	K168A	+++
4539/4540	A233T/E343G	+++
4541/4542	D143E	+++
4543/4544	E284S	+++
4545/4546	P151R	+++
4547/4548	D143V	+++
4549/4550	E156W	+++
4551/4552	A233R	+++
4553/4554	A192S	+++
4555/4556	I172L	+++
4557/4558	E284L	+++
4559/4560	H167T	+++
4561/4562	S261R	+++
4563/4564	I172R	+++
4565/4566	P239T	+++
4567/4568	E284R	+++
4569/4570	S261T	+++
4571/4572	Q195T	+++
4573/4574	K101S	+++
4575/4576	E223Q	+++
4577/4578	S118R	+++

Table 25.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4278)	FIOP Percent Product (Relative to SEQ ID NO: 4278) ¹
4579/4580	A233W	+++
4581/4582	K115R	+++
4583/4584	T235C	+++
4585/4586	I172V	+++
4587/4588	A163L	+++
4589/4590	I236T	++
4591/4592	V244S	++
4593/4594	K249I	++
4595/4596	A163R	++
4597/4598	G243Q	++
4599/4600	M292G	++
4601/4602	H167Q	++
4603/4604	S259L	++
4605/4606	Q191H	++
4607/4608	I190F	++
4609/4610	H150M	++
4611/4612	K257T	++
4613/4614	K296C	++
4615/4616	P254A	++
4617/4618	S261V	++
4619/4620	Q73L	++
4621/4622	Q191V	++
4623/4624	T286V	++
4625/4626	T286G	++
4627/4628	K161R	++
4629/4630	Q195L	++
4631/4632	V262L	++
4633/4634	A160R	++
4635/4636	Q195V	++
4637/4638	P239A	++
4639/4640	T286E	++
4641/4642	K115Q	++
4643/4644	T235E	++
4645/4646	H150T	++
4647/4648	E284I	++

Table 25.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4278)	FIOP Percent Product (Relative to SEQ ID NO: 4278) ¹
4649/4650	A258V	++
4651/4652	P239N/R250H	++
4653/4654	A233G	++
4655/4656	V244W	++
4657/4658	A160T	++
4659/4660	H167R	++
4661/4662	V262I/A283V	+
4663/4664	K168N	+
4665/4666	K213R	+
4667/4668	K296R	+
4669/4670	D143Q	+
4671/4672	A192R	+
4673/4674	T235D	+
4675/4676	K251H	+
4677/4678	S261K	+
4679/4680	K157R	+
4681/4682	K249L	+
4683/4684	K157L	+
4685/4686	S166T	+
4687/4688	L225V	+
4689/4690	T235S	+
4691/4692	K249S	+
4693/4694	P239E	+
4695/4696	A160V	+
4697/4698	D143L	+
4699/4700	K154S	+
4701/4702	D237Q	+
4703/4704	A255S	+
4705/4706	M292E	+
4707/4708	K157P	+
4709/4710	P277S/M292V	+
4711/4712	A233Q	+
4713/4714	A163K	+
4715/4716	Q171I	+
4717/4718	H150I	+

Table 25.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4278)	FIOP Percent Product (Relative to SEQ ID NO: 4278) ¹
4719/4720	A163Q	+
4721/4722	V244A	+
4723/4724	K249T	+
4725/4726	E156L	+
4727/4728	C169L	+
4729/4730	K251Q	+
4731/4732	S259N	+
4733/4734	W145Y	+
4735/4736	A192E	+
4737/4738	S261A	+
4739/4740	V244F	+
4741/4742	P289L	+
4743/4744	Q171M	+
4745/4746	K249R	+
4747/4748	V244T	+
4749/4750	K115S	+
4751/4752	T286I	+
4753/4754	Q191L	+
4755/4756	H150L	+
4757/4758	Y234L	+
4759/4760	C169V	+
4761/4762	A233F	+
4763/4764	L219R	+
4765/4766	T152L	+
4767/4768	I238L	+
4769/4770	I295V	+
4771/4772	P151H	+
4773/4774	K249G	+
4775/4776	E223D	+
4777/4778	K115A	+
4779/4780	A233L	+
4781/4782	Y234V	+
4783/4784	Q195M	+
4785/4786	G243V	+
4787/4788	P151D	+

Table 25.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4278)	FIOP Percent Product (Relative to SEQ ID NO: 4278) ¹
4789/4790	G243I	+
4791/4792	E223A	+
4793/4794	K154P	+
4795/4796	K154Y	+
4797/4798	E156D	+
4799/4800	V114S	+
4801/4802	I214L	+
4803/4804	A255G	+
4805/4806	P151A	+
4807/4808	I236M	+
4809/4810	K168V	+
4811/4812	P151F	+
4813/4814	Q73R	+
4815/4816	Q171L	+
4817/4818	A283G	+
4819/4820	V244L	+
4821/4822	A255W	+
4823/4824	I214V	+
4825/4826	P254E	+
4827/4828	A192G	+
4829/4830	E218P	+
4831/4832	T286L	+
4833/4834	K257L	+
4835/4836	D153G	+
4837/4838	E156Q	+
4839/4840	A283T	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4278, and defined as follows: "+" 1.01 to 1.45; "++" 1.45 to 1.81; and "+++" > 1.81.		

EXAMPLE 26

**Improvements Over SEQ ID NO: 4642 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5

HTP Screening for Improved Transposase Variants

SEQ ID NO: 4642 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation

mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 26.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 26.1**, and analyzed on CE as described in Example 4.

Table 26.1 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 2 μ M; Target substrate: annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, at 0.035 μ M; Final dilution before CE analysis: 12x

- Activity relative to SEQ ID NO: 4642 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, observed by the reaction with SEQ ID NO: 4642 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 26.2**.

Table 26.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4642)	FIOP Percent Product (Relative to SEQ ID NO: 4642) ¹
4841/4842	A163V/A255K/I315L	+++
4843/4844	D93S	+++
4845/4846	D93T	+++
4847/4848	D93A	+++
4849/4850	A163V/K168M/C169L	+++
4851/4852	P148R	+++
4853/4854	D143Q/K154L/A255K	+++
4855/4856	A163V/K168M/Q195E/I214L/A255K/I315L	+++
4857/4858	T134R	+++
4859/4860	A320Y	+++

Table 26.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4642)	FIOP Percent Product (Relative to SEQ ID NO: 4642) ¹
4861/4862	S118T/D143Q/A163V/K168N/Q195E/A255K/I315L	+++
4863/4864	Q115K/S118C	+++
4865/4866	L333G	+++
4867/4868	M317C	+++
4869/4870	S118T/A163V/K168N/I214L/A255K/I315L	+++
4871/4872	E113S/D143R	+++
4873/4874	A320I	+++
4875/4876	D93G	++
4877/4878	A163V/K168N/C169T/A255K/I315L	++
4879/4880	E113S	++
4881/4882	S118T/A255K	++
4883/4884	A163V/K168M	++
4885/4886	A163V	++
4887/4888	A255K/I315L	++
4889/4890	R366K	++
4891/4892	E113S/Q115E	++
4893/4894	S118T/D143Q/A255K	++
4895/4896	V352G	++
4897/4898	M317L	++
4899/4900	E293P	++
4901/4902	A320V	++
4903/4904	K213R	++
4905/4906	M317G	++
4907/4908	V206M	++
4909/4910	M317S	++
4911/4912	E293G	++
4913/4914	M317A	++
4915/4916	H328F	++
4917/4918	P277W	++
4919/4920	R366A	++
4921/4922	P277D	++
4923/4924	Q115R/M177V	+
4925/4926	A163V/A255K	+
4927/4928	V206L	+

Table 26.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4642)	FIOP Percent Product (Relative to SEQ ID NO: 4642) ¹
4929/4930	K329L	+
4931/4932	P294R	+
4933/4934	H210R	+
4935/4936	A320C	+
4937/4938	E293T	+
4939/4940	I315L	+
4941/4942	R186K	+
4943/4944	W322F	+
4945/4946	E113C/Q115K	+
4947/4948	D143Q/A163V/A255K	+
4949/4950	D93F	+
4951/4952	W445R	+
4953/4954	L194M	+
4955/4956	E156D/I214L/I315L	+
4957/4958	K168N/I315L	+
4959/4960	P277S	+
4961/4962	D143Q/K168N/C169L/A255K/I315L	+
4963/4964	E113S/V114G	+
4965/4966	M317Q	+
4967/4968	W445Y	+
4969/4970	D143Q/A255K	+
4971/4972	A163V/I214L	+
4973/4974	T134V	+
4975/4976	E113S/D143K	+
4977/4978	T126V	+
4979/4980	A208G	+
4981/4982	D143Q/E156D/K168N/D216G/A255K	+
4983/4984	A163V/K168M/C169V	+
4985/4986	P294V	+
4987/4988	A163V/K168N/C169V/I315L	+
4989/4990	L129M	+
4991/4992	E113G/Q115K	+
4993/4994	P294G	+
4995/4996	D143Q/K168N	+
4997/4998	H124T	+

Table 26.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4642)	FIOP Percent Product (Relative to SEQ ID NO: 4642) ¹
4999/5000	S118T/K168N/I214L/I315L	+
5001/5002	A330S	+
5003/5004	A163V/K168M/C169T	+
5005/5006	R207M	+
5007/5008	K197L	+
5009/5010	P148Q	+
5011/5012	C392A	+
5013/5014	V352I	+
5015/5016	D143Q/A163V	+
5017/5018	K329S	+
5019/5020	P148L	+
5021/5022	L333Q	+
5023/5024	S159T	+
5025/5026	E293R	+
5027/5028	D93M	+
5029/5030	P277G	+
5031/5032	A330T	+
5033/5034	E293S	+
5035/5036	D93H	+
5037/5038	Q115K/S118G	+
5039/5040	L333R	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4642, and defined as follows: "+" 1.01 to 1.27; "++" 1.27 to 1.51; and "+++" > 1.51.		

EXAMPLE 27**Improvements Over SEQ ID NO: 4642 in the Transposition of Double-Stranded DNA (dsDNA)****Adapters into dsDNA Target Sequences****5 HTP Screening for Improved Transposase Variants**

SEQ ID NO: 4642 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96-well format BioRad PCR plates. 100 nL of 100% HTP lysate variants in lysis buffer (**Table 27.1**) were combined in 1:1 volume ratio with 2.5 μ M seqWell i5 tagging reagents and incubated at room temperature for 1 hour with mild shaking. 100 nL of the same set of samples were also combined with 2.5 μ M seqWell i7 oligos in 1:1 volume ratio in a separate set of

- reactions and incubated at room temperature for one hour with mild shaking. The loaded i5 and i7 transposases were then combined in equal volumes and diluted with 12.5 μ L MTBS prior to library preparation. To prepare the library for sequencing 10 ng of *E.coli* genomic DNA was mixed in equal volumes (5 μ L) with i5/i7 transposase mix and 3x coding buffer. The reactions were incubated at 55°C for 15 minutes and terminated with 7.5 μ L of seqWell X-solution at 68°C for 10 minutes. The reactions were then purified with (1x) MAGwise beads and eluted in 25 μ L of 10 mM Tris pH 8.0. This was followed by 12 cycle amplification (**Table 27.1**), sample pooling and MAGwise bead purification. An equal volume of the samples (10 μ L each) was pooled together and purified with (1x) MAGwise beads. The resulting library pool was evaluated using NanoDrop Spectrophotometer or Qubit and the Illumina MiSeq assay was performed.

Table 27.1 Reaction Conditions	
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes;	
Library Amplification conditions	
Amplification Reaction Set-up	PCR program
10 μ L purified tag product	72C, 10 mins
2.5 μ L 25 μ M Library Primer Mix (LPM)	95C, 3 mins
12.5 μ L 2x Library Amp Mix	98C, 0:30 mins 64C, 0:15 mins 12X 72C, 0:30 mins
	72C, 3 mins
	8C, Hold

- Relative target bias for each variant was calculated as the mean RMSD₁₅ of SEQ ID NO: 4642 divided by the mean RMSD₁₅ of the transposase variant. Variants with target bias mean RMSD₁₅ less than SEQ ID NO: 4642 will have a ratio greater than 1. The results are shown in **Table 27.2**.

Table 27.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4642)	FIOP RMSD15 of SEQ ID NO: 4642 Relative to Variant ¹
4859/4860	A320Y	+
4873/4874	A320I	+
4889/4890	R366K	+
4899/4900	E293P	+
4909/4910	M317S	+
4911/4912	E293G	++
4913/4914	M317A	++
4953/4954	L194M	+++

Table 27.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4642)	FIOP RMSD15 of SEQ ID NO: 4642 Relative to Variant ¹
4959/4960	P277S	+
4989/4990	L129M	+
4993/4994	P294G	++
5029/5030	P277G	+
5033/5034	E293S	+
5041/5042	L128V	+++
5043/5044	S182T	+++
5045/5046	L299T	+++
5047/5048	E293K	++
5049/5050	S182Q	++
5051/5052	M317I	++
5053/5054	I127L	++
5055/5056	M317F	++
5057/5058	P277A	++
5059/5060	V324I	++
5061/5062	P294Q	++
5063/5064	A356G	+
5065/5066	I140V	+
5067/5068	P294L	+
5069/5070	T319A	+
5071/5072	V180L	+
5073/5074	P277H	+
5075/5076	L333D	+
5077/5078	L300T	+
5079/5080	L333T	+
5081/5082	C392M	+
5083/5084	C392H	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4642, and defined as follows: "+" 1.01 to 1.02; "++" > 1.02 to 1.04; and "+++" > 1.04.		

EXAMPLE 28**Improvements Over SEQ ID NO: 4984 in the Transposition of Double-Stranded DNA (dsDNA)****Adapters into dsDNA Target Sequences**

5

HTP Screening for Improved Transposase Variants

SEQ ID NO: 4984 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation

mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 28.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 28.1**, and analyzed on CE as described in Example 4.

Table 28.1 Reaction Conditions

Transposase expression: 30°C in *E. coli* W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl₂, 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 2 μ M; Target substrate: annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, at 0.035 μ M; Final dilution before CE analysis: 12x

- Activity relative to SEQ ID NO: 4984 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, observed by the reaction with SEQ ID NO: 4984 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 28.2**.

Table 28.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4984)	FIOP Percent Product (Relative to SEQ ID NO: 4984) ¹
5085/5086	P151K	+++
5087/5088	P151R	+++
5089/5090	D196F	+++
5091/5092	P151A	+++
5093/5094	P151N	+++
5095/5096	Q191K	+++
5097/5098	D196L	+++
5099/5100	P151G	+++
5101/5102	P151S	++
5103/5104	D196Y	++
5105/5106	P151I	++

Table 28.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4984)	FIOP Percent Product (Relative to SEQ ID NO: 4984) ¹
5107/5108	Q191V	++
5109/5110	D196S	++
5111/5112	P151V	++
5113/5114	P151L	++
5115/5116	S118R	++
5117/5118	Q191L	++
5119/5120	D196V	++
5121/5122	Q191R	++
5123/5124	Q191S	++
5125/5126	D196W	++
5127/5128	P151T	++
5129/5130	S118K	+
5131/5132	D196I	+
5133/5134	D196A	+
5135/5136	E113R	+
5137/5138	P151E	+
5139/5140	V205T	+
5141/5142	P151H	+
5143/5144	P151C	+
5145/5146	E113A	+
5147/5148	E113S	+
5149/5150	Q191M	+
5151/5152	P151W	+
5153/5154	P151M	+
5155/5156	D196K	+
5157/5158	P151Q	+
5159/5160	Q191G	+
5161/5162	D196M	+
5163/5164	E113C	+
5165/5166	P151Y	+
5167/5168	D196R	+
5169/5170	P151F	+
5171/5172	E113T	+
5173/5174	D196G	+
5175/5176	E113K	+

Table 28.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4984)	FIOP Percent Product (Relative to SEQ ID NO: 4984) ¹
5177/5178	D196H	+
5179/5180	S118A	+
5181/5182	K117R	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4984, and defined as follows: "+" 1.02 to 1.94; "++" 1.94 to 2.60; and "+++" > 2.60.		

EXAMPLE 29

**Improvements Over SEQ ID NO: 4984 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 4984 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 29.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 29.1**, and analyzed on CE as described in Example 4.

Table 29.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 20% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 41°C for 30 minutes; Adapter concentration: 2 μ M; Target substrate: annealed SEQ ID NO: 5697/SEQ ID NO: 5698, at 0.035 μ M; Final dilution before CE analysis: 12x

20 Activity relative to SEQ ID NO: 4984 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, observed by the reaction with SEQ ID NO: 4984 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 29.2**.

25

Table 29.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4984)	FIOP Percent Product (Relative to SEQ ID NO: 4984) ¹
5183/5184	P151R/K154S/K213R/E343G	+++
5185/5186	D143E	+++
5187/5188	Q115E/D143L/E343G	+++
5189/5190	Q115E/D143L/P151R/K154S/A233Q/E343G	+++
5191/5192	Q115E/D143L/A233R	+++
5193/5194	P151R/K213R/A233F	+++
5195/5196	Q115E	+++
5197/5198	D143V/K213R/A233R	++
5199/5200	A233T/I238V/E343G	++
5201/5202	D143L/K213R/A233R	++
5203/5204	D143E/P151R/K154S	++
5205/5206	D143V/K154S	++
5207/5208	P151R/K213R	++
5209/5210	P151R/K154S/E343G	++
5211/5212	Q115E/D143L/A233Q	++
5213/5214	V114G/Q115E/A233T	++
5215/5216	Q115E/D143L/K154S/K213R/A233Q/I238V/S261R	+
5217/5218	D143E/K213R/A233F/I238V/S261R/E343G	+
5219/5220	D143L/S261R	+
5221/5222	K213R/A233R/E343G	+
5223/5224	D143V/K213R/S261R/E343G	+
5225/5226	Q115E/D143E/K213R	+
5227/5228	Q115E/E343G	+
5229/5230	V114G/E343G	+
5231/5232	Q115E/D143L/P151R/K154S/K213R/S261R	+
5233/5234	Q115E/P151R/K154S/A233F	+
5235/5236	P151R/K154S	+
5237/5238	V114G/D143V	+
5239/5240	V114G/K213R/I238V	+
5241/5242	D143E/A233F/S261R	+
5243/5244	Q115E/D143L/P151R/A233F/E343G	+
5245/5246	Q115E/K213R/A233R	+
5247/5248	Q115E/K213R/A233F/I238V	+
5249/5250	V114G/Q115E/D143E/A233T	+

Table 29.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 4984)	FIOP Percent Product (Relative to SEQ ID NO: 4984) ¹
5251/5252	D143E/P151R/K154S/K213R	+
5253/5254	V114G/D143E/A233Q	+
5255/5256	V114G/P151R/S261R/E343G	+
5257/5258	Q115E/K213R/S261R	+
5259/5260	Q115E/I238V	+
5261/5262	Q115E/D143E	+
5263/5264	Q115E/L348F	+
5265/5266	Q115E/P151R/S261R	+
5267/5268	V114G/D143V/P151R/A233T/I238V/S261R	+
5269/5270	V114G/Q115E/D143V	+
5271/5272	Q115E/D143V/P151R/A233F/S261R	+
5273/5274	V114G/Q115E/E343G	+
5275/5276	V114G/D143L/A233F/I238V	+
5277/5278	V114G/D143V/K213R/S261R/E343G	+
5279/5280	Q115E/A233F/I238V	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 4984, and defined as follows: "+" 1.04 to 3.27; "++" 3.27 to 5.88; and "+++" > 5.88.		

EXAMPLE 30

**Improvements Over SEQ ID NO: 5196 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 5196 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- 10 Reactions were performed in 96-well format BioRad PCR plates. 100 nL of 100% HTP lysate variants in lysis buffer (**Table 30.1**) were combined in 1:1 volume ratio with 2.5 μ M seqWell i5 tagging reagents and incubated at room temperature for 1 hour with mild shaking. 100 nL of the same set of samples were also combined with 2.5 μ M seqWell i7 oligos in 1:1 volume ratio in a separate set of reactions and incubated at room temperature for one hour with mild shaking. The loaded i5 and i7
- 15 transposases were then combined in equal volumes and diluted with 12.5 μ L MTSB prior to library preparation. To prepare the library for sequencing 10 ng of *E.coli* genomic DNA was mixed in equal volumes (5 μ L) with i5/i7 transposase mix and 3x coding buffer. The reactions were incubated at 55°C for 15 minutes and terminated with 7.5 μ L of seqWell X-solution at 68°C for 10 minutes. The reactions were then purified with (1x) MAGwise beads and eluted in 25 μ L of 10 mM Tris pH 8.0. This was followed by 12

cycle amplification (**Table 30.1**), sample pooling and MAGwise bead purification. An equal volume of the samples (10 μ L each) was pooled together and purified with (1x) MAGwise beads. The resulting library pool was evaluated using NanoDrop Spectrophotometer or Qubit and the Illumina MiSeq assay was performed.

5

Table 30.1 Reaction Conditions	
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes;	
Library Amplification conditions	
Amplification Reaction Set-up	PCR program
10 μ L purified tag product	72C, 10 mins
2.5 μ L 25 μ M Library Primer Mix (LPM)	95C, 3 mins
12.5 μ L 2x Library Amp Mix	98C, 0:30 mins 64C, 0:15 mins 12X 72C, 0:30 mins
	72C, 3 mins
	8C, Hold

Relative target bias for each variant was calculated as the mean RMSD₁₅ of SEQ ID NO: 5196 divided by the mean RMSD₁₅ of the transposase variant. Variants with target bias mean RMSD₁₅ less than SEQ ID NO: 5196 will have a ratio greater than 1. The results are shown in **Table 30.2**.

10

Table 30.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5196)	FIOP RMSD15 of SEQ ID NO: 5196 Relative to Variant ¹
5281/5282	K117R/S118R/D143Q/M168K/V169T/C392A	+++
5283/5284	K117R/D143Q/M168K/V169T/C392A	+++
5285/5286	K117R/S118R/M168K/V169T/C392A	+++
5287/5288	K117R/S118R/D143Q/C392A	+++
5289/5290	K117R/S118R/M168K/V169M	+++
5291/5292	S118R/D143Q/M168K/V169T	+++
5293/5294	K117R/S118R/M168K/V169T/C392P	++
5295/5296	S118R/V169T/C392A	++
5297/5298	D143Q/V169T	++
5299/5300	S118R/D143Q/C392A	++
5301/5302	K117R/V169T	++
5303/5304	S118R/D143Q	++
5305/5306	K117R/M168K/C392A	++
5307/5308	K117R/M168K/V169T/C392P	++

Table 30.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5196)	FIOP RMSD15 of SEQ ID NO: 5196 Relative to Variant ¹
5309/5310	S118R/C392A	++
5311/5312	K117R/M168K/V169T	++
5313/5314	K117R	+
5315/5316	S118R/M168K	+
5317/5318	M168K/C392A	+
5319/5320	K117R/D143Q/M168K/V169T/R170H/C392A	+
5321/5322	S118R/D143Q/C392P	+
5323/5324	M168K/V169T/C392A	+
5325/5326	S118R/C392P	+
5327/5328	D143Q/C392P	+
5329/5330	D143Q/M168K/V169T/C392P	+
5331/5332	M168K/C392P	+
5333/5334	K117R/C392P	+
5335/5336	S118R/M168K/C392P	+
5337/5338	M168K/V169T	+
5339/5340	S118R/D143Q/V169T	+
5341/5342	C392P	+
5343/5344	M168K/V169T/C392P	+
5345/5346	C392A	+
5347/5348	V169T/C392S	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 5196, and defined as follows: "+" 1.03 to 1.11; "++" > 1.11 to 1.17; and "+++" > 1.17.		

EXAMPLE 31

**Improvements Over SEQ ID NO: 5286 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 5286 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

10 Reactions were performed in 96-well format BioRad PCR plates. 100 nL of 100% HTP lysate variants in lysis buffer (**Table 31.1**) were combined in 1:1 volume ratio with 2.5 μ M seqWell i5 tagging reagents and incubated at room temperature for 1 hour with mild shaking. 100 nL of the same set of samples were also combined with 2.5 μ M seqWell i7 oligos in 1:1 volume ratio in a separate set of reactions and incubated at room temperature for one hour with mild shaking. The loaded i5 and i7

transposases were then combined in equal volumes and diluted with 12.5 μ L MTSB prior to library preparation. To prepare the library for sequencing 10 ng of *E.coli* genomic DNA was mixed in equal volumes (5 μ L) with i5/i7 transposase mix and 3x coding buffer. The reactions were incubated at 55°C for 15 minutes and terminated with 7.5 μ L of seqWell X-solution at 68°C for 10 minutes. The reactions were then purified with (1x) MAGwise beads and eluted in 25 μ L of 10 mM Tris pH 8.0. This was followed by 12 cycle amplification (**Table 31.1**), sample pooling and MAGwise bead purification. An equal volume of the samples (10 μ L each) was pooled together and purified with (1x) MAGwise beads. The resulting library pool was evaluated using NanoDrop Spectrophotometer or Qubit and the Illumina MiSeq assay was performed.

10

Table 31.1 Reaction Conditions	
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes;	
Library Amplification conditions	
Amplification Reaction Set-up	PCR program
10 μ L purified tag product	72C, 10 mins
2.5 μ L 25 μ M Library Primer Mix (LPM)	95C, 3 mins
12.5 μ L 2x Library Amp Mix	98C, 0:30 mins 64C, 0:15 mins 12X 72C, 0:30 mins
	72C, 3 mins
	8C, Hold

Relative target bias for each variant was calculated as the mean RMSD₁₅ of SEQ ID NO: 5286 divided by the mean RMSD₁₅ of the transposase variant. Variants with target bias mean RMSD₁₅ less than SEQ ID NO: 5286 will have a ratio greater than 1. The results are shown in **Table 31.2**.

15

Table 31.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5286)	FIOP RMSD15 of SEQ ID NO: 5286 Relative to Variant ¹
5349/5350	T134V/S182T/L194M/P294Q	++++
5351/5352	T134R/L194M/E293P	++++
5353/5354	T134R/L194M/P294Q	++++
5355/5356	L194M/E293P/A320Y	+++
5357/5358	T134R/L194M/E293P/P294Q	+++
5359/5360	T134R/L194M/P277W	+++
5361/5362	T134V/L194M/E293P	+++
5363/5364	A320Y	+++
5365/5366	L194M	+++

Table 31.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5286)	FIOP RMSD15 of SEQ ID NO: 5286 Relative to Variant ¹
5367/5368	S182T	+++
5369/5370	S182T/L194M/E293G/A320Y	+++
5371/5372	T134R/L194M/M317S	+++
5373/5374	T134R/L194M/M317S/A320Y	+++
5375/5376	T134R/S182T/P294Q	+++
5377/5378	T134V/L194M/P277W	+++
5379/5380	S182T/L194M	++
5381/5382	T134R/L194M/P277W/M317S	++
5383/5384	T134V	++
5385/5386	E293P	++
5387/5388	L194M/E293P/P294Q/M317S	++
5389/5390	M317S	++
5391/5392	T134R/L194M/E293P/M317S	++
5393/5394	L194M/P277W/A320Y	++
5395/5396	L194M/P277W/E293G/M317S	++
5397/5398	L194M/P277W/M317S/A320Y	++
5399/5400	M317S/A320Y	++
5401/5402	P277W	++
5403/5404	S182T/E293G/P294Q	++
5405/5406	T134R/M317S	++
5407/5408	T134R/S182T/P277W/E293G	++
5409/5410	T134V/S182T/E293P/A320Y	++
5411/5412	T134P	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 5286, and defined as follows: "+" 0.92 to 1.00; "++" 1.02 to 1.05; "+++" > 1.05 to 1.10; and "++++" > 1.10.		

EXAMPLE 32

**Improvements Over SEQ ID NO: 5412 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5

HTP Screening for Improved Transposase Variants

SEQ ID NO: 5412 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (e.g., saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP
- 5 plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 32.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 32.1**, and analyzed on CE as described in Example 4.

Table 32.1 Reaction Conditions

Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150 mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes; Adapter concentration: 2 μ M; Target substrate: annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698 at 0.035 μ M; Final dilution before CE analysis: 12x
--

10

- Activity relative to SEQ ID NO: 5412 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, observed by the reaction with SEQ ID NO: 5412 (where the percent product may be set as the average of replicates or else the highest single
- 15 sample as appropriate). The results are shown in **Table 32.2**.

Table 32.2

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5412)	FIOP Percent Product (Relative to SEQ ID NO: 5412) ¹
5497/5498	P134R/E473G	+++
5499/5500	P134R/S457M	+++
5501/5502	P134R/Q390V	+++
5503/5504	P134R/V464A	+++
5505/5506	H13T/P134R	+++
5507/5508	P134R/A393T	+++
5509/5510	P134R/Q456T	+++
5511/5512	H13L/P134R	+++
5513/5514	P134R/E467T	+++
5515/5516	P134R/L387G	+++
5517/5518	P134R/L421V	+++
5519/5520	P134R/V395I	+++
5521/5522	P134R/V459L	++

Table 32.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5412)	FIOP Percent Product (Relative to SEQ ID NO: 5412) ¹
5523/5524	P134R/K402L	++
5525/5526	H13V/P134R	++
5527/5528	P87G/P134R	++
5529/5530	P134R/K378R	++
5531/5532	P134R/K378V	++
5533/5534	P134R/V381D	++
5535/5536	P134R/H210R/K402I	++
5537/5538	P134R/L475G	++
5539/5540	P134R/L474T	+
5541/5542	P134R/S391T	+
5543/5544	I89C/P134R	+
5545/5546	P134R/L387I	+
5547/5548	P134R/E388F	+
5549/5550	P134R/M407V	+
5551/5552	P134R/L380V	+
5553/5554	P64R/D93T/P134R/T235S	+
5555/5556	P134R/E398S	+
5557/5558	D93T/P134R/E344T	+
5559/5560	I89L/P134R	+
5561/5562	P134R/Q456D	+
5563/5564	H13R/P134R	+
5565/5566	E12N/P134R	+
5567/5568	P134T/H210R/D382R	+
5569/5570	P64R/P134R/E344Q/N347R	+
5571/5572	P134R/A384F	+
5573/5574	P134R/V460E	+
5575/5576	P134R/K402V	+
5577/5578	P134R/D471S	+
5579/5580	P134R/A470L	+
5581/5582	P134R/E386W	+
5583/5584	P134R/D382W	+
5585/5586	D93T/P134R/V244L/E344T	+
5587/5588	P134R/V395M	+
5589/5590	H13V/P134T/H210R	+
5591/5592	P134R/K377I	+

Table 32.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5412)	FIOP Percent Product (Relative to SEQ ID NO: 5412) ¹
5593/5594	P134R/M468S	+
5595/5596	D93T/P134R/T152P/T235S/E344T	+
5597/5598	P134R/A470P	+
5599/5600	P134R/E467R	+
5601/5602	H13C/P134T/H210R	+
5603/5604	P134R/W401V	+
5605/5606	P134R/T152P/E344Q	+
5607/5608	H13L/P134T	+
5609/5610	D93T/P134R/T235S/E344T	+
5611/5612	E77Y/P134R	+
5613/5614	T79V/P134T/H210R	+
5615/5616	P22R/P134R/T235S/E344T	+
5617/5618	P87M/P134R	+
5619/5620	P134R/D175C	+
5621/5622	P134R/D382K	+
5623/5624	Y86S/P134R	+
5625/5626	H13S/P134R	+
5627/5628	P134R/D471M	+
5629/5630	P134R/S391H	+
5631/5632	P134R/E467L	+
5633/5634	P134T/H210R/Q456C	+
5635/5636	P134R/D382G	+
5637/5638	P134T/H210R/S478D	+
5639/5640	P134R/E473D/L474I	+
5641/5642	Q84V/P134R	+
5643/5644	P22R/P134R/T152P	+
5645/5646	P134R/S391R	+
5647/5648	P134R/E344T	+
5649/5650	G131I	+
5651/5652	P134R/K406C	+
5653/5654	P134T/L380V	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 5412, and defined as follows: "+" 1.01 to 2.20; "++" 2.20 to 3.66; and "+++" > 3.66.		

EXAMPLE 33**Improvements Over SEQ ID NO: 5412 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

HTP Screening for Improved Transposase Variants

5 SEQ ID NO: 5412 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

Reactions were performed in 96-well format BioRad PCR plates. 100 nL of 100% HTP lysate
10 variants in lysis buffer (**Table 33.1**) were combined in 1:1 volume ratio with 2.5 μ M seqWell i5 tagging reagents and incubated at room temperature for 1 hour with mild shaking. 100 nL of the same set of samples were also combined with 2.5 μ M seqWell i7 oligos in 1:1 volume ratio in a separate set of reactions and incubated at room temperature for one hour with mild shaking. The loaded i5 and i7 transposases were then combined in equal volumes and diluted with 12.5 μ L MTSB prior to library
15 preparation. To prepare the library for sequencing 10 ng of *E.coli* genomic DNA was mixed in equal volumes (5 μ L) with i5/i7 transposase mix and 3x coding buffer. The reactions were incubated at 55°C for 15 minutes and terminated with 7.5 μ L of seqWell X-solution at 68°C for 10 minutes. The reactions were then purified with (1x) MAGwise beads and eluted in 25 μ L of 10 mM Tris pH 8.0. This was followed by 12 cycle amplification (**Table 33.1**), sample pooling and MAGwise bead purification. An equal volume of the
20 samples (10 μ L each) was pooled together and purified with (1x) MAGwise beads. The resulting library pool was evaluated using NanoDrop Spectrophotometer or Qubit and the Illumina MiSeq assay was performed.

Table 33.1 Reaction Conditions

Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 40% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 30 minute resuspension at room temperature followed by incubation at 46°C for 30 minutes;	
Library Amplification conditions	
Amplification Reaction Set-up	PCR program
10 μ L purified tag product	72C, 10 mins
2.5 μ L 25 μ M Library Primer Mix (LPM)	95C, 3 mins
12.5 μ L 2x Library Amp Mix	98C, 0:30 mins 64C, 0:15 mins 12X 72C, 0:30 mins
	72C, 3 mins
	8C, Hold

25 Relative target bias for each variant was calculated as the mean RMSD₁₅ of SEQ ID NO: 5412 divided by the mean RMSD₁₅ of the transposase variant. Variants with target bias mean RMSD₁₅ less than SEQ ID NO: 5412 will have a ratio greater than 1. The results are shown in **Table 33.2**.

Table 33.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5412)	FIOP RMSD15 of SEQ ID NO: 5412 Relative to Variant ¹
5413/5414	I127L/P134R	+++
5415/5416	P134R/L333T	+++
5417/5418	Q73K/I127L/P134T/K290R/L333T	+++
5419/5420	Q73K/P134R/A160G/I172R/K290R	+++
5421/5422	P134R	+++
5423/5424	P134R/A192S/R198Q	+++
5425/5426	P134R/R198Q	+++
5427/5428	P134R/R198Q/V266A/L333T	+++
5429/5430	Q73K/I127L/P134R/I172R/A192S/V266A/L333T	+++
5431/5432	Q73K/P134R/I172R/H210R/L333T	+++
5433/5434	Q73K/P134R	++
5435/5436	Q73K/P134R/I172R	++
5437/5438	Q73K/I127L/P134T/H210R/K290R	++
5439/5440	Q73K/I127L/P134T/H210R/L333T	++
5441/5442	Q73K/P134T/A160G/I172R/R198Q/L333T	++
5443/5444	I127L/P134R/I172R/R198Q/V266A/L333T	++
5445/5446	I127L/P134T	++
5447/5448	P134R/I172R/A192S/K290R/L333T	++
5449/5450	P134T/A160G/R198Q/V266A/K290R	++
5451/5452	Q73K/I127L/P134R/A160G/I172R/R198Q/K290R	++
5453/5454	Q73K/I127L/P134R/I172R/R198Q	++
5455/5456	Q73K/P134T/H210R/L333T	++
5457/5458	P134R/A160G/I172R/L333T	+
5459/5460	P134T/H210R	+
5461/5462	Q73K/I127L/P134R/L333T	+
5463/5464	Q73K/P134R/I172R/V266A/L333T	+
5465/5466	Q73K/P134R/L333T	+
5467/5468	I127L/P134T/H210R/L333T	+
5469/5470	Q73K/I127L/P134R/A160G/R198Q/L333T	+
5471/5472	Q73K/P134R/A160G/L333T	+
5473/5474	Q73K/P134T/A160G/I172R/A192S/K290R	+
5475/5476	Q73K/P134T/R198Q/H210R	+
5477/5478	Q73K/P134T/R198Q/H210R/V266A	+
5479/5480	I127L/P134T/R198Q/H210R/K290R	+
5481/5482	Q73K/P134T/A160G/V266A	+
5483/5484	I127L/P134R/A160G/I172R/A192S/L333T	+

Table 33.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5412)	FIOP RMSD15 of SEQ ID NO: 5412 Relative to Variant ¹
5485/5486	I127L/P134R/H210R/V266A/K290R	+
5487/5488	I127L/P134T/A160G/I172R/H210R/V266A/K290R	+
5489/5490	Q73K/I127L/P134T/A160G/H210R/V266A	+
5491/5492	I127L/P134T/A192S/H210R	+
5493/5494	P134T/H210R/K290R	+
5495/5496	Q73K/P134R/R366H	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 5412, and defined as follows: "+" 1.01 to 1.04; "++" 1.04 to 1.07; and "+++" > 1.09.		

EXAMPLE 34

**Improvements Over SEQ ID NO: 5422 in the Transposition of Double-Stranded DNA (dsDNA)
Adapters into dsDNA Target Sequences**

5 HTP Screening for Improved Transposase Variants

SEQ ID NO: 5422 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

- 10 Reactions were performed in 96 or 384-well format BioRad PCR plates. Bound transposomes or synaptic complexes were prepared by incubating 67% HTP lysate with 2 - 17 μ M seqWell UR1 adapter at room-temperature for 45 mins. After 45 mins incubation, annealed target substrate(s) in seqWell 3x coding buffer was added to the bound HTP transposomes to a final concentration of 0.03-0.3 μ M. HTP plates were centrifuged, and samples were heated in a thermocycler at 55°C for 5-15 mins. The exact
- 15 concentration of UR1 adapter, target substrate identity and concentration is shown in **Table 34.1**. Samples were quenched by heating at 75°C with 0.04% SDS. After heat-quench, samples were diluted with water to a final dilution as described in **Table 34.1**, and analyzed on CE as described in Example 4.

Table 34.1 Reaction Conditions
Transposase expression: 30°C in <i>E. coli</i> W3110, Lysis buffer: 25 mM Tris-HCL, 150mM NaCl, 5 mM MgCl ₂ , 0% glycerol, pH 8.0, 1.0 g/L lysozyme; Lysis conditions: 60 minute resuspension at room temperature followed by incubation at 39°C for 30 minutes; Adapter concentration: 2 μ M; Target substrate: annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, at 0.035 μ M; Final dilution before CE analysis: 12x

- 20 Activity relative to SEQ ID NO: 5422 (Activity FIOP) was calculated as the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, produced by the variant compared with the percent product of annealed oligos SEQ ID NO: 5697/SEQ ID NO: 5698, observed by the reaction with SEQ ID

NO: 5422 (where the percent product may be set as the average of replicates or else the highest single sample as appropriate). The results are shown in **Table 34.2**.

Table 34.2		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5422)	FIOP Percent Product (Relative to SEQ ID NO: 5422)¹
5655/5656	H13L/S178R/S457M	+++
5657/5658	S178R	++
5659/5660	S178R/E473G	++
5661/5662	V381W/A393T/S457M	++
5663/5664	S178R/V381W/S457M	++
5665/5666	H13L/K402L/S457M	++
5667/5668	H13L/V381D/S457M	++
5669/5670	V381W/E473G	++
5671/5672	H13L/S457M	+
5673/5674	A393T/S457M	+
5675/5676	S178R/K402L/S457M	+
5677/5678	V381W	+
5679/5680	H13L/V381W	+
5681/5682	S178R/V381D/S457M/E473G	+
5683/5684	H13L	+
5685/5686	H13L/S178R/V381D/E473G	+
5687/5688	H13L/S178R/V381W	+
5689/5690	V381D/S457M/E473G	+
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 5422, and defined as follows: "+" 1.32 to 2.16; "++" 2.16 to 3.02; and "+++" > 3.02.		

5

EXAMPLE 35**Improvements Over SEQ ID NO: 972 in the Transposition of Double-Stranded DNA (dsDNA)****Adapters into dsDNA Target Sequences**

HTP Screening for Improved Transposase Variants

SEQ ID NO: 972 was selected as the parent transposase enzyme. Libraries of engineered genes were produced from the parent gene using well-established techniques (*e.g.*, saturation mutagenesis and recombination of previously identified beneficial mutations). The polypeptides encoded by each gene were produced in HTP and prepared as described in Example 2.

Reactions were performed in 96-well format BioRad PCR plates. 100 nL of 2.5 μ M shake-flask variants prepared in Seqwell MTSB buffer were combined in 1:1 volume ratio with 2.5 μ M seqWell i5 tagging reagents and incubated at room temperature for 1 hour with mild shaking. 100 nL of the same set

of samples were also combined with 2.5 μ M seqWell i7 oligos in 1:1 volume ratio in a separate set of reactions and incubated at room temperature for one hour with mild shaking. The loaded i5 and i7 transposases were then combined in equal volumes and diluted with 12.5 μ L MTSB prior to library preparation. To prepare the library for sequencing 10 ng of *E.coli* genomic DNA was mixed in equal volumes (5 μ L) with i5/i7 transposase mix and 3x coding buffer. The reactions were incubated at 55°C for 15 minutes and terminated with 7.5 μ L of seqWell X-solution at 68°C for 10 minutes. The reactions were then purified with (1x) MAGwise beads and eluted in 25 μ L of 10 mM Tris pH 8.0. This was followed by 12 cycle amplification (**Table 33.1**), sample pooling and MAGwise bead purification. An equal volume of the samples (10 μ L each) was pooled together and purified with (1x) MAGwise beads. The resulting library pool was evaluated using NanoDrop Spectrophotometer or Qubit and the Illumina MiSeq assay was performed.

Relative target bias for each variant was calculated as the mean RMSD₁₅ of SEQ ID NO: 972 divided by the mean RMSD₁₅ of the transposase variant. Variants with target bias mean RMSD₁₅ less than SEQ ID NO: 972 will have a ratio greater than 1. The results are shown in **Table 35.1**.

Table 35.1		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP RMSD15 of SEQ ID NO: 972 Relative to Variant¹
3787/3788	C38F/K73Q/S113E/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198R/P220R/A266V/K270P/R317M/T333L/W423D	++
4149/4150	C38F/K73Q/S113E/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198R/P220R/V243G/A266V/K270P/R317M/T333L/W423D	++
4277/4278	C38F/K73Q/S113E/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D	++
4641/4642	C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D	++
4933/4934	C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/S192A/Q198R/H210R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D	+
4983/4984	C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/A163V/D164R/F168M/C169V/R172I/C	+

Table 35.1		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP RMSD15 of SEQ ID NO: 972 Relative to Variant ¹
	184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D	
5195/5196	C38F/K73Q/S113E/K115E/A118S/L127I/E152T/A154K/G160A/A163V/D164R/F168M/C169V/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D	+
5285/5286	C38F/K73Q/S113E/K115E/K117R/A118R/L127I/E152T/A154K/G160A/A163V/D164R/F168K/C169T/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/C392A/W423D	++
5421/5422	C38F/K73Q/S113E/K115E/K117R/A118R/L127I/T134R/E152T/A154K/G160A/A163V/D164R/F168K/C169T/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/C392A/W423D	++
5499/5500	C38F/K73Q/S113E/K115E/K117R/A118R/L127I/T134R/E152T/A154K/G160A/A163V/D164R/F168K/C169T/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/C392A/W423D/S457M	++
2979/2980	C38F/S113E/L127I/E152T/A154K/G160A/D164R/F168K/R172I/Q198R/P220R/K290R/R317M/W423D	++
5691/5692	C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/D164R/F168K/R172I/C184A/D189S/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D	+
5693/5694	C38F/K73Q/S113E/K115Q/A118S/L127I/E152T/A154K/G160A/W162L/D164R/F168K/R172I/C184A/S192A/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/W423D	+
5701/5702	C38F/K73Q/S113E/K115E/K117R/A118R/L127I/T134V/E152T/A154K/G160A/A163V/D164R/F168K/C169T/R172I/C184A/S192A/L194M/Q198R/P220R/V243G/L260I/A266V/K270P/L295I/R317M/T333L/C392A/W423D	+

Table 35.1		
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 972)	FIOP RMSD15 of SEQ ID NO: 972 Relative to Variant ¹
¹ Levels of increased activity were determined relative to the reference polypeptide of SEQ ID NO: 972, and defined as follows: "+" 1.46 to 2.04; and "++" > 2.04 to 2.37.		

EXAMPLE 36

Improvements Over SEQ ID NO: 5704 in the Transposase Activity
and Library Construction for NGS

5 Adapter Loading of Transposases

Transposases (e.g., engineered transposases) were loaded with oligonucleotide adapters. Each oligonucleotide adapter was diluted to a concentration of 100 μM and combined in a 1:1 ratio of top strand to bottom strand. The oligonucleotides were then incubated at 95°C for two minutes and then the temperature was reduced in a stepwise manner by 1°C per minute until the temperature reached 25°C.

10

Adapter Sequences:

A14 Adapter:

A14: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG (SEQ ID NO: 5757)
20TT: /5Phos/CTGTCTCTTATACACATCT/3InvdT/ (SEQ ID NO: 5758)

15

B15 Adapter:

B15: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG (SEQ ID NO: 5759)
20TT: /5Phos/CTGTCTCTTATACACATCT/3InvdT/ (SEQ ID NO: 5758)

20 The molar binding ratio (MBR) of an engineered transposase (e.g., an engineered Tn5 variant) was defined by the ratio of picomoles of transposase required to bind 250 picomoles of B15 adapter. For instance, 500 pmol of a Tn5 variant required to bind 250 pmol of B15 adapter is an MBR of 2.

To determine the MBR, 2.5 μL of 50 μM B15 adapter was combined with each Tn5 variant (with the respective concentration shown below in **Table 36.1**) at varying volumes of 0 μL, 2.4 μL, 3.6 μL, 4.8 μL, 6 μL, 7.2 μL, and 9.6 μL. The adapter was loaded into the Tn5 variant by incubating the mixture for 1 hour at 25°C on a thermocycler in which the lid was removed. The volumes of the transposomes were then raised to 16.9 μL with modified Tn5 buffer (MTSB). A sample from each tube (3.4 μL) was mixed with 19.5 μL of 10 mM Tris and were then run on a 4% agarose gel for 15 minutes. The MBR was calculated from the gel lane in which the B15 adapter was completely consumed by the Tn5 variant (i.e., a band corresponding to unbound B15 adapter was not observed).

30 The Tn5 variants were loaded with an adapter based on the results from the MBR assay. To ensure that all transposases were loaded with an adapter, an excess amount of an adapter was included at two-times the amount of the determined MBR concentration for each reaction. Each transposase was incubated with 2 μL of 50 μM adapter at 25°C for 1 hour. Following this incubation step, the loaded transposases were diluted to 1 μM via adding MTSB. Each transposase was loaded twice: once with an

35

A14 adapter and again with a B15 adapter. The loaded A14 and B15 transposomes were then combined to make a 1 μ M A14/B15 tagging reagent for each enzyme.

The MBR and transposase loading for each Tn5 variant is summarized in **Table 36.1**. The Tn5 enzyme was obtained from a commercial source.

5

SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5704)	Concentration (μM)	MBR	Volume Transposase (μL)	Volume of 50 μM Adapter (μL)	Volume MTSB (μL)
5703/5704	N/A	46.5	2	2.15	2	95.85
5705/5706	D192S	64.1	3.5	2.73	2	95.27
5707/5708	K273P	67.3	3	2.23	2	95.77
5709/5710	K120R	47.4	2.75	2.90	2	95.10
5711/5712	S181R	55	3.25	2.95	2	95.05
5713/5714	K412N	50.8	3	2.95	2	95.05
5715/5716	Q341L	54.5	3.25	2.98	2	95.02
5717/5718	L263I	46.3	2.75	2.97	2	95.03
5719/5720	L296I	58.9	2.5	2.12	2	95.88
5721/5722	I48V	56.8	2.75	2.42	2	95.58
5723/5724	E454R	37.4	3.5	4.68	2	93.32
5725/5726	L130I	51.5	3	2.91	2	95.09
5727/5728	I318M	43.6	2.5	2.87	2	95.13
5729/5730	R175I	56.5	2.5	2.21	2	95.79
5731/5732	V246G	55	3.25	2.95	2	95.05
5733/5734	A157K	56.1	2.5	2.23	2	95.77
5735/5736	Q118E	28.1	2.5	4.45	2	93.55
5737/5738	S116E	28	2.25	4.02	2	93.98
5739/5740	C187A	35.3	2.25	3.19	2	94.81
5741/5742	S458M	40.8	2	2.45	2	95.55
5743/5744	D153H	35.6	2.25	3.16	2	94.84
5745/5746	E394A	23.8	2	4.20	2	93.80
5747/5748	F137R	21.7	2.25	5.18	2	92.82
5749/5750	G163A	46.6	2.25	2.41	2	95.59
5751/5752	T334L	47	2	2.13	2	95.87
5753/5754	H213R	30.9	2	3.24	2	94.76
5755/5756	W424D	14.6	3	10.27	2	87.73

Table 36.1						
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5704)	Concentration (μ M)	MBR	Volume Transposase (μ L)	Volume of 50 μ M Adapter (μ L)	Volume MTSB (μ L)
Tn5*	N/A	74	1	0.68	2	97.32
*This enzyme was purchased from a commercial supplier.						

Transposase Activity and Performance

The activity of each transposase variant was determined using 2.5 μ M B15/Tn5 variant at the selected MBR. The transposome was serially diluted 2-fold in MTSB from 1:1 to 1:64 volumetric dilutions. A negative control containing only MTSB without a transposome was included. Each dilution was used to linearize a pUC-19 plasmid, in which each reaction contained 100 ng of the pUC-19 plasmid, 4 μ L of the transposome dilution, and 1X coding buffer. The reaction mixtures were incubated at 55°C for 15 minutes and then the reaction was stopped by adding 1.2 μ L of 0.2% SDS and incubating the reaction mixtures at 75°C for ten minutes.

After quenching the reaction, 30 μ L of 10 mM Tris pH 8.0 was mixed into each reaction tube and then loaded to a 2% agarose gel (E-GEL™ Ex Agarose Gel, Invitrogen) for ten minutes. The gel lane in which all of the pUC-19 plasmid was linearized was used to calculate transposase activity. An activity score between 0-10 was assigned to each transposase variant based on the relative activity as compared to SEQ ID NO: 5704 (**Table 36.3**).

Exemplary NGS sequencing libraries were generated via tagmentation reactions with the transposase variants followed by amplification reactions. Each tagmentation reaction contained 1 μ M of an engineered transposome (e.g., a transposome that included a Tn5 variant loaded with A14 and B15 adapters as described herein), 50 ng of human genomic DNA (NA12878), and 1X coding buffer, which were incubated at 55°C for fifteen minutes. After the incubation, the tagmentation reactions were quenched with 7.5 μ L of X-Solution (transposase stop solution; seqWell, Inc) and then heated to 68°C for ten minutes.

Oligonucleotide fragments generated via tagmentation were then cleaned up with 1.5x solid-phase reversible immobilization (SPRI) magnetic beads, which were washed twice with an 80% ethanol solution. The oligonucleotide fragments were then eluted with 12 μ L of 10 mM Tris pH 8.0.

Following elution, 10 μ L of the oligonucleotide fragments were combined with 6 μ L of Illumina UD Adapter, and 16 μ L of 2x Kapa HiFi Readymix. The oligonucleotide fragments were then amplified via PCR using the thermocycler settings summarized in **Table 36.2**, in which the thermocycler lid was held at a temperature of 105°C.

Table 36.2		
Temperature	Duration	
72°C	10 minutes	
95°C	3 minutes	
98°C	20 seconds	9 Cycles
64°C	45 seconds	
72°C	30 seconds	
72°C	2 minutes	

The resulting sequencing libraries were cleaned up via 1.5x SPRI magnetic beads, which were washed twice with an 80% ethanol solution and then eluted in 32 μ L of 10 mM Tris pH 8.0. After elution of the purified libraries from the beads, the DNA concentration was determined by electrophoresis on a TapeStation system (Agilent), as shown in **Figure 3** and **Table 36.3**.

The sequencing libraries were then pooled and loaded at 14 μ M on a MiSeq sequencing system (Illumina) using the MiSeq v2 reagents. The data was demultiplexed and aligned using Picard tools. Insertion bias was measured by calculating the RMSE of the first 25 bases of the insert using the following equation:

$$RMSE = \sqrt{\sum_{i=1}^n \frac{(y_i - y_p)^2}{n}}; y_i = \text{actual value}, y_p = \text{predicted value}.$$

The predicted values were 29.5% for A and T and 20.5% for C and G based on the average base composition of the human genome. For each Tn5 variant, the RMSE for nucleobase positions 1-20 were added, and the calculated insertion bias values for the transposases were compared (**Figures 4A-4D** and **Table 36.3**).

Table 36.3						
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5704)	Starting Transposase Concentration (μ M)	MBR	Activity Score	Insertion Bias Metric (sum(RMSE) 1-20 bases from insert site)	Average Library Size (bp)
5703/5704	N/A	46.5	2	5	109	1687
5705/5706	D192S	64.1	3.5	4	139	1488
5707/5708	K273P	67.3	3	6	110	1114
5709/5710	K120R	47.4	2.75	5	105	1390
5711/5712	S181R	55	3.25	7	116	1060

Table 36.3						
SEQ ID NO: (nt/aa)	Amino Acid Differences (Relative to SEQ ID NO: 5704)	Starting Transposase Concentration (μ M)	MBR	Activity Score	Insertion Bias Metric (sum(RMSE) 1-20 bases from insert site)	Average Library Size (bp)
5713/5714	K412N	50.8	3	5		1200
5715/5716	Q341L	54.5	3.25	7	117	1397
5717/5718	L263I	46.3	2.75	7	115	1391
5719/5720	L296I	58.9	2.5	7	109	1143
5721/5722	I48V	56.8	2.75	6	115	1129
5723/5724	E454R	37.4	3	6	112	1434
5725/5726	L130I	51.5	3.5	6	109	1216
5727/5728	I318M	43.6	2.5	6	115	1331
5729/5730	R175I	56.5	2.5	6	119	1442
5731/5732	V246G	55	3.25	6	106	1387
5733/5734	A157K	56.1	2.5	7	113	1212
5735/5736	Q118E	28.1	2.5	6	100	1263
5737/5738	S116E	28	2.25	7	87	1109
5739/5740	C187A	35.3	2.25	8	114	1237
5741/5742	S458M	40.8	2	7	116	1029
5743/5744	D153H	35.6	2.25	6	114	1263
5745/5746	E394A	23.8	2	6	111	1163
5747/5748	F137R	21.7	2.25	3	117	1422
5749/5750	G163A	46.6	2.25	5	118	1004
5751/5752	T334L	47	2	6	115	1112
5753/5754	H213R	30.9	2	2	159	1739
5755/5756	W424D	14.6		1		
Tn5*			1		117	1030
*This enzyme was purchased from a commercial supplier.						

These data highlight Tn5 transposase variants that have improved function over the SEQ ID NO: 5704. Some variants, including the one set forth in SEQ ID NO: 5738, exhibited a lower insertion bias as shown in **Table 36.3**.

5

EXAMPLE 37**Loading of Transposase Reagent with Unique Dual-Indexed Adapters**

A method was developed for loading transposases, i.e., engineered transposases, with unique dual-indexed adapters.

10

Sequences Used

i7 Index Adapters (Top Strand Sequence Shown)

Identifier: i7-1437

Sequence:

CAAGCAGAAGACGGCATACGAGATTGGACTTGACGTCTCGTGGGCTCGGAGATGTGTATAAGAGAC

15

AG (SEQ ID NO: 5760)

Identifier: i7-1610

CAAGCAGAAGACGGCATACGAGATGGAAGAGATAGTCTCGTGGGCTCGGAGATGTGTATAA

GAGACAG (SEQ ID NO: 5761)

Identifier: i7-0790

20

CAAGCAGAAGACGGCATACGAGATTTCTTCGCGGGTCTCGTGGGCTCGGAGATGTGTATAA

GAGACAG (SEQ ID NO: 5762)

Identifier: i7-0822

CAAGCAGAAGACGGCATACGAGATTCCGAGTAGGGTCTCGTGGGCTCGGAGATGTGTATAA

GAGACAG (SEQ ID NO: 5763)

25

Identifier: i7-1872

CAAGCAGAAGACGGCATACGAGATGTGATACGAAGTCTCGTGGGCTCGGAGATGTGTATAA

GAGACAG (SEQ ID NO: 5764)

Identifier: i7-1666

CAAGCAGAAGACGGCATACGAGATGGATGGAGGAGTCTCGTGGGCTCGGAGATGTGTATAA

30

GAGACAG (SEQ ID NO: 5765)

Identifier: i7-0474

CAAGCAGAAGACGGCATACGAGATGCGATGTTATGTCTCGTGGGCTCGGAGATGTGTATAA

GAGACAG (SEQ ID NO: 5766)

Identifier: i7-1201

35

CAAGCAGAAGACGGCATACGAGATAACGCATATCGTCTCGTGGGCTCGGAGATGTGTATAA

GAGACAG (SEQ ID NO: 5767)

i5 Index Adapters (Top Strand Sequence Shown)

Identifier: i5-1714

AATGATACGGCGACCACCGAGATCTACACGTAACACAGATCGTCGGCAGCGTCAGATGTGT
ATAAGAGACAG (SEQ ID NO: 5768)

5 Identifier: i5-0744

AATGATACGGCGACCACCGAGATCTACACTAAGTTGTGGTCGTCGGCAGCGTCAGATGTGTA
TAAGAGACAG (SEQ ID NO: 5769)

Identifier: i5-1180

AATGATACGGCGACCACCGAGATCTACACCACCTACCTCTCGTCGGCAGCGTCAGATGTGTA
10 TAAGAGACAG (SEQ ID NO: 5770)

Identifier: i5-1347

AATGATACGGCGACCACCGAGATCTACACATATCAGTCCTCGTCGGCAGCGTCAGATGTGTA
TAAGAGACAG (SEQ ID NO: 5771)

Identifier: i5-0995

AATGATACGGCGACCACCGAGATCTACACGGTTAAGTAGTCGTCGGCAGCGTCAGATGTGT
15 ATAAGAGACAG (SEQ ID NO: 5772)

Identifier: i5-0500

AATGATACGGCGACCACCGAGATCTACACTGGTGGATATTCGTCGGCAGCGTCAGATGTGTA
TAAGAGACAG (SEQ ID NO: 5773)

20 Identifier: i5-1482

AATGATACGGCGACCACCGAGATCTACACTAGCTGGCACTCGTCGGCAGCGTCAGATGTGT
ATAAGAGACAG (SEQ ID NO: 5774)

Identifier: i5-0732

AATGATACGGCGACCACCGAGATCTACACTTACCGTTGGTCGTCGGCAGCGTCAGATGTGTA
25 TAAGAGACAG (SEQ ID NO: 5775)

Identifier: 20TT

/5Phos/CTGTCTCTTATACACATCTG (SEQ ID NO: 5776)

1 Adapter Annealing

30 Mix top strand oligo and mosaic end bottom strand oligo (20TT) as equimolar oligo dilutions
using oligo annealing buffer

2 min at 95 °C

Thermocycler Protocol (steps (i) through (iii))

(i) 2 min at 95 °C

35 (ii) 1 min at 94 °C

(iii) 1 min ramp down: -1 °C each cycle

hold 20 °C

2 Transposase + Adapter complex formation

Mix equimolar Adapter and transposase

40 Incubate 1 hr at RT

- 3 Mix in Modified Transposase Storage Buffer (MTSB) to bring synaptic complex to $\leq 5\mu\text{M}$
- 4 Repeat above for both the i7 and i5 barcodes
- 5 Mix i7 and i5 barcode equally and dilute to working concentration in MTSB

5

EXAMPLE 38**Activity Assay for Loaded Transposase**

Engineered transposases were assayed for relative activity by measuring the molar amount of adapter-loaded transposase required to convert specific molar amount of closed circular plasmid DNA standard to a singly cut linear DNA molecule. Specifically, 2.5 pmol of each adapter-loaded transposase was used to convert a series of pUC19 dilutions centered at 200 ng to linear form after incubation at 55°C for 15 minutes in 1x Coding Buffer. The tagmentation reaction was stopped by adding SDS and incubating at 75°C for 10 minutes. Samples were diluted and loaded in a 2% E-Gel for 10 minutes. In the following example, 2.5 pmol of engineered transposase completely converted 100 ng of pUC-19 to the linear form (performed with SequenceK (SEQ ID NO: 5499/5500 for nt/aa); see **Figure 5A**), while 2.5 pmol of commercially available transposase was only able to completely convert all the 50 ng input of pUC-19 to linear form (performed with commercial hyperactive Tn5; see **Figure 5B**). The fold improvement in this case was 2x (100 ng / 50ng = 2). In particular, the linear form is shown as the band and/or smear that starts at ~3000bp in **Figure 5A** and **Figure 5B**.

20

EXAMPLE 39.

**Unique Dual-Indexed Transposition for next-generation
sequencing (NGS) Library Generation**

The engineered transposases were used to generate an exemplary NGS library using the protocol described below. Reactions may be performed, for example, in a 96-well plate.

- 25 1. Tagging Reaction (in each reaction vessel, e.g., each well of a 96-well plate)
 - DNA 8 μL (50-200 ng total) (purified human genomic DNA from Coriell NA12878)
 - i7+i5 transposase 8 μL
 - 3x Coding Buff 8 μL
 - TOTAL: 24 μL
 - 30 Tag Thermocycler Protocol:
 - 15min at 55 °C
 - Hold at 25 °C
- 35 2. Stop Reaction
 - Mix in 12 μL X-Solution (transposase stop solution) (seqWell, Inc)
 - Stop Reaction Thermocycler Protocol:
 - 10 min at 68 °C
 - Hold at 25 °C
- 40 3. Pool samples
 - Pipette 32 μL from each stopped reaction
 - Samples are pooled by row of a 96-well plate into 8 wells

4. SPRI Cleanup

0.75x - 1x solid-phase reversible immobilization (SPRI) cleanup (magnetic bead cleanup)
based on desired library size

2 x 80% EtOH Wash

5 Elute in 48 μ L 10mM Tris

5. Library Amplification

Library – 46 μ L total volume from elution above

Add 4 μ L 50 μ M Library Primer Mix

Add 50 μ L 2x Kapa HiFi

10 6. Post Amplification SPRI Cleanup

1x SPRI cleanup

2x 80% EtOH Wash

Elute in 30 μ L 10 mM Tris

15 The exemplary NGS library was analyzed using Agilent 2100 Bioanalyzer automated electrophoresis to analyze library fragment size. Results are shown in **Figure 6**. Similar to results shown above in Example 38, transposition using the engineered transposase completely converted input DNA sample into linear form, regardless of input amount (50-200 ng). The resulting electropherogram describes the composition of the library.

20

EXAMPLE 40

**Comparison of Sequencing Performance of Different
Engineered Transposases in Bacterial WGS**

25 Engineered transposases were compared based on their performance in generating libraries for bacterial whole genome sequencing. Transposase activity of the different engineered transposases were determined according to the protocol in Example 38. Loaded transposase library reagents were prepared according to the protocol described in Example 37 using a distinct combination of Illumina-compatible i5/p5 and i7/p7 barcoded adapters for transposase loading.

30 Transposition reactions were performed, and Illumina-compatible libraries were generated according to the protocol described in Example 39 from genomic DNA isolates of three microbial species: *Staphylococcus epidermidis*, *Bacillus subtilis*, and *Pseudomonas aeruginosa*. These three microbial species were selected to represent a range of GC-containing genomic sequence content.

35 The resulting libraries were pooled and sequenced together on an Illumina NextSeq2000 instrument, and the resulting data were demultiplexed and analyzed by reference-based mapping and analysis using bwa-mem and Picard. Resulting summary statistics were collected for each transposase (e.g., engineered transposase or hyperactive Tn5 control) and are reported below.

Metrics Collected:

40 Norm GC (slope) – The slope of normalized coverage by GC bin in 1000 bp windows of genomic sequence

Fold80 Penalty – The required excess of sequencing depth that would be required to bring 80% of covered bases to the median level of coverage

CV of Coverage – The standard deviation of coverage divided by the mean of coverage

% Duplication – The percentage of mapped reads having identical starting and ending genomic coordinates

Unique Insert Positions – The percentage of reads having a unique insert starting or ending genomic coordinate

Performance of Engineered Transposases:

The metrics selected for analysis enable performance characterization of each variant relative to each other and to hyperactive Tn5 control (Tn5).

To obtain a global representation of each transposases' effect on sequence integrity, transposase-induced GC bias was examined across three bacterial genomes representing a range of GC-containing genomic sequence (*Staphylococcus epidermidis* – 32% GC, *Bacillus subtilis* – 44% GC and *Pseudomonas aeruginosa* – 66% GC) using Picard GC bias plots. Low GC bias would produce a normalized coverage slope of one, or a horizontal line over windows of GC content. Alternatively, high GC bias would produce a normalized coverage slope greater or less than one, over windows of GC content.

Quantified performance of the engineered transposases in *Staphylococcus epidermidis*, *Bacillus subtilis*, and *Pseudomonas aeruginosa* are provided below in **Tables 40.1-40.3**.

Table 40.1. Performance metrics of engineered transposases in <i>Staphylococcus epidermidis</i> (32% GC content)					
Transposase SEQ ID NO: (nt/aa)	Norm GC (slope)	fold80 Penalty	CV of Coverage	% Duplication	Unique Insert Positions
Tn5	1.69	1.469	38.8%	2.8%	71.8%
SequenceA (971/972)	0.59	1.498	48.5%	9.0%	51.1%
SequenceB (2367/2368)	0.38	1.183	30.4%	3.6%	69.1%
SequenceC (2979/2980)	0.66	1.194	30.6%	3.2%	79.2%
SequenceD (4143/4144)	0.49	1.221	32.1%	0.9%	84.2%
SequenceE (4641/4642)	0.94	1.173	30.3%	0.7%	84.0%
SequenceF (4933/4934)	0.52	1.188	30.4%	1.1%	76.2%

SequenceG (5691/5692)	0.38	1.185	30.0%	0.9%	77.9%
SequenceM (5195/5196)	1.56	1.256	31.8%	2.7%	80.8%
SequenceH (5285/5286)	1.08	1.253	31.4%	2.8%	82.5%
SequenceI (5299/5300)	0.37	1.243	31.1%	4.0%	80.7%
SequenceJ (5701/5702)	1.99	1.320	34.2%	3.0%	82.8%

Table 40.2. Performance metrics of engineered transposases in <i>Bacillus subtilis</i> (44% GC Content)					
Transposase SEQ ID NO: (nt/aa)	Norm GC (slope)	fold80 Penalty	CV of Coverage	% Duplication	Unique Insert Positions
Tn5	0.77	1.307	32.2%	4.2%	70.2%
SequenceA (971/972)	1.49	1.561	46.9%	10.3%	49.7%
SequenceB (2367/2368)	0.61	1.182	24.4%	4.2%	69.0%
SequenceC (2979/2980)	0.89	1.192	24.8%	5.1%	76.4%
SequenceD (4143/4144)	2.33	1.239	26.9%	4.3%	79.1%
SequenceE (4641/4642)	2.13	1.187	25.1%	1.0%	84.5%
SequenceF (4933/4934)	0.24	1.197	25.3%	1.3%	76.0%
SequenceG (5691/5692)	1.31	1.177	24.0%	1.2%	78.1%
SequenceM (5195/5196)	2.14	1.239	25.3%	3.8%	81.1%
SequenceH (5285/5286)	1.52	1.238	24.6%	3.2%	83.3%
SequenceI (5299/5300)	2.42	1.234	25.9%	3.3%	82.4%

SequenceJ (5701/5702)	1.13	1.218	25.4%	3.5%	83.4%
--------------------------	------	-------	-------	------	-------

Table 40.3. Performance metrics of engineered transposases in <i>Pseudomonas aeruginosa</i> (66% GC Content)					
Transposase SEQ ID NO: (nt/aa)	Norm GC (slope)	fold80 Penalty	CV of Coverage	% Duplication	Unique Insert Positions
Tn5	1.75	1.277	30.9%	5.40%	69.2%
SequenceA (971/972)	7.34	1.664	53.4%	10.60%	49.7%
SequenceB (2367/2368)	0.32	1.195	26.0%	5.10%	67.1%
SequenceC (2979/2980)	0.76	1.195	25.5%	5.70%	75.4%
SequenceD (4143/4144)	0.68	1.212	26.4%	5.40%	78.0%
SequenceE (4641/4642)	2.07	1.183	23.7%	1.70%	83.9%
SequenceF (4933/4934)	0.34	1.217	28.2%	3.40%	72.2%
SequenceG (5691/5692)	0.53	1.191	26.3%	1.70%	76.8%
SequenceM (5195/5196)	1.08	1.276	32.3%	6.36%	84.1%
SequenceH (5285/5286)	0.39	1.232	32.3%	5.77%	81.9%
SequenceI (5299/5300)	1.31	1.218	33.0%	5.68%	81.5%
SequenceJ (5701/5702)	-0.84	1.238	33.5%	6.79%	82.1%

- Results from libraries generated from the low GC *Staphylococcus epidermidis* (32% GC) genome (Table 40.1) indicated that SequenceE and SequenceH have the least bias for high AT sequences. SequenceC and SequenceJ performed best from libraries generated from the more balanced *Bacillus subtilis* (44% GC) genome (Table 40.2), and SequenceM performed best from libraries generated from the high GC *Pseudomonas aeruginosa* (66% GC) (Table 40.3).

The fold-80 penalty measures the amount of excess sequencing required to bring 80% of covered bases to the median level of coverage. A completely unbiased system would return a score of one. For all

3 bacterial genomes (**Tables 40.1-40.3**), all variants with the exception of SequenceA performed better than commercial hyperactive Tn5, indicating generally higher complexity and more uniform coverage libraries from these variants.

Another uniformity of coverage metric, CV of coverage, confirmed the fold-80 penalty results.

- 5 Here, all variants with the exception of SequenceA for all bacterial genomes (**Tables 40.1-40.3**), and SequenceM, SequenceH, SequenceI, and SequenceJ for libraries generated from the high GC *Pseudomonas aeruginosa* (66% GC) genome performed better than hyperactive Tn5.

Library complexity can be measured by calculating the percentage of reads having a unique insert starting or ending genomic coordinate.

10

EXAMPLE 41

Comparison of Sequencing Performance of Different Engineered Transposases in Human WGS

- 15 Engineered transposases were compared based on their performance in generating libraries for human whole genome sequencing. Transposase activity of the different engineered transposases was determined according to the protocol in Example 38. Loaded transposase library reagents were prepared according to the protocol described in Example 37 using a distinct combination of Illumina-compatible i5/p5 and i7/p7 barcoded adapters for transposase loading.

- 20 Transposition reactions were performed in triplicate for each of the engineered transposases and hyperactive Tn5 control, and Illumina-compatible libraries were generated in the manner described from 100 ng of human NA12878 genomic DNA for each condition tested. Human DNA was selected to represent a high complexity genomic content to enable analysis of transposition insertion sites and to measure the impact on coverage uniformity and library complexity metrics.

- 25 The resulting libraries were pooled and sequenced together on an Illumina NovaSeq X instrument, and the resulting data were demultiplexed and analyzed by reference-based mapping and analysis. Resulting summary statistics were collected for each transposase (e.g., engineered transposase or hyperactive Tn5 control) and are reported below.

- 30 Metrics Collected after down sampling to 620 million reads per sample:

- Library size – The estimated library size after de-duplication.

Picard: GC Coverage Bias – The relative proportions of guanine (G) and cytosine (C) nucleotides in a sample

Picard WGS Coverage – The coverage in bases of the genome territory, after all filters are applied.

- 35 Per base sequence content – Heatmap of the per base sequence content at the transposition insertion site

Transposase Performance:

- 40 The estimated sizes of the NGS libraries prepared by the different commercial transposases and engineered transposases are shown in **Figure 7**.

The GC coverage bias of the NGS libraries prepared by the different commercial transposases and engineered transposases are shown in **Figure 8**.

The WGS coverage of the NGS libraries prepared by the different commercial transposases and engineered transposases are shown in **Figure 9**.

5 The per base sequence content of the transposition sites of the different commercial transposases and engineered transposases are shown in **Figure 10**.

Each nucleotide (A, T, G, or C) in **Figure 10** is represented by a single line trace, representing the percentage of reads having that nucleotide at the corresponding position relative to the read start site (transposition site) of the DNA molecules in the sequenced library. Engineered Transposases SequenceK and SequenceH showed improvements in the uniform representation of nucleotides at positions where alternate transposases (Illumina DNA Prep and commercial hyperactive Tn5 Transposase) have specific nucleotides that are overrepresented. For example, commercial hyperactive Tn5 shows a higher bias and percentage of reads with "C" at position 9 of sequenced reads, than either enzyme SequenceK or SequenceH.

15

EXAMPLE 42

Evaluation of Engineered Transposases in ExpressPlex for Bacterial WGS

20 To compare the performance of engineered transposases in alternate buffer systems that are less optimal for transposase activity, engineered transposases were tested in the context of the ExpressPlex assay (seqWell, Inc.). Experimental tagging reagents were prepared by binding full-length sequencing adapters to transposase proteins (engineered transposases and control hyperactive Tn5 transposase). Then, libraries were prepared from different bacterial genomic DNA isolates using these tagging reagents, as follows:

25 1. 4 μ L of DNA and 4 μ L of the tagging reagent were added to 8 μ L of standard Ready Reaction reagent (seqWell, Inc.) in 96-well PCR plates.

2. Next, the 96-well plate was sealed and placed in a thermal cycler. The following thermal cycling program (**Table 42.1**) was used to tag the DNA and amplify the libraries:

Table 42.1. Thermocycler Programming		
5 min	55 °C	Tagging
5 min	75 °C	Fill-in/Heat-kill
3 min	68 °C	
5 min	79 °C	
3 min	68 °C	

5 min	83 °C	
3 min	98 °C	Denaturation
15 sec	98 °C	Amplification 12 cycles
30 sec	64 °C	
1 min	72 °C	
5 min	72 °C	Final extension
Hold	4 °C	

3. After completion of the thermal cycling program, 10 µL of each reaction was pooled together in a single tube and purified with 0.75X MAGwise paramagnetic beads (seqWell, Inc.).

4. After elution of the purified library from the beads, the DNA concentration was determined by electrophoresis on a TapeStation system (Agilent).

5. The library was denatured and loaded onto a MiSeq sequencing system (Illumina), and run in the 2 x 151 base read configuration.

6. The output sequencing data were demultiplexed and aligned to their respective DNA reference sequences, and assessed for uniformity of coverage.

GC bias was evaluated for NGS libraries prepared by ExpressPlex using hyperactive Tn5 transposase and the SequenceH engineered transposase from *Staphylococcus epidermidis*, *Bacillus subtilis*, *Escherichia coli*, and *Pseudomonas aeruginosa*. Results are shown in **Figure 11**.

EXAMPLE 43

Evaluation of Mutant Transposase Variants in ExpressPlex for Plasmid Sequencing

To compare the performance of transposase variants in alternate buffer systems that are less optimal for transposase activity, engineered transposases were evaluated in the context of the ExpressPlex assay (seqWell, Inc.). Experimental tagging reagents were prepared by binding full-length sequencing adapters to transposase proteins (engineered transposases and control hyperactive Tn5 transposase). Then, libraries were prepared from purified plasmid samples and PCR products derived from lambda DNA, using these tagging reagents, as follows:

1. 4 µL of DNA and 4 µL of the tagging reagent were added to 8 µL of standard Ready Reaction reagent (seqWell, Inc.) in 96-well PCR plates.

2. Next, the 96-well plate was sealed and placed in a thermal cycler. The following thermal cycling program (**Table 43.1**) was used to tag the DNA and amplify the libraries:

Table 43.1. Thermocycler Programming		
5 min	55°C	Tagging
5 min	75°C	Fill-in/Heat-kill
3 min	68°C	
5 min	79°C	
3 min	68°C	
5 min	83°C	
3 min	98°C	Denaturation
15 sec	98°C	Amplification 12 cycles
30 sec	64°C	
1 min	72°C	
5 min	72°C	Final extension
Hold	4°C	

3. After completion of the thermal cycling program, 10 μ L of each reaction was pooled together in a single tube and purified with 0.75X MAGwise paramagnetic beads (seqWell, Inc.).

4. After elution of the purified library from the beads, the DNA concentration was determined by electrophoresis on a TapeStation system (Agilent).

5. The library was denatured and loaded onto a MiSeq sequencing system (Illumina), and run in the 2 x 151 base read configuration.

6. The output sequencing data were demultiplexed and aligned to their respective DNA reference sequences, and assessed for uniformity of coverage and GC bias.

10

Result of the uniformity of coverage and the GC bias analysis in raw data and analyzed as the coefficient of variation (CV) is found in **Figure 12** (Lambda amplicon 18) and **Figure 13** (pUC19). As measured by CV, the variability in coverage produced by engineered transposases SequenceE, SequenceH, and SequenceK produced superior uniformity of coverage (lower CV) versus commercial hyperactive Tn5 transposase. Because the ExpressPlex reaction conditions were optimized for both transposase and PCR polymerase activity, these results also demonstrated the improved robustness of engineered transposases SequenceE, SequenceH, and SequenceK when tested under altered reaction conditions.

15

OTHER EMBODIMENTS

While the invention has been described with reference to the specific embodiments, various changes can be made and equivalents can be substituted to adapt to a particular situation, material, composition of matter, process, process step or steps, thereby achieving benefits of the invention without
5 departing from the scope of what is claimed.

For all purposes in the United States of America, each and every publication and patent document cited in this disclosure is incorporated herein by reference as if each such publication or document was specifically and individually indicated to be incorporated herein by reference. Citation of publications and patent documents is not intended as an indication that any such document is pertinent
10 prior art, nor does it constitute an admission as to its contents or date.

WHAT IS CLAIMED IS:

1. An engineered transposase comprising a polypeptide sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, or a functional fragment thereof, wherein the engineered transposase comprises at least one substitution or substitution set in its polypeptide sequence, and wherein the amino acid positions of the polypeptide sequence are numbered with reference to SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

2. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 10, 13, 24, 35, 39, 44, 46, 51, 54, 55, 62/361, 72, 75, 82, 89, 92, 107, 128, 137, 164, 167, 177/359, 191, 192, 193, 194, 201, 217, 281, 289, 291, 291/340, 305, 309, 310, 317, 331, 337, 340, 346, 352, 353, 357, 358, 359, 362, 371, 396, 414/435, 415, 428, 437, 443, 447, and 448, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2.

3. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 4, 5, 6, 8, 10, 10/317, 10/317/358, 10/317/359, 10/346, 10/346/359, 10/358, 10/359, 12, 12/373, 13, 15, 16, 18, 19, 26, 30, 30/154, 31, 34, 35, 37, 38, 68, 70, 73, 74, 93, 106, 118, 126, 134, 135, 143, 146, 154/453, 163, 182, 207/286, 208, 226, 237, 246, 247, 259, 262, 263, 264, 280, 282, 286, 287, 288, 289, 291, 292, 293, 298, 305, 309, 310, 313, 317, 317/346, 317/359, 317/445, 320, 346, 351, 355, 356, 358, 359, 360, 368, 369, 410, 411, 422, 426, 427, 428, 446, 448, 453, and 454, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4.

4. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18/73/74/259, 18/73/286/289/411, 18/74/118/259/286/289/411, 18/74/259, 18/74/259/286/289, 18/74/259/286/411, 18/118/259, 18/259/286/289/411, 18/259/289/411, 18/289/411, 73/74/259/286/289, 73/74/259/289, 73/259/286, 74/259, 74/259/289, 74/286/289/411, 259, 259/286/289, and 289, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 156.

5. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 16/19/355/427, 16/37/146/286/287/289/310/313/355, 16/146/226/287/289/355, 19/37/146/286/287/310/355,

19/135/286/287/289/355, 19/182/355, 37/70/226/310/313/355, 37/74/310/355/427, 37/135/146/182/226/286/287/355/368, 37/146/226/289/427, 37/146/287/289/427, 38, 41, 42, 43, 44, 45, 47, 48, 50, 53, 54, 55, 70/74/226/286/287/289/368/427, 70/74/355/368, 70/135/146/150/182/286/289/313/427, 74/310/313/355, 74/310/355, 74/355/427, 101, 103, 117, 135/226/286/289/355/368, 135/355, 146, 146/150/287/289/310/355, 146/182/355, 146/226/286/355, 146/286/287/289/355/427, 146/289, 146/427, 150/226/310/313, 182/226/355, 182/355, 244, 247, 249, 251, 287/355/427, 300, 317/411, 355, 393, 397, 398, 399, 401, 405, 407, 408, 409, 411, 413, 426, 429, 431, 439, 444, 446, 450, 453, 454, 457, 459, 460, 464, 466, 467, 469, 470, 471, 472, 474, 475, 476, and 478, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 302.

6. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 16, 16/313, 16/409/411/427, 37/150/408, 37/313/317/408/411, 37/317/405/408, 37/317/405/426, 37/405/409/427, 150, 150/313/405/411, 150/405/406, 150/408, 150/408/411, 226/405, 317/460, and 408/426, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 338.

7. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 12, 13, 18, 19, 22, 26, 31/262, 34, 60, 79, 80, 81, 82, 92, 106, 107, 113, 115, 132, 134, 136, 140, 143, 152, 160, 163, 167, 168, 175, 179, 183, 195, 199, 201, 202, 215, 216, 218, 220, 223, 226, 233, 235, 236, 237, 243, 255, 259, 261, 262, 266, 268, 270, 273, 282, 290, 292, 293, 309, 319, 320, 326, 343, 346, 365, 370, 374, 380, 382, 386, 389, 390, 408, 412, 414, 415, 424, and 442, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

8. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 2, 12, 13, 15, 16, 18, 19, 22, 26, 30, 31/262, 34, 51, 60, 69, 79, 82, 84, 87, 88, 106, 107, 113, 114, 115, 128, 132, 134, 135, 136, 140, 143, 146, 150, 152, 160, 163, 167, 168, 175, 179, 183, 195, 198, 199, 201, 215, 217, 218, 219, 220, 223, 224, 226, 227, 231, 233, 235, 236, 243, 254, 255, 259, 261, 266, 267, 268, 270, 272, 273, 275, 277, 279, 282, 287, 289, 290, 291, 292, 293, 304, 307, 309, 319, 320, 343, 344, 346, 355, 365, 370, 373, 380, 382, 386, 389, 390, 408, 412, 414, 415, 418, 423, 424, and 468, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

9. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 44/45/47/408/453, 44/45/47/453, 44/45/47/453/454, 44/45/103/453/457, 44/45/453/454/457, 44/47/103/393/408/453/454, 44/47/103/453/454/457, 44/47/393/408/453/457, 44/47/393/453/457, 44/103/408/453/454, 45/47/103/393/453/454/457, 45/47/393/453/454, 45/47/408/423/453/457, 45/47/453, 45/103/393/453/454, 45/453/454, 47/103/408/453/454, 47/393/408/453/454/457, 47/393/453/454, 47/408/453/457,

53/54/244/450, 103/393/408/453/457, 244/450, 408/453/454, 446/450/464, and 453/457, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

10. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 38/45/53, 38/45/53/54/476, 38/45/53/244, 38/45/53/470/472, 38/45/54/244, 38/45/54/470/474/478, 38/45/472/476/478, 38/45/476, 38/53/54, 38/53/54/244, 38/53/54/244/472/478, 38/53/472, 38/54/244/470/474/476/478, 38/244/470/472/474/478, 38/478, 45/53/54/244/470/474/476/478, 45/53/54/470/474/476, 45/53/54/472/478, 45/54/244/474/478, 53, 53/54/470/476/478, 53/244/474/478, 54, 54/244, 54/244/472/476, 54/472, 244, and 474/478, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

11. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 22/38/45/107/259/317/412, 22/38/45/259/317/412, 22/107, 22/107/317/412, 22/134, 22/134/220, 22/134/220/317, 22/134/317/412, 22/317, 38, 38/45/81/143/244, 38/53/243, 38/81, 38/81/244, 38/107/134/220, 38/132, 38/220/317, 45/81, 45/81/243, 45/81/243/343, 45/143/243/270/343, 45/244/270, 53/115/163/243/270/343, 53/143/243/270/343, 53/270, 64/134/220/317/412, 81/132/243, 107, 107/134/317, 115/244/270, 134/259, 143/243/270, 243, 243/270, 244, 244/270, 259, 270, and 317, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

12. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 22/38/45/107/259/317/412, 22/38/45/259/317/412, 22/38/64/134/220, 22/45/220/259, 22/64/107/259/317/412, 22/64/134/220, 22/107, 22/134, 22/134/317, 22/134/317/412, 22/220, 22/220/412/415, 22/259/317, 22/259/317/412, 22/317, 38, 38/45/81/143/244, 38/45/143/270, 38/53/243, 38/81, 38/81/143/243/270, 38/81/183/243/270, 38/81/243, 38/81/244, 38/107/134/220, 38/132, 38/220/317, 38/270/343, 45/81, 45/81/132/243/270, 45/81/243, 45/143/243/270/343, 45/244/270, 53/143/243/270/343, 53/244/270, 53/270, 64, 64/107/134, 64/107/317, 64/134, 64/134/220/317/412, 81/132, 81/132/243, 107, 107/134/317, 107/412, 115/244/270, 143/243/270, 220/259, 243, 243/270, 244/270, 255, 259, and 270, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

13. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 143, 143/175/272/273/415, 152/168/216/237, 152/168/423, 160, 160/201/272, 160/201/273/415, 160/226, 160/244/272/273/415, 160/415, 168, 168/187/216/255/423, 168/216/237, 168/255/423, 168/290/423, 187, 187/423, 199/201/226, 199/201/415, 201, 216, 216/423, 226/244/415, 226/415, 237, 237/290, 244, 255, 272, 290, 290/423, 415, and 423, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

14. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 143, 143/160/226/415, 143/175/272/273/415, 152/168/216/237, 152/168/423, 160, 160/201/272, 160/201/273/415, 160/226, 160/244/272/273/415, 160/415, 168, 168/187/216/255/423, 168/216/237, 168/216/237/255/343, 168/255/423, 168/290/423, 187, 199/201/226, 199/201/415, 201, 216, 216/423, 226/244/415, 226/415, 237, 237/290, 244, 255, 255/343, 272, 290, 290/423, 415, and 423, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

15. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 72, 90, 91, 92, 93, 95, 122, 129, 141, 142, 143, 144, 145, 150, 151, 151/155, 152, 154, 155, 157, 160, 161, 161/313, 164, 165, 168, 180, 181, 184, 193, 197, 204, 205, 205/472, 206, 299, and 320, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

16. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 90, 92, 122, 129, 141, 143, 144, 145, 150, 151, 151/155, 152, 154, 155, 155/265, 157, 160, 161, 164, 165, 168, 180, 181, 184, 193, 197, 204, 205, 205/472, 206, 299, and 320, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

17. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 129, 129/160, 129/160/164, 129/160/184, 129/164, 129/164/184, 129/205, 146/152/244, 150/155/226, 150/155/244, 152, 154, 154/160/184, 154/164, 154/205, 155, 160, 160/164, 160/164/184, 160/184, 164, 164/204, 164/205, 204, 205, 226, 226/244, and 244, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1218.

18. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 129, 129/143, 129/154/164, 129/154/164/205, 129/154/205, 129/160, 129/160/164, 129/160/205, 129/164/205, 143, 143/150/155/226/244, 143/150/155/244, 143/150/226, 143/150/226/244/343, 143/150/226/343, 143/150/244, 143/150/343, 143/152, 143/152/226, 143/152/226/244, 143/152/226/343, 143/152/343, 143/155, 143/155/226/244, 143/155/226/244/343, 143/155/343, 143/226, 143/226/244/343, 143/226/343, 143/244, 143/244/343, 143/343, 146/152/244, 150/155/226, 150/155/244, 150/155/244/343, 150/226/244/343, 152, 152/174/244/343, 152/226, 152/226/343, 152/343, 154, 154/160/184, 154/164, 154/164/204, 154/164/204/205, 154/205, 155, 160, 160/164, 160/164/184, 164, 205, 226, 226/244, and 244, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1218.

19. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 98/129/157/168, 129/157/168, 151, 151/152/255, 151/152/273, 151/255, 151/343, 157, 157/164, 157/164/168, 157/168, 164, 164/255, 168, 183, 255, 255/343, and 343, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

20. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 129, 129/157, 129/157/168, 129/164, 129/164/168/183, 129/164/183, 143, 143/151, 143/151/152, 143/151/164, 143/164, 143/164/255, 143/164/343, 143/255, 143/343, 151, 151/152, 151/152/164/343, 151/152/343, 151/255, 151/255/343, 151/343, 152, 152/255, 157, 157/164/168, 157/168, 164, 164/168, 164/255, 164/447, 168, 183, 255, 255/343, 298/343, and 343, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

21. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 69, 69/164, 72, 73, 91, 92/164, 95, 96, 97, 98, 110, 112, 113, 121, 121/164, 122, 122/164, 124, 126, 127, 139, 140, 141, 144, 144/164, 145, 146, 147, 148, 150, 150/164, 151, 151/164, 153, 153/164, 155, 156, 157, 159, 160, 161, 163, 164, 164/172, 164/173, 164/182, 164/188, 164/191, 164/192, 164/195, 164/197, 164/199, 164/202, 164/205/249, 164/209, 165, 166, 167, 169, 170, 171, 172, 173, 177, 184, 187, 188, 190, 191, 192, 193, 194, 195, 197, 197/266, 198, 199, 200, 202, 203, 204, 205, 208, 209, 210, 321, 324, 331, and 349, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

22. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 69, 69/164, 72, 73, 91, 92/164, 95, 97, 98, 121/164, 126, 127, 139, 140, 142, 144, 144/164, 145, 146, 146/164, 148, 150, 150/164, 151, 153, 153/164, 155, 156, 157, 160, 163, 164/173, 164/188, 164/191, 164/192, 164/195, 164/199, 164/205/249, 164/209, 165, 166, 167, 169, 171, 172, 173, 177, 182, 187, 188, 190, 191, 192, 194, 195, 197/266, 198, 199, 200, 202, 203, 204, 205, 208, 209, 210, 324, 331, and 349, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

23. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73, 91, 93, 95, 97, 98, 110, 112, 122, 122/164, 124, 127, 139, 144, 145, 146, 146/164, 147, 148, 150, 150/164, 151, 151/164, 153, 153/164, 155, 156, 157, 159, 160, 161, 162, 163, 164, 164/182, 164/188, 164/191, 164/192, 164/195, 164/199, 164/205/249, 164/209, 165, 166, 167, 169, 171, 172, 177, 184, 187, 188, 190, 191, 192, 193, 195, 197/266, 198, 199, 200, 204, 205, 206, 207, 208, 209, 210, and 324, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

24. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 69, 73, 91, 127, 140, 144, 145, 146, 146/164, 150, 150/164, 151, 153, 155, 156, 157, 160, 163, 164, 164/172, 164/188, 164/191, 164/192, 164/195, 164/199, 164/209, 165, 167, 169, 171, 172, 182, 183, 188, 191, 192, 193, 195, 196, 198, 199, 200, 204, 205, 208, 210, and 324, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

25. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 38/143/168/220/290/317/423, 38/152/154/164/168/220/290/317/423, 38/152/168/220/290/317/343/423, 38/152/168/220/290/317/423, 38/154/164/168/220/290/317/423, 38/168/220/290/317/423, and 38/220/317, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

26. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73/143/169/191/193, 73/156, 73/156/192/435, 73/156/210, 73/163, 73/163/210, 73/192/193, 127, 127/160/172/198, 127/172, 127/198, 146, 146/151/188, 146/324, 151/171/324, 151/324, 156/169, 156/192, 157/171, 163, 163/169, 163/169/191/192/193, 165/171/324, 165/188/324, 165/324, 169, 171/188/324, 171/324, 172, 188, 192, 192/193, 198, 210, and 324 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2368.

27. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73, 73/163/191/192/193, 127/160/172, 127/160/172/198, 144, 160, 160/172, 163/191/192, and 192 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2368.

28. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 8, 13, 17, 18, 20/146/157/167, 26, 34/38, 38/39/45, 67, 70, 73, 73/150/155/169, 73/150/171/192, 73/155, 73/155/169/199, 73/155/171, 73/192, 73/199, 78, 82, 84, 86, 86/384, 89, 90, 91, 95, 102, 110, 111, 113, 114, 115, 122/127, 126/127, 127/131/165, 127/133, 143, 143/146, 143/146/153, 143/146/157, 143/146/167, 143/146/167/188, 143/146/188, 143/153/167/188/246, 143/156/157/167/195, 143/156/188, 143/188, 143/188/195, 143/195, 144, 145, 145/151, 146, 146/152, 146/153/157, 146/153/157/167, 146/153/167, 146/157, 146/157/167, 146/167, 150, 150/155/169, 150/155/169/192, 150/155/192, 150/169, 150/192, 150/199, 151/152/154, 152/154, 152/154/155/160, 153/157/167, 155/171, 155/192, 156/270, 160/371, 167/188, 168/172, 169, 184, 190, 192/198, 198, 206, 208, 209, 210, 214/220, 215/220, 216/220, 217/220, 219/220, 220, 220/222, 220/225, 220/226, 231, 235, 242, 257, 259, 261, 262, 267, 268, 269, 270, 272, 274, 275, 276, 277, 278, 282, 283/290, 284/290, 287/290, 288/290, 289/290, 290, 290/291, 290/292, 290/293, 290/294, 304, 305, 305/373, 306, 307, 309, 317, 329, 333, 334, 339, 341, 344, 345, 346, 352, 353, 355, 358, 366, 368, 370, 372, 373, 375, 376, 378, 379, 380, 381, 382, 383, 384, 385, 386/393, 387/393, 389/393, 390/393,

391/393, 392/393, 393/394, 414, 415, 420/423, 422/423, 423, 423/425, 432, 440, 441, 443, 469, and 472 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2426.

29. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 8, 18, 25, 29, 30, 67, 73, 73/150/155/169, 73/150/155/171, 73/150/171/192, 73/171, 73/192, 73/199, 74, 82, 86, 86/384, 113, 115, 118, 127/128, 143, 150, 150/155/169, 150/155/169/192, 150/155/171, 150/155/171/199, 150/155/192, 150/155/199, 150/192, 150/199, 155/192, 168/171/172, 195/198, 197/198, 198/201, 198/203, 198/205, 235, 252, 256, 257, 262, 263, 267/343, 268, 269, 270, 272, 273, 276, 277, 288/290, 289/290, 290, 290/291, 290/292, 290/294, 298, 304, 305, 307, 333, 346, 355, 366, 370, 386/393, 389/393, 393/395, 440, and 472 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2426.

30. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 66, 67, 70, 101, 102, 103, 104, 106, 107, 108, 113, 114, 115, 117, 118, 143, 152, 154, 235, 236/251, 237, 238, 242, 243, 246, 247, 249, 250, 251, 256/352, 257, 259, 285, 286, 287, 288, 289, 290, 292, 293, 294, 295, 296, 322, 323, 326, 329, 330, 333, 339, 340, 343, 344, 345, and 352. and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2540.

31. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18, 18/113/270, 18/113/270/329/370, 18/113/270/370/384, 18/113/270/370/412, 18/113/329/384, 18/259/370/384, 18/270, 18/270/370, 18/270/370/384, 18/370, 18/370/384, 73, 73/115, 73/115/184/333, 73/115/184/341, 73/115/256, 73/184/256, 73/184/273/333, 73/184/333, 73/256, 73/256/375, 73/273, 73/273/333/375, 73/273/375, 73/333, 73/333/375, 73/375, 73/379, 77/233/238/376, 77/233/334/376, 77/262/376, 77/276, 77/334/376, 82, 82/127, 82/127/144/235/290, 82/127/144/266, 82/127/144/266/290, 82/127/144/378, 82/127/235, 82/127/290, 82/144/266/290/378, 82/144/266/378, 82/235/266, 82/235/266/378/383, 82/235/290, 82/235/378, 82/266, 82/266/378, 82/378, 113/270/370/384, 115/256/333/375, 115/256/375, 115/273/375, 115/375, 127/144/290, 127/149/150, 127/266/290/378, 127/378/383, 144/235/290, 144/266/290, 233/334, 233/376, 235, 235/290, 256/333, 262, 266/378, 270/370, 273, 276/352, 290, 290/378, 333, 358, 370, 375, and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2540.

32. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18, 18/113, 18/113/252, 18/113/252/336/384, 18/113/252/370, 18/113/270, 18/113/270/370/384, 18/259/370/384, 18/270,

18/270/370, 18/270/370/384, 18/370, 18/370/384, 73, 73/115/184/341, 73/115/341, 73/131/256/273/333/341/375, 73/184/256, 73/184/256/341, 73/184/256/375, 73/184/273/333, 73/184/333, 73/184/341/375, 73/256, 73/256/273, 73/256/273/375, 73/256/333/375, 73/256/375, 73/273, 73/273/333/341, 73/273/333/375, 73/273/375, 73/333, 73/333/341/375, 73/333/375, 73/341, 73/375, 73/379, 77, 77/118, 77/118/143/262/276/334/376, 77/118/143/334/352/376, 77/118/143/376, 77/118/233, 77/118/233/352/376, 77/118/262/334/376, 77/118/334, 77/143/262/276, 77/143/262/334, 77/143/376, 77/233/238/376, 77/233/262, 77/233/276/376, 77/233/334/376, 77/262/376, 77/276, 77/334/376, 77/352, 82, 82/127, 82/127/144/235/290, 82/127/144/266, 82/127/144/266/290, 82/127/144/378, 82/127/235, 82/127/235/266/358, 82/127/290, 82/127/290/358, 82/127/290/378, 82/144/266/290/378, 82/144/266/378, 82/150/235/266, 82/235/266, 82/235/266/290, 82/235/266/378/383, 82/235/290, 82/235/378, 82/266, 82/266/378, 82/290, 82/378, 113, 113/252, 113/252/270/370, 113/270/370/384, 113/336, 118/233/276/334, 118/334, 127/235/266, 127/266/290/378, 127/378/383, 143, 143/276, 143/276/334, 144/235/290, 144/266/290, 144/290/358, 233/334, 233/376, 235, 235/266/290/358/383, 235/358, 252/329/336/370, 252/384, 256/333, 256/375, 262, 262/276, 262/334, 266/290, 266/378, 270/370, 273, 276/352, 290, 290/378, 333, 334, 334/376, 336/370, 341, 358, 370, 375, and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2540.

33. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18, 18/77/118/169/233/238/256/266/270, 18/77/150/256/376, 18/118, 18/118/143/169/238/256/266/333, 18/118/150/169/184/238/256/270/378, 18/118/169, 18/118/233/238, 18/143, 18/143/256, 18/150/169, 18/150/266, 18/169/184/290/333/376, 18/169/233/333, 18/169/392, 18/184/270, 18/233/256/334, 18/256/266/290/378, 77/113/118/256/266/290, 77/118/290, 77/169/270/290, 77/233, 77/233/256, 77/238, 113/118/169/184/256/376, 113/169, 113/233, 118, 118/143/238, 118/233/238/256/378, 118/233/238/270, 143/169/233/270, 143/233/266/290/376, 143/238/270/290/378, 143/256/334, 143/290, 169, 169/256/266, 183/270/333, 233/238, 233/238/376, 233/392, 238/256, 256/290/303, 270, 290, 376, and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3402.

34. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18, 18/77/118/169/266/334/376, 18/77/150/256/376, 18/77/256/290/334, 18/113/238, 18/118, 18/118/143/169/238/256/266/333, 18/118/143/169/376, 18/118/169, 18/118/233/238, 18/118/266, 18/143, 18/143/169/233/238/358, 18/143/233/238/290/376, 18/143/256, 18/143/290, 18/143/333/334/378, 18/150/169, 18/150/266, 18/169/184/290/333/376, 18/169/233/333, 18/169/238/270/334/376, 18/169/392, 18/233/256/334, 18/233/266/290/376, 18/256/266/290/378, 18/266/270/378, 18/378, 77/113/121/143/233/334, 77/118/169/233/266/270, 77/118/238/376, 77/118/290, 77/169/270/290, 77/233, 77/238, 113/238/376, 118, 118/143/238, 118/143/392, 118/169/184/233/266/270/333, 118/233/238/256/378,

118/256/290/333/334/376/378, 118/256/334/376, 118/266/290, 143/169/233/270, 143/233/266/290/376, 143/238/270/290/378, 143/256/334, 143/290, 143/364, 150/233/333, 169, 169/256/266, 183/270/333, 233/238, 233/238/376, 233/392, 238/256, 238/266/270/378, 256/266/334, 266/270/376, 266/333, and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3402.

35. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 77/156, 77/156/270, 77/184/244, 77/184/333, 77/244/270, 77/270, 143, 143/169/238/257/344, 143/169/255, 143/169/255/257, 143/169/343, 143/255, 143/255/257, 143/255/344, 143/257, 156, 156/270, 169/238, 169/238/255/257/344, 169/257/344, 169/344, 184, 184/244/324, 238, 238/255/344, 244/270, 257/344, and 270 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3662.

36. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 77, 77/156/184/244/324, 77/184/270, 143, 143/238, 143/238/255, 143/238/255/343, 143/238/255/392, 143/238/344, 143/255, 143/257/392, 143/344, 169, 169/238/255/344, 169/238/344, 169/255, 169/343, 169/392, 238/344, 255, 255/257, 255/257/392, 257, and 344 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3662.

37. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 20, 22, 26, 42, 43, 44, 45, 50, 51, 52, 53, 54, 55, 60, 62, 64, 65, 95, 97, 99, 103, 104, 111, 112, 115, 115/143/169, 115/169/343, 115/343, 116, 116/343, 143, 143/169, 154/439, 169, 169/324, 169/324/343, 169/343, 186, 234, 236, 238, 238/246, 239, 239/244, 241, 242, 243, 244, 250, 251, 254, 255, 260, 283, 284, 295, 296, 298/326, 319, 322, 323, 324, 324/343, 326, 328, 329, 333, 334, 338, 340, 341, 343, 344, 347, 366, 434, 439, 442, 443, 444, 445, 446, and 447 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3788.

38. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 26, 26/236, 26/236/260/333, 26/236/295, 26/260, 26/260/295, 26/260/295/333, 26/295/319, 26/333, 64, 64/244, 64/251/260/347, 64/260/295/333, 64/260/347, 103/115/255, 103/115/343, 103/115/343/344, 103/143/255/284/343/344, 103/169/343/344, 103/255/284, 103/343/344, 115, 115/143/255, 115/143/255/343, 115/143/284, 115/143/343, 115/143/344, 115/169, 115/169/343, 115/255, 115/255/344, 115/284, 115/284/343/344, 115/284/344, 115/343, 115/343/344, 115/344, 143, 143/169/255/284/343/344, 143/169/343/344, 143/255,

143/255/284/343/344, 143/255/343, 143/255/343/344, 143/255/344, 143/343/344, 169, 169/255, 169/255/284, 169/255/343, 169/284, 169/343, 169/344, 236, 236/244/333, 236/260, 236/347, 244, 244/251, 244/251/319, 244/260, 244/295, 244/347, 251, 251/260/295/347, 251/333, 255, 255/284, 255/343, 255/343/344, 255/344, 259, 260, 260/295, 260/333, 284, 284/343/344, 284/344, 295, 295/333, 333, 343, 343/344, 344, and 347 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4150.

39. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 115, 115/117, 115/117/118, 115/117/118/152, 115/117/118/169, 115/117/118/246, 115/117/152, 115/117/152/169, 115/117/152/169/246, 115/117/152/246, 115/117/169, 115/117/246, 115/118, 115/118/152, 115/118/152/169, 115/118/152/169/246, 115/118/169, 115/118/246, 115/152, 115/152/169, 115/152/169/246, 115/152/246, 115/152/246/374, 115/169, 115/169/246, 115/246, 117, 117/118, 117/118/152, 117/118/152/169, 117/118/152/246, 117/118/169, 117/118/246, 117/152, 117/152/169, 117/152/246/288, 117/169, 117/246, 118/246, 152, 152/169, 152/169/246, 152/246, 169, 169/246, and 246 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4278.

40. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73, 101, 114, 115, 118, 143, 145, 150, 151, 152, 153, 154, 156, 157, 160, 161, 163, 166, 167, 168, 169, 171, 172, 190, 191, 192, 195, 213, 214, 218, 219, 223, 225, 233, 233/343, 234, 235, 236, 237, 238, 239, 239/250, 243, 244, 249, 251, 254, 255, 257, 258, 259, 261, 262, 262/283, 277/292, 283, 284, 286, 289, 292, 295, and 296 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4278.

41. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 93, 113, 113/114, 113/115, 113/143, 115/118, 115/177, 118/143/163/168/195/255/315, 118/143/255, 118/163/168/214/255/315, 118/168/214/315, 118/255, 124, 126, 129, 134, 143/154/255, 143/156/168/216/255, 143/163, 143/163/255, 143/168, 143/168/169/255/315, 143/255, 148, 156/214/315, 159, 163, 163/168, 163/168/169, 163/168/169/255/315, 163/168/169/315, 163/168/195/214/255/315, 163/214, 163/255, 163/255/315, 168/315, 186, 194, 197, 206, 207, 208, 210, 213, 255/315, 277, 293, 294, 315, 317, 320, 322, 328, 329, 330, 333, 352, 366, 392, and 445 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4642.

42. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 127, 128, 129, 140, 180, 182, 194, 277, 293, 294, 299, 300, 317, 319, 320, 324, 333, 356, 366, and 392 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4642.

43. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 113, 117, 118, 151, 191, 196, and 205 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4984.

44. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 114/115/143, 114/115/143/233, 114/115/233, 114/115/343, 114/143, 114/143/151/233/238/261, 114/143/213/261/343, 114/143/233, 114/143/233/238, 114/151/261/343, 114/213/238, 114/343, 115, 115/143, 115/143/151/154/213/261, 115/143/151/154/233/343, 115/143/151/233/261, 115/143/151/233/343, 115/143/154/213/233/238/261, 115/143/213, 115/143/233, 115/143/343, 115/151/154/233, 115/151/261, 115/213/233, 115/213/233/238, 115/213/261, 115/233/238, 115/238, 115/343, 115/348, 143, 143/151/154, 143/151/154/213, 143/154, 143/213/233, 143/213/233/238/261/343, 143/213/261/343, 143/233/261, 143/261, 151/154, 151/154/213/343, 151/154/343, 151/213, 151/213/233, 213/233/343, and 233/238/343 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4984.

45. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 117, 117/118/143/168/169/392, 117/118/143/392, 117/118/168/169, 117/118/168/169/392, 117/143/168/169/170/392, 117/143/168/169/392, 117/168/169, 117/168/169/392, 117/168/392, 117/169, 117/392, 118/143, 118/143/168/169, 118/143/169, 118/143/392, 118/168, 118/168/392, 118/169/392, 118/392, 143/168/169/392, 143/169, 143/392, 168/169, 168/169/392, 168/392, 169/392, and 392 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5196.

46. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 134, 134/182/194/294, 134/182/277/293, 134/182/293/320, 134/182/294, 134/194/277, 134/194/277/317, 134/194/293, 134/194/293/294, 134/194/293/317, 134/194/294, 134/194/317, 134/194/317/320, 134/317, 182, 182/194, 182/194/293/320, 182/293/294, 194, 194/277/293/317, 194/277/317/320, 194/277/320, 194/293/294/317,

194/293/320, 277, 293, 317, 317/320, and 320 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5286.

47. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 12/134, 13/134, 13/134/210, 22/134/152, 22/134/235/344, 64/93/134/235, 64/134/344/347, 77/134, 79/134/210, 84/134, 86/134, 87/134, 89/134, 93/134/152/235/344, 93/134/235/344, 93/134/244/344, 93/134/344, 131, 134/152/344, 134/175, 134/210/382, 134/210/402, 134/210/456, 134/210/478, 134/344, 134/377, 134/378, 134/380, 134/381, 134/382, 134/384, 134/386, 134/387, 134/388, 134/390, 134/391, 134/393, 134/395, 134/398, 134/401, 134/402, 134/406, 134/407, 134/421, 134/456, 134/457, 134/459, 134/460, 134/464, 134/467, 134/468, 134/470, 134/471, 134/473, 134/473/474, 134/474, and 134/475 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5412.

48. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73/127/134/160/172/198/290, 73/127/134/160/198/333, 73/127/134/160/210/266, 73/127/134/172/192/266/333, 73/127/134/172/198, 73/127/134/210/290, 73/127/134/210/333, 73/127/134/290/333, 73/127/134/333, 73/134, 73/134/160/172/192/290, 73/134/160/172/198/333, 73/134/160/172/290, 73/134/160/266, 73/134/160/333, 73/134/172, 73/134/172/210/333, 73/134/172/266/333, 73/134/198/210, 73/134/198/210/266, 73/134/210/333, 73/134/333, 73/134/366, 127/134, 127/134/160/172/192/333, 127/134/160/172/210/266/290, 127/134/172/198/266/333, 127/134/192/210, 127/134/198/210/290, 127/134/210/266/290, 127/134/210/333, 134, 134/160/172/333, 134/160/198/266/290, 134/172/192/290/333, 134/192/198, 134/198, 134/198/266/333, 134/210, 134/210/290, and 134/333 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5412.

49. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 13, 13/178/381, 13/178/381/473, 13/178/457, 13/381, 13/381/457, 13/402/457, 13/457, 178, 178/381/457, 178/381/457/473, 178/402/457, 178/473, 381, 381/393/457, 381/457/473, 381/473, and 393/457 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5422.

50. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/194/198/220/243/260/266/270/295/317/333/392/423, 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/29

5/317/333/392/423,
 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/29
 5/317/333/392/423/457,
 38/73/113/115/117/118/127/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/31
 7/333/392/423,
 38/73/113/115/118/127/152/154/160/162/164/168/172/184/192/198/220/243/260/266/270/295/317/333/42
 3,
 38/73/113/115/118/127/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/33
 3/423,
 38/73/113/115/118/127/152/154/160/164/168/172/184/189/192/198/220/243/260/266/270/295/317/333/42
 3,
 38/73/113/115/118/127/152/154/160/164/168/172/184/192/198/210/220/243/260/266/270/295/317/333/42
 3, 38/73/113/115/118/127/152/154/160/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423,
 38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423,
 38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/243/266/270/317/333/423,
 38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/266/270/317/333/423, and
 38/113/127/152/154/160/164/168/172/198/220/290/317/423, and/or any combinations thereof, wherein
 the amino acid positions are numbered with reference to SEQ ID NO: 972.

51. The engineered transposase of Claim 1, wherein at least one substitution or substitution set comprises substitutions at amino acid positions selected from 48, 116, 118, 120, 130, 137, 153, 157, 163, 175, 181, 187, 192, 213, 246, 263, 273, 296, 318, 334, 341, 394, 412, 424, 454, and 458, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5704.

52. The engineered transposase of any of Claims 1-51, wherein said engineered transposase is capable of binding to adapters or donor DNA to form a transposome, cleaving or fragmenting target DNA, and/or ligating the adapters or donor DNA to the target DNA.

53. The engineered transposase of any of Claims 1-52, having at least one improved property, as compared to a wild-type or reference transposase.

54. The engineered transposase of Claim 53, wherein said improved property is selected from increased polynucleotide fragmentation or cleavage activity, increased thermostability, increased activity at elevated temperatures, increased soluble expression, decreased inhibition, increased adapter loading or adapter ligation (i.e. tagging), increased binding to a target polynucleotide, and/or decreased sequence insertion bias.

55. The engineered transposase of any of Claims 1-54, wherein said engineered transposase comprises increased polynucleotide fragmentation or cleavage activity as compared to a wild-type or reference transposase.

56. The engineered transposase of any of Claims 1-54, wherein said engineered transposase comprises increased thermal stability as compared to a wild-type or reference transposase.

57. The engineered transposase of any of Claims 1-54, wherein said engineered transposase comprises increased activity at elevated temperatures as compared to a wild-type or reference transposase.

58. The engineered transposase of any of Claims 1-54, wherein said engineered transposase comprises increased adapter loading or adapter ligation (i.e. tagging) as compared to a wild-type or reference transposase.

59. The engineered transposase of any of Claims 1-54, wherein said engineered transposase comprises increased binding to a target polynucleotide, as compared to a wild-type or reference transposase.

60. The engineered transposase of any of Claims 1-54, wherein said engineered transposase comprises decreased sequence insertion bias, as compared to a wild-type or reference transposase.

61. The engineered transposase of any of Claims 1-60, wherein said engineered transposase comprises a polypeptide sequence selected from the polypeptide sequences between SEQ ID NOs: 2-2368, 2387-5694, and 5706-5756.

62. The engineered transposase of any of Claims 1-61, wherein said transposase is purified.

63. A composition comprising at least one engineered transposase of any of Claims 1-62.

64. A polynucleotide sequence encoding at least one engineered transposase of any of Claims 1-61.

65. A polynucleotide sequence comprising at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 1, 3, 155, 301, 337, 605, 971, 1081, 1217, 1579, 1597, 1599, 1603, 1719, 2367, 2425, 2539, 3401, 3661, 3787, 4149, 4277, 4641, 4983, 5195, 5285, 5411, 5421, and/or 5703, and/or or a functional fragment thereof, wherein said polynucleotide sequence encodes an engineered polypeptide comprising at least one substitution at one or more amino acid positions.

66. The polynucleotide sequence of Claim 65, wherein said polynucleotide sequence encodes at least one engineered transposase comprising a sequence having at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704.

67. The polynucleotide sequence of Claim 65, wherein said sequence comprises SEQ ID NOs: 1, 3, 155, 301, 337, 605, 971, 1081, 1217, 1579, 1597, 1599, 1603, 1719, 2367, 2425, 2539, 3401, 3661, 3787, 4149, 4277, 4641, 4983, 5195, 5285, 5411, 5421, and/or 5703.

68. The polynucleotide sequence of Claim 64, wherein said polynucleotide sequence is operably linked to a control sequence.

69. The polynucleotide sequence of Claim 64, wherein said polynucleotide sequence is codon-optimized.

70. An expression vector comprising at least one polynucleotide sequence of any of Claims 64-69.

71. A host cell comprising at least one expression vector of Claim 70.

72. A method of producing an engineered transposase polypeptide in a host cell comprising culturing a host cell of Claim 71, under suitable culture conditions, such that at least one engineered transposase is produced.

73. The method of Claim 72, further comprising recovering at least one engineered transposase from the culture and/or host cells.

74. The method of Claim 72 or 73, further comprising the step of purifying said at least one engineered transposase.

75. A method for generating a mixture of altered nucleic acids comprising exposing a target nucleic acid to an engineered transposase under conditions and for a time sufficient for the engineered transposase to carry out a transposition event, wherein the engineered transposase comprises a polypeptide sequence comprising at least 85% sequence identity to a reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, or a functional fragment thereof, wherein the engineered transposase comprises at least one substitution or substitution set in its polypeptide sequence relative to the reference sequence.

76. The method of Claim 75, wherein the altered nucleic acids are fragmented from the target nucleic acid.

77. A method of preparing a nucleic acid library comprising:

(a) contacting a target nucleic acid with an engineered transposase under conditions and for a time sufficient for the engineered transposase to carry out a transposition event, wherein the engineered transposase comprises a polypeptide sequence comprising at least 85% sequence identity to a reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, or a functional fragment thereof, wherein the engineered transposase comprises at least one substitution or substitution set in its polypeptide sequence relative to the reference sequences;

(b) fragmenting the target nucleic acid into a plurality of nucleic acid fragments;

(c) selecting a subset of nucleic acid fragments from the plurality of nucleic acid fragments;

and

(d) amplifying the subset of nucleic acid fragments to generate the nucleic acid library.

78. A method of sequencing a target nucleic acid comprising:

(a) contacting the target nucleic acid with an engineered transposase under conditions and for a time sufficient for the engineered transposase to carry out a transposition event, wherein the engineered transposase comprises a polypeptide sequence comprising at least 85% sequence identity to a reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, and/or 5704, or a functional fragment thereof, wherein the engineered transposase comprises at least one substitution or substitution set in its polypeptide sequence relative to the reference sequence;

(b) fragmenting the target nucleic acid into a plurality of nucleic acid fragments;

(c) selecting a subset of nucleic acid fragments from the plurality of nucleic acid fragments;

(d) amplifying the subset of nucleic acid fragments to generate a nucleic acid library; and

(e) sequencing the nucleic acid library.

79. The method of Claim 77 or 78, further comprising adding a polynucleotide to the 5' and/or 3' end of each nucleic acid fragment of the plurality of nucleic acid fragments.

80. The method of any one of Claims 75-78, wherein the engineered transposase comprises a polypeptide sequence comprising at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more sequence identity to the reference sequence of SEQ ID NOs: 2, 4, 156, 302, 338, 606, 972, 1082, 1218, 1580, 1598, 1600, 1604, 1720, 2368, 2426, 2540, 3402, 3662, 3788, 4150, 4278, 4642, 4984, 5196, 5286, 5412, 5422, or 5704.

81. The method of any one of Claims 75--80, wherein the reference sequence is SEQ ID NO. 2 and the at least one substitution or substitution set comprises substitutions at amino acid positions

selected from 10, 13, 24, 35, 39, 44, 46, 51, 54, 55, 62/361, 72, 75, 82, 89, 92, 107, 128, 137, 164, 167, 177/359, 191, 192, 193, 194, 201, 217, 281, 289, 291, 291/340, 305, 309, 310, 317, 331, 337, 340, 346, 352, 353, 357, 358, 359, 362, 371, 396, 414/435, 415, 428, 437, 443, 447, and 448, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2.

82. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 4 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 4, 5, 6, 8, 10, 10/317, 10/317/358, 10/317/359, 10/346, 10/346/359, 10/358, 10/359, 12, 12/373, 13, 15, 16, 18, 19, 26, 30, 30/154, 31, 34, 35, 37, 38, 68, 70, 73, 74, 93, 106, 118, 126, 134, 135, 143, 146, 154/453, 163, 182, 207/286, 208, 226, 237, 246, 247, 259, 262, 263, 264, 280, 282, 286, 287, 288, 289, 291, 292, 293, 298, 305, 309, 310, 313, 317, 317/346, 317/359, 317/445, 320, 346, 351, 355, 356, 358, 359, 360, 368, 369, 410, 411, 422, 426, 427, 428, 446, 448, 453, and 454, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4.

83. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 156 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18/73/74/259, 18/73/286/289/411, 18/74/118/259/286/289/411, 18/74/259, 18/74/259/286/289, 18/74/259/286/411, 18/118/259, 18/259/286/289/411, 18/259/289/411, 18/289/411, 73/74/259/286/289, 73/74/259/289, 73/259/286, 74/259, 74/259/289, 74/286/289/411, 259, 259/286/289, and 289, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 156.

84. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 302 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 16/19/355/427, 16/37/146/286/287/289/310/313/355, 16/146/226/287/289/355, 19/37/146/286/287/310/355, 19/135/286/287/289/355, 19/182/355, 37/70/226/310/313/355, 37/74/310/355/427, 37/135/146/182/226/286/287/355/368, 37/146/226/289/427, 37/146/287/289/427, 38, 41, 42, 43, 44, 45, 47, 48, 50, 53, 54, 55, 70/74/226/286/287/289/368/427, 70/74/355/368, 70/135/146/150/182/286/289/313/427, 74/310/313/355, 74/310/355, 74/355/427, 101, 103, 117, 135/226/286/289/355/368, 135/355, 146, 146/150/287/289/310/355, 146/182/355, 146/226/286/355, 146/286/287/289/355/427, 146/289, 146/427, 150/226/310/313, 182/226/355, 182/355, 244, 247, 249, 251, 287/355/427, 300, 317/411, 355, 393, 397, 398, 399, 401, 405, 407, 408, 409, 411, 413, 426, 429, 431, 439, 444, 446, 450, 453, 454, 457, 459, 460, 464, 466, 467, 469, 470, 471, 472, 474, 475, 476, and 478, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 302.

85. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 338 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 16, 16/313, 16/409/411/427, 37/150/408, 37/313/317/408/411, 37/317/405/408, 37/317/405/426, 37/405/409/427, 150, 150/313/405/411, 150/405/406, 150/408, 150/408/411, 226/405,

317/460, and 408/426, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 338.

86. The method of any one of claims 75-80, wherein the reference sequence is SEQ ID NO. 606 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 12, 13, 18, 19, 22, 26, 31/262, 34, 60, 79, 80, 81, 82, 92, 106, 107, 113, 115, 132, 134, 136, 140, 143, 152, 160, 163, 167, 168, 175, 179, 183, 195, 199, 201, 202, 215, 216, 218, 220, 223, 226, 233, 235, 236, 237, 243, 255, 259, 261, 262, 266, 268, 270, 273, 282, 290, 292, 293, 309, 319, 320, 326, 343, 346, 365, 370, 374, 380, 382, 386, 389, 390, 408, 412, 414, 415, 424, and 442, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

87. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 606 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 2, 12, 13, 15, 16, 18, 19, 22, 26, 30, 31/262, 34, 51, 60, 69, 79, 82, 84, 87, 88, 106, 107, 113, 114, 115, 128, 132, 134, 135, 136, 140, 143, 146, 150, 152, 160, 163, 167, 168, 175, 179, 183, 195, 198, 199, 201, 215, 217, 218, 219, 220, 223, 224, 226, 227, 231, 233, 235, 236, 243, 254, 255, 259, 261, 266, 267, 268, 270, 272, 273, 275, 277, 279, 282, 287, 289, 290, 291, 292, 293, 304, 307, 309, 319, 320, 343, 344, 346, 355, 365, 370, 373, 380, 382, 386, 389, 390, 408, 412, 414, 415, 418, 423, 424, and 468, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

88. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 606 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 44/45/47/408/453, 44/45/47/453, 44/45/47/453/454, 44/45/103/453/457, 44/45/453/454/457, 44/47/103/393/408/453/454, 44/47/103/453/454/457, 44/47/393/408/453/457, 44/47/393/453/457, 44/103/408/453/454, 45/47/103/393/453/454/457, 45/47/393/453/454, 45/47/408/423/453/457, 45/47/453, 45/103/393/453/454, 45/453/454, 47/103/408/453/454, 47/393/408/453/454/457, 47/393/453/454, 47/408/453/457, 53/54/244/450, 103/393/408/453/457, 244/450, 408/453/454, 446/450/464, and 453/457, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 606.

89. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 972 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 38/45/53, 38/45/53/54/476, 38/45/53/244, 38/45/53/470/472, 38/45/54/244, 38/45/54/470/474/478, 38/45/472/476/478, 38/45/476, 38/53/54, 38/53/54/244, 38/53/54/244/472/478, 38/53/472, 38/54/244/470/474/476/478, 38/244/470/472/474/478, 38/478, 45/53/54/244/470/474/476/478, 45/53/54/470/474/476, 45/53/54/472/478, 45/54/244/474/478, 53, 53/54/470/476/478, 53/244/474/478, 54, 54/244, 54/244/472/476, 54/472, 244, and 474/478, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

90. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 972 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 22/38/45/107/259/317/412, 22/38/45/259/317/412, 22/107, 22/107/317/412, 22/134, 22/134/220, 22/134/220/317, 22/134/317/412, 22/317, 38, 38/45/81/143/244, 38/53/243, 38/81, 38/81/244, 38/107/134/220, 38/132, 38/220/317, 45/81, 45/81/243, 45/81/243/343, 45/143/243/270/343, 45/244/270, 53/115/163/243/270/343, 53/143/243/270/343, 53/270, 64/134/220/317/412, 81/132/243, 107, 107/134/317, 115/244/270, 134/259, 143/243/270, 243, 243/270, 244, 244/270, 259, 270, and 317, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

91. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 972 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 22/38/45/107/259/317/412, 22/38/45/259/317/412, 22/38/64/134/220, 22/45/220/259, 22/64/107/259/317/412, 22/64/134/220, 22/107, 22/134, 22/134/317, 22/134/317/412, 22/220, 22/220/412/415, 22/259/317, 22/259/317/412, 22/317, 38, 38/45/81/143/244, 38/45/143/270, 38/53/243, 38/81, 38/81/143/243/270, 38/81/183/243/270, 38/81/243, 38/81/244, 38/107/134/220, 38/132, 38/220/317, 38/270/343, 45/81, 45/81/132/243/270, 45/81/243, 45/143/243/270/343, 45/244/270, 53/143/243/270/343, 53/244/270, 53/270, 64, 64/107/134, 64/107/317, 64/134, 64/134/220/317/412, 81/132, 81/132/243, 107, 107/134/317, 107/412, 115/244/270, 143/243/270, 220/259, 243, 243/270, 244/270, 255, 259, and 270, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

92. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1082 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 143, 143/175/272/273/415, 152/168/216/237, 152/168/423, 160, 160/201/272, 160/201/273/415, 160/226, 160/244/272/273/415, 160/415, 168, 168/187/216/255/423, 168/216/237, 168/255/423, 168/290/423, 187, 187/423, 199/201/226, 199/201/415, 201, 216, 216/423, 226/244/415, 226/415, 237, 237/290, 244, 255, 272, 290, 290/423, 415, and 423, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

93. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1082 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 143, 143/160/226/415, 143/175/272/273/415, 152/168/216/237, 152/168/423, 160, 160/201/272, 160/201/273/415, 160/226, 160/244/272/273/415, 160/415, 168, 168/187/216/255/423, 168/216/237, 168/216/237/255/343, 168/255/423, 168/290/423, 187, 199/201/226, 199/201/415, 201, 216, 216/423, 226/244/415, 226/415, 237, 237/290, 244, 255, 255/343, 272, 290, 290/423, 415, and 423, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

94. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1082 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 72, 90, 91, 92, 93, 95, 122, 129, 141, 142, 143, 144, 145, 150, 151, 151/155, 152, 154, 155, 157, 160, 161, 161/313, 164, 165, 168, 180, 181, 184, 193, 197, 204, 205, 205/472, 206, 299, and 320, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

95. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1082 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 90, 92, 122, 129, 141, 143, 144, 145, 150, 151, 151/155, 152, 154, 155, 155/265, 157, 160, 161, 164, 165, 168, 180, 181, 184, 193, 197, 204, 205, 205/472, 206, 299, and 320, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1082.

96. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1218 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 129, 129/160, 129/160/164, 129/160/184, 129/164, 129/164/184, 129/205, 146/152/244, 150/155/226, 150/155/244, 152, 154, 154/160/184, 154/164, 154/205, 155, 160, 160/164, 160/164/184, 160/184, 164, 164/204, 164/205, 204, 205, 226, 226/244, and 244, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1218.

97. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1218 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 129, 129/143, 129/154/164, 129/154/164/205, 129/154/205, 129/160, 129/160/164, 129/160/205, 129/164/205, 143, 143/150/155/226/244, 143/150/155/244, 143/150/226, 143/150/226/244/343, 143/150/226/343, 143/150/244, 143/150/343, 143/152, 143/152/226, 143/152/226/244, 143/152/226/343, 143/152/343, 143/155, 143/155/226/244, 143/155/226/244/343, 143/155/343, 143/226, 143/226/244/343, 143/226/343, 143/244, 143/244/343, 143/343, 146/152/244, 150/155/226, 150/155/244, 150/155/244/343, 150/226/244/343, 152, 152/174/244/343, 152/226, 152/226/343, 152/343, 154, 154/160/184, 154/164, 154/164/204, 154/164/204/205, 154/205, 155, 160, 160/164, 160/164/184, 164, 205, 226, 226/244, and 244, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1218.

98. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1580 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 98/129/157/168, 129/157/168, 151, 151/152/255, 151/152/273, 151/255, 151/343, 157, 157/164, 157/164/168, 157/168, 164, 164/255, 168, 183, 255, 255/343, and 343, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

99. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1580 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 129, 129/157, 129/157/168, 129/164, 129/164/168/183, 129/164/183, 143, 143/151, 143/151/152, 143/151/164, 143/164, 143/164/255, 143/164/343, 143/255, 143/343, 151, 151/152, 151/152/164/343, 151/152/343, 151/255, 151/255/343, 151/343, 152, 152/255, 157, 157/164/168, 157/168, 164, 164/168, 164/255, 164/447, 168, 183, 255, 255/343, 298/343, and 343, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

100. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1580 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 69, 69/164, 72, 73, 91, 92/164, 95, 96, 97, 98, 110, 112, 113, 121, 121/164, 122, 122/164, 124, 126, 127, 139, 140, 141, 144, 144/164, 145, 146, 147, 148, 150, 150/164, 151, 151/164, 153, 153/164, 155, 156, 157, 159, 160, 161, 163, 164, 164/172, 164/173, 164/182, 164/188, 164/191, 164/192, 164/195, 164/197, 164/199, 164/202, 164/205/249, 164/209, 165, 166, 167, 169, 170, 171, 172, 173, 177, 184, 187, 188, 190, 191, 192, 193, 194, 195, 197, 197/266, 198, 199, 200, 202, 203, 204, 205, 208, 209, 210, 321, 324, 331, and 349, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

101. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1580 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 69, 69/164, 72, 73, 91, 92/164, 95, 97, 98, 121/164, 126, 127, 139, 140, 142, 144, 144/164, 145, 146, 146/164, 148, 150, 150/164, 151, 153, 153/164, 155, 156, 157, 160, 163, 164/173, 164/188, 164/191, 164/192, 164/195, 164/199, 164/205/249, 164/209, 165, 166, 167, 169, 171, 172, 173, 177, 182, 187, 188, 190, 191, 192, 194, 195, 197/266, 198, 199, 200, 202, 203, 204, 205, 208, 209, 210, 324, 331, and 349, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

102. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1580 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73, 91, 93, 95, 97, 98, 110, 112, 122, 122/164, 124, 127, 139, 144, 145, 146, 146/164, 147, 148, 150, 150/164, 151, 151/164, 153, 153/164, 155, 156, 157, 159, 160, 161, 162, 163, 164, 164/182, 164/188, 164/191, 164/192, 164/195, 164/199, 164/205/249, 164/209, 165, 166, 167, 169, 171, 172, 177, 184, 187, 188, 190, 191, 192, 193, 195, 197/266, 198, 199, 200, 204, 205, 206, 207, 208, 209, 210, and 324, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

103. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 1580 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 69, 73, 91, 127, 140, 144, 145, 146, 146/164, 150, 150/164, 151, 153, 155, 156, 157, 160,

163, 164, 164/172, 164/188, 164/191, 164/192, 164/195, 164/199, 164/209, 165, 167, 169, 171, 172, 182, 183, 188, 191, 192, 193, 195, 196, 198, 199, 200, 204, 205, 208, 210, and 324, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 1580.

104. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 972 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 38/143/168/220/290/317/423, 38/152/154/164/168/220/290/317/423, 38/152/168/220/290/317/343/423, 38/152/168/220/290/317/423, 38/154/164/168/220/290/317/423, 38/168/220/290/317/423, and 38/220/317, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

105. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 2368 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73/143/169/191/193, 73/156, 73/156/192/435, 73/156/210, 73/163, 73/163/210, 73/192/193, 127, 127/160/172/198, 127/172, 127/198, 146, 146/151/188, 146/324, 151/171/324, 151/324, 156/169, 156/192, 157/171, 163, 163/169, 163/169/191/192/193, 165/171/324, 165/188/324, 165/324, 169, 171/188/324, 171/324, 172, 188, 192, 192/193, 198, 210, and 324 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2368.

106. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 2368 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73, 73/163/191/192/193, 127/160/172, 127/160/172/198, 144, 160, 160/172, 163/191/192, and 192 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2368.

107. The method of any one of claims 75-80, wherein the reference sequence is SEQ ID NO. 2426 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 8, 13, 17, 18, 20/146/157/167, 26, 34/38, 38/39/45, 67, 70, 73, 73/150/155/169, 73/150/171/192, 73/155, 73/155/169/199, 73/155/171, 73/192, 73/199, 78, 82, 84, 86, 86/384, 89, 90, 91, 95, 102, 110, 111, 113, 114, 115, 122/127, 126/127, 127/131/165, 127/133, 143, 143/146, 143/146/153, 143/146/157, 143/146/167, 143/146/167/188, 143/146/188, 143/153/167/188/246, 143/156/157/167/195, 143/156/188, 143/188, 143/188/195, 143/195, 144, 145, 145/151, 146, 146/152, 146/153/157, 146/153/157/167, 146/153/167, 146/157, 146/157/167, 146/167, 150, 150/155/169, 150/155/169/192, 150/155/192, 150/169, 150/192, 150/199, 151/152/154, 152/154, 152/154/155/160, 153/157/167, 155/171, 155/192, 156/270, 160/371, 167/188, 168/172, 169, 184, 190, 192/198, 198, 206, 208, 209, 210, 214/220, 215/220, 216/220, 217/220, 219/220, 220, 220/222, 220/225, 220/226, 231, 235, 242, 257, 259, 261, 262, 267, 268, 269, 270, 272, 274, 275, 276, 277, 278, 282, 283/290, 284/290, 287/290, 288/290, 289/290, 290, 290/291, 290/292, 290/293, 290/294, 304, 305, 305/373, 306, 307, 309, 317, 329, 333, 334, 339, 341, 344, 345, 346, 352, 353, 355, 358, 366, 368, 370, 372, 373, 375, 376, 378, 379, 380, 381, 382, 383, 384, 385, 386/393, 387/393, 389/393, 390/393, 391/393, 392/393, 393/394, 414, 415, 420/423,

422/423, 423, 423/425, 432, 440, 441, 443, 469, and 472 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2426.

108. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 2426 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 8, 18, 25, 29, 30, 67, 73, 73/150/155/169, 73/150/155/171, 73/150/171/192, 73/171, 73/192, 73/199, 74, 82, 86, 86/384, 113, 115, 118, 127/128, 143, 150, 150/155/169, 150/155/169/192, 150/155/171, 150/155/171/199, 150/155/192, 150/155/199, 150/192, 150/199, 155/192, 168/171/172, 195/198, 197/198, 198/201, 198/203, 198/205, 235, 252, 256, 257, 262, 263, 267/343, 268, 269, 270, 272, 273, 276, 277, 288/290, 289/290, 290, 290/291, 290/292, 290/294, 298, 304, 305, 307, 333, 346, 355, 366, 370, 386/393, 389/393, 393/395, 440, and 472 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2426.

109. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 2540 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 66, 67, 70, 101, 102, 103, 104, 106, 107, 108, 113, 114, 115, 117, 118, 143, 152, 154, 235, 236/251, 237, 238, 242, 243, 246, 247, 249, 250, 251, 256/352, 257, 259, 285, 286, 287, 288, 289, 290, 292, 293, 294, 295, 296, 322, 323, 326, 329, 330, 333, 339, 340, 343, 344, 345, and 352. and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2540.

110. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 2540 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18, 18/113/270, 18/113/270/329/370, 18/113/270/370/384, 18/113/270/370/412, 18/113/329/384, 18/259/370/384, 18/270, 18/270/370, 18/270/370/384, 18/370, 18/370/384, 73, 73/115, 73/115/184/333, 73/115/184/341, 73/115/256, 73/184/256, 73/184/273/333, 73/184/333, 73/256, 73/256/375, 73/273, 73/273/333/375, 73/273/375, 73/333, 73/333/375, 73/375, 73/379, 77/233/238/376, 77/233/334/376, 77/262/376, 77/276, 77/334/376, 82, 82/127, 82/127/144/235/290, 82/127/144/266, 82/127/144/266/290, 82/127/144/378, 82/127/235, 82/127/290, 82/144/266/290/378, 82/144/266/378, 82/235/266, 82/235/266/378/383, 82/235/290, 82/235/378, 82/266, 82/266/378, 82/378, 113/270/370/384, 115/256/333/375, 115/256/375, 115/273/375, 115/375, 127/144/290, 127/149/150, 127/266/290/378, 127/378/383, 144/235/290, 144/266/290, 233/334, 233/376, 235, 235/290, 256/333, 262, 266/378, 270/370, 273, 276/352, 290, 290/378, 333, 358, 370, 375, and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2540.

111. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 2540 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18, 18/113, 18/113/252, 18/113/252/336/384, 18/113/252/370, 18/113/270, 18/113/270/370/384, 18/259/370/384, 18/270, 18/270/370, 18/270/370/384, 18/370, 18/370/384, 73, 73/115/184/341, 73/115/341, 73/131/256/273/333/341/375, 73/184/256, 73/184/256/341, 73/184/256/375,

73/184/273/333, 73/184/333, 73/184/341/375, 73/256, 73/256/273, 73/256/273/375, 73/256/333/375, 73/256/375, 73/273, 73/273/333/341, 73/273/333/375, 73/273/375, 73/333, 73/333/341/375, 73/333/375, 73/341, 73/375, 73/379, 77, 77/118, 77/118/143/262/276/334/376, 77/118/143/334/352/376, 77/118/143/376, 77/118/233, 77/118/233/352/376, 77/118/262/334/376, 77/118/334, 77/143/262/276, 77/143/262/334, 77/143/376, 77/233/238/376, 77/233/262, 77/233/276/376, 77/233/334/376, 77/262/376, 77/276, 77/334/376, 77/352, 82, 82/127, 82/127/144/235/290, 82/127/144/266, 82/127/144/266/290, 82/127/144/378, 82/127/235, 82/127/235/266/358, 82/127/290, 82/127/290/358, 82/127/290/378, 82/144/266/290/378, 82/144/266/378, 82/150/235/266, 82/235/266, 82/235/266/290, 82/235/266/378/383, 82/235/290, 82/235/378, 82/266, 82/266/378, 82/290, 82/378, 113, 113/252, 113/252/270/370, 113/270/370/384, 113/336, 118/233/276/334, 118/334, 127/235/266, 127/266/290/378, 127/378/383, 143, 143/276, 143/276/334, 144/235/290, 144/266/290, 144/290/358, 233/334, 233/376, 235, 235/266/290/358/383, 235/358, 252/329/336/370, 252/384, 256/333, 256/375, 262, 262/276, 262/334, 266/290, 266/378, 270/370, 273, 276/352, 290, 290/378, 333, 334, 334/376, 336/370, 341, 358, 370, 375, and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 2540.

112. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 3402 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18, 18/77/118/169/233/238/256/266/270, 18/77/150/256/376, 18/118, 18/118/143/169/238/256/266/333, 18/118/150/169/184/238/256/270/378, 18/118/169, 18/118/233/238, 18/143, 18/143/256, 18/150/169, 18/150/266, 18/169/184/290/333/376, 18/169/233/333, 18/169/392, 18/184/270, 18/233/256/334, 18/256/266/290/378, 77/113/118/256/266/290, 77/118/290, 77/169/270/290, 77/233, 77/233/256, 77/238, 113/118/169/184/256/376, 113/169, 113/233, 118, 118/143/238, 118/233/238/256/378, 118/233/238/270, 143/169/233/270, 143/233/266/290/376, 143/238/270/290/378, 143/256/334, 143/290, 169, 169/256/266, 183/270/333, 233/238, 233/238/376, 233/392, 238/256, 256/290/303, 270, 290, 376, and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3402.

113. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 3402 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 18, 18/77/118/169/266/334/376, 18/77/150/256/376, 18/77/256/290/334, 18/113/238, 18/118, 18/118/143/169/238/256/266/333, 18/118/143/169/376, 18/118/169, 18/118/233/238, 18/118/266, 18/143, 18/143/169/233/238/358, 18/143/233/238/290/376, 18/143/256, 18/143/290, 18/143/333/334/378, 18/150/169, 18/150/266, 18/169/184/290/333/376, 18/169/233/333, 18/169/238/270/334/376, 18/169/392, 18/233/256/334, 18/233/266/290/376, 18/256/266/290/378, 18/266/270/378, 18/378, 77/113/121/143/233/334, 77/118/169/233/266/270, 77/118/238/376, 77/118/290, 77/169/270/290, 77/233, 77/238, 113/238/376, 118, 118/143/238, 118/143/392, 118/169/184/233/266/270/333, 118/233/238/256/378, 118/256/290/333/334/376/378, 118/256/334/376, 118/266/290, 143/169/233/270, 143/233/266/290/376, 143/238/270/290/378, 143/256/334, 143/290, 143/364, 150/233/333, 169, 169/256/266, 183/270/333, 233/238, 233/238/376, 233/392, 238/256, 238/266/270/378, 256/266/334,

266/270/376, 266/333, and 378 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3402.

114. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 3662 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 77/156, 77/156/270, 77/184/244, 77/184/333, 77/244/270, 77/270, 143, 143/169/238/257/344, 143/169/255, 143/169/255/257, 143/169/343, 143/255, 143/255/257, 143/255/344, 143/257, 156, 156/270, 169/238, 169/238/255/257/344, 169/257/344, 169/344, 184, 184/244/324, 238, 238/255/344, 244/270, 257/344, and 270 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3662.

115. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 3662 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 77, 77/156/184/244/324, 77/184/270, 143, 143/238, 143/238/255, 143/238/255/343, 143/238/255/392, 143/238/344, 143/255, 143/257/392, 143/344, 169, 169/238/255/344, 169/238/344, 169/255, 169/343, 169/392, 238/344, 255, 255/257, 255/257/392, 257, and 344 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3662.

116. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 3788 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 20, 22, 26, 42, 43, 44, 45, 50, 51, 52, 53, 54, 55, 60, 62, 64, 65, 95, 97, 99, 103, 104, 111, 112, 115, 115/143/169, 115/169/343, 115/343, 116, 116/343, 143, 143/169, 154/439, 169, 169/324, 169/324/343, 169/343, 186, 234, 236, 238, 238/246, 239, 239/244, 241, 242, 243, 244, 250, 251, 254, 255, 260, 283, 284, 295, 296, 298/326, 319, 322, 323, 324, 324/343, 326, 328, 329, 333, 334, 338, 340, 341, 343, 344, 347, 366, 434, 439, 442, 443, 444, 445, 446, and 447 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 3788.

117. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 4150 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 26, 26/236, 26/236/260/333, 26/236/295, 26/260, 26/260/295, 26/260/295/333, 26/295/319, 26/333, 64, 64/244, 64/251/260/347, 64/260/295/333, 64/260/347, 103/115/255, 103/115/343, 103/115/343/344, 103/143/255/284/343/344, 103/169/343/344, 103/255/284, 103/343/344, 115, 115/143/255, 115/143/255/343, 115/143/284, 115/143/343, 115/143/344, 115/169, 115/169/343, 115/255, 115/255/344, 115/284, 115/284/343/344, 115/284/344, 115/343, 115/343/344, 115/344, 143, 143/169/255/284/343/344, 143/169/343/344, 143/255, 143/255/284/343/344, 143/255/343, 143/255/343/344, 143/255/344, 143/343/344, 169, 169/255, 169/255/284, 169/255/343, 169/284, 169/343, 169/344, 236, 236/244/333, 236/260, 236/347, 244, 244/251, 244/251/319, 244/260, 244/295, 244/347, 251, 251/260/295/347, 251/333, 255, 255/284, 255/343, 255/343/344, 255/344, 259, 260, 260/295, 260/333, 284, 284/343/344, 284/344, 295, 295/333, 333, 343, 343/344, 344, and 347 and/or any

combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4150.

118. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 4278 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 115, 115/117, 115/117/118, 115/117/118/152, 115/117/118/169, 115/117/118/246, 115/117/152, 115/117/152/169, 115/117/152/169/246, 115/117/152/246, 115/117/169, 115/117/246, 115/118, 115/118/152, 115/118/152/169, 115/118/152/169/246, 115/118/169, 115/118/246, 115/152, 115/152/169, 115/152/169/246, 115/152/246, 115/152/246/374, 115/169, 115/169/246, 115/246, 117, 117/118, 117/118/152, 117/118/152/169, 117/118/152/246, 117/118/169, 117/118/246, 117/152, 117/152/169, 117/152/246/288, 117/169, 117/246, 118/246, 152, 152/169, 152/169/246, 152/246, 169, 169/246, and 246 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4278.

119. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 4278 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73, 101, 114, 115, 118, 143, 145, 150, 151, 152, 153, 154, 156, 157, 160, 161, 163, 166, 167, 168, 169, 171, 172, 190, 191, 192, 195, 213, 214, 218, 219, 223, 225, 233, 233/343, 234, 235, 236, 237, 238, 239, 239/250, 243, 244, 249, 251, 254, 255, 257, 258, 259, 261, 262, 262/283, 277/292, 283, 284, 286, 289, 292, 295, and 296 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4278.

120. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 4642 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 93, 113, 113/114, 113/115, 113/143, 115/118, 115/177, 118/143/163/168/195/255/315, 118/143/255, 118/163/168/214/255/315, 118/168/214/315, 118/255, 124, 126, 129, 134, 143/154/255, 143/156/168/216/255, 143/163, 143/163/255, 143/168, 143/168/169/255/315, 143/255, 148, 156/214/315, 159, 163, 163/168, 163/168/169, 163/168/169/255/315, 163/168/169/315, 163/168/195/214/255/315, 163/214, 163/255, 163/255/315, 168/315, 186, 194, 197, 206, 207, 208, 210, 213, 255/315, 277, 293, 294, 315, 317, 320, 322, 328, 329, 330, 333, 352, 366, 392, and 445 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4642.

121. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 4642 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 127, 128, 129, 140, 180, 182, 194, 277, 293, 294, 299, 300, 317, 319, 320, 324, 333, 356, 366, and 392 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4642.

122. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 4984 and the at least one substitution or substitution set comprises substitutions at amino acid positions

selected from 113, 117, 118, 151, 191, 196, and 205 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4984.

123. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 4984 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 114/115/143, 114/115/143/233, 114/115/233, 114/115/343, 114/143, 114/143/151/233/238/261, 114/143/213/261/343, 114/143/233, 114/143/233/238, 114/151/261/343, 114/213/238, 114/343, 115, 115/143, 115/143/151/154/213/261, 115/143/151/154/233/343, 115/143/151/233/261, 115/143/151/233/343, 115/143/154/213/233/238/261, 115/143/213, 115/143/233, 115/143/343, 115/151/154/233, 115/151/261, 115/213/233, 115/213/233/238, 115/213/261, 115/233/238, 115/238, 115/343, 115/348, 143, 143/151/154, 143/151/154/213, 143/154, 143/213/233, 143/213/233/238/261/343, 143/213/261/343, 143/233/261, 143/261, 151/154, 151/154/213/343, 151/154/343, 151/213, 151/213/233, 213/233/343, and 233/238/343 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 4984.

124. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 5196 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 117, 117/118/143/168/169/392, 117/118/143/392, 117/118/168/169, 117/118/168/169/392, 117/143/168/169/170/392, 117/143/168/169/392, 117/168/169, 117/168/169/392, 117/168/392, 117/169, 117/392, 118/143, 118/143/168/169, 118/143/169, 118/143/392, 118/168, 118/168/392, 118/169/392, 118/392, 143/168/169/392, 143/169, 143/392, 168/169, 168/169/392, 168/392, 169/392, and 392 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5196.

125. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 5286 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 134, 134/182/194/294, 134/182/277/293, 134/182/293/320, 134/182/294, 134/194/277, 134/194/277/317, 134/194/293, 134/194/293/294, 134/194/293/317, 134/194/294, 134/194/317, 134/194/317/320, 134/317, 182, 182/194, 182/194/293/320, 182/293/294, 194, 194/277/293/317, 194/277/317/320, 194/277/320, 194/293/294/317, 194/293/320, 277, 293, 317, 317/320, and 320 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5286.

126. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 5412 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 12/134, 13/134, 13/134/210, 22/134/152, 22/134/235/344, 64/93/134/235, 64/134/344/347, 77/134, 79/134/210, 84/134, 86/134, 87/134, 89/134, 93/134/152/235/344, 93/134/235/344, 93/134/244/344, 93/134/344, 131, 134/152/344, 134/175, 134/210/382, 134/210/402, 134/210/456, 134/210/478, 134/344, 134/377, 134/378, 134/380, 134/381, 134/382, 134/384, 134/386, 134/387, 134/388, 134/390, 134/391, 134/393, 134/395, 134/398, 134/401, 134/402, 134/406, 134/407, 134/421,

134/456, 134/457, 134/459, 134/460, 134/464, 134/467, 134/468, 134/470, 134/471, 134/473, 134/473/474, 134/474, and 134/475 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5412.

127. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 5412 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 73/127/134/160/172/198/290, 73/127/134/160/198/333, 73/127/134/160/210/266, 73/127/134/172/192/266/333, 73/127/134/172/198, 73/127/134/210/290, 73/127/134/210/333, 73/127/134/290/333, 73/127/134/333, 73/134, 73/134/160/172/192/290, 73/134/160/172/198/333, 73/134/160/172/290, 73/134/160/266, 73/134/160/333, 73/134/172, 73/134/172/210/333, 73/134/172/266/333, 73/134/198/210, 73/134/198/210/266, 73/134/210/333, 73/134/333, 73/134/366, 127/134, 127/134/160/172/192/333, 127/134/160/172/210/266/290, 127/134/172/198/266/333, 127/134/192/210, 127/134/198/210/290, 127/134/210/266/290, 127/134/210/333, 134, 134/160/172/333, 134/160/198/266/290, 134/172/192/290/333, 134/192/198, 134/198, 134/198/266/333, 134/210, 134/210/290, and 134/333 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5412.

128. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 5422 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 13, 13/178/381, 13/178/381/473, 13/178/457, 13/381, 13/381/457, 13/402/457, 13/457, 178, 178/381/457, 178/381/457/473, 178/402/457, 178/473, 381, 381/393/457, 381/457/473, 381/473, and 393/457 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5422.

129. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 972 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/392/423, 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/194/198/220/243/260/266/270/295/317/333/392/423, 38/73/113/115/117/118/127/134/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/392/423/457, 38/73/113/115/117/118/127/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/392/423, 38/73/113/115/118/127/152/154/160/162/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423, 38/73/113/115/118/127/152/154/160/163/164/168/169/172/184/192/198/220/243/260/266/270/295/317/333/423, 38/73/113/115/118/127/152/154/160/164/168/172/184/189/192/198/220/243/260/266/270/295/317/333/423

3,
38/73/113/115/118/127/152/154/160/164/168/172/184/192/198/210/220/243/260/266/270/295/317/333/423, 38/73/113/115/118/127/152/154/160/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423, 38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/243/260/266/270/295/317/333/423, 38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/243/266/270/317/333/423, 38/73/113/118/127/152/154/160/164/168/172/184/192/198/220/266/270/317/333/423, and 38/113/127/152/154/160/164/168/172/198/220/290/317/423 and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 972.

130. The method of any one of Claims 75-80, wherein the reference sequence is SEQ ID NO. 5704 and the at least one substitution or substitution set comprises substitutions at amino acid positions selected from 48, 116, 118, 120, 130, 137, 153, 157, 163, 175, 181, 187, 192, 213, 246, 263, 273, 296, 318, 334, 341, 394, 412, 424, 454, and 458, and/or any combinations thereof, wherein the amino acid positions are numbered with reference to SEQ ID NO: 5704.

131. The method of any of Claims 75-130, wherein the engineered transposase comprises a polypeptide sequence selected from the even-numbered sequences selected from SEQ ID NOs: 2-2368, 2388-5694, 5702, and 5706-5756.

132. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 2396.

133. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 4642.

134. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 4872.

135. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 4954.

136. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 5186.

137. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 5354.

138. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 5500.

139. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 5520.

140. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 5702.

141. The method of any one of Claims 75-80 and 131, wherein the engineered transposase comprises a polypeptide sequence of SEQ ID NO: 5738.

142. The method of any one of Claims 75-141, wherein the engineered transposase is purified.

143. The method of any one of Claims 76-141, wherein fragmenting the target nucleic acid comprises tagmentation or random sheering and adapter ligation.

144. The method of any one of Claims 76-141, wherein fragmenting the target nucleic acid comprises tagmentation.

145. The method of Claim 75, wherein each altered nucleic acid of the mixture of altered nucleic acids comprises an identifiable sequence tag (IST).

146. The method of any one of Claims 75-79, wherein each nucleic acid fragment of the subset of nucleic acid fragments comprises an identifiable sequence tag (IST).

147. The method of Claim 155 or 156, wherein the IST is between 6 and 30 nucleotides in length.

148. The method of Claim 75, wherein each altered nucleic acid of the mixture of altered nucleic acids is between 30 and 30000 nucleotides in length.

149. The method of any one of Claims 76-79, wherein each nucleic acid fragment of the plurality of nucleic acid fragments is between 30 and 30000 nucleotides in length.

150. The method of any one of Claims 75-134, wherein the transposition event occurs in reaction conditions additionally comprising a terminal deoxynucleotide transferase, dNTP, and/or buffer components suitable for the addition of deoxynucleotides to the 3' terminus of each of the plurality of nucleic acid fragments.

151. The method of any one of Claims 75-144, wherein the transposition event occurs in reaction conditions additionally comprising a DNA ligase and/or buffer components suitable for a ligation reaction.

152. The method of Claim 75, wherein each altered nucleic acid of the mixture of altered nucleic acids further comprises a sample tag and/or a unique molecular identifier (UMI).

153. The method of any one of Claims 76-79, wherein each nucleic acid fragment of the subset of nucleic acid fragments further comprises a sample tag and/or a unique molecular identifier (UMI).

154. The method of any one of Claims 77-144, wherein amplification of the subset of nucleic acid fragments is performed through polymerase chain reaction (PCR), multiple displacement amplification (MDA), ligase chain reaction (LCR), loop mediated isothermal amplification (LAMP), rolling circle amplification (RCA), or strand displacement amplification (SDA).

155. The method of any one of Claims 78-144, wherein the sequencing comprises next-generation sequencing (NGS).

156. The method of any one of Claims 78-144, wherein sequencing comprises sequencing by synthesis, sequencing by ligation, or nanopore sequencing.

157. The method of Claim 156, wherein the sequencing by synthesis comprises Illumina™ dye sequencing, single-molecule real-time (SMRT™) sequencing, or pyrosequencing.

158. The method of Claim 156, wherein the sequencing by ligation comprises polony-based sequencing or SOLiD™ sequencing.

159. The method of any one of Claims 75-158, wherein the target nucleic acid comprises genomic DNA or cDNAs from a single cell.

160. The method of any one of Claims 75-158, wherein the target nucleic acid comprises nucleic acids from a plurality of haplotypes.

161. The method of any one of Claims 75-158, wherein the target nucleic acid is crosslinked via histones or chromatin from single or multiple cells.

162. The method of any one of Claims 75-158, wherein the target nucleic acid has been condensed or optionally treated with one or more condensing agents.