



US 20220315971A1

(19) **United States**

(12) **Patent Application Publication**

WU et al.

(10) **Pub. No.: US 2022/0315971 A1**

(43) **Pub. Date: Oct. 6, 2022**

(54) **METHODS FOR SEAMLESS NUCLEIC ACID ASSEMBLY**

(60) Provisional application No. 62/518,489, filed on Jun. 12, 2017.

(71) Applicant: **Twist Bioscience Corporation**, South San Francisco, CA (US)

Publication Classification

(72) Inventors: **Cheng-Hsien WU**, Burlingame, CA (US); **Sebastian TREUSCH**, San Francisco, CA (US)

(51) **Int. Cl.**
C12P 19/34 (2006.01)

(52) **U.S. Cl.**
CPC *C12P 19/34* (2013.01)

(21) Appl. No.: **17/825,863**

(57) **ABSTRACT**

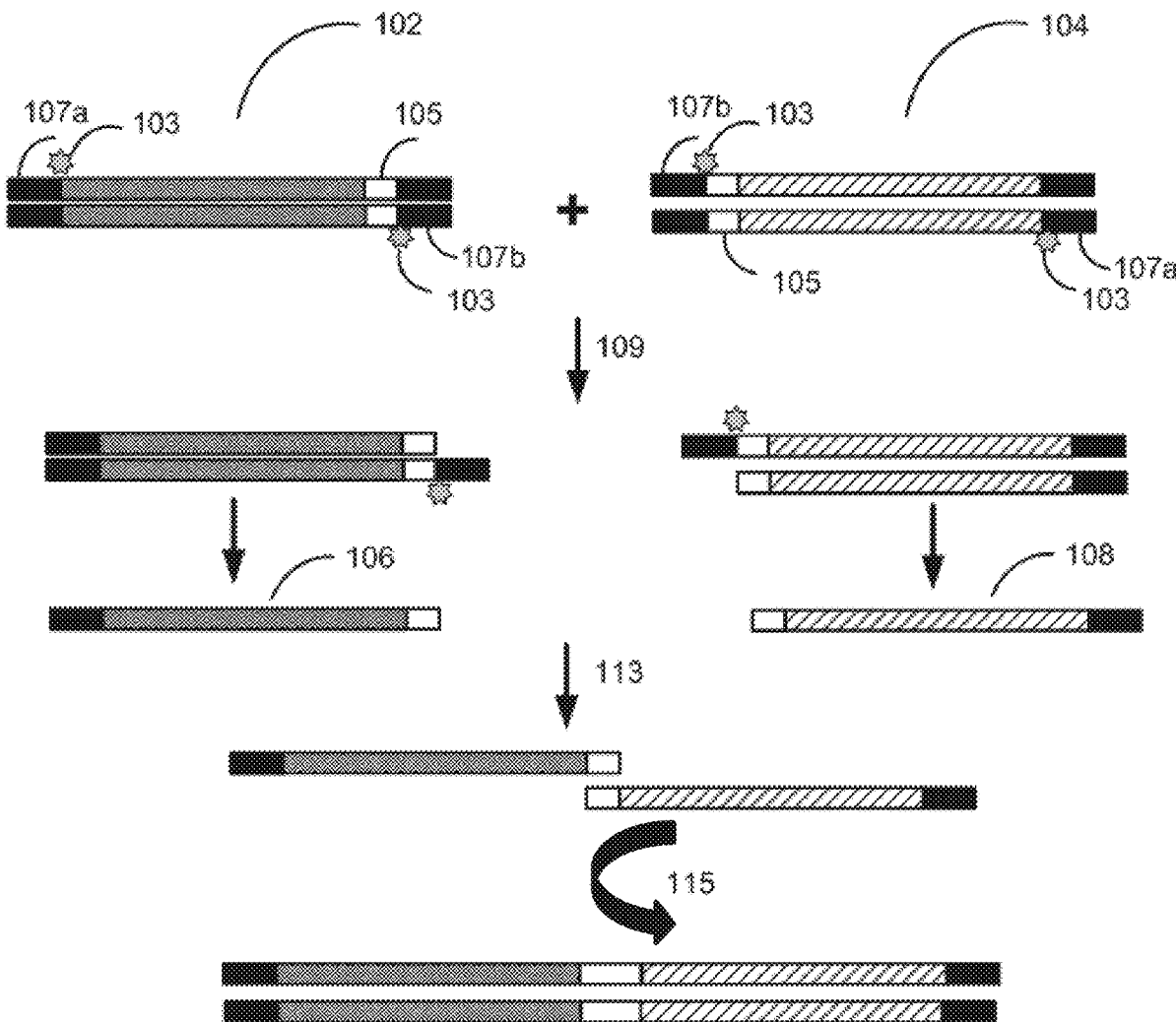
(22) Filed: **May 26, 2022**

Provided herein are methods, systems, and compositions for seamless nucleic acid assembly. Such methods, systems, and compositions for seamless nucleic acid assembly include those for in vitro recombination cloning, single-stranded hierarchal DNA assembly, or overlap extension PCR without primer removal.

Related U.S. Application Data

Specification includes a Sequence Listing.

(60) Division of application No. 16/712,678, filed on Dec. 12, 2019, now Pat. No. 11,377,676, which is a continuation of application No. PCT/US2018/037152, filed on Jun. 12, 2018.



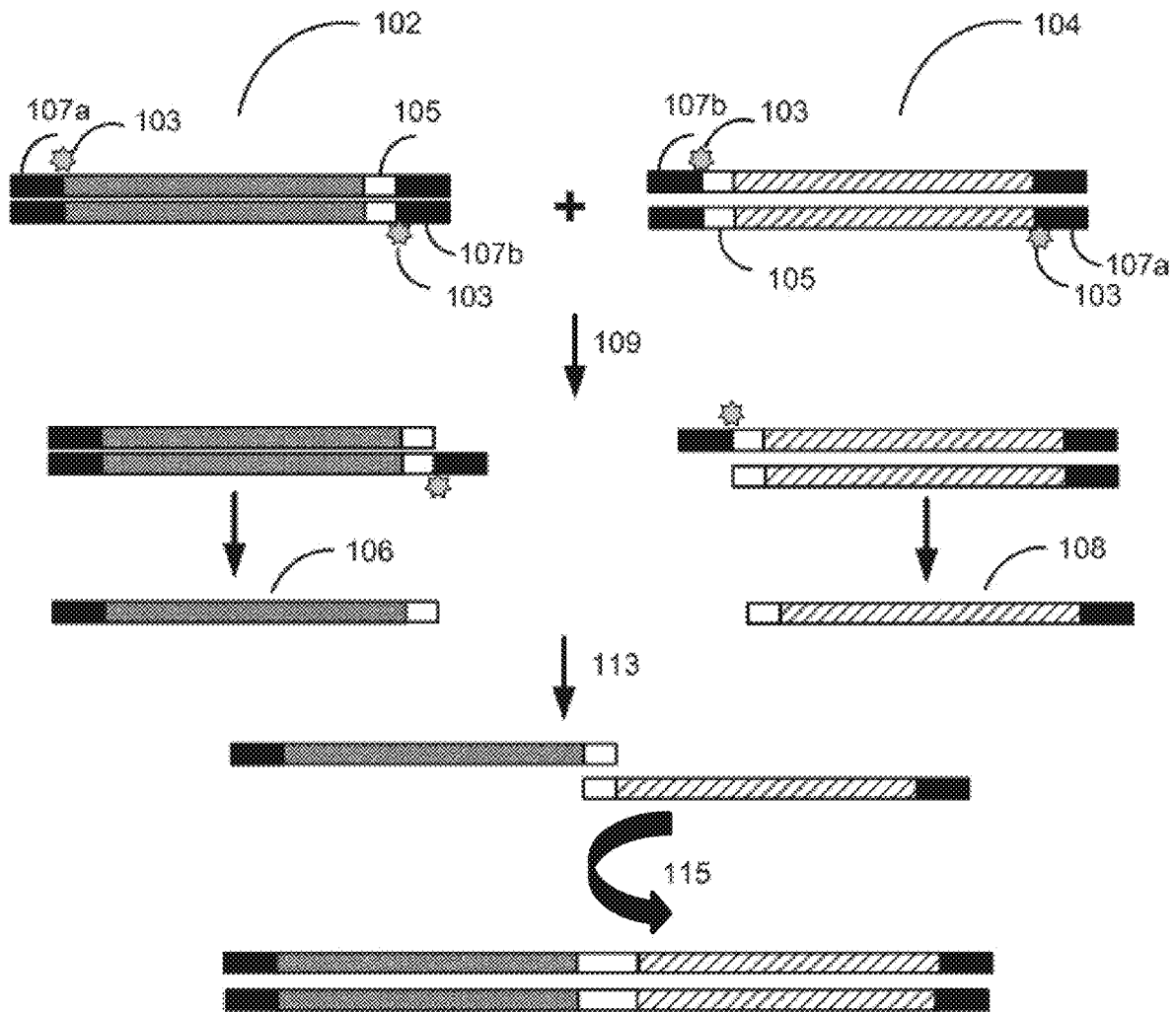


FIG. 1

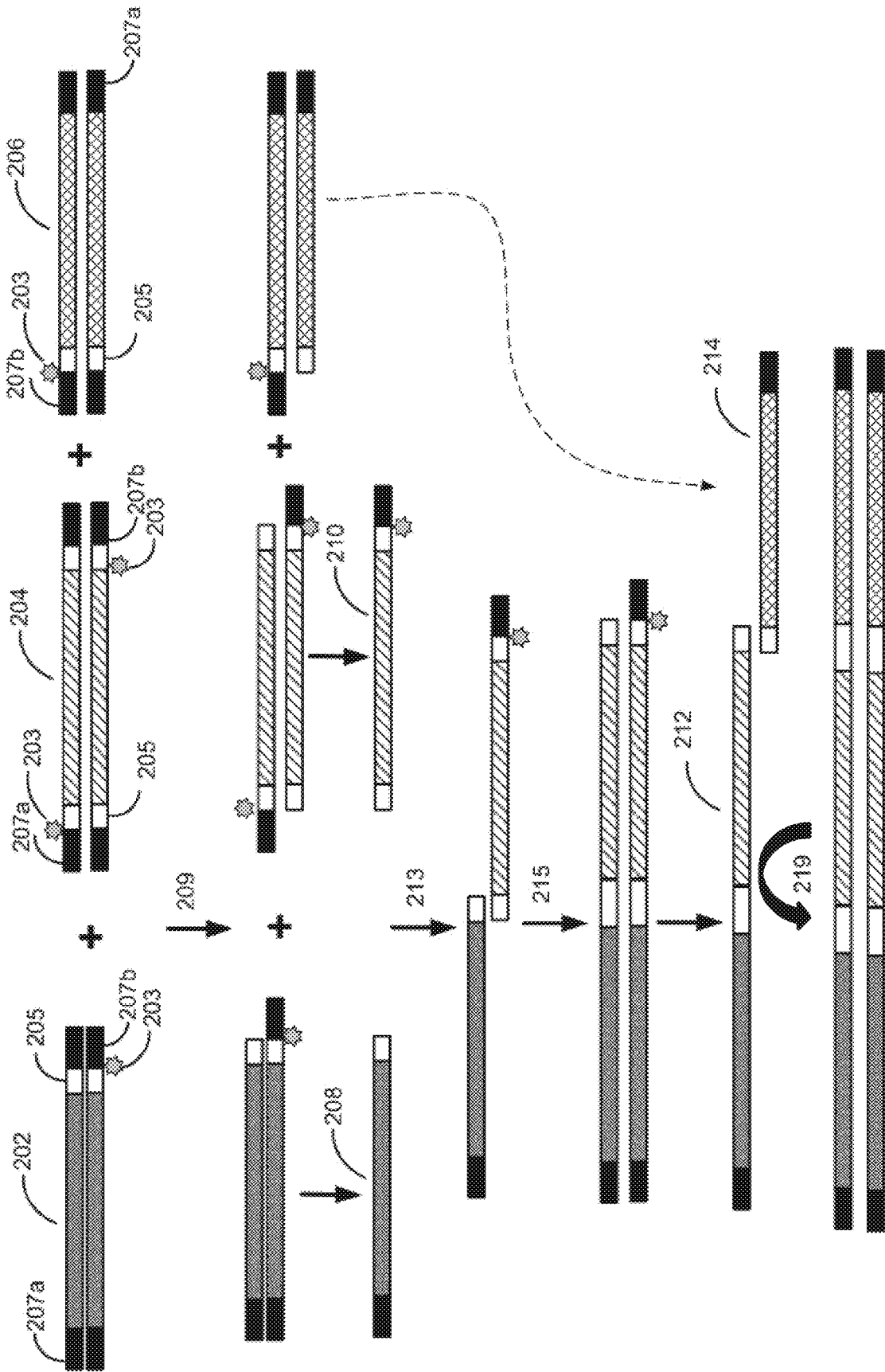


FIG. 2

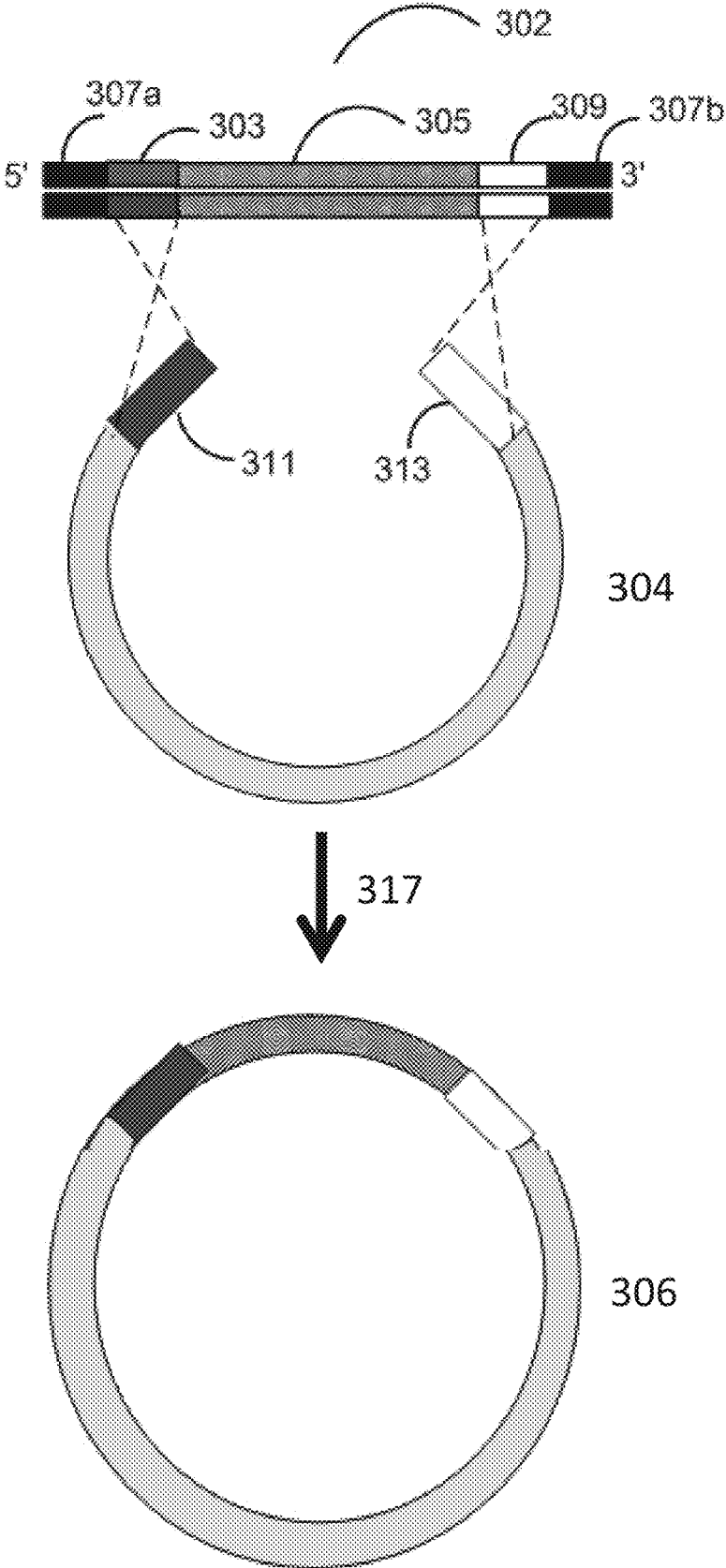


FIG. 3

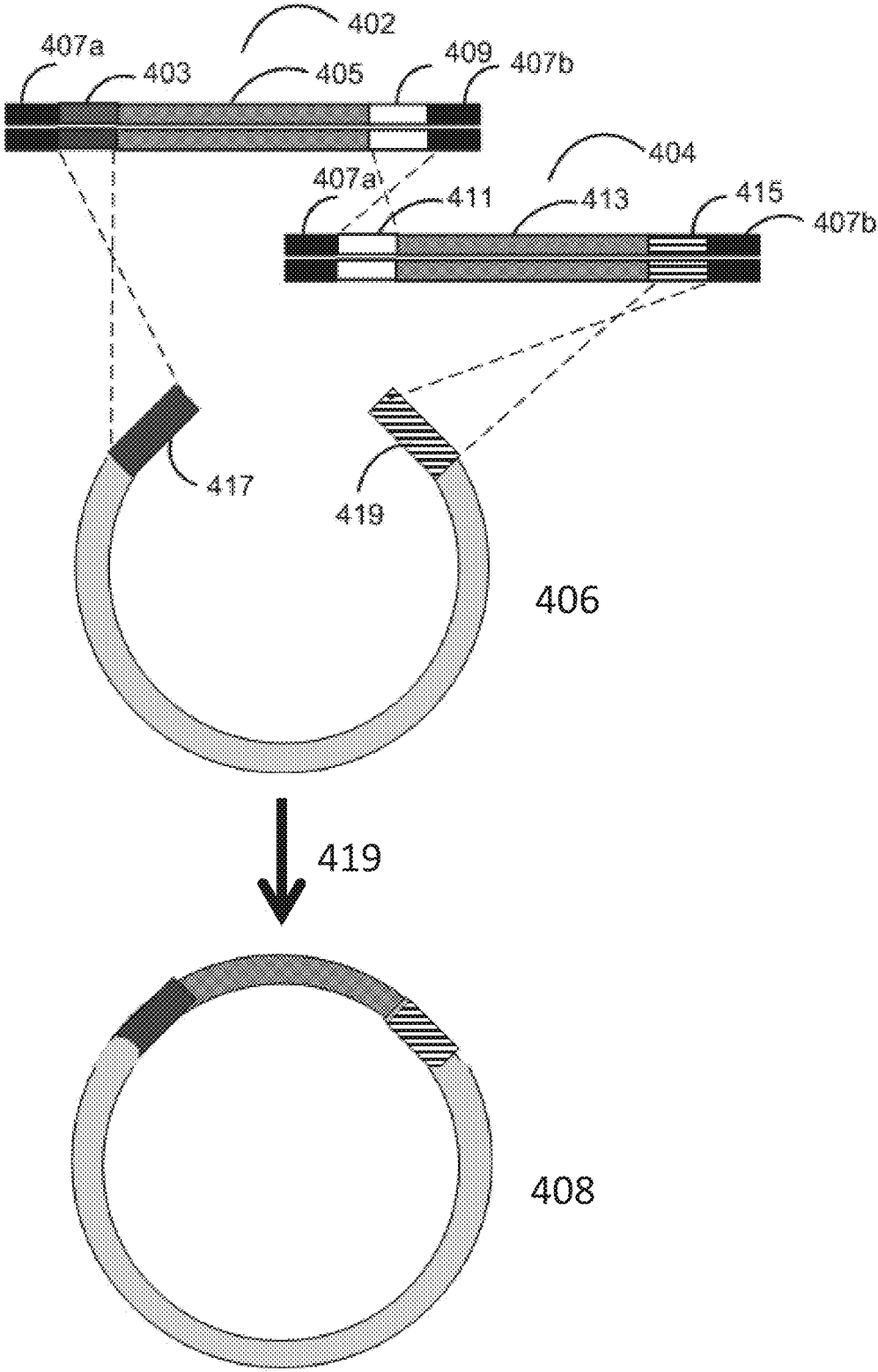


FIG. 4

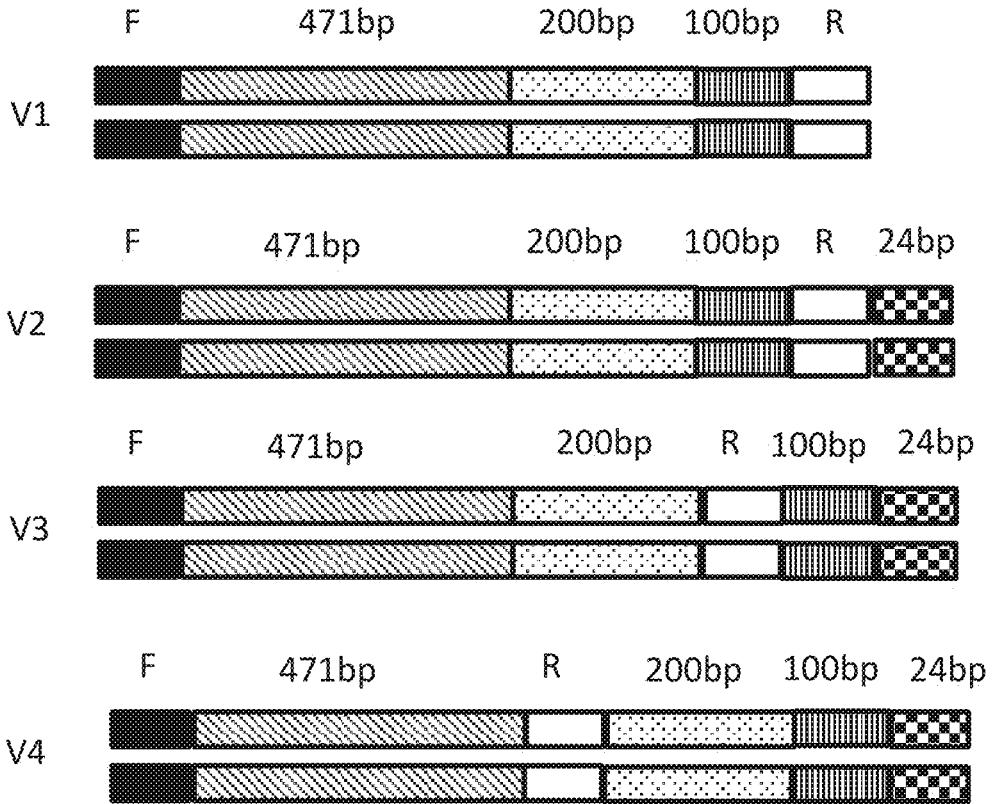


FIG. 5

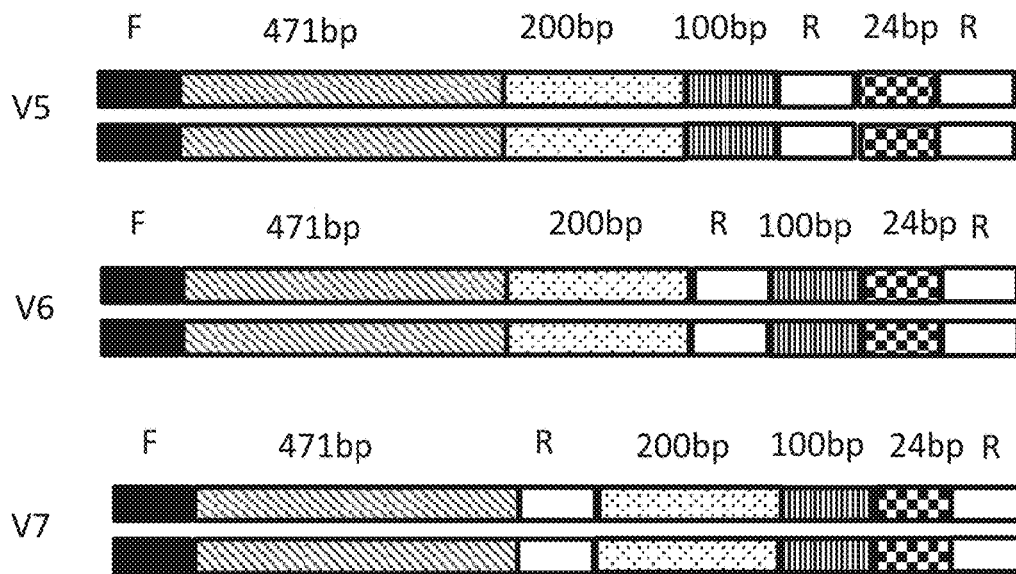


FIG. 6A

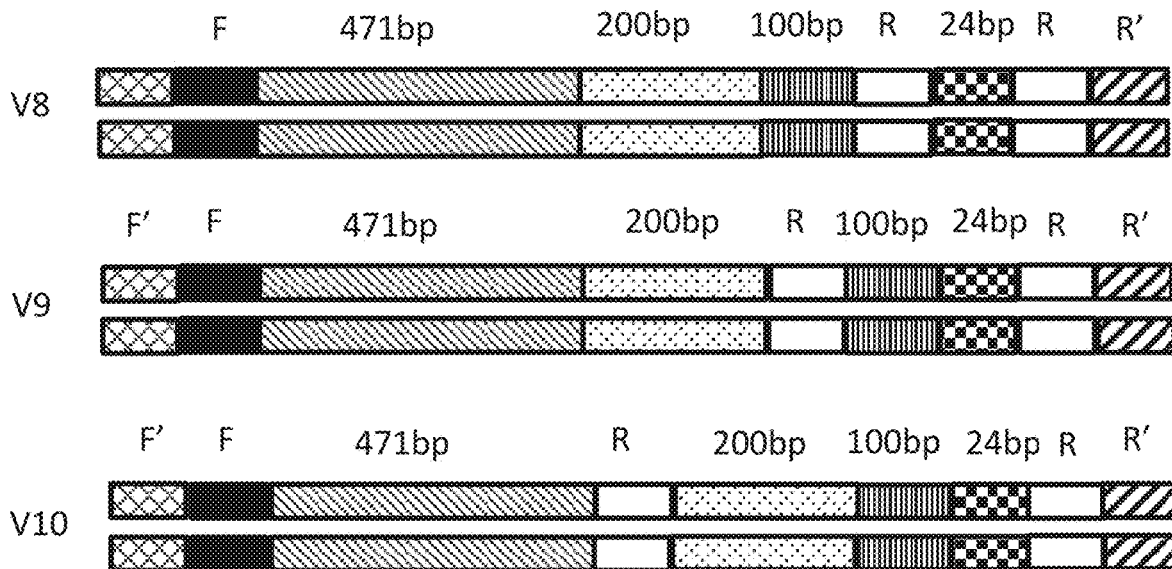


FIG. 6B

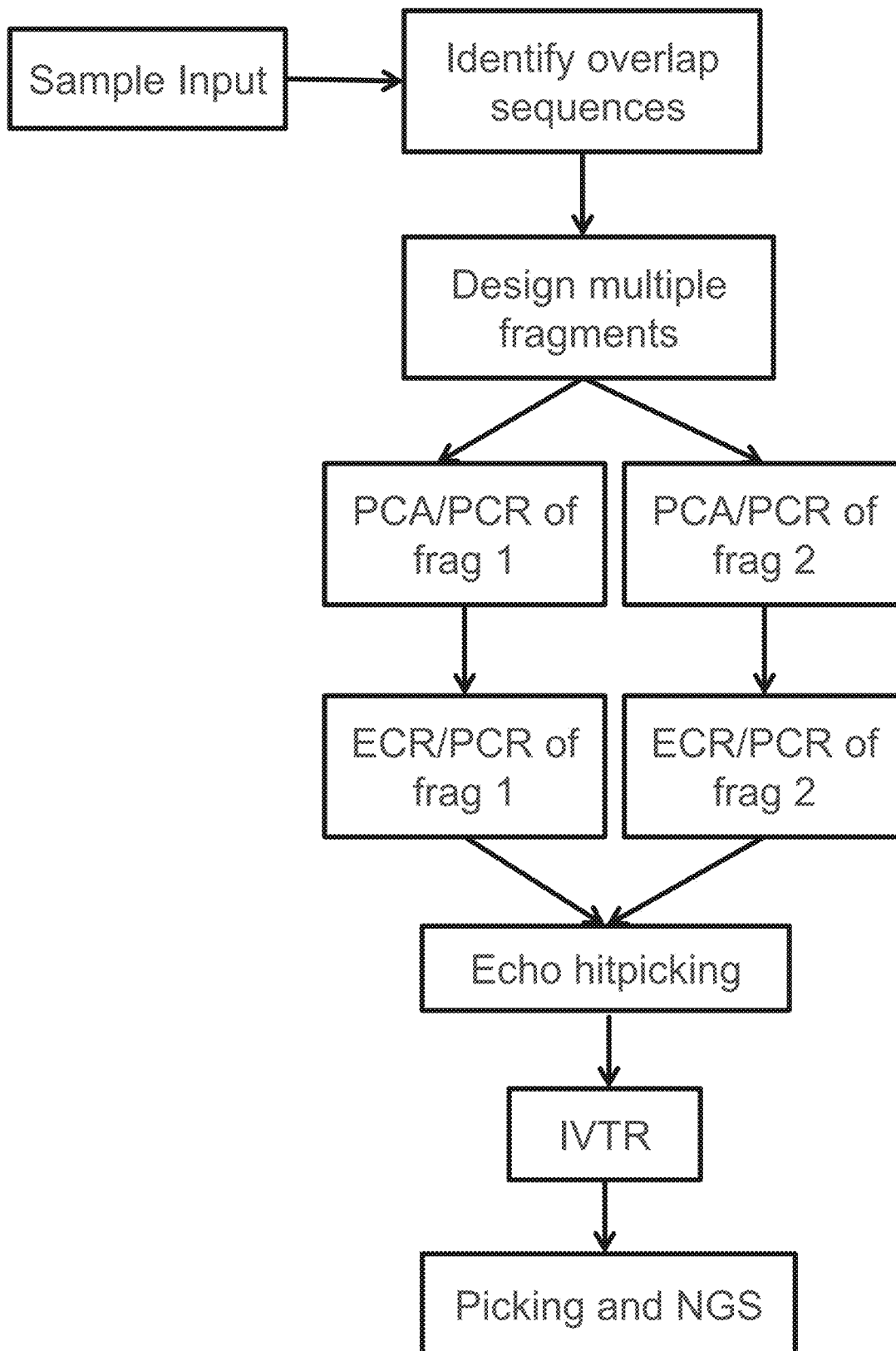


FIG. 7

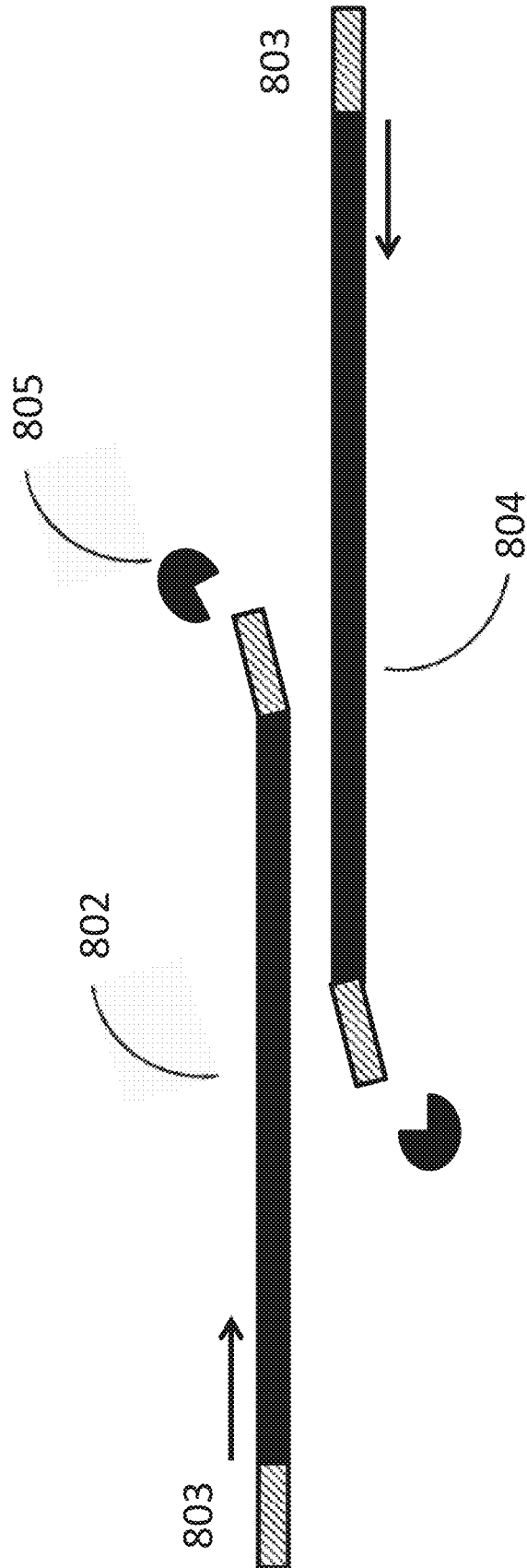


FIG. 8

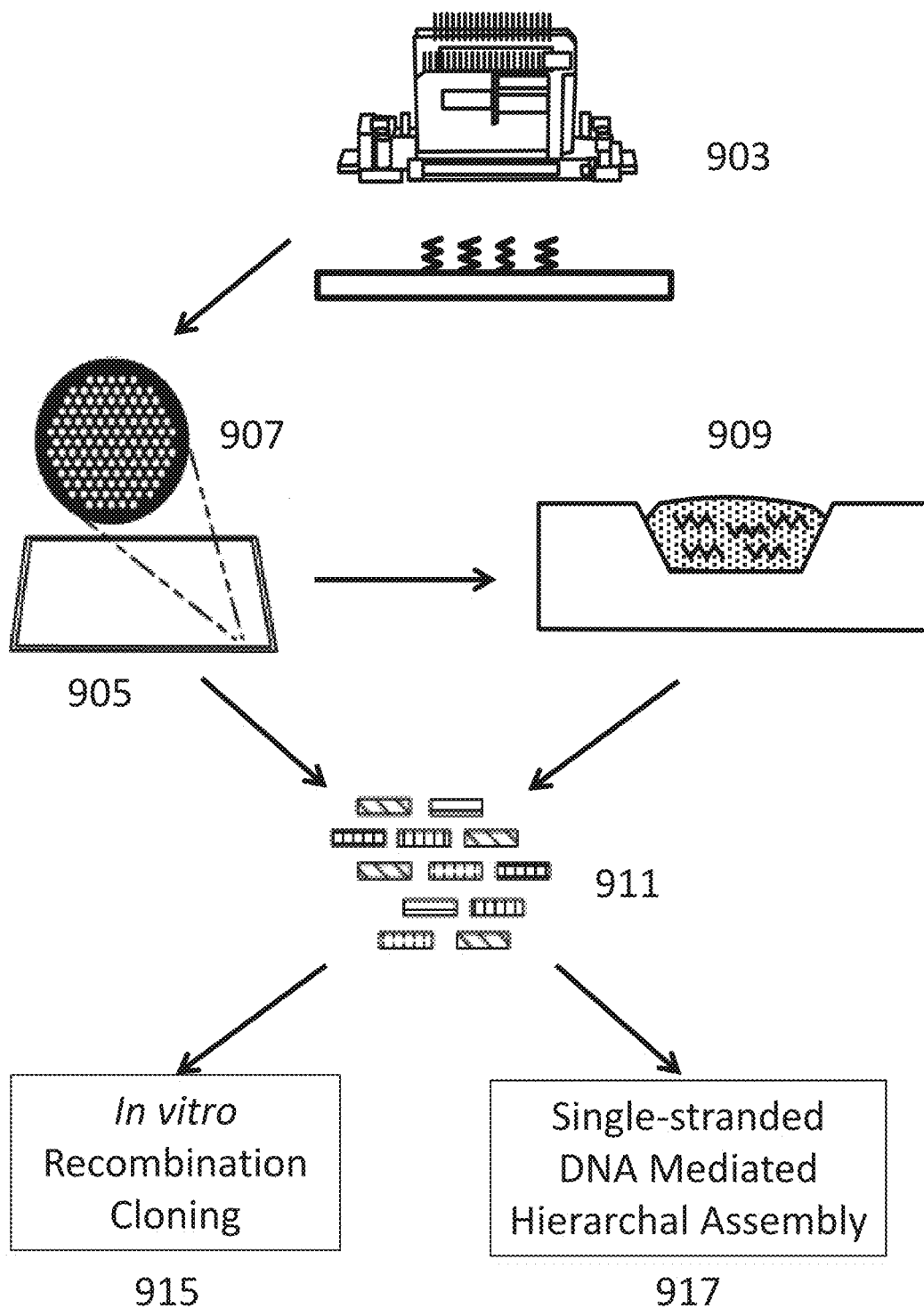


FIG. 9

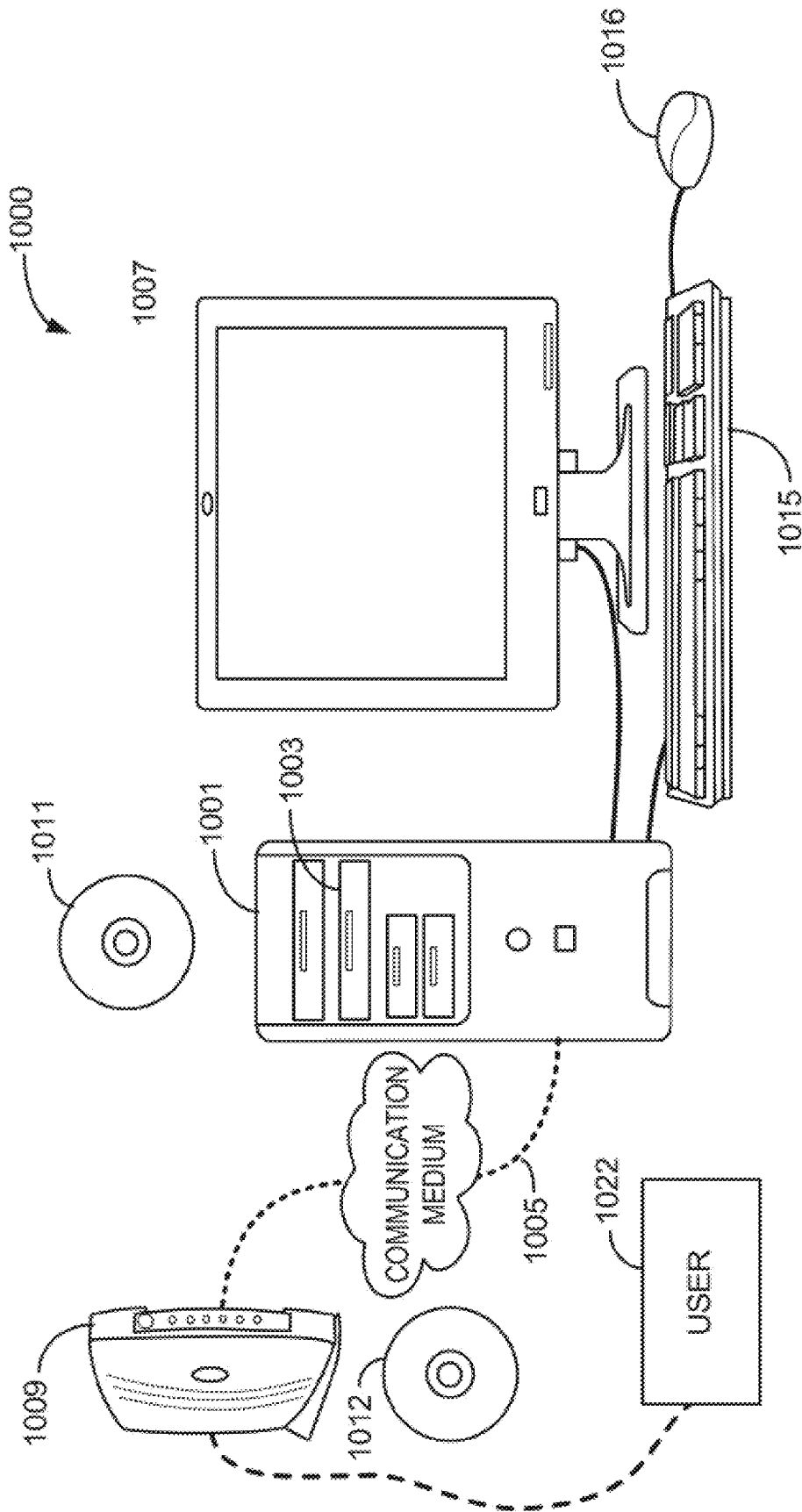


FIG. 10

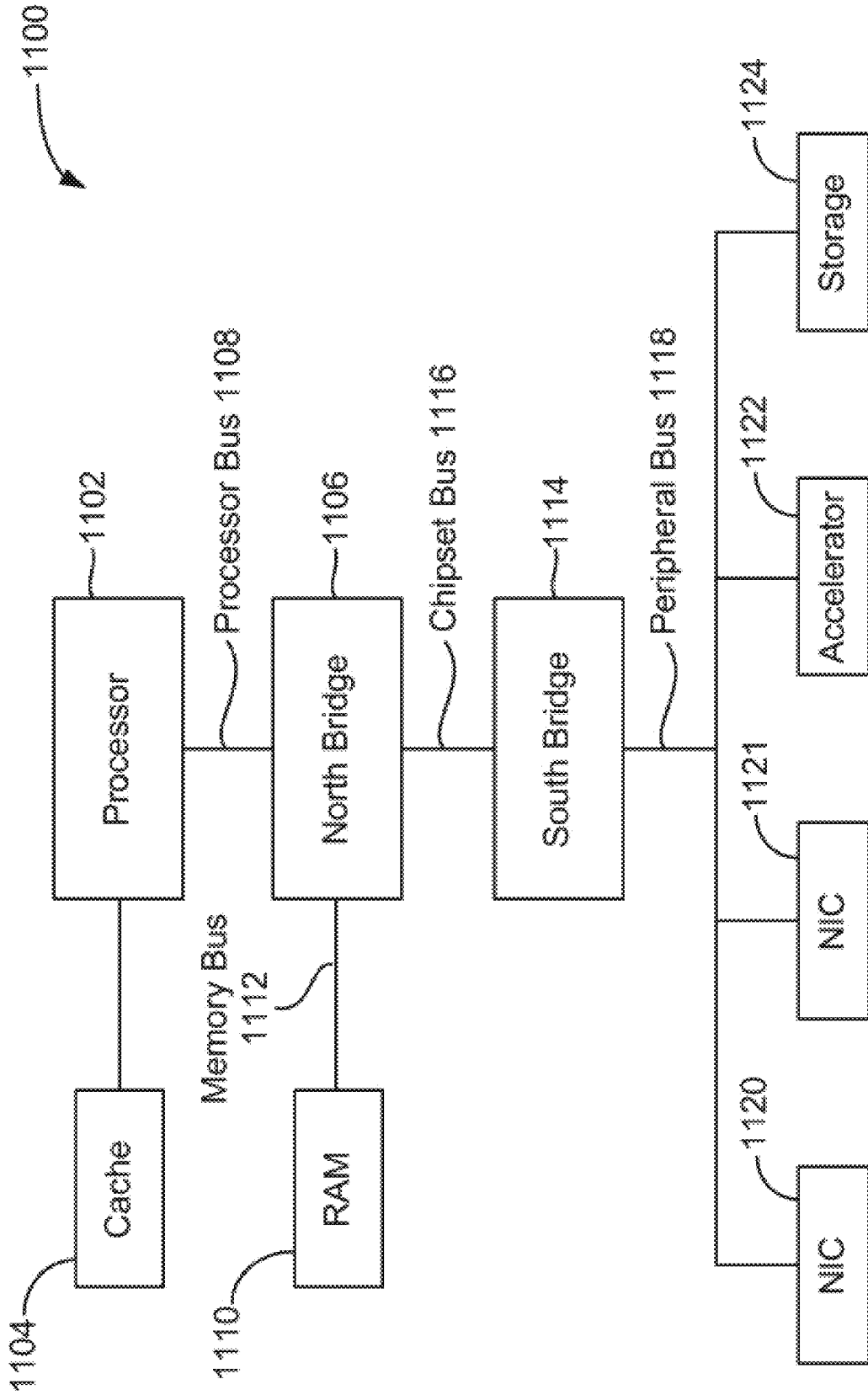


FIG. 11

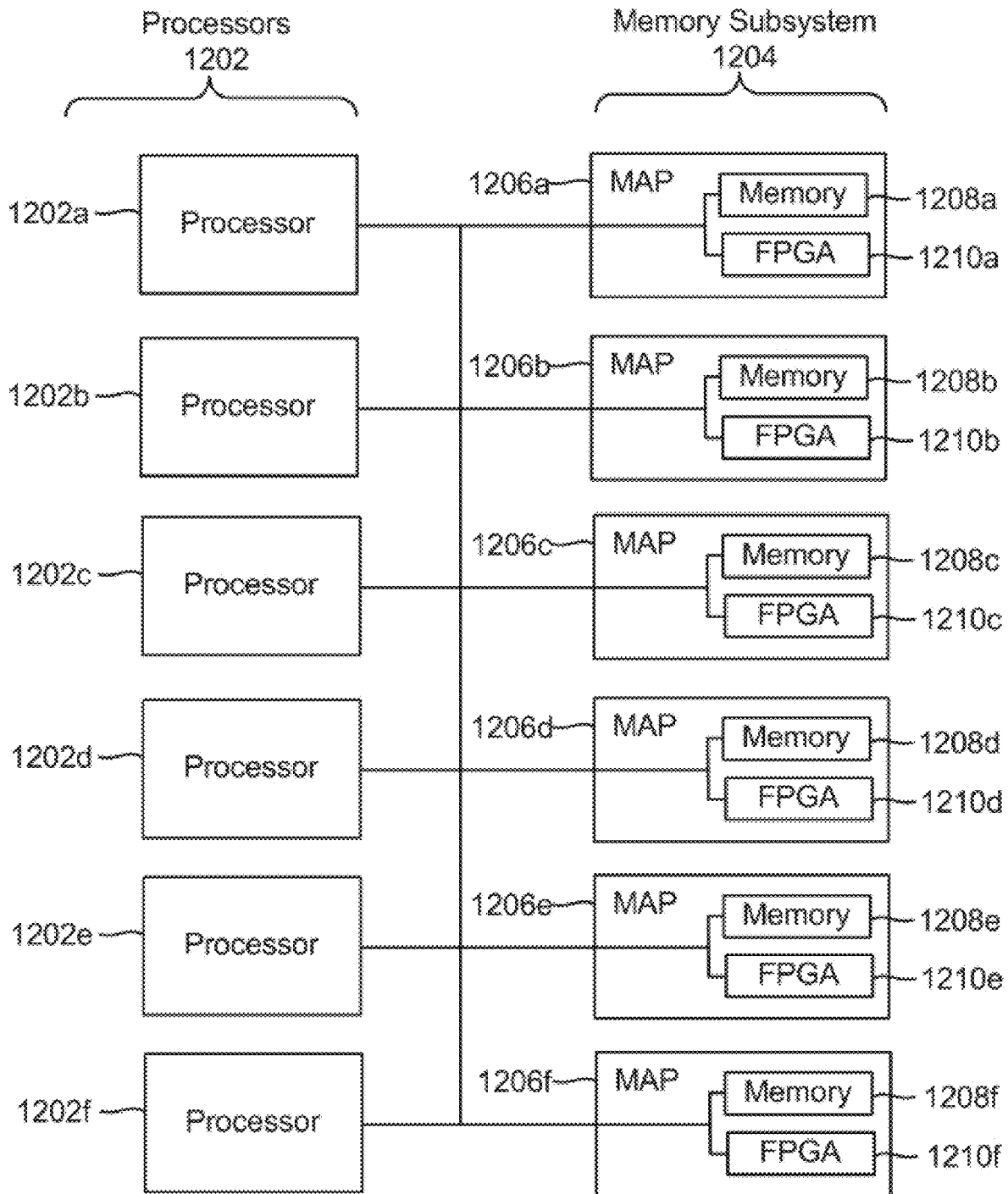


FIG. 12

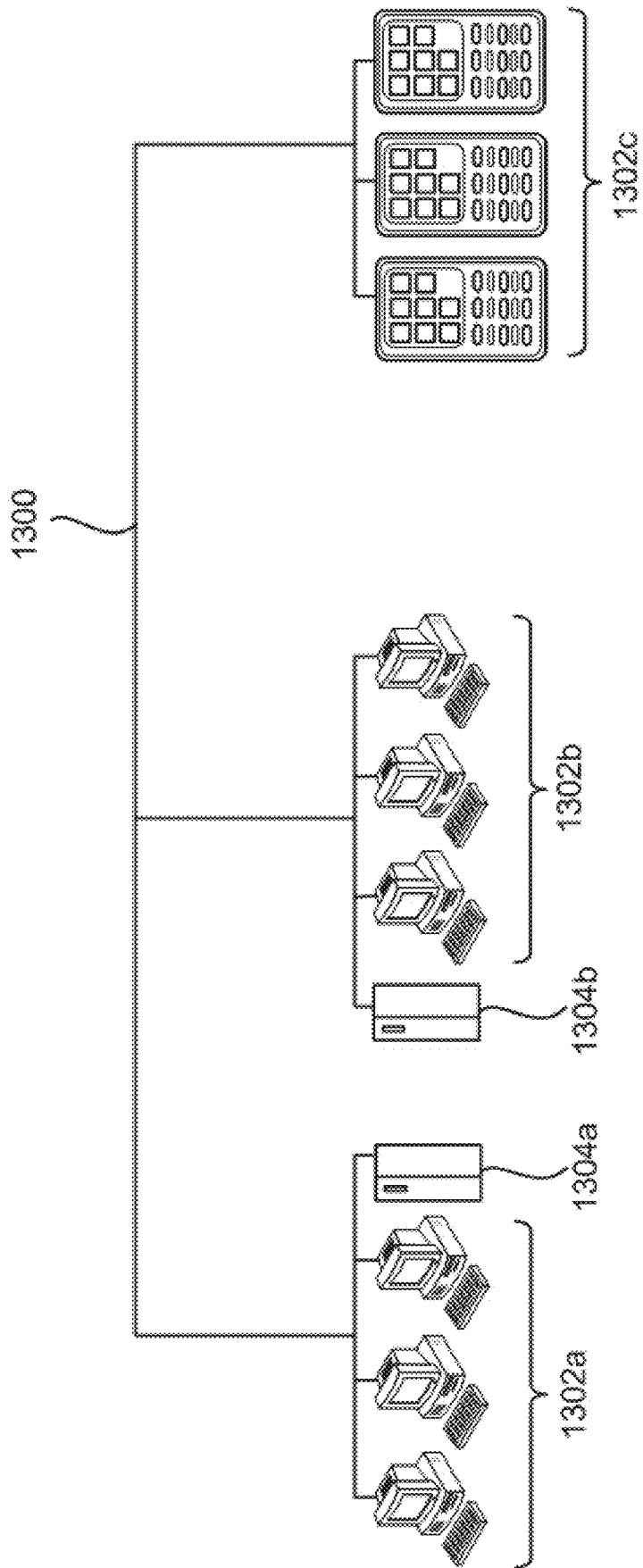


FIG. 13

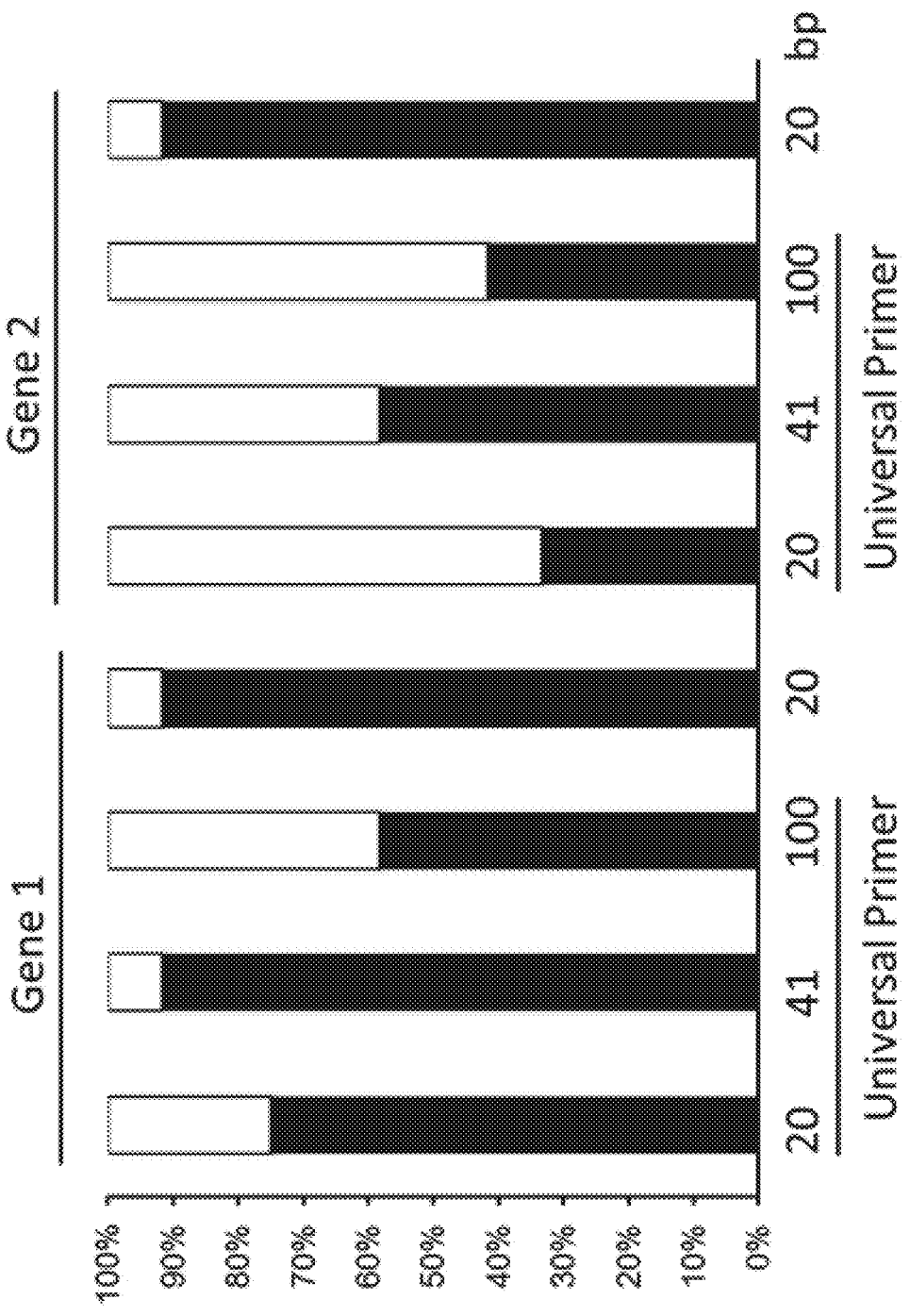


FIG. 14

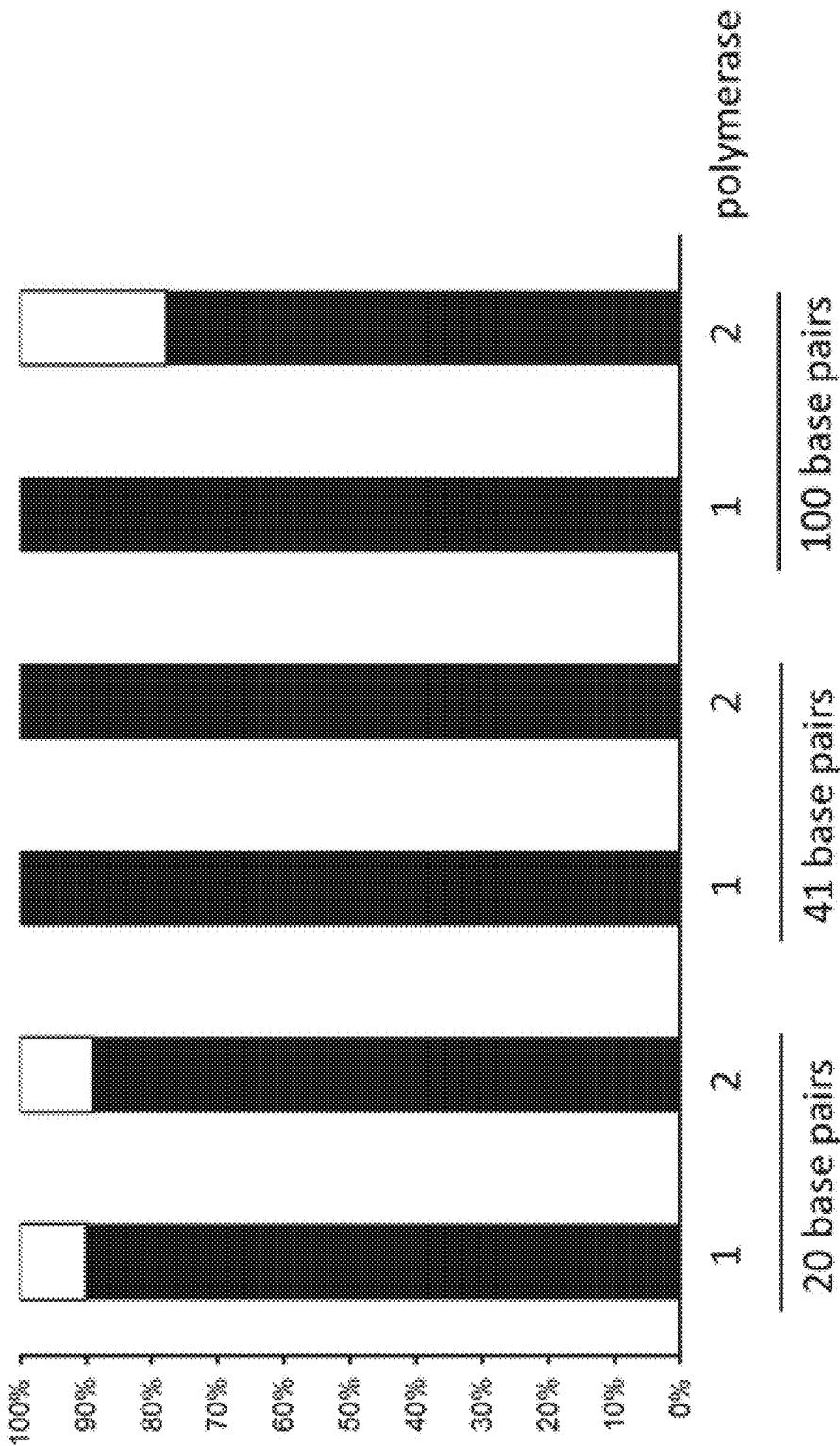
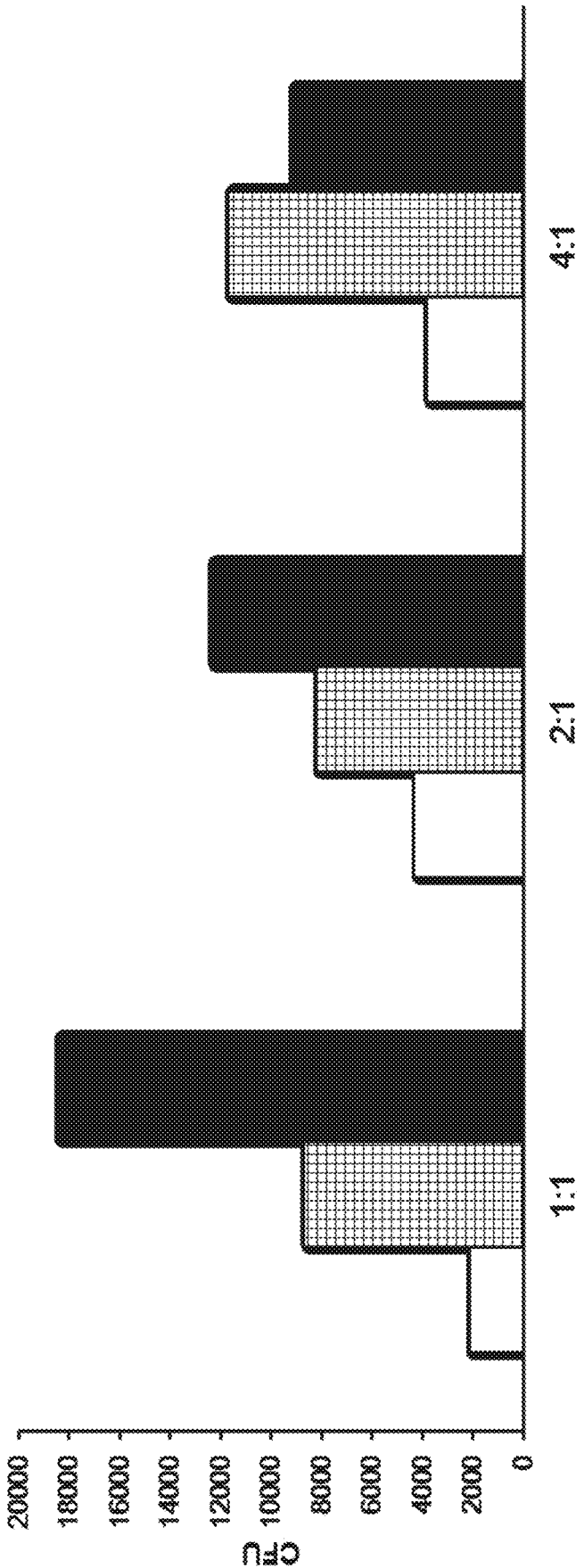


FIG. 15



Insert : Vector Ratio

FIG. 16

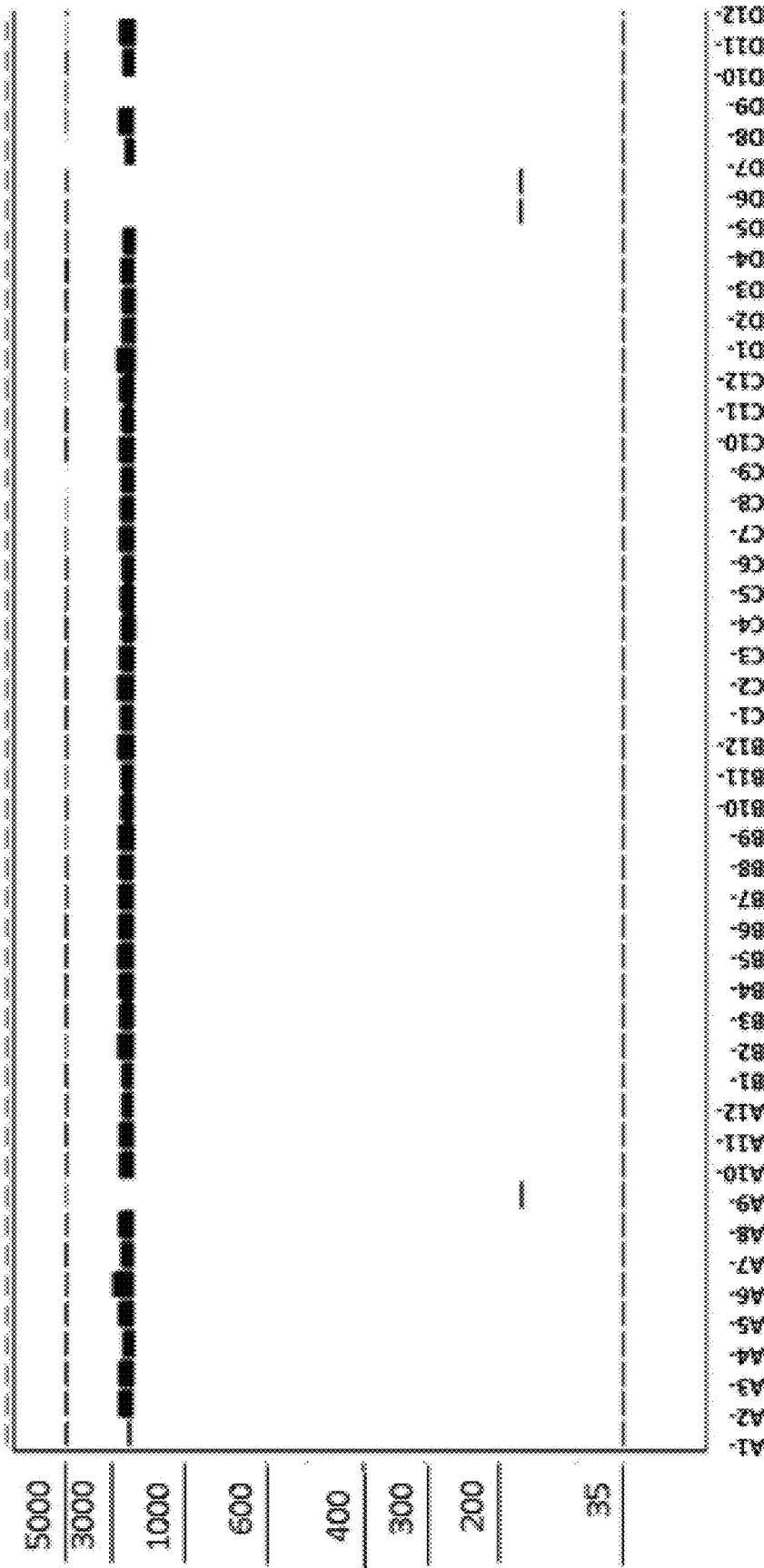


FIG. 17

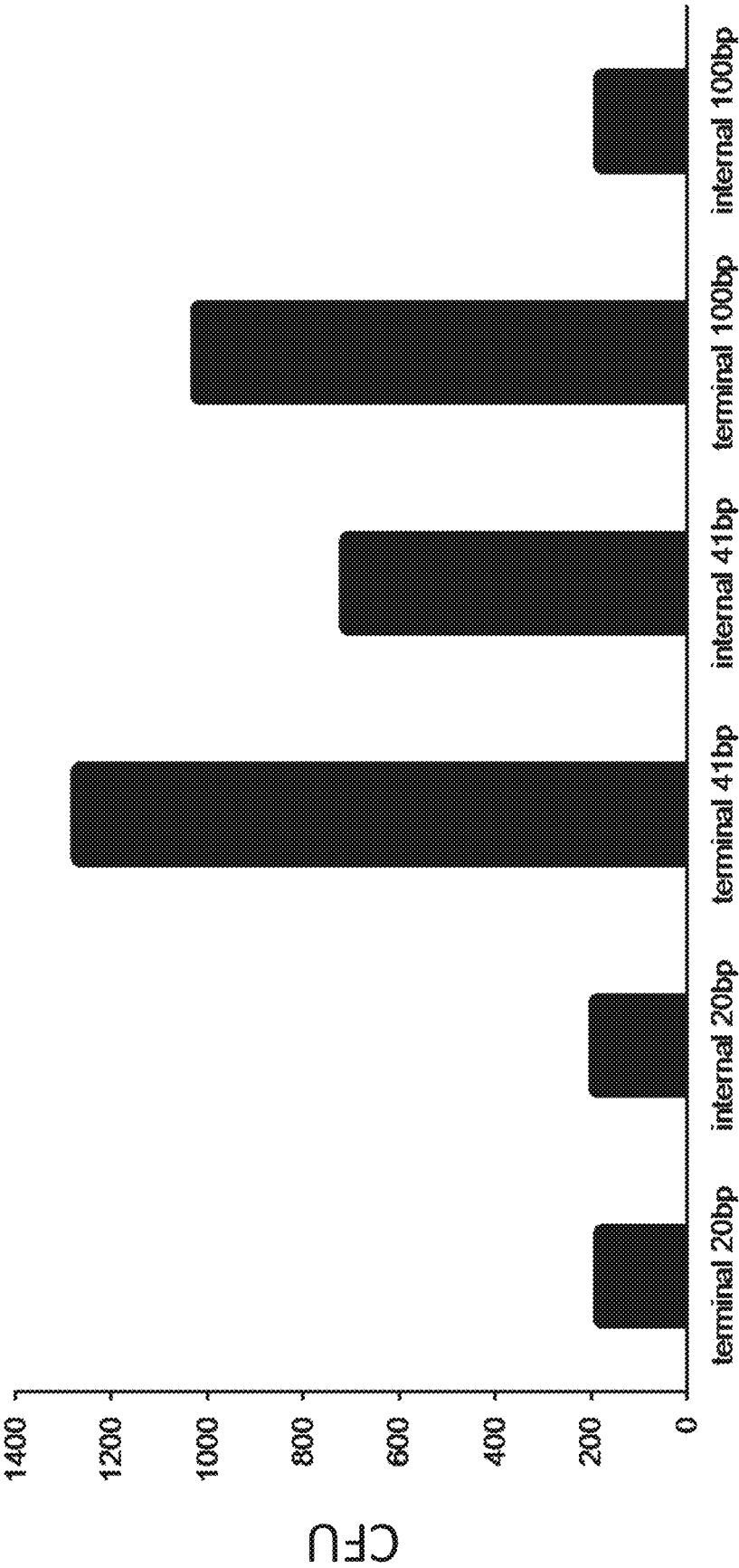


FIG.18

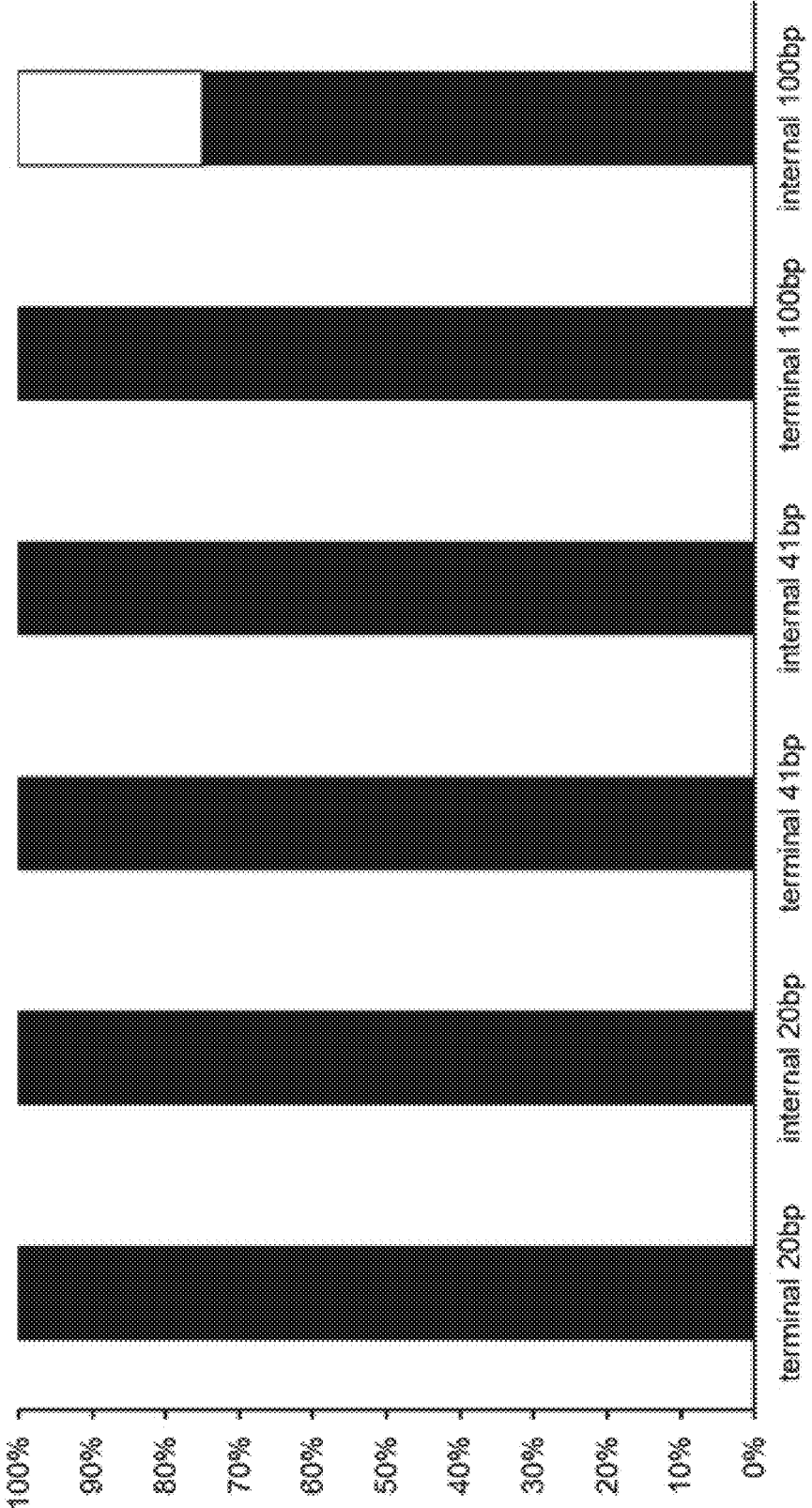


FIG. 19

Correct HA Rate

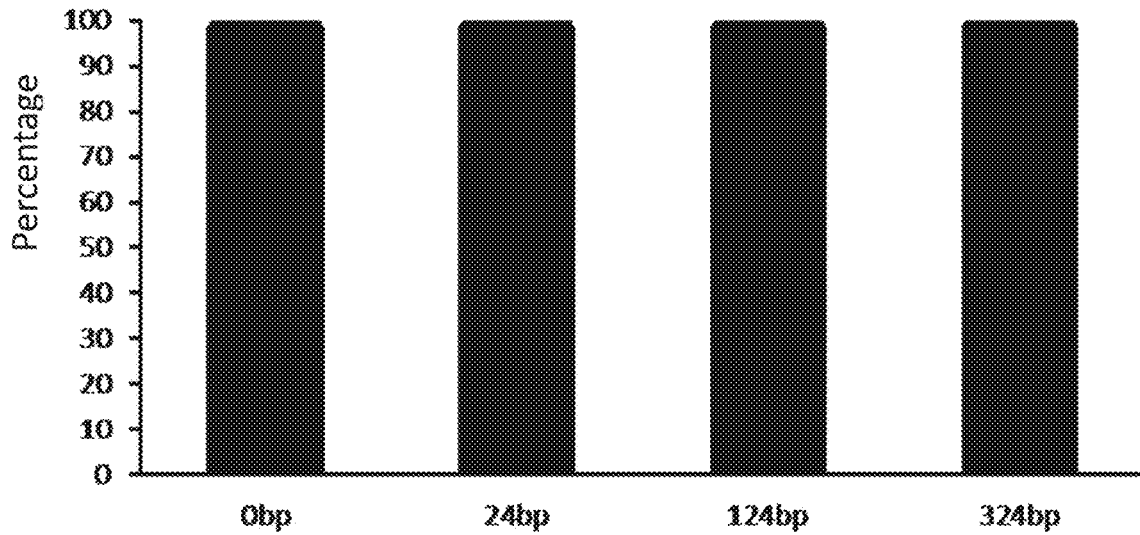


FIG. 20A

Total CFU

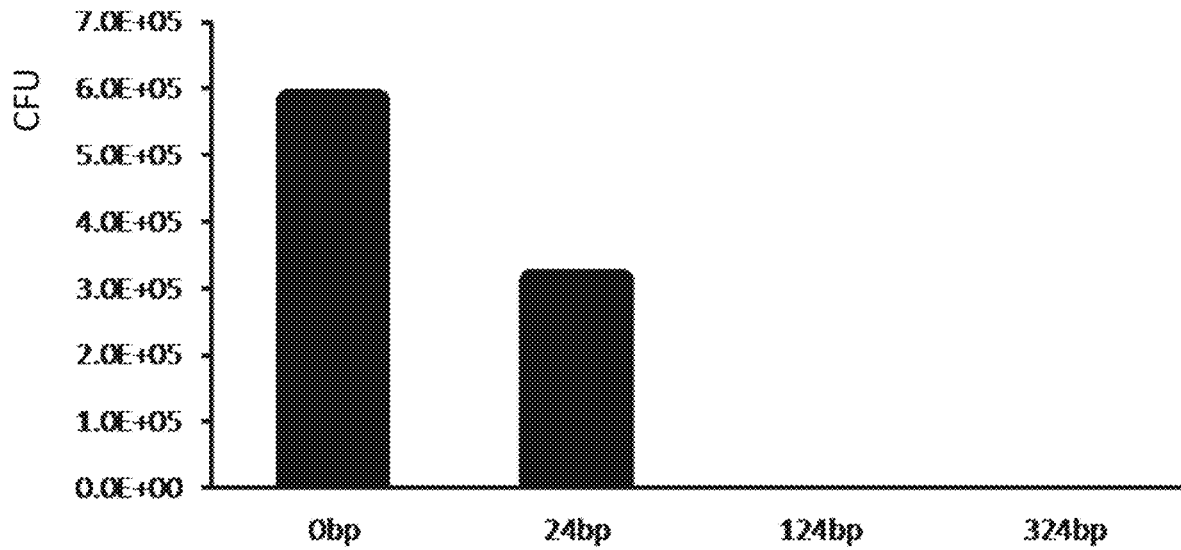


FIG. 20B

Correct HA Rate

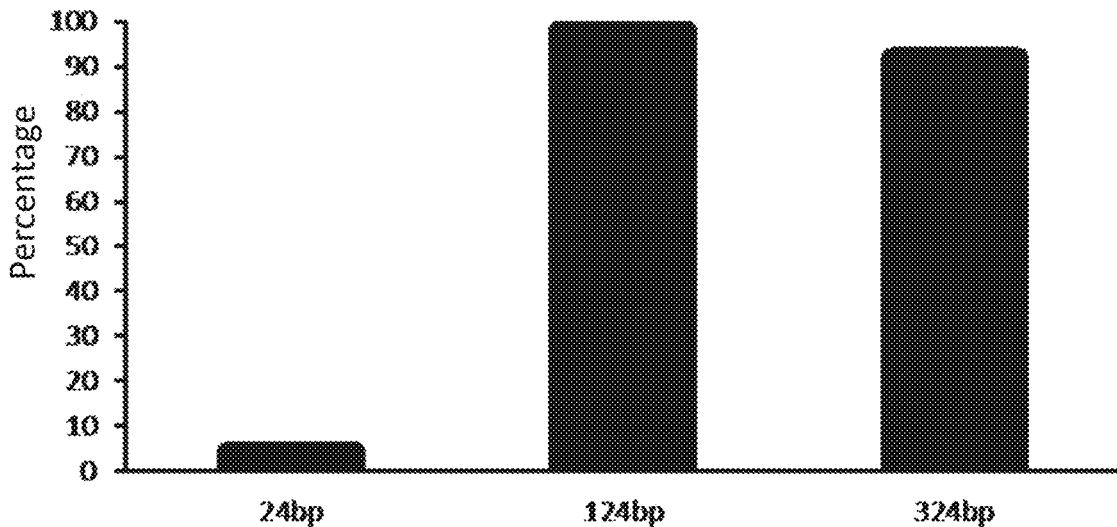


FIG. 21A

Total CFU

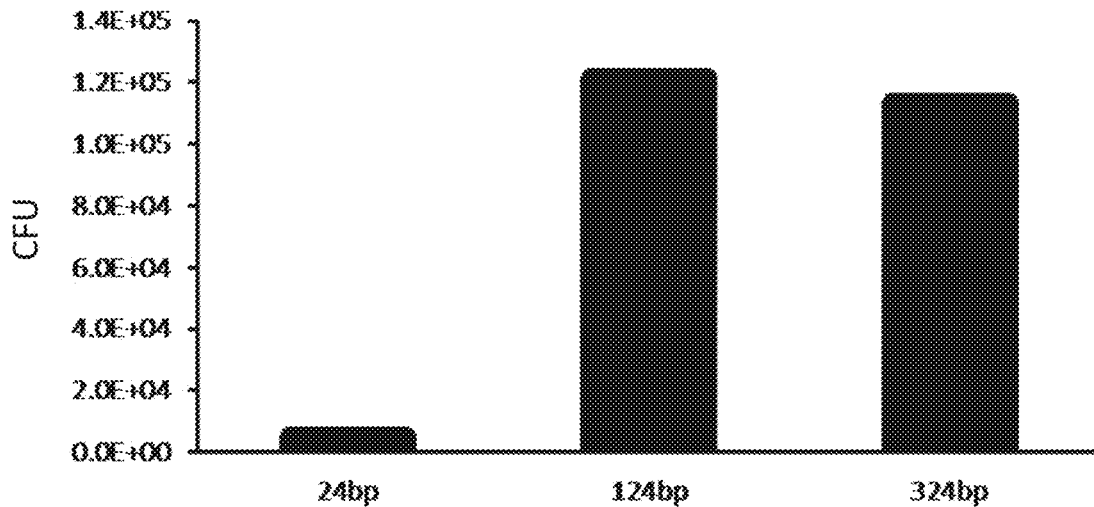


FIG. 21B

METHODS FOR SEAMLESS NUCLEIC ACID ASSEMBLY

CROSS-REFERENCE

[0001] This application is a divisional of U.S. patent application Ser. No. 16/712,678, filed Dec. 12, 2019, which is a continuation of PCT/US2018/37152 filed Jun. 12, 2018, which claims the benefit of U.S. Provisional Patent Application No. 62/518,489 filed on Jun. 12, 2017, which are incorporated herein by reference in their entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on Jun. 6, 2018, is named 44854-743_401_SL.txt and is 1,591 bytes in size.

BACKGROUND

[0003] De novo nucleic acid synthesis is a powerful tool for basic biological research and biotechnology applications. While various methods are known for the synthesis of relatively short fragments of nucleic acids in a small scale, these techniques suffer from scalability, automation, speed, accuracy, and cost. Thus, a need remains for efficient methods of seamless nucleic acid assembly.

BRIEF SUMMARY

[0004] Provided herein is a method for nucleic acid assembly, comprising: (a) providing at least one double stranded nucleic acid comprising in 5' to 3' order: a 5' flanking adapter sequence, a first homology sequence, an insert sequence, a second homology sequence, and a 3' flanking adapter sequence, wherein the first homology sequence and the second homology sequence comprises about 20 to about 100 base pairs in length; (b) providing a vector comprising the first homology sequence and the second homology sequence; and (c) mixing the at least one double stranded nucleic acid and the vector with a bacterial lysate. Further provided herein is a method, wherein the bacterial lysate comprises a nuclease or a recombinase. Further provided herein is a method, wherein the bacterial lysate comprises a nuclease and a recombinase. Further provided herein is a method, wherein the first homology sequence and the second homology sequence each comprises about 20 base pairs. Further provided herein is a method, wherein the first homology sequence and the second homology sequence each comprises about 41 base pairs. Further provided herein is a method, wherein the first homology sequence and the second homology sequence each comprises 30 to 50 base pairs. Further provided herein is a method, wherein the first homology sequence and the second homology sequence each comprises 35 to 45 base pairs. Further provided herein is a method, wherein the first homology sequence and the second homology sequence each comprises about 100 base pairs. Further provided herein is a method, wherein the first homology sequence or the second homology sequence is flanked by the 5' flanking adapter sequence and the 3' flanking adapter sequence. Further provided herein is a method, wherein the first homology sequence or the second homology sequence is at a terminal end. Further provided herein is a method, wherein a percentage of correct assembly is at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.

[0005] Provided herein is a method for nucleic acid assembly comprising: (a) de novo synthesizing a plurality of polynucleotides, wherein each polynucleotide comprises a first homology region that comprises in 5' to 3' order: a 5' flanking adapter sequence, a first homology sequence, an insert sequence, a second homology sequence, and a 3' flanking adapter sequence, wherein the first homology sequence and the second homology sequence each comprises about 20 to about 100 base pairs in length, and wherein each polynucleotide comprises a homology sequence identical to that of another polynucleotide of the plurality of polynucleotides; and (b) mixing of the plurality of polynucleotides with a bacterial lysate to processively form nucleic acids each having a predetermined sequence. Further provided herein is a method, wherein the bacterial lysate comprises a nuclease. Further provided herein is a method, wherein the first homology sequence and the second homology sequence each comprises about 20 base pairs. Further provided herein is a method, wherein the first homology sequence and the second homology sequence each comprises about 41 base pairs. Further provided herein is a method, wherein the first homology sequence and the second homology sequence each comprises about 100 base pairs. Further provided herein is a method, wherein the first homology sequence or the second homology sequence is flanked by the 5' flanking adapter sequence and the 3' flanking adapter sequence. Further provided herein is a method, wherein the first homology sequence or the second homology sequence is at a terminal end. Further provided herein is a method wherein a percentage of correct assembly is at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.

[0006] Provided herein is a method for nucleic acid assembly comprising: (a) providing a first double stranded nucleic acid and a second double stranded nucleic acid, wherein the first double stranded nucleic acid comprises in 5' to 3' order: a 5' flanking adapter sequence, an insert sequence, a homology sequence, and a 3' flanking adapter sequence, and wherein the second double stranded nucleic acid comprises in 5' to 3' order: a 5' flanking adapter sequence, an insert sequence, a homology sequence, and a 3' flanking adapter sequence; (b) annealing a uracil separately to each of (i) a 5' end and a 3' end of the first double stranded nucleic acid and (ii) a 5' end and a 3' end of the second double stranded nucleic acid; (c) amplifying the first double stranded nucleic acid and the second double stranded nucleic acid using a uracil compatible polymerase to form amplification products; (d) mixing the amplification products to form a mixture; and (e) amplifying the mixture using a uracil incompatible polymerase to generate the nucleic acid. Further provided herein is a method, wherein the uracil incompatible polymerase is a DNA polymerase. Further provided herein is a method, wherein a plurality of double stranded nucleic acids is provided. Further provided herein is a method, wherein the homology sequence comprises about 20 to about 100 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 20 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 41 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 100 base pairs. Further provided herein is a method, wherein a percentage of correct assembly is at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.

[0007] Provided herein is a method for nucleic acid assembly comprising: (a) providing predetermined sequences for

a first double stranded nucleic acid and a second double stranded nucleic acid, wherein the first double stranded nucleic acid comprises in 5' to 3' order: a 5' flanking adapter sequence, an insert sequence, a homology sequence, and a 3' flanking adapter sequence, and wherein the second double stranded nucleic acid comprises in 5' to 3' order: a 5' flanking adapter sequence, an insert sequence, a homology sequence, and a 3' flanking adapter sequence; (b) synthesizing a plurality of polynucleotides encoding for the predetermined sequences; (c) annealing a universal primer comprising uracil at a terminal end of the first double stranded nucleic acid and the second double stranded nucleic acid; (d) amplifying the first double stranded nucleic acid and the second double stranded nucleic acid using a uracil incompatible polymerase to form amplification products; (e) mixing the amplification products to form a mixture; and (f) amplifying the mixture to generate the nucleic acid. Further provided herein is a method, wherein the uracil incompatible polymerase is a DNA polymerase. Further provided herein is a method, wherein a plurality of double stranded nucleic acids is provided. Further provided herein is a method, wherein the homology sequence comprises about 20 to about 100 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 20 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 41 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 100 base pairs. Further provided herein is a method, wherein a percentage of correct assembly is at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.

[0008] Provided herein is a method for nucleic acid assembly comprising: (a) providing a plurality of double stranded nucleic acids; (b) annealing a uracil at a 5' end and a 3' end of at least two of the double stranded nucleic acids; (c) amplifying the double stranded nucleic acids using a uracil compatible polymerase to form amplification products; (d) mixing the amplification products from step (c) to form a mixture; and (e) amplifying the mixture from step (d) using a uracil incompatible polymerase to generate a single-stranded nucleic acid. Further provided herein is a method, wherein the uracil incompatible polymerase is a DNA polymerase. Further provided herein is a method, wherein a plurality of double stranded nucleic acids is provided. Further provided herein is a method, wherein the homology sequence comprises about 20 to about 100 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 20 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 41 base pairs. Further provided herein is a method, wherein the homology sequence comprises about 100 base pairs. Further provided herein is a method, wherein a percentage of correct assembly is at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.

INCORPORATION BY REFERENCE

[0009] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] FIG. 1 depicts a schematic of single-stranded DNA mediated hierarchal assembly with two fragments.

[0011] FIG. 2 depicts a schematic of single-stranded DNA mediated hierarchal assembly with three fragments.

[0012] FIG. 3 depicts a schematic for in vitro recombination cloning with a single gene fragment.

[0013] FIG. 4 depicts a schematic for in vitro recombination cloning with two gene fragments.

[0014] FIG. 5 depicts gene fragment designs with varying lengths of non-homologous sequences.

[0015] FIGS. 6A-6B depict gene fragment designs of internal homology sequences.

[0016] FIG. 7 depicts a workflow for in vitro recombination cloning.

[0017] FIG. 8 depicts a schematic of overlap extension polymerase chain reaction without primer removal.

[0018] FIG. 9 depicts systems for polynucleotide synthesis and seamless nucleic acid assembly.

[0019] FIG. 10 illustrates a computer system.

[0020] FIG. 11 is a block diagram illustrating architecture of a computer system.

[0021] FIG. 12 is a block diagram of a multiprocessor computer system using a shared virtual address memory space.

[0022] FIG. 13 is a diagram demonstrating a network configured to incorporate a plurality of computer systems, a plurality of cell phones and personal data assistants, and Network Attached Storage (NAS).

[0023] FIG. 14 is a plot of correct assembly (black bars) and incorrect assembly (white bars) following overlap extension polymerase chain reaction without primer removal for two genes (Gene 1 and Gene 2). Homology sequence length includes 20, 41, and 100 base pairs.

[0024] FIG. 15 is a plot of correct assembly (black bars) and incorrect assembly (white bars) following single-stranded DNA mediated hierarchal assembly using Q5 DNA polymerase and KapaHiFi polymerase enzymes. Homology sequence length includes 20, 41, and 100 base pairs.

[0025] FIG. 16 is a plot of colony forming units (CFU, Y-axis) versus insert: vector ratio (X-axis). An amount of insert includes 13 fmol (white bars), 26 fmol (hashed bars), and 40 fmol (black bars).

[0026] FIG. 17 is an image capture of a capillary gel electrophoresis following in vitro recombination cloning.

[0027] FIG. 18 is a plot of colony forming units (CFU) of homology sequences comprising 20, 41, or 100 base pairs. Homology sequences are flanked by universal primers (internal) or at a 5' or 3' end of an insert (terminal).

[0028] FIG. 19 is a plot of correct assembly (black bars) and incorrect assembly (white bars) following in vitro recombination cloning. Homology sequences comprise 20, 41, or 100 base pairs and are flanked by universal primers (internal) or at a 5' or 3' end of an insert (terminal).

[0029] FIG. 20A is a plot of percentage of hierarchal assembly (HA) for non-homologous sequences comprising 0, 24, 124, or 324 base pair lengths.

[0030] FIG. 20B is a plot of colony forming units (CFU) for non-homologous sequences comprising 0, 24, 124, or 324 base pair lengths.

[0031] FIG. 21A is a plot of percentage of hierarchal assembly (HA) for internal sequences comprising 24, 124, or 324 base pair lengths.

[0032] FIG. 21B is a plot of colony forming units (CFU) for internal sequences comprising 24, 124, or 324 base pair lengths.

DETAILED DESCRIPTION

Definitions

[0033] Throughout this disclosure, various embodiments are presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of any embodiments. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range to the tenth of the unit of the lower limit unless the context clearly dictates otherwise. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual values within that range, for example, 1.1, 2, 2.3, 5, and 5.9. This applies regardless of the breadth of the range. The upper and lower limits of these intervening ranges may independently be included in the smaller ranges, and are also encompassed within the invention, subject to any specifically excluded limit in the stated range. Where the stated range includes one or both of the limits, ranges excluding either or both of those included limits are also included in the invention, unless the context clearly dictates otherwise.

[0034] The terminology used herein is for the purpose of describing particular embodiments only and is not intended to be limiting of any embodiment. As used herein, the singular forms “a,” “an” and “the” are intended to include the plural forms as well, unless the context clearly indicates otherwise. It will be further understood that the terms “comprises” and/or “comprising,” when used in this specification, specify the presence of stated features, integers, steps, operations, elements, and/or components, but do not preclude the presence or addition of one or more other features, integers, steps, operations, elements, components, and/or groups thereof. As used herein, the term “and/or” includes any and all combinations of one or more of the associated listed items.

[0035] Unless specifically stated or obvious from context, as used herein, the term “nucleic acid” as used herein encompass double- or triple-stranded nucleic acids, as well as single-stranded molecules. In double- or triple-stranded nucleic acids, the nucleic acid strands need not be coextensive (i.e., a double-stranded nucleic acid need not be double-stranded along the entire length of both strands). Nucleic acid sequences, when provided, are listed in the 5' to 3' direction, unless stated otherwise. Methods described herein provide for the generation of isolated nucleic acids. Methods described herein additionally provide for the generation of isolated and purified nucleic acids. A “nucleic acid” as referred to herein can comprise at least 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, or more bases in length. Moreover, provided herein are methods for the synthesis of any number of polypeptide-segments encoding nucleotide sequences, including sequences encoding non-ribosomal peptides (NRPs), sequences encoding non-ribosomal peptide-synthetase (NRPS) modules and synthetic variants, polypeptide segments of other modular proteins, such as antibodies, polypeptide segments from other protein families, including non-coding DNA or RNA, such as regulatory sequences e.g.

promoters, transcription factors, enhancers, siRNA, shRNA, RNAi, miRNA, small nucleolar RNA derived from microRNA, or any functional or structural DNA or RNA unit of interest. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, intergenic DNA, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), small nucleolar RNA, ribozymes, complementary DNA (cDNA), which is a DNA representation of mRNA, usually obtained by reverse transcription of messenger RNA (mRNA) or by amplification; DNA molecules produced synthetically or by amplification, genomic DNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. cDNA encoding for a gene or gene fragment referred herein may comprise at least one region encoding for exon sequences without an intervening intron sequence in the genomic equivalent sequence.

[0036] Unless specifically stated or obvious from context, as used herein, the term “about” in reference to a number or range of numbers is understood to mean the stated number and numbers $\pm 10\%$ thereof, or 10% below the lower listed limit and 10% above the higher listed limit for the values listed for a range.

[0037] Seamless Assembly of Nucleic Acids

[0038] Provided herein are methods for assembly of nucleic acids with increased efficiency and accuracy. Further provided herein are methods of assembly of nucleic acids into long genes. De novo synthesized polynucleotides as described herein are assembled into nucleic acids by in vitro recombination cloning, single-stranded DNA mediated hierarchal assembly, or overlap extension. Generally, methods for nucleic acid assembly as described herein do not require primer removal.

[0039] A first exemplary process for seamless assembly of nucleic acids is depicted in FIG. 1. Single-stranded DNA mediated hierarchal assembly is performed with a first gene fragment **102** and second gene fragment **104** comprising a homology sequence **105**. In this workflow, the first gene fragment **102** and the second gene fragment **104** are double-stranded and comprise a 5' flanking adapter sequence **107a** and a 3' flanking adapter sequence **107b** comprising uracil **103**. The first gene fragment **102** and the second gene fragment **104** are amplified with primers and a uracil compatible polymerase. In some instances, the uracil compatible polymerase is Phusion U or Kapa Uracil. The resultant PCR product comprises a uracil at the end of the 3' flanking adapter sequence **107b**. The first gene fragment **102** and the second gene fragment **104** are diluted, mixed, and amplified **109** with a primer and a uracil incompatible polymerase that stalls at a uracil. In some instances, the uracil incompatible polymerase is Q5 DNA polymerase. The resultant fragments **106**, **108** that do not comprise uracil, serve as primers for each other and are combined **113** and amplified **115** to generate a single-stranded DNA molecule. Single-stranded DNA mediated hierarchal assembly can be performed with multiple gene fragments as seen in FIG. 2. Single-stranded DNA mediated hierarchal assembly is performed with a first gene fragment **202**, a second gene fragment **204**, and a third gene fragment **206** comprising a homology sequence **205**. The first gene fragment **202**, the second gene fragment **204**,

and the third gene fragment **206** are double-stranded and comprise a 5' flanking adapter sequence **207a** and a 3' flanking adapter sequence **207b**. The 3' flanking adapter sequence **207b** of the first gene fragment **202** comprises uracil. The 5' flanking adapter sequence **207a** and the 3' flanking adapter sequence **207b** of the second gene fragment **204** comprise a uracil. The 3' flanking adapter sequence **207b** of the third gene fragment **206** comprises uracil. The first gene fragment **202**, the second gene fragment **204**, and the third gene fragment **206** are amplified with universal primers (primers that are complementary to a region of each of the gene fragments) and a uracil compatible polymerase. In some instances, the uracil compatible polymerase is Phusion U or Kapa Uracil. The resultant PCR product comprises a uracil at the end of the at least one of the 5' flanking adapter sequence **207a** and the 3' flanking adapter sequence **207b**. The first gene fragment **202**, the second gene fragment **204**, and the third gene fragment **206** are diluted, mixed, and amplified **209** with universal primers and a uracil incompatible polymerase that stalls at a uracil or is inefficient when interacting with a uracil. In some instances, the uracil incompatible polymerase is Q5 DNA polymerase. The resultant fragment without 208 uracil and fragment comprising uracil **210** serve as primers for each other. The resultant fragment without 208 uracil and fragment comprising uracil **210** are then combined **213** and diluted, mixed, and then amplified **215** with universal primers and DNA polymerase that stalls at uracil (such as Q5 DNA polymerase) to generate an intermediate fragment **212**. Intermediate fragment **212** and an additional fragment **214** serve as primers for each other and are combined and amplified **219** to generate a single-stranded DNA molecule.

[0040] A second exemplary process for seamless assembly of nucleic acids is depicted in FIG. 3. In vitro recombination cloning is performed with a first gene fragment **302** comprising from 5' to 3': a 5' flanking adapter sequence **307a**, a first homology sequence **303**, an insert sequence **305**, a second homology sequence **309**, and a 3' flanking adapter sequence **307b**. The first homology sequence **303** is homologous to sequence **311** of vector **304**. The second homology sequence is homologous to sequence **313** of vector **304**. The first gene fragment **302** and vector **304** are incubated **317** with bacterial cell lysate to generate assembled construct **306**.

[0041] In vitro recombination cloning can be performed with multiple gene fragments as seen in FIG. 4. In vitro recombination cloning is performed using two gene fragments. A first gene fragment **402** comprises from 5' to 3': a 5' flanking adapter sequence **407a**, a first homology sequence **403**, an insert sequence **405**, a second homology sequence **409**, and a 3' flanking adapter sequence **407b**. A second gene fragment **404** comprises from 5' to 3': a 5' flanking adapter sequence **407a**, a first homology sequence **411**, an insert sequence **413**, a second homology sequence **415**, and a 3' flanking adapter sequence **407b**. The first homology sequence **403** of the first gene fragment **402** is homologous to sequence **417** on vector **406**. The second homology sequence **409** of the first gene fragment **402** is homologous to the first homology sequence **411** of the second gene fragment **404**. The second homology sequence **415** of the second gene fragment **404** is homologous to the sequence **419** of vector **406**. The first gene fragment **402**, the

second gene fragment **404**, and vector **406** are incubated **419** with bacterial cell lysate to generate assembled construct **408**.

[0042] A third exemplary process for seamless assembly of nucleic acids is depicted in FIG. 8. Overlap extension PCR is performed using a nucleic acid **802** comprising a universal primer binding site **803** and a region complementary to nucleic acid **804**. Nucleic acid **804** comprises a universal primer binding site **803**. An enzyme **805** cleaves a terminal end of nucleic acid **802** and nucleic acid **804**. Nucleic acid **802** and nucleic acid **804** are then amplified and serve as a template for each other.

[0043] Primers referred to in the exemplary workflows mentioned herein as "universal primers" are short polynucleotides that recognize a primer binding site common to multiple DNA fragments. However, these workflows are not limited to only use of universal primers, and fragment-specific primers may be incorporated in addition or alternatively. In addition, while exemplary workflows described herein refer to assembly of gene fragments, they are not limited as such and are applicable to the assembly of longer nucleic acids in general.

[0044] In Vitro Recombination Cloning

[0045] Provided herein are methods for seamless assembly of nucleic acids, comprising in vitro recombination cloning. In some instances, in vitro recombination cloning comprises a gene or fragment thereof for insertion into a vector using a bacterial lysate. In some instances, the gene fragment comprises at least one universal primer. In some instances, the gene fragment comprises a vector homology sequence.

[0046] Provided herein are methods for in vitro recombination cloning, wherein a bacterial lysate is used. The bacterial lysate may be derived from *Escherichia coli*. In some instances, the bacterial lysate is derived from RecA⁻ bacteria. In some instances, the bacterial strain is from JM109 cells. In some instances, the bacterial lysate comprises a nuclease or a recombinase. In some instances, the bacterial lysate comprises a nuclease and a recombinase.

[0047] Provided herein are methods for in vitro recombination cloning, wherein a gene fragment is de novo synthesized and comprise a flanking adapter sequence and a homology sequence. The homology sequence may be at a 5' or 3' end of the gene fragment. In some instances, the homology sequence is flanked by a pair of flanking adapter sequences. In some instances, the gene fragment comprises a least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 homology sequences.

[0048] Homology sequences described herein for in vitro recombination cloning may vary in length. Exemplary lengths for homology sequences include, but are not limited to, at least or about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, or more than 200 base pairs. In some instances, the length of the homology sequence is 20 base pairs. In some instances, the length of the homology sequence is 41 base pairs. In some instances the length of the homology sequence is 100 base pairs. In some instances, the length of the homology sequence has a range of about 10 to 20, 10 to 30, 10 to 40, 10 to 50, 10 to 60, 10 to 70, 10 to 80, 10 to 100, 10 to 125, 10 to 150, 10 to 200, 20 to 30, 20 to 40, 20 to 50, 20 to 60, 20 to 70, 20 to 80, 20 to 100, 20 to 125, 20 to 150, 20 to 200, 30 to 40, 30 to 50, 30 to 60, 30 to 70, 30 to 80, 30 to 100, 30 to 125, 30 to 150, 30 to 200, 40 to 50, 40 to 60, 40 to 70, 40 to 80, 40 to 100, 40 to 125,

40 to 150, 40 to 200, 50 to 60, 50 to 70, 50 to 80, 50 to 100, 50 to 125, 50 to 150, 50 to 200, 60 to 70, 60 to 80, 60 to 100, 60 to 125, 60 to 150, 60 to 200, 70 to 80, 70 to 100, 70 to 125, 70 to 150, 70 to 200, 80 to 100, 80 to 125, 80 to 150, 80 to 200, 100 to 125, 100 to 150, 100 to 200, 125 to 150, 125 to 200, or 150 to 200 base pairs.

[0049] Provided herein are methods for in vitro recombination cloning, wherein a number of gene fragments are inserted into a vector. In some instances, the number of gene fragments that are inserted is at least or about 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 gene fragments. In some instances, the number of gene fragments that are inserted has a range of about 1 to 2, 1 to 3, 1 to 4, 1 to 5, 1 to 6, 1 to 7, 1 to 8, 1 to 9, 1 to 10, 2 to 3, 2 to 4, 2 to 5, 2 to 6, 2 to 7, 2 to 8, 2 to 9, 2 to 10, 3 to 4, 3 to 5, 3 to 6, 3 to 7, 3 to 8, 3 to 9, 3 to 10, 4 to 5, 4 to 6, 4 to 7, 4 to 8, 4 to 9, 4 to 10, 5 to 6, 5 to 7, 5 to 8, 5 to 9, 5 to 10, 6 to 7, 6 to 8, 6 to 9, 6 to 10, 7 to 8, 7 to 9, 7 to 10, 8 to 9, 8 to 10, or 9 to 10.

[0050] Provided herein are methods for in vitro recombination cloning, wherein a gene fragment comprises a non-homologous sequence. In some instances, the non-homologous sequence comprises at least 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 140, 160, 180, 200, 220, 240, 260, 280, 300 or more than 300 base pairs in length. In some instances, the number of base pairs is 24 base pairs. In some instances, the number of base pairs is 124 base pairs. In some instances the number of base pairs is 324 base pairs. In some instances, the gene fragment does not comprise a non-homologous sequence.

[0051] Provided herein are methods for in vitro recombination cloning, wherein the amount of gene fragment or the amount of vector varies. In some instances, the amount of gene fragment is at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, or more than 100 femtomoles. In some instances, the amount of vector is at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, or more than 100 femtomoles. In some instances, a ratio of gene fragment to vector varies. In some instances, the molar ratio of gene fragment to vector is at least or about 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1, or more.

[0052] Provided herein are methods for in vitro recombination cloning, wherein a reaction for in vitro recombination cloning occurs at an optimal temperature. In some instances, the reaction occurs at a temperature optimal for enzymatic activity, for example, a temperature in a range of about 25-80° C. 25-70° C. 25-60° C. 25-50° C. or 25-40° C. In some instances, the temperature is at least or about 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. or more than 80° C. In some instances, the temperature is about 65° C. In some instances, the enzymatic activity is a nuclease activity. In some instances, the enzymatic activity is a recombinase activity.

[0053] Methods described herein for in vitro recombination cloning result in a high percentage of correct assembly. In some instances, the percentage of correct assembly is at least or about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 99%, or more than 99%. In some instances, the percentage of correct assembly is 100%. In some instances, the percentage of incorrect assembly is at most 5%, 10%, 15%, 20%, 25%, or 30%, or more than 30%.

[0054] Methods described herein comprising in vitro recombination cloning result in increased efficiency. In some

instances, efficiency is measured by number of colony forming units. In some instances, methods described herein result in at least or about 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 12000, 14000, 16000, 18000, 20000, 25000, 30000, 35000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 200000, 300000, 400000, 500000, 600000, 700000, 800000, 900000, or more than 900000 colony forming units.

[0055] Single-stranded DNA Mediated Hierarchical Assembly

[0056] Provided herein are methods for seamless assembly of nucleic acids, wherein methods comprise single-stranded DNA mediated hierarchical assembly. In some instances, the single-stranded DNA mediated hierarchical assembly comprises assembly of a nucleic acid de novo synthesized by methods described herein. In some instances, the assembly comprises amplification of the nucleic acid with a primer, wherein the primer is not removed after amplification. In some instances, assembly results in increased percentage of correctly assembled nucleic acids and improved efficiency.

[0057] Provided herein are methods for single-stranded DNA mediated hierarchical assembly, wherein methods comprise an amplification reaction. In some instances, the amplification reaction comprises a polymerase. In some instances, the polymerase is a high fidelity polymerase. In some instances, the polymerase is a DNA polymerase. The DNA polymerase may be from any family of DNA polymerases including, but not limited to, Family A polymerase, Family B polymerase, Family C polymerase, Family D polymerase, Family X polymerase, and Family Y polymerase. In some instances, the DNA polymerase may be a Family B polymerase. Exemplary Family B polymerase is from a species of, but not limited to, *Pyrococcus furiosus*, *Thermococcus gorgonarius*, *Desulfurococcus* strain Tok, *Thermococcus* sp. 9°N-7, *Pyrococcus kodakaraensis*, *Thermococcus litoralis*, *Methanococcus voltae*, *Pyrobaculum islandicum*, *Archaeoglobus fulgidus*, *Cenarchaeum symbiosum*, *Sulfolobus acidocaldarius*, *Sulfurisphaera ohwakuensis*, *Sulfolobus solfataricus*, *Pyrodictium occultum*, and *Aeropyrum pernix*. In some instances, the Family B polymerase is a polymerase or derivative thereof (e.g., mutants, chimeras) from *Pyrococcus furiosus*.

[0058] Polymerases described herein for use in an amplification reaction may comprise various enzymatic activities. Polymerases are used in the methods of the invention, for example, to extend primers to produce extension products. In some instances, the DNA polymerase has 5' to 3' polymerase activity. In some instances, the DNA polymerase comprises 3' to 5' exonuclease activity. In some instances, the DNA polymerase comprises proofreading activity. Exemplary polymerases include, but are not limited to, DNA polymerase (I, II, or III), T4 DNA polymerase, T7 DNA polymerase, Bst DNA polymerase, Bca polymerase, Vent DNA polymerase, Pfu DNA polymerase, and Taq DNA polymerase.

[0059] Polymerases described herein for use in an amplification reaction may recognize a modified base. In some instances, the modified base is a variation in nucleic acid composition or a chemical modification. In some instances, a modified base comprises a base other than adenine, guanine, cytosine or thymine in DNA or a base other than adenine, guanine, cytosine or uracil in RNA. Modified bases described herein include, without limitation, oxidized bases,

alkylated bases, deaminated bases, pyrimidine derivatives, purine derivatives, ring-fragmented bases, and methylated bases. Exemplary modified bases include, but are not limited to, uracil, 3-meA (3-methyladenine), hypoxanthine, 8-oxoG (7,8-dihydro-8-oxoguanine), FapyG, FapyA, Tg (thymine glycol), hoU (hydroxyuracil), hmU (hydroxymethyluracil), fU (formyluracil), hoC (hydroxycytosine), fC (formylcytosine), 5-meC (5-methylcytosine), 6-meG (6-methylguanine), 7-meG (N7-methylguanine), εC (ethenocytosine), 5-caC (5-carboxylcytosine), 2-hA, EA (ethenoadenine), 5-fU (5-fluorouracil), 3-meG (3-methylguanine), and isodialuric acid. In some instances, a modified base in DNA is a uracil. Non-limiting examples of uracil compatible DNA polymerases include Pfu polymerase, Pfu Turbo Cx and KAPA HiFi Uracil+. In some instances, the polymerase selected for the amplification reaction is not capable of recognizing a modified base. For example, the polymerase is incompatible with uracil. Exemplary polymerases that are incompatible with uracil include, but are not limited to, KAPA HiFi polymerase, KAPA HiFi, Phusion®, and Q5® High Fidelity DNA polymerase.

[0060] In some instances, a single DNA polymerase or a plurality of DNA polymerases are used. In some instances, the same DNA polymerase or set of DNA polymerases are used at different stages of the present methods. For example, in a first amplification reaction a DNA polymerase that is compatible with uracil is used, and in a second amplification reaction a DNA polymerase that is incompatible with uracil is used. In some instances, the DNA polymerases are varied. For example, the DNA polymerases are varied based on enzymatic activities. In some instances, additional polymerases are added during various steps.

[0061] Described herein are methods for nucleic acid assembly comprising an amplification reaction, wherein the amplification reaction comprises a universal primer binding sequence. In some instances, the universal primer binding sequence is capable of binding the same 5' or 3' primer. In some instances, the universal primer binding sequence is shared among a plurality of target nucleic acids in the amplification reaction.

[0062] Provided herein are methods for single-stranded DNA mediated hierarchal assembly, wherein a reaction for single-stranded DNA mediated hierarchal assembly occurs at an optimal temperature. In some instances, the reaction occurs at a temperature optimal for polymerase activity. In some instances, the reaction occurs at a temperature optimal for enzymatic activity. In some instances, the reaction occurs at a temperature in a range of about 25-80° C. 25-70° C. 25-60° C. 25-50° C. or 25-40° C. In some instances, the temperature is at least or about 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. or more than 80° C.

[0063] Provided herein are methods for single-stranded DNA mediated hierarchal assembly, wherein a gene fragment to be assembled comprises a homology sequence. In some instances, the homology sequence is complementary to a homology sequence in another gene fragment to be assembled. The homology sequence may comprise a number of base pairs. In some instances, the number of base pairs is at least or about 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 125, 150, 175, 200, or more than 200 base pairs. In some instances, the number of base pairs is 20 base pairs. In some instances, the number of base pairs is 41 base pairs. In some instances the number of base pairs is 100 base pairs. In some

instances, the number of base pairs is 10 to 20, 10 to 30, 10 to 40, 10 to 50, 10 to 60, 10 to 70, 10 to 80, 10 to 100, 10 to 125, 10 to 150, 10 to 200, 20 to 30, 20 to 40, 20 to 50, 20 to 60, 20 to 70, 20 to 80, 20 to 100, 20 to 125, 20 to 150, 20 to 200, 30 to 40, 30 to 50, 30 to 60, 30 to 70, 30 to 80, 30 to 100, 30 to 125, 30 to 150, 30 to 200, 40 to 50, 40 to 60, 40 to 70, 40 to 80, 40 to 100, 40 to 125, 40 to 150, 40 to 200, 50 to 60, 50 to 70, 50 to 80, 50 to 100, 50 to 125, 50 to 150, 50 to 200, 60 to 70, 60 to 80, 60 to 100, 60 to 125, 60 to 150, 60 to 200, 70 to 80, 70 to 100, 70 to 125, 70 to 150, 70 to 200, 80 to 100, 80 to 125, 80 to 150, 80 to 200, 100 to 125, 100 to 150, 100 to 200, 125 to 150, 125 to 200, or 150 to 200 base pairs.

[0064] Provided herein are methods for single-stranded DNA mediated hierarchal assembly, wherein a plurality of gene fragments are assembled. In some instances, the number of gene fragments that are assembled is at least or about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more than 10 gene fragments. In some instances, the number of gene fragments is 1 to 2, 1 to 3, 1 to 4, 1 to 5, 1 to 6, 1 to 7, 1 to 8, 1 to 9, 1 to 10, 2 to 3, 2 to 4, 2 to 5, 2 to 6, 2 to 7, 2 to 8, 2 to 9, 2 to 10, 3 to 4, 3 to 5, 3 to 6, 3 to 7, 3 to 8, 3 to 9, 3 to 10, 4 to 5, 4 to 6, 4 to 7, 4 to 8, 4 to 9, 4 to 10, 5 to 6, 5 to 7, 5 to 8, 5 to 9, 5 to 10, 6 to 7, 6 to 8, 6 to 9, 6 to 10, 7 to 8, 7 to 9, 7 to 10, 8 to 9, 8 to 10, or 9 to 10.

[0065] Methods described herein for single-stranded DNA mediated hierarchal assembly result in a high percentage of correct assembly. In some instances, the percentage of correct assembly is at least or about 70%, 75%, 80%, 85%, 90%, 95%, 97%, 99%, or more than 99%. In some instances, the percentage of correct assembly is 100%. In some instances, the percentage of incorrect assembly is at most 5%, 10%, 15%, 20%, 25%, or 30%, or more than 30%.

[0066] Systems for Synthesis of Nucleic Acids and Seamless Assembly

[0067] Polynucleotide Synthesis

[0068] Provided herein are methods for seamless assembly of nucleic acids following generation of polynucleotides by de novo synthesis by methods described herein. An exemplary workflow is seen in FIG. 9. A computer readable input file comprising a nucleic acid sequence is received. A computer processes the nucleic acid sequence to generate instructions for synthesis of the polynucleotide sequence or a plurality of polynucleotide sequences collectively encoding the nucleic acid sequence. Instructions are transmitted to a material deposition device **903** for synthesis of the plurality of polynucleotides based on the plurality of nucleic acid sequences. The material deposition device **903**, such as a polynucleotide acid synthesizer, is designed to release reagents in a step wise fashion such that multiple polynucleotides extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence. The material deposition device **903** generates oligomers on an array **905** that includes multiple clusters **907** of loci for polynucleotide acid synthesis and extension. However, the array need not have loci organized in clusters. For example, the loci can be uniformly spread across the array. De novo polynucleotides are synthesized and removed from the plate and an assembly reaction commenced in a collection chamber **909** followed by formation population of longer polynucleotides **911**. The collection chamber may comprise a sandwich of multiple surfaces (e.g., a top and bottom surface) or well or channel in containing transferred material from the synthesis surface. De novo polynucleotides can

also be synthesized and removed from the plate to form a population of longer polynucleotides **911**. The population of longer polynucleotides **911** can then be partitioned into droplets or subject to PCR. The population of longer polynucleotides **911** is then subject to nucleic acid assembly by either in vitro recombination cloning **915**, or single-stranded DNA hierarchical assembly **917**.

[0069] Provided herein are systems for seamless assembly of nucleic acids following generation of polynucleotides by de novo synthesis by methods described herein. In some instances, the system comprises a computer, a material deposition device, a surface, and a nucleic acid assembly surface. In some instances, the computer comprises a readable input file with a nucleic acid sequence. In some instances, the computer processes the nucleic acid sequence to generate instructions for synthesis of the polynucleotide sequence or a plurality of polynucleotide sequences collectively encoding for the nucleic acid sequence. In some instances, the computer provides instructions to the material deposition device for the synthesis of the plurality of polynucleotide acid sequences. In some instances, the material deposition device deposits nucleosides on the surface for an extension reaction. In some instances, the surface comprises a locus for the extension reaction. In some instances, the locus is a spot, well, microwell, channel, or post. In some instances, the plurality of polynucleotide acid sequences is synthesized following the extension reaction. In some instances, the plurality of polynucleotide acid sequences are removed from the surface and prepared for nucleic acid assembly. In some instances, nucleic acid assembly comprises in vitro recombination cloning. In some instances, nucleic acid assembly comprises single-stranded hierarchical DNA assembly. In some instances, nucleic acid assembly comprises overlap extension PCR without primer removal.

[0070] Provided herein are methods for polynucleotide synthesis involving phosphoramidite chemistry. In some instances, polynucleotide synthesis comprises coupling a base with phosphoramidite. In some instances, polynucleotide synthesis comprises coupling a base by deposition of phosphoramidite under coupling conditions, wherein the same base is optionally deposited with phosphoramidite more than once, i.e., double coupling. In some instances, polynucleotide synthesis comprises capping of unreacted sites. In some cases, capping is optional. In some instances, polynucleotide synthesis comprises oxidation. In some instances, polynucleotide synthesis comprises deblocking or detritylation. In some instances, polynucleotide synthesis comprises sulfurization. In some cases, polynucleotide synthesis comprises either oxidation or sulfurization. In some instances, between one or each step during a polynucleotide synthesis reaction, the substrate is washed, for example, using tetrazole or acetonitrile. Time frames for any one step in a phosphoramidite synthesis method include less than about 2 min, 1 min, 50 sec, 40 sec, 30 sec, 20 sec and 10 sec.

[0071] Polynucleotide synthesis using a phosphoramidite method comprises the subsequent addition of a phosphoramidite building block (e.g., nucleoside phosphoramidite) to a growing polynucleotide chain for the formation of a phosphite triester linkage. Phosphoramidite polynucleotide synthesis proceeds in the 3' to 5' direction. Phosphoramidite polynucleotide synthesis allows for the controlled addition of one nucleotide to a growing nucleic acid chain per synthesis cycle. In some instances, each synthesis cycle comprises a coupling step. Phosphoramidite coupling

involves the formation of a phosphite triester linkage between an activated nucleoside phosphoramidite and a nucleoside bound to the substrate, for example, via a linker. In some instances, the nucleoside phosphoramidite is provided to the substrate activated. In some instances, the nucleoside phosphoramidite is provided to the substrate with an activator. In some instances, nucleoside phosphoramidites are provided to the substrate in a 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100-fold excess or more over the substrate-bound nucleosides. In some instances, the addition of nucleoside phosphoramidite is performed in an anhydrous environment, for example, in anhydrous acetonitrile. Following addition of a nucleoside phosphoramidite, the substrate is optionally washed. In some instances, the coupling step is repeated one or more additional times, optionally with a wash step between nucleoside phosphoramidite additions to the substrate. In some instances, a polynucleotide synthesis method used herein comprises 1, 2, 3 or more sequential coupling steps. Prior to coupling, in many cases, the nucleoside bound to the substrate is de-protected by removal of a protecting group, where the protecting group functions to prevent polymerization. A common protecting group is 4,4'-dimethoxytrityl (DMT).

[0072] Following coupling, phosphoramidite polynucleotide synthesis methods optionally comprise a capping step. In a capping step, the growing polynucleotide is treated with a capping agent. A capping step is useful to block unreacted substrate-bound 5'-OH groups after coupling from further chain elongation, preventing the formation of polynucleotides with internal base deletions. Further, phosphoramidites activated with 1H-tetrazole may react, to a small extent, with the O6 position of guanosine. Without being bound by theory, upon oxidation with I₂/water, this side product, possibly via O6-N7 migration, may undergo depurination. The apurinic sites may end up being cleaved in the course of the final deprotection of the polynucleotide thus reducing the yield of the full-length product. The O6 modifications may be removed by treatment with the capping reagent prior to oxidation with I₂/water. In some instances, inclusion of a capping step during polynucleotide synthesis decreases the error rate as compared to synthesis without capping. As an example, the capping step comprises treating the substrate-bound polynucleotide with a mixture of acetic anhydride and 1-methylimidazole. Following a capping step, the substrate is optionally washed.

[0073] In some instances, following addition of a nucleoside phosphoramidite, and optionally after capping and one or more wash steps, the substrate bound growing nucleic acid is oxidized. The oxidation step comprises the phosphite triester is oxidized into a tetracoordinated phosphate triester, a protected precursor of the naturally occurring phosphate diester internucleoside linkage. In some cases, oxidation of the growing polynucleotide is achieved by treatment with iodine and water, optionally in the presence of a weak base (e.g., pyridine, lutidine, collidine). Oxidation may be carried out under anhydrous conditions using, e.g. tert-Butyl hydroperoxide or (1S)-(+)-(10-camphorsulfonyl)-oxaziridine (CSO). In some methods, a capping step is performed following oxidation. A second capping step allows for substrate drying, as residual water from oxidation that may persist can inhibit subsequent coupling. Following oxidation, the substrate and growing polynucleotide is optionally washed. In some instances, the step of oxidation is substi-

tuted with a sulfurization step to obtain polynucleotide phosphorothioates, wherein any capping steps can be performed after the sulfurization. Many reagents are capable of the efficient sulfur transfer, including but not limited to 3-(Dimethylaminomethylidene)amino)-3H-1,2,4-dithiazole-3-thione, DDTT, 3H-1,2-benzodithiol-3-one 1,1-dioxide, also known as Beaucage reagent, and N,N,N'-Tetraethylthiuram disulfide (TETD).

[0074] In order for a subsequent cycle of nucleoside incorporation to occur through coupling, the protected 5' end of the substrate bound growing polynucleotide is removed so that the primary hydroxyl group is reactive with a next nucleoside phosphoramidite. In some instances, the protecting group is DMT and deblocking occurs with trichloroacetic acid in dichloromethane. Conducting detritylation for an extended time or with stronger than recommended solutions of acids may lead to increased depurination of solid support-bound polynucleotide and thus reduces the yield of the desired full-length product. Methods and compositions of the invention described herein provide for controlled deblocking conditions limiting undesired depurination reactions. In some cases, the substrate bound polynucleotide is washed after deblocking. In some cases, efficient washing after deblocking contributes to synthesized polynucleotides having a low error rate.

[0075] Methods for the synthesis of polynucleotides typically involve an iterating sequence of the following steps: application of a protected monomer to an actively functionalized surface (e.g., locus) to link with either the activated surface, a linker or with a previously deprotected monomer; deprotection of the applied monomer so that it is reactive with a subsequently applied protected monomer; and application of another protected monomer for linking. One or more intermediate steps include oxidation or sulfurization. In some cases, one or more wash steps precede or follow one or all of the steps.

[0076] Methods for phosphoramidite based polynucleotide synthesis comprise a series of chemical steps. In some instances, one or more steps of a synthesis method involve reagent cycling, where one or more steps of the method comprise application to the substrate of a reagent useful for the step. For example, reagents are cycled by a series of liquid deposition and vacuum drying steps. For substrates comprising three-dimensional features such as wells, microwells, channels and the like, reagents are optionally passed through one or more regions of the substrate via the wells and/or channels.

[0077] Polynucleotides synthesized using the methods and/or substrates described herein comprise at least about 20, 30, 40, 50, 60, 70, 75, 80, 90, 100, 120, 150, 200, 500 or more bases in length. In some instances, at least about 1 μ mol, 10 pmol, 20 pmol, 30 pmol, 40 pmol, 50 pmol, 60 pmol, 70 pmol, 80 pmol, 90 pmol, 100 pmol, 150 pmol, 200 pmol, 300 pmol, 400 pmol, 500 pmol, 600 pmol, 700 pmol, 800 pmol, 900 pmol, 1 nmol, 5 nmol, 10 nmol, 100 nmol or more of a polynucleotide is synthesized within a locus. Methods for polynucleotide synthesis on a surface provided herein allow for synthesis at a fast rate. As an example, at least 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, 35, 40, 45, 50, 55, 60, 70, 80, 90, 100, 125, 150, 175, 200 nucleotides per hour, or more are synthesized. Nucleotides include adenine, guanine, thymine, cytosine, uridine building blocks, or analogs/modified versions thereof. In some instances, libraries of

polynucleotides are synthesized in parallel on a substrate. For example, a substrate comprising about or at least about 100; 1,000; 10,000; 100,000; 1,000,000; 2,000,000; 3,000,000; 4,000,000; or 5,000,000 resolved loci is able to support the synthesis of at least the same number of distinct polynucleotides, wherein polynucleotide encoding a distinct sequence is synthesized on a resolved locus.

[0078] Various suitable methods are known for generating high density polynucleotide arrays. In an exemplary workflow, a substrate surface layer is provided. In the example, chemistry of the surface is altered in order to improve the polynucleotide synthesis process. Areas of low surface energy are generated to repel liquid while areas of high surface energy are generated to attract liquids. The surface itself may be in the form of a planar surface or contain variations in shape, such as protrusions or microwells which increase surface area. In the workflow example, high surface energy molecules selected serve a dual function of supporting DNA chemistry, as disclosed in International Patent Application Publication WO/2015/021080, which is herein incorporated by reference in its entirety.

[0079] In situ preparation of polynucleotide arrays is generated on a solid support and utilizes single nucleotide extension process to extend multiple oligomers in parallel. A deposition device, such as a polynucleotide synthesizer, is designed to release reagents in a step wise fashion such that multiple polynucleotides extend, in parallel, one residue at a time to generate oligomers with a predetermined nucleic acid sequence. In some cases, polynucleotides are cleaved from the surface at this stage. Cleavage includes gas cleavage, e.g., with ammonia or methylamine.

[0080] Substrates

[0081] Devices used as a surface for polynucleotide synthesis may be in the form of substrates which include, without limitation, homogenous array surfaces, patterned array surfaces, channels, beads, gels, and the like. Provided herein are substrates comprising a plurality of clusters, wherein each cluster comprises a plurality of loci that support the attachment and synthesis of polynucleotides. The term "locus" as used herein refers to a discrete region on a structure which provides support for polynucleotides encoding for a single predetermined sequence to extend from the surface. In some instances, a locus is on a two dimensional surface, e.g., a substantially planar surface. In some instances, a locus is on a three-dimensional surface, e.g., a well, microwell, channel, or post. In some instances, a surface of a locus comprises a material that is actively functionalized to attach to at least one nucleotide for polynucleotide synthesis, or preferably, a population of identical nucleotides for synthesis of a population of polynucleotides. In some instances, polynucleotide refers to a population of polynucleotides encoding for the same nucleic acid sequence. In some cases, a surface of a substrate is inclusive of one or a plurality of surfaces of a substrate. The average error rates for polynucleotides synthesized within a library described herein using the systems and methods provided are often less than 1 in 1000, less than about 1 in 2000, less than about 1 in 3000 or less often without error correction.

[0082] Provided herein are surfaces that support the parallel synthesis of a plurality of polynucleotides having different predetermined sequences at addressable locations on a common support. In some instances, a substrate provides support for the synthesis of more than 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2,000; 5,000;

10,000; 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more non-identical polynucleotides. In some cases, the surfaces provides support for the synthesis of more than 50, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2,000; 5,000; 10,000; 20,000; 50,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; 10,000,000 or more polynucleotides encoding for distinct sequences. In some instances, at least a portion of the polynucleotides have an identical sequence or are configured to be synthesized with an identical sequence. In some instances, the substrate provides a surface environment for the growth of polynucleotides having at least 80, 90, 100, 120, 150, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, 500 bases or more.

[0083] Provided herein are methods for polynucleotide synthesis on distinct loci of a substrate, wherein each locus supports the synthesis of a population of polynucleotides. In some cases, each locus supports the synthesis of a population of polynucleotides having a different sequence than a population of polynucleotides grown on another locus. In some instances, each polynucleotide sequence is synthesized with 1, 2, 3, 4, 5, 6, 7, 8, 9 or more redundancy across different loci within the same cluster of loci on a surface for polynucleotide synthesis. In some instances, the loci of a substrate are located within a plurality of clusters. In some instances, a substrate comprises at least 10, 500, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 11000, 12000, 13000, 14000, 15000, 20000, 30000, 40000, 50000 or more clusters. In some instances, a substrate comprises more than 2,000; 5,000; 10,000; 100,000; 200,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,100,000; 1,200,000; 1,300,000; 1,400,000; 1,500,000; 1,600,000; 1,700,000; 1,800,000; 1,900,000; 2,000,000; 300,000; 400,000; 500,000; 600,000; 700,000; 800,000; 900,000; 1,000,000; 1,200,000; 1,400,000; 1,600,000; 1,800,000; 2,000,000; 2,500,000; 3,000,000; 3,500,000; 4,000,000; 4,500,000; 5,000,000; or 10,000,000 or more distinct loci. In some instances, a substrate comprises about 10,000 distinct loci. The amount of loci within a single cluster is varied in different instances. In some cases, each cluster includes 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 120, 130, 150, 200, 300, 400, 500 or more loci. In some instances, each cluster includes about 50-500 loci. In some instances, each cluster includes about 100-200 loci. In some instances, each cluster includes about 100-150 loci. In some instances, each cluster includes about 109, 121, 130 or 137 loci. In some instances, each cluster includes about 19, 20, 61, 64 or more loci.

[0084] In some instances, the number of distinct polynucleotides synthesized on a substrate is dependent on the number of distinct loci available in the substrate. In some instances, the density of loci within a cluster of a substrate is at least or about 1, 10, 25, 50, 65, 75, 100, 130, 150, 175, 200, 300, 400, 500, 1,000 or more loci per mm². In some cases, a substrate comprises 10-500, 25-400, 50-500, 100-500, 150-500, 10-250, 50-250, 10-200, or 50-200 mm². In some instances, the distance between the centers of two adjacent loci within a cluster is from about 10-500, from

about 10-200, or from about 10-100 um. In some instances, the distance between two centers of adjacent loci is greater than about 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 um. In some instances, the distance between the centers of two adjacent loci is less than about 200, 150, 100, 80, 70, 60, 50, 40, 30, 20 or 10 um. In some instances, each locus has a width of about 0.5, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90 or 100 um. In some cases, the each locus is has a width of about 0.5-100, 0.5-50, 10-75, or 0.5-50 um.

[0085] In some instances, the density of clusters within a substrate is at least or about 1 cluster per 100 mm², 1 cluster per 10 mm², 1 cluster per 5 mm², 1 cluster per 4 mm², 1 cluster per 3 mm², 1 cluster per 2 mm², 1 cluster per 1 mm², 2 clusters per 1 mm², 3 clusters per 1 mm², 4 clusters per 1 mm², 5 clusters per 1 mm², 10 clusters per 1 mm², 50 clusters per 1 mm² or more. In some instances, a substrate comprises from about 1 cluster per 10 mm² to about 10 clusters per 1 mm². In some instances, the distance between the centers of two adjacent clusters is at least or about 50, 100, 200, 500, 1000, 2000, or 5000 um. In some cases, the distance between the centers of two adjacent clusters is between about 50-100, 50-200, 50-300, 50-500, and 100-2000 um. In some cases, the distance between the centers of two adjacent clusters is between about 0.05-50, 0.05-10, 0.05-5, 0.05-4, 0.05-3, 0.05-2, 0.1-10, 0.2-10, 0.3-10, 0.4-10, 0.5-10, 0.5-5, or 0.5-2 mm. In some cases, each cluster has a cross section of about 0.5 to about 2, about 0.5 to about 1, or about 1 to about 2 mm. In some cases, each cluster has a cross section of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 mm. In some cases, each cluster has an interior cross section of about 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.15, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9 or 2 mm.

[0086] In some instances, a substrate is about the size of a standard 96 well plate, for example between about 100 and about 200 mm by between about 50 and about 150 mm. In some instances, a substrate has a diameter less than or equal to about 1000, 500, 450, 400, 300, 250, 200, 150, 100 or 50 mm. In some instances, the diameter of a substrate is between about 25-1000, 25-800, 25-600, 25-500, 25-400, 25-300, or 25-200 mm. In some instances, a substrate has a planar surface area of at least about 100; 200; 500; 1,000; 2,000; 5,000; 10,000; 12,000; 15,000; 20,000; 30,000; 40,000; 50,000 mm² or more. In some instances, the thickness of a substrate is between about 50-2000, 50-1000, 100-1000, 200-1000, or 250-1000 mm.

[0087] Surface Materials

[0088] Substrates, devices, and reactors provided herein are fabricated from any variety of materials suitable for the methods, compositions, and systems described herein. In certain instances, substrate materials are fabricated to exhibit a low level of nucleotide binding. In some instances, substrate materials are modified to generate distinct surfaces that exhibit a high level of nucleotide binding. In some instances, substrate materials are transparent to visible and/or UV light. In some instances, substrate materials are sufficiently conductive, e.g., are able to form uniform electric fields across all or a portion of a substrate. In some instances, conductive materials are connected to an electric ground. In some instances, the substrate is heat conductive or insulated. In some instances, the materials are chemical resistant and heat resistant to support chemical or biochemical reactions, for example polynucleotide synthesis reaction processes. In some instances, a substrate comprises flexible

materials. For flexible materials, materials can include, without limitation: nylon, both modified and unmodified, nitrocellulose, polypropylene, and the like. In some instances, a substrate comprises rigid materials. For rigid materials, materials can include, without limitation: glass, fuse silica, silicon, plastics (for example polytetrafluoroethylene, polypropylene, polystyrene, polycarbonate, and blends thereof, and the like), and metals (for example, gold, platinum, and the like). The substrate, solid support or reactors can be fabricated from a material selected from the group consisting of silicon, polystyrene, agarose, dextran, cellulosic polymers, polyacrylamides, polydimethylsiloxane (PDMS), and glass. The substrates/solid supports or the microstructures, reactors therein may be manufactured with a combination of materials listed herein or any other suitable material known in the art.

[0089] Surface Architecture

[0090] Provided herein are substrates for the methods, compositions, and systems described herein, wherein the substrates have a surface architecture suitable for the methods, compositions, and systems described herein. In some instances, a substrate comprises raised and/or lowered features. One benefit of having such features is an increase in surface area to support polynucleotide synthesis. In some instances, a substrate having raised and/or lowered features is referred to as a three-dimensional substrate. In some cases, a three-dimensional substrate comprises one or more channels. In some cases, one or more loci comprise a channel. In some cases, the channels are accessible to reagent deposition via a deposition device such as a polynucleotide synthesizer. In some cases, reagents and/or fluids collect in a larger well in fluid communication one or more channels. For example, a substrate comprises a plurality of channels corresponding to a plurality of loci with a cluster, and the plurality of channels are in fluid communication with one well of the cluster. In some methods, a library of polynucleotides is synthesized in a plurality of loci of a cluster.

[0091] Provided herein are substrates for the methods, compositions, and systems described herein, wherein the substrates are configured for polynucleotide synthesis. In some instances, the structure is configured to allow for controlled flow and mass transfer paths for polynucleotide synthesis on a surface. In some instances, the configuration of a substrate allows for the controlled and even distribution of mass transfer paths, chemical exposure times, and/or wash efficacy during polynucleotide synthesis. In some instances, the configuration of a substrate allows for increased sweep efficiency, for example by providing sufficient volume for a growing a polynucleotide such that the excluded volume by the growing polynucleotide does not take up more than 50, 45, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1%, or less of the initially available volume that is available or suitable for growing the polynucleotide. In some instances, a three-dimensional structure allows for managed flow of fluid to allow for the rapid exchange of chemical exposure.

[0092] Provided herein are substrates for the methods, compositions, and systems described herein, wherein the substrates comprise structures suitable for the methods, compositions, and systems described herein. In some instances, segregation is achieved by physical structure. In some instances, segregation is achieved by differential functionalization of the surface generating active and passive

regions for polynucleotide synthesis. In some instances, differential functionalization is achieved by alternating the hydrophobicity across the substrate surface, thereby creating water contact angle effects that cause beading or wetting of the deposited reagents. Employing larger structures can decrease splashing and cross-contamination of distinct polynucleotide synthesis locations with reagents of the neighboring spots. In some cases, a device, such as a material deposition device, is used to deposit reagents to distinct polynucleotide synthesis locations. Substrates having three-dimensional features are configured in a manner that allows for the synthesis of a large number of polynucleotides (e.g., more than about 10,000) with a low error rate (e.g., less than about 1:500, 1:1000, 1:1500, 1:2,000, 1:3,000, 1:5,000, or 1:10,000). In some cases, a substrate comprises features with a density of about or greater than about 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400 or 500 features per mm².

[0093] A well of a substrate may have the same or different width, height, and/or volume as another well of the substrate. A channel of a substrate may have the same or different width, height, and/or volume as another channel of the substrate. In some instances, the diameter of a cluster or the diameter of a well comprising a cluster, or both, is between about 0.05-50, 0.05-10, 0.05-5, 0.05-4, 0.05-3, 0.05-2, 0.05-1, 0.05-0.5, 0.05-0.1, 0.1-10, 0.2-10, 0.3-10, 0.4-10, 0.5-10, 0.5-5, or 0.5-2 mm. In some instances, the diameter of a cluster or well or both is less than or about 5, 4, 3, 2, 1, 0.5, 0.1, 0.09, 0.08, 0.07, 0.06, or 0.05 mm. In some instances, the diameter of a cluster or well or both is between about 1.0 and about 1.3 mm. In some instances, the diameter of a cluster or well, or both is about 1.150 mm. In some instances, the diameter of a cluster or well, or both is about 0.08 mm. The diameter of a cluster refers to clusters within a two-dimensional or three-dimensional substrate.

[0094] In some instances, the height of a well is from about 20-1000, 50-1000, 100-1000, 200-1000, 300-1000, 400-1000, or 500-1000 um. In some cases, the height of a well is less than about 1000, 900, 800, 700, or 600 um.

[0095] In some instances, a substrate comprises a plurality of channels corresponding to a plurality of loci within a cluster, wherein the height or depth of a channel is 5-500, 5-400, 5-300, 5-200, 5-100, 5-50, or 10-50 um. In some cases, the height of a channel is less than 100, 80, 60, 40, or 20 um.

[0096] In some instances, the diameter of a channel, locus (e.g., in a substantially planar substrate) or both channel and locus (e.g., in a three-dimensional substrate wherein a locus corresponds to a channel) is from about 1-1000, 1-500, 1-200, 1-100, 5-100, or 10-100 um, for example, about 90, 80, 70, 60, 50, 40, 30, 20 or 10 um. In some instances, the diameter of a channel, locus, or both channel and locus is less than about 100, 90, 80, 70, 60, 50, 40, 30, 20 or 10 um. In some instances, the distance between the center of two adjacent channels, loci, or channels and loci is from about 1-500, 1-200, 1-100, 5-200, 5-100, 5-50, or 5-30, for example, about 20 um.

[0097] Surface Modifications

[0098] Provided herein are methods for polynucleotide synthesis on a surface, wherein the surface comprises various surface modifications. In some instances, the surface modifications are employed for the chemical and/or physical alteration of a surface by an additive or subtractive process to change one or more chemical and/or physical properties

of a substrate surface or a selected site or region of a substrate surface. For example, surface modifications include, without limitation, (1) changing the wetting properties of a surface, (2) functionalizing a surface, i.e., providing, modifying or substituting surface functional groups, (3) defunctionalizing a surface, i.e., removing surface functional groups, (4) otherwise altering the chemical composition of a surface, e.g., through etching, (5) increasing or decreasing surface roughness, (6) providing a coating on a surface, e.g., a coating that exhibits wetting properties that are different from the wetting properties of the surface, and/or (7) depositing particulates on a surface.

[0099] In some cases, the addition of a chemical layer on top of a surface (referred to as adhesion promoter) facilitates structured patterning of loci on a surface of a substrate. Exemplary surfaces for application of adhesion promotion include, without limitation, glass, silicon, silicon dioxide, and silicon nitride. In some cases, the adhesion promoter is a chemical with a high surface energy. In some instances, a second chemical layer is deposited on a surface of a substrate. In some cases, the second chemical layer has a low surface energy. In some cases, surface energy of a chemical layer coated on a surface supports localization of droplets on the surface. Depending on the patterning arrangement selected, the proximity of loci and/or area of fluid contact at the loci are alterable.

[0100] In some instances, a substrate surface, or resolved loci, onto which nucleic acids or other moieties are deposited, e.g., for polynucleotide synthesis, are smooth or substantially planar (e.g., two-dimensional) or have irregularities, such as raised or lowered features (e.g., three-dimensional features). In some instances, a substrate surface is modified with one or more different layers of compounds. Such modification layers of interest include, without limitation, inorganic and organic layers such as metals, metal oxides, polymers, small organic molecules and the like.

[0101] In some instances, resolved loci of a substrate are functionalized with one or more moieties that increase and/or decrease surface energy. In some cases, a moiety is chemically inert. In some cases, a moiety is configured to support a desired chemical reaction, for example, one or more processes in a polynucleotide acid synthesis reaction. The surface energy, or hydrophobicity, of a surface is a factor for determining the affinity of a nucleotide to attach onto the surface. In some instances, a method for substrate functionalization comprises: (a) providing a substrate having a surface that comprises silicon dioxide; and (b) silanizing the surface using, a suitable silanizing agent described herein or otherwise known in the art, for example, an organofunctional alkoxysilane molecule. Methods and functionalizing agents are described in U.S. Pat. No. 5,474,796, which is herein incorporated by reference in its entirety.

[0102] In some instances, a substrate surface is functionalized by contact with a derivatizing composition that contains a mixture of silanes, under reaction conditions effective to couple the silanes to the substrate surface, typically via reactive hydrophilic moieties present on the substrate surface. Silanization generally covers a surface through self-assembly with organofunctional alkoxysilane molecules. A variety of siloxane functionalizing reagents can further be used as currently known in the art, e.g., for lowering or increasing surface energy. The organofunctional alkoxysilanes are classified according to their organic functions.

[0103] Computer Systems

[0104] Any of the systems described herein, may be operably linked to a computer and may be automated through a computer either locally or remotely. In some instances, the methods and systems of the invention further comprise software programs on computer systems and use thereof. Accordingly, computerized control for the synchronization of the dispense/vacuum/refill functions such as orchestrating and synchronizing the material deposition device movement, dispense action and vacuum actuation are within the bounds of the invention. The computer systems may be programmed to interface between the user specified base sequence and the position of a material deposition device to deliver the correct reagents to specified regions of the substrate.

[0105] The computer system **1000** illustrated in FIG. **10** may be understood as a logical apparatus that can read instructions from media **1011** and/or a network port **1005**, which can optionally be connected to server **1009** having fixed media **1012**. The system, such as shown in FIG. **10** can include a CPU **1001**, disk drives **1003**, optional input devices such as keyboard **1015** and/or mouse **1016** and optional monitor **1007**. Data communication can be achieved through the indicated communication medium to a server at a local or a remote location. The communication medium can include any means of transmitting and/or receiving data. For example, the communication medium can be a network connection, a wireless connection or an internet connection. Such a connection can provide for communication over the World Wide Web. It is envisioned that data relating to the present disclosure can be transmitted over such networks or connections for reception and/or review by a party **1022** as illustrated in FIG. **10**.

[0106] FIG. **11** is a block diagram illustrating architecture of a computer system **1100** that can be used in connection with example embodiments of the present invention. As depicted in FIG. **11**, the example computer system can include a processor **1102** for processing instructions. Non-limiting examples of processors include: Intel® Xeon® processor, AMD Opteron™ processor, Samsung 32-bit RISC ARM 1176JZ(F)-S v1.0 processor, ARM Cortex-A8 Samsung S5PC100 processor, ARM Cortex-A8 Apple A4 processor, Marvell PXA 930 processor, or a functionally-equivalent processor. Multiple threads of execution can be used for parallel processing. In some instances, multiple processors or processors with multiple cores can also be used, whether in a single computer system, in a cluster, or distributed across systems over a network comprising a plurality of computers, cell phones, and/or personal data assistant devices.

[0107] As illustrated in FIG. **11**, a high speed cache **1104** can be connected to, or incorporated in, the processor **1102** to provide a high speed memory for instructions or data that have been recently, or are frequently, used by processor **1102**. The processor **1102** is connected to a north bridge **1106** by a processor bus **1108**. The north bridge **1106** is connected to random access memory (RAM) **1110** by a memory bus **1112** and manages access to the RAM **1110** by the processor **1102**. The north bridge **1106** is also connected to a south bridge **1114** by a chipset bus **1116**. The south bridge **1114** is, in turn, connected to a peripheral bus **1118**. The peripheral bus can be, for example, PCI, PCI-X, PCI Express, or other peripheral bus. The north bridge and south bridge are often referred to as a processor chipset and manage data transfer between the processor, RAM, and

peripheral components on the peripheral bus **1118**. In some alternative architectures, the functionality of the north bridge can be incorporated into the processor instead of using a separate north bridge chip. In some instances, system **1100** can include an accelerator card **1122** attached to the peripheral bus **1118**. The accelerator can include field programmable gate arrays (FPGAs) or other hardware for accelerating certain processing. For example, an accelerator can be used for adaptive data restructuring or to evaluate algebraic expressions used in extended set processing.

[0108] Software and data are stored in external storage **1124** and can be loaded into RAM **1110** and/or cache **1104** for use by the processor. The system **1100** includes an operating system for managing system resources; non-limiting examples of operating systems include: Linux, Windows™, MACOS™, BlackBerry OST™, iOS™, and other functionally-equivalent operating systems, as well as application software running on top of the operating system for managing data storage and optimization in accordance with example embodiments of the present invention. In this example, system **1100** also includes network interface cards (NICs) **1120** and **1121** connected to the peripheral bus for providing network interfaces to external storage, such as Network Attached Storage (NAS) and other computer systems that can be used for distributed parallel processing.

[0109] FIG. 12 is a block diagram of a multiprocessor computer system using a shared virtual address memory space in accordance with an example embodiment. The system includes a plurality of processors **1202a-f** that can access a shared memory subsystem **1204**. The system incorporates a plurality of programmable hardware memory algorithm processors (MAPs) **1206a-f** in the memory subsystem **1204**. Each MAP **1206a-f** can comprise a memory **1208a-f** and one or more field programmable gate arrays (FPGAs) **1210a-f**. The MAP provides a configurable functional unit and particular algorithms or portions of algorithms can be provided to the FPGAs **1210a-f** for processing in close coordination with a respective processor. For example, the MAPs can be used to evaluate algebraic expressions regarding the data model and to perform adaptive data restructuring in example embodiments. In this example, each MAP is globally accessible by all of the processors for these purposes. In one configuration, each MAP can use Direct Memory Access (DMA) to access an associated memory **1208a-f**, allowing it to execute tasks independently of, and asynchronously from, the respective microprocessor **1202a-f**. In this configuration, a MAP can feed results directly to another MAP for pipelining and parallel execution of algorithms.

[0110] FIG. 13 is a diagram showing a network with a plurality of computer systems **1302a** and **1302b**, a plurality of cell phones and personal data assistants **1302c**, and Network Attached Storage (NAS) **1304a** and **1304b**. In example embodiments, systems **1302a**, **1302b**, and **1302c** can manage data storage and optimize data access for data stored in Network Attached Storage (NAS) **1304a** and **1304b**. A mathematical model can be used for the data and be evaluated using distributed parallel processing across computer systems **1302a**, and **1302b**, and cell phone and personal data assistant systems **1302c**. Computer systems **1302a**, and **1302b**, and cell phone and personal data assistant systems **1302c** can also provide parallel processing for adaptive data restructuring of the data stored in Network Attached Storage (NAS) **1304a** and **1304b**. FIG. 13 illus-

trates an example only, and a wide variety of other computer architectures and systems can be used in conjunction with the various embodiments of the present invention. For example, a blade server can be used to provide parallel processing. Processor blades can be connected through a back plane to provide parallel processing. Storage can also be connected to the back plane or as Network Attached Storage (NAS) through a separate network interface. In some instances, processors can maintain separate memory spaces and transmit data through network interfaces, back plane or other connectors for parallel processing by other processors. In some instances, some or all of the processors can use a shared virtual address memory space.

[0111] Any of the systems described herein may comprise sequence information stored on non-transitory computer readable storage media. In some instances, any of the systems described herein comprise a computer input file. In some instances, the computer input file comprises sequence information. In some instances, the computer input file comprises instructions for synthesis of a plurality of polynucleotide sequence. In some instances, the instructions are received by a computer. In some instances, the instructions are processed by the computer. In some instances, the instructions are transmitted to a material deposition device. In some instances, the non-transitory computer readable storage media is encoded with a program including instructions executable by the operating system of an optionally networked digital processing device. In some instances, a computer readable storage medium is a tangible component of a digital processing device. In some instances, a computer readable storage medium is optionally removable from a digital processing device. In some instances, a computer readable storage medium includes, by way of non-limiting examples, CD-ROMs, DVDs, flash memory devices, solid state memory, magnetic disk drives, magnetic tape drives, optical disk drives, cloud computing systems and services, and the like. In some instances, the program and instructions are permanently, substantially permanently, semi-permanently, or non-transitorily encoded on the media.

EXAMPLES

[0112] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

Example 1: Functionalization of a Substrate Surface

[0113] A substrate was functionalized to support the attachment and synthesis of a library of polynucleotides. The substrate surface was first wet cleaned using a piranha solution comprising 90% H₂SO₄ and 10% H₂O₂ for 20 minutes. The substrate was rinsed in several beakers with deionized water, held under a deionized water gooseneck faucet for 5 min, and dried with N₂. The substrate was subsequently soaked in NH₄OH (1:100; 3 mL:300 mL) for 5 min, rinsed with DI water using a handgun, soaked in three successive beakers with deionized water for 1 min each, and

then rinsed again with deionized water using the handgun. The substrate was then plasma cleaned by exposing the substrate surface to O₂. A SAMCO PC-300 instrument was used to plasma etch O₂ at 250 watts for 1 min in downstream mode.

[0114] The cleaned substrate surface was actively functionalized with a solution comprising N-(3-triethoxysilylpropyl)-4-hydroxybutyramide using a YES-1224P vapor deposition oven system with the following parameters: 0.5 to 1 torr, 60 min, 70° C. 135° C. vaporizer. The substrate surface was resist coated using a Brewer Science 200X spin coater. SPR™ 3612 photoresist was spin coated on the substrate at 2500 rpm for 40 sec. The substrate was pre-baked for 30 min at 90° C. on a Brewer hot plate. The substrate was subjected to photolithography using a Karl Suss MA6 mask aligner instrument. The substrate was exposed for 2.2 sec and developed for 1 min in MSF 26A. Remaining developer was rinsed with the handgun and the substrate soaked in water for 5 min. The substrate was baked for 30 min at 100° C. in the oven, followed by visual inspection for lithography defects using a Nikon L200. A cleaning process was used to remove residual resist using the SAMCO PC-300 instrument to O₂ plasma etch at 250 watts for 1 min.

[0115] The substrate surface was passively functionalized with a 100 μL solution of perfluorooctyltrichlorosilane mixed with 10 μL light mineral oil. The substrate was placed in a chamber, pumped for 10 min, and then the valve was closed to the pump and left to stand for 10 min. The chamber was vented to air. The substrate was resist stripped by performing two soaks for 5 min in 500 mL NMP at 70° C. with ultrasonication at maximum power (9 on Crest system). The substrate was then soaked for 5 min in 500 mL isopropanol at room temperature with ultrasonication at maximum power. The substrate was dipped in 300 mL of 200 proof ethanol and blown dry with N₂. The functionalized surface was activated to serve as a support for polynucleotide synthesis.

Example 2: Synthesis of a 50-Mer Sequence on an Oligonucleotide Synthesis Device

[0116] A two dimensional oligonucleotide synthesis device was assembled into a flowcell, which was connected to a flowcell (Applied Biosystems (“ABI394 DNA Synthesizer”). The two-dimensional oligonucleotide synthesis device was uniformly functionalized with N-(3-TRIETHOXYSILYLPROPYL)-4-HYDROXYBUTYRAMIDE (Gelest) was used to synthesize an exemplary polynucleotide of 50 bp (“50-mer polynucleotide”) using polynucleotide synthesis methods described herein.

[0117] The sequence of the 50-mer was as described in SEQ ID NO.: 1. 5'AGACAATCAACCATTGGGGTGGACAGCCTTGACCTCTAGACTTCGGCAT##TTTTTT TTT3' (SEQ ID NO.: 1), where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes), which is a cleavable linker enabling the release of polynucleotides from the surface during deprotection.

[0118] The synthesis was done using standard DNA synthesis chemistry (coupling, capping, oxidation, and deblocking) according to the protocol in Table 1 and ABI394 DNA Synthesizer.

TABLE 1

Synthesis Protocol			
General DNA Synthesis	Table 1		
Process Name	Process Step	Time (sec)	
General DNA Synthesis	WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
		Acetonitrile to Flowcell	23
		N2 System Flush	4
		Acetonitrile System Flush	4
	DNA BASE ADDITION (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2
		Activator to Flowcell	6
		Activator + Phosphoramidite to Flowcell	6
		Activator to Flowcell	0,5
		Activator + Phosphoramidite to Flowcell	5
		Activator to Flowcell	0,5
		Activator + Phosphoramidite to Flowcell	5
		Activator to Flowcell	0,5
		Activator + Phosphoramidite to Flowcell	5
		Incubate for 25 sec	25
	WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4
	Acetonitrile to Flowcell	15	
	N2 System Flush	4	
	Acetonitrile System Flush	4	
DNA BASE ADDITION (Phosphoramidite + Activator Flow)	Activator Manifold Flush	2	
	Activator to Flowcell	5	
	Activator + Phosphoramidite to Flowcell	18	
	Incubate for 25 sec	25	
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4	
	Acetonitrile to Flowcell	15	
	N2 System Flush	4	
	Acetonitrile System Flush	4	
CAPPING (CapA + B, 1:1, Flow)	CapA + B to Flowcell	15	
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4	
	Acetonitrile to Flowcell	15	
	Acetonitrile System Flush	4	
OXIDATION (Oxidizer Flow)	Oxidizer to Flowcell	18	
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4	
	N2 System Flush	4	
	Acetonitrile System Flush	4	
	Acetonitrile to Flowcell	15	
	Acetonitrile System Flush	4	
	Acetonitrile to Flowcell	15	
	N2 System Flush	4	
	Acetonitrile System Flush	4	
	Acetonitrile to Flowcell	23	
	N2 System Flush	4	
	Acetonitrile System Flush	4	
DEBLOCKING (Deblock Flow)	Deblock to Flowcell	36	
WASH (Acetonitrile Wash Flow)	Acetonitrile System Flush	4	
	N2 System Flush	4	
	Acetonitrile System Flush	4	
	Acetonitrile to Flowcell	18	
	N2 System Flush	4,13	
	Acetonitrile System Flush	4,13	
	Acetonitrile to Flowcell	15	

[0119] The phosphoramidite/activator combination was delivered similar to the delivery of bulk reagents through the flowcell. No drying steps were performed as the environment stays “wet” with reagent the entire time.

[0120] The flow restrictor was removed from the ABI394 DNA Synthesizer to enable faster flow. Without flow restric-

tor, flow rates for amidites (0.1M in ACN), Activator, (0.25M Benzoylthiotetrazole (“BTT”; 30-3070-xx from GlenResearch) in ACN), and Ox (0.02M 12 in 20% pyridine, 10% water, and 70% THF) were roughly ~100 uL/sec, for acetonitrile (“ACN”) and capping reagents (1:1 mix of CapA and CapB, wherein CapA is acetic anhydride in THF/Pyridine and CapB is 16% 1-methylimidazole in THF), roughly ~200 uL/sec, and for Deblock (3% dichloroacetic acid in toluene), roughly ~300 uL/sec (compared to ~50 uL/sec for all reagents with flow restrictor). The time to completely push out Oxidizer was observed, the timing for chemical flow times was adjusted accordingly and an extra ACN wash was introduced between different chemicals. After polynucleotide synthesis, the chip was deprotected in gaseous ammonia overnight at 75 psi. Five drops of water were applied to the surface to recover polynucleotides. The recovered polynucleotides were then analyzed on a BioAnalyzer small RNA chip (data not shown).

Example 3: Synthesis of a 100-Mer Sequence on an Oligonucleotide Synthesis Device

[0121] The same process as described in Example 2 for the synthesis of the 50-mer sequence was used for the synthesis of a 100-mer polynucleotide (“100-mer polynucleotide”; 5' CGGGATCCTTATCGTCATCGTCGTACA-GATCCCGACCCATTTGCTGTCCACCAGTCATGCTAGCCATACCATGATGATGATGATGATGAGA ACCCCGCAT##TTTTTTTTTT3', where # denotes Thymidine-succinyl hexamide CED phosphoramidite (CLP-2244 from ChemGenes); SEQ ID NO.: 2) on two different silicon chips, the first one uniformly functionalized with N-(3-TRIETHOXSILYLPROPYL)-4-HYDROXYBUTYRAMIDE and the second one functionalized with 5/95 mix of 11-acetoxyundecyltriethoxysilane and n-decyltriethoxysilane, and the polynucleotides extracted from the surface were analyzed on a BioAnalyzer instrument (data not shown).

[0122] All ten samples from the two chips were further PCR amplified using a forward (5'ATGCCGGGTTCTCATCATC3'; SEQ ID NO.: 3) and a reverse (5'CGGGATCCTTATCGTCATCG3'; SEQ ID NO.: 4) primer in a 50 uL PCR mix (25 uL NEB Q5 mastermix, 2.5 uL 10 uM Forward primer, 2.5 uL 10 uM Reverse primer, 1 uL polynucleotide

extracted from the surface, and water up to 50 uL) using the following thermalcycling program:

[0123] 98° C. 30 sec

[0124] 98° C. 10 sec; 63° C. 10 sec; 72° C. 10 sec; repeat 12 cycles

[0125] 72° C. 2 min

[0126] The PCR products were also run on a BioAnalyzer (data not shown), demonstrating sharp peaks at the 100-mer position. Next, the PCR amplified samples were cloned, and Sanger sequenced. Table 2 summarizes the results from the Sanger sequencing for samples taken from spots 1-5 from chip 1 and for samples taken from spots 6-10 from chip 2.

TABLE 2

Sequencing Results		
Spot	Error rate	Cycle efficiency
1	1/763 bp	99.87%
2	1/824 bp	99.88%
3	1/780 bp	99.87%
4	1/429 bp	99.77%
5	1/1525 bp	99.93%
6	1/1615 bp	99.94%
7	1/531 bp	99.81%
8	1/1769 bp	99.94%
9	1/854 bp	99.88%
10	1/1451 bp	99.93%

[0127] Thus, the high quality and uniformity of the synthesized polynucleotides were repeated on two chips with different surface chemistries. Overall, 89%, corresponding to 233 out of 262 of the 100-mers that were sequenced were perfect sequences with no errors.

[0128] Finally, Table 3 summarizes key error characteristics for the sequences obtained from the polynucleotides samples from spots 1-10.

TABLE 3

Error Characteristics										
Sample ID/ Spot no.	OSA_ 0046/1	OSA_ 0047/2	OSA_ 0048/3	OSA_ 0049/4	OSA_ 0050/5	OSA_ 0051/6	OSA_ 0052/7	OSA_ 0053/8	OSA_ 0054/9	OSA_ 0055/10
Total Sequences	32	32	32	32	32	32	32	32	32	32
Sequencing Quality	25 of 28	27 of 27	26 of 30	21 of 23	25 of 26	29 of 30	27 of 31	29 of 31	28 of 29	25 of 28
Oligo Quality	23 of 25	25 of 27	22 of 26	18 of 21	24 of 25	25 of 29	22 of 27	28 of 29	26 of 28	20 of 25
ROI Match Count	2500	2698	2561	2122	2499	2666	2625	2899	2798	2348
ROI Mutation	2	2	1	3	1	0	2	1	2	1
ROI Multi	0	0	0	0	0	0	0	0	0	0
Base Deletion										
ROI Small	1	0	0	0	0	0	0	0	0	0
Insertion										
ROI Single	0	0	0	0	0	0	0	0	0	0
Base Deletion										
Large Deletion	0	0	1	0	0	1	1	0	0	0
Count										
Mutation: G > A	2	2	1	2	1	0	2	1	2	1
Mutation: T > C	0	0	0	1	0	0	0	0	0	0

TABLE 3-continued

Sample ID/ Spot no.	Error Characteristics									
	OSA_ 0046/1	OSA_ 0047/2	OSA_ 0048/3	OSA_ 0049/4	OSA_ 0050/5	OSA_ 0051/6	OSA_ 0052/7	OSA_ 0053/8	OSA_ 0054/9	OSA_ 0055/10
ROI Error Count	3	2	2	3	1	1	3	1	2	1
ROI Error Rate	Err: ~1 in 834	Err: ~1 in 1350	Err: ~1 in 1282	Err: ~1 in 708	Err: ~1 in 2500	Err: ~1 in 2667	Err: ~1 in 876	Err: ~1 in 2900	Err: ~1 in 1400	Err: ~1 in 2349
ROI Minus	MP	MP	MP	MP	MP	MP	MP	MP	MP	MP
Primer Error Rate	Err: ~1 in 763	Err: ~1 in 824	Err: ~1 in 780	Err: ~1 in 429	Err: ~1 in 1525	Err: ~1 in 1615	Err: ~1 in 531	Err: ~1 in 1769	Err: ~1 in 854	Err: ~1 in 1451

Example 4. Overlap Extension PCR without Primer Removal

[0129] Two genes (Gene 1 and Gene 2) were selected to perform overlap extension PCR without primer removal. Varying lengths of base pair overlap were tested for Gene 1 and Gene 2, including 20, 41, and 100 base pairs. PCR amplification was performed using universal primers. Referring to FIG. 14, incorrect assembly (white) and correct assembly (black) were determined for Gene 1 and Gene 2 amplified with universal primers and having a base pair overlap of 20, 41, and 100 base pairs. Assembly of Gene 1 and Gene 2 without amplification by universal primers with a 20 base pair overlap was also determined. As seen in FIG. 14, universal primers having a base pair overlap of 41 base pairs resulted in improved assembly for Gene 1 and Gene 2. For Gene 1, assembly was about 91%.

Example 5. Single-Stranded DNA Mediated Hierarchical Assembly of Two Gene Fragments

[0130] A gene was assembled into a 3 kb gene from two gene fragments. The 5' and 3' of each fragment end was appended with uracils and amplified using a uracil compatible polymerase such as KapaU polymerase or PhusionU polymerase. Each of the two fragments further comprised a homology sequence. See FIG. 1. The two fragments were then mixed and amplified with universal primers and Q5 DNA polymerase. Q5 DNA polymerase is incompatible with uracil and resulted in stalling at a uracil base. Single stranded fragments that do not comprise uracil were thus generated. The two fragments were then combined and amplified to generate the 3 kb gene.

Example 6. Single-Stranded DNA Mediated Hierarchical Assembly of Three Gene Fragments

[0131] A gene was assembled into a 3 kb gene from three gene fragments. The 5' and 3' of each fragment end was appended with flanking adapter sequences comprising uracils and amplified using a uracil compatible polymerase such as KapaU polymerase or PhusionU polymerase. Each of the three fragments comprised a homology sequence. See FIG. 2. The three fragments were mixed and amplified with Q5 DNA polymerase. Q5 DNA polymerase is incompatible with uracil and resulted in stalling at a uracil base. Single stranded fragments that do not comprise uracil were thus generated. Two of the three fragments were then combined and amplified to generate a single fragment. The third fragment was combined with the synthesized fragment and amplified to generate the 3 kb gene.

Example 7. Single-Stranded DNA Mediated Hierarchical Assembly with Varying Base Pair Overlap Sequence Length

[0132] Two genes (Gene 1 and Gene 2) were assembled similar to Examples 5-6. The 5' and 3' of each fragment end was appended with flanking adapter sequences comprising uracils and amplified using a uracil compatible polymerase such as KapaU polymerase or PhusionU polymerase. Each of the two fragments comprised a homology sequence. The length of the homology sequence was 20, 41, or 100 base pairs. The two fragments were mixed and amplified with universal primers and Q5 DNA polymerase or KapaHiFi polymerase. Q5 DNA polymerase and KapaHiFi polymerase are incompatible with uracil and resulted in stalling at a uracil base. Single stranded fragments that do not comprise uracil were thus generated. The two fragments are combined and amplified to generate Gene 1 and Gene 2.

[0133] Incorrect assembly (white) and correct assembly (black) were determined for Gene 1 and Gene 2 amplified with Q5 DNA polymerase and KapaHiFi polymerase and having a base pair overlap of 20, 41, and 100 base pair (FIG. 15). The 20, 41, and 100 base pair overlap resulted in over 70% correct assembly with both Q5 DNA polymerase and KapaHiFi polymerase (FIG. 15). The 41 base pair overlap resulted in further improved assembly at 100% of correct assembly with both Q5 DNA polymerase and KapaHiFi polymerase (FIG. 15).

Example 8. In Vitro Recombination Cloning

[0134] In Vitro Recombination (IVTR) Lysate Preparation
[0135] JM109 cell lysate was prepared. JM109 was streaked on a M9 agar plate and incubated at 37° C. for about 24 hours. Colonies were picked and incubated in 15 mL tubes comprising 3 mL of M9 broth for a seed culture. The tubes were shaken at 250 rpm at 37° C. for about 19 hours. When OD600 was between about 2-3, 0.016 mL of the seed culture was removed and added to flasks comprising 100 mL of TB broth. The flasks were shaken at 300 rpm at 37° C. for about 4.5 hours. The cells were spun down at 4,816×g for 10 minutes and washed with 100 mL of ice-cold water. The cells were lysed with 2.4 mL CelLytic B Cell Lysis Reagent (Sigma-Aldrich) for 10 minutes at room-temperature and centrifuged at 18,615×g for 5 minutes to collect the supernatant (about 2 mL). The supernatant was mixed with 2 mL of ice cold 80% glycerol and snap frozen in a dry ice-ethanol bath followed by storage at -80° C.

[0136] IVTR Cloning

[0137] A gene fragment comprised from 5' to 3': a 5' flanking adapter sequence, a first homology sequence, an insert sequence, a second homology sequence, and a 3'

flanking adapter sequence. The gene fragment and a vector comprising homologous sequences to the first homology sequence and the second homology sequence were incubated with JM109 cell lysate. See FIG. 4. A reaction was set up similar to Table 4. Varying amounts of insert were tested: 13 fmol (white bars), 26 fmol (hashed bars), and 40 fmol (black bars) (FIG. 16). The ratio of insert:vector (X-axis) was also varied including ratios of 1:1, 2:1, and 4:1.

[0138] Tens of thousands colonies were retrieved. Colony forming unit (CFU, Y-axis) was then determined as a measure of efficiency. More than 99% correct assembly of any selected condition (1 misassembly out of 437 valid clones) was observed (FIG. 16). Referring to FIG. 17, colony PCR results of 48 colonies showed only 4 samples failed, which were background plasmids.

TABLE 4

Reaction Conditions	
	Final Concentration
10X IVTR buffer	1X
JM109	
VS208 ASP	4 nM
Fragment 1	4 nM
Fragment 2	4 nM

Example 9. In Vitro Recombination Cloning with Two Fragments

[0139] A first gene fragment comprised from 5' to 3': a 5' flanking adapter sequence, a first homology sequence, an insert sequence, a second homology sequence, and a 3' flanking adapter sequence. A second gene fragment comprised from 5' to 3': a 5' flanking adapter sequence, a first homology sequence, an insert sequence, a second homology sequence, and a 3' flanking adapter sequence. The first homology sequence of the first gene fragment was homologous to a sequence on the vector. See FIG. 4. The first gene fragment, second gene fragment, and vector were incubated with JM109 cell lysate.

[0140] The effects of the length of the homology sequence (20, 41, or 100 base pairs) and location of the homology sequence on colony forming unit (CFU, Y-axis) were determined (FIG. 18). Data is summarized in Table 5. The homology sequence was either flanked by the universal primers (internal) or at the 5' or 3' end of the gene fragment (terminal). There was increased CFUs with a homology sequence of 41 base pairs. A terminal location of the homology sequence of 41 base pairs resulted in a greater number of CFUs as compared to an internal location. A terminal location of the homology sequence of 100 base pairs also resulted in a greater number of CFUs as compared to an internal location.

[0141] Referring to FIG. 19, the percentage of misassembly (white bars) and assembly (black bars) was determined. A terminal location and homology sequence length of 20, 41, and 100 base pairs resulted in 100% assembly. An internal location and homology sequence length of 20 and 41 base pairs also resulted in 100% assembly.

TABLE 5

Colony Forming Unit Results		
Homology Sequence Length (number of base pairs)	Terminal or Internal	CFU
20	Terminal	190
20	Internal	200
41	Terminal	1280
41	Internal	720
100	Terminal	1030
100	Internal	190

Example 10. In Vitro Recombination Cloning with Varying Lengths of Non-Homologous Sequences

[0142] The effect of varying lengths of the non-homologous sequence was determined. Four fragments were designed: V1, V2, V3, and V4 (FIG. 5). V1 comprised an overlapping sequence at the terminal end. V2 comprised an internal overlapping sequence followed by a 24 base pair sequence at the 3' end. V3 comprised an overlapping sequence followed by a 124 base pair sequence at the end 3' end. V4 comprised an overlapping sequence followed by a 324 base pair at the 3' end.

[0143] Percentage of assembly (FIG. 20A) and total CFU (FIG. 20B) were determined for sequences comprising 0 (V1), 24 (V2), 124 (V3), and 324 (V4) base pairs at the 3' end. Referring to FIG. 20A, 100% correct assembly was observed for sequences comprising 0, 24, 124, and 324 base pairs at the 3' end. Referring to FIG. 20B, sequences comprising 0 and 24 base pairs at the 3' end resulted in increased CFUs.

Example 11. In Vitro Recombination Cloning with Varying Internal Homology

[0144] The effect of internal homology was determined. Three fragments were designed: V5, V6, and V7 (FIG. 6A). V5 comprised a 24 base pair sequence between two overlapping sequences. V6 comprised a 124 base pair sequence between two overlapping sequences. V7 comprised a 324 base pair sequence between two overlapping sequences.

[0145] Percentage of assembly (FIG. 21A) and Total CFU (FIG. 21B) were determined for internal sequences comprising 24 (V5), 124 (V6), and 324 (V7) base pairs. Referring to FIG. 21A, more than 80% correct assembly was observed for internal sequences comprising 124 and 324 base pairs. Referring to FIG. 21B, internal sequences comprising 124 and 324 base pairs also resulted in increased CFUs.

Example 12. Single-Stranded DNA Mediated Hierarchical Assembly of Multiple Gene Fragments

[0146] A gene is assembled into a 3 kb gene from five gene fragments. The 5' and 3' of each fragment end is appended with flanking adapter sequences comprising uracils. Each of the five fragments comprises a homology sequence. The five fragments are mixed and amplified with universal primers and Q5 DNA polymerase. Following amplification using DNA polymerase, single stranded fragments that do not comprise uracil are generated. Two of the five fragments are combined and amplified with DNA polymerase to generate a single fragment. A third fragment is combined with the synthesized fragment and amplified to generate a single

fragment. A fourth fragment is combined with the synthesized fragment and amplified to generate a single fragment. A fifth fragment is combined with the synthesized fragment and amplified to generate the assembled 3 kb gene.

[0147] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations,

changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention. It is intended that the following claims define the scope of the invention and that methods and structures within the scope of these claims and their equivalents be covered thereby.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 4

<210> SEQ ID NO 1

<211> LENGTH: 62

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<220> FEATURE:

<221> NAME/KEY: modified_base

<222> LOCATION: (51)..(52)

<223> OTHER INFORMATION: Thymidine-succinyl hexamide CED phosphoramidite

<400> SEQUENCE: 1

agacaatcaa ccatttgggg tggacagcct tgacctctag acttcggcat tttttttttt 60

tt 62

<210> SEQ ID NO 2

<211> LENGTH: 112

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<220> FEATURE:

<221> NAME/KEY: modified_base

<222> LOCATION: (101)..(102)

<223> OTHER INFORMATION: Thymidine-succinyl hexamide CED phosphoramidite

<400> SEQUENCE: 2

cgggatcctt atcgatcgcg tcgtacagat cccgacccat ttgctgtcca ccagtcgatgc 60

tagccatacc atgatgatga tgatgatgag aaccccgcac tttttttttt tt 112

<210> SEQ ID NO 3

<211> LENGTH: 19

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 3

atgcgggggtt ctcatcatc 19

<210> SEQ ID NO 4

<211> LENGTH: 20

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 4

cgggatcctt atcgatcgcg 20

What we claim is:

1. A method for nucleic acid assembly, comprising:
 - (a) providing at least one double stranded nucleic acid comprising in 5' to 3' order: a 5' flanking adapter sequence, a first homology sequence, an insert sequence, a second homology sequence, and a 3' flanking adapter sequence, wherein the first homology sequence and the second homology sequence each comprises about 20 to about 100 base pairs in length;
 - (b) providing a vector comprising the first homology sequence and the second homology sequence; and
 - (c) mixing the at least one double stranded nucleic acid and the vector with a bacterial lysate.
2. The method of claim 1, wherein the bacterial lysate comprises a nuclease or a recombinase.
3. The method of claim 1, wherein the bacterial lysate comprises a nuclease and a recombinase.
4. The method of claim 1, wherein the first homology sequence and the second homology sequence each comprises about 20 base pairs.
5. The method of claim 1, wherein the first homology sequence and the second homology sequence each comprises about 41 base pairs.
6. The method of claim 1, wherein the first homology sequence and the second homology sequence each comprises about 100 base pairs.
7. The method of claim 1, wherein the first homology sequence or the second homology sequence is flanked by the 5' flanking adapter sequence and the 3' flanking adapter sequence.
8. The method of claim 1, wherein a percentage of correct assembly is at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.
9. A method for nucleic acid synthesis and assembly, comprising:
 - (a) de novo synthesizing a plurality of polynucleotides, wherein each polynucleotide comprises a first homology region that comprises in 5' to 3' order: a 5' flanking adapter sequence, a first homology sequence, an insert sequence, a second homology sequence, and a 3' flanking adapter sequence, wherein the first homology sequence and the second homology sequence each comprises about 20 to about 100 base pairs in length, and wherein each polynucleotide comprises a homology sequence identical to that of another polynucleotide of the plurality of polynucleotides; and
 - (b) mixing of the plurality of polynucleotides with a bacterial lysate to processively form nucleic acids each having a predetermined sequence.
10. The method of claim 9, wherein the bacterial lysate comprises a nuclease or a recombinase.
11. The method of claim 9, wherein the bacterial lysate comprises a nuclease and a recombinase.
12. The method of claim 9, wherein the first homology sequence and the second homology sequence each comprises about 20 base pairs.
13. The method of claim 9, wherein the first homology sequence and the second homology sequence each comprises about 41 base pairs.
14. The method of claim 9, wherein the first homology sequence and the second homology sequence each comprises about 100 base pairs.
15. The method of claim 9, wherein the first homology sequence or the second homology sequence is flanked by the 5' flanking adapter sequence and the 3' flanking adapter sequence.
16. The method of claim 9, wherein a percentage of correct assembly is at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.
17. A method for nucleic acid assembly, comprising:
 - (a) providing a plurality of double stranded nucleic acids;
 - (b) annealing a uracil at a 5' end and a 3' end of at least two of the double stranded nucleic acids;
 - (c) amplifying the double stranded nucleic acids using a uracil compatible polymerase to form amplification products;
 - (d) mixing the amplification products from step (c) to form a mixture; and
 - (e) amplifying the mixture from step (d) using a uracil incompatible polymerase to generate a single-stranded nucleic acid.
18. The method of claim 17, wherein the uracil incompatible polymerase is a DNA polymerase.
19. The method of claim 17, wherein the homology sequence comprises about 20 to about 100 base pairs.
20. The method of claim 17, wherein a percentage of correct assembly is at least 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 99%.

* * * * *