(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 January 2014 (03.01.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/004393 A1

(51) International Patent Classification:
C40B 50/06 (2006.01)

(21) International Application Number:
PCT/US2013/047370

(22) International Filing Date:
24 June 2013 (24.06.2013)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/664,118 25 June 2012 (25.06.2012) US
61/731,627 30 November 2012 (30.11.2012) US

(71) Applicant: **GEN9, INC.** [US/US]; 500 Technology Square, Suite 130, Cambridge, MA 02139 (US).

(72) Inventors: **HUDSON, Michael, E.**; 21 Crestwood Drive, Framingham, MA 01701 (US). **KUNG, Li-jun, A.**; 4 Knowles Farm Road, Arlington, MA 02474 (US). **SCHINDLER, Daniel**; 31 Williams Street, Newton, MA 02464 (US). **ARCHER, Stephen**; 126 Charles Street, Apt. 4, Cambridge, MA 02141 (US). **SAAEM, Ishtiaq**; 100 Stockton Street, Apt. 205, Chelsea, MA 02150 (US).

(74) Agent: **CAMACHO, Jennifer, A.**; Greenberg Traurig, LLP, One International Place, Boston, MA 02110 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: METHODS FOR NUCLEIC ACID ASSEMBLY AND HIGH THROUGHPUT SEQUENCING

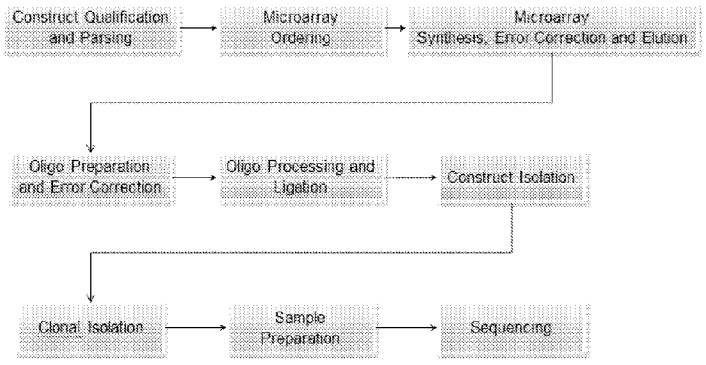


FIG. 1

(57) Abstract: Methods and apparatus of some aspects of the invention relate to the synthesis of high fidelity polynucleotides. In particular, aspects of the invention relate to concurrent enzymatic removal of amplification sequences and ligation of processed oligonucleotides into nucleic acid assemblies. According to some embodiments, the invention provides a method for producing a target nucleic acid having a predefined sequence. In some embodiments, the method comprises the step of providing a plurality of oligonucleotides, wherein each oligonucleotides comprises (i) an internal sequence identical to a different portion of a sequence of a target nucleic acid, (ii) a 5' sequence flanking the 5' end of the internal sequence and a 3' flanking sequence flanking the 3' end of the internal sequence, each of the flanking sequence comprising a primer recognition site for a primer pair and a restriction enzyme recognition site.

WO 2014/004393 A1

METHODS FOR NUCLEIC ACID ASSEMBLY AND HIGH THROUGHPUT SEQUENCING

RELATED APPLICATIONS

[0001] This application claims the benefit of and priority to United States Provisional Application No. 61/664,118, filed June 25, 2012, and United States Provisional Application No. 61/731,627, filed November 30, 2012, each of which is incorporated herein by reference in its entirety.

FIELD OF THE INVENTION

[0002] Methods and apparatuses provided herein relate to the synthesis and assembly of high fidelity nucleic acids and nucleic acid libraries having a predefined sequence. More particularly, methods and apparatuses are provided for polynucleotide synthesis, error reduction, and/or high throughput sequence verification.

BACKGROUND

[0003] Using the techniques of recombinant DNA chemistry, it is now common for DNA sequences to be replicated and amplified from nature and then disassembled into component parts. As component parts, the sequences are then recombined or reassembled into new DNA sequences. However, reliance on naturally available sequences significantly limits the possibilities that may be explored by researchers. While it is now possible for short DNA sequences to be directly synthesized from individual nucleosides, it has been generally impractical to directly construct large segments or assemblies of polynucleotides, i.e., polynucleotide sequences longer than about 400 base pairs.

[0004] Oligonucleotide synthesis can be performed through massively parallel custom syntheses on microchips (Zhou et al. (2004) Nucleic Acids Res. 32:5409; Fodor et al. (1991) Science 251:767). However, current microchips have very low surface areas and hence only small amounts of oligonucleotides can be produced. When released into solution, the oligonucleotides are present at picomolar or lower concentrations per sequence, concentrations that are insufficiently high to drive bimolecular priming reactions efficiently. Current methods for assembling small numbers of variant nucleic acids cannot be scaled up in a cost-effective

manner to generate large numbers of specified variants. As such, a need remains for improved methods and devices for high-fidelity gene assembly and the like.

[0005] Furthermore, oligonucleotides on microchips are generally synthesized via chemical reactions. Spurious chemical reactions cause random base errors in oligonucleotides. One of the critical limitations in chemical nucleic acid synthesis is the error-rate. The error rate of chemically-synthesized oligonucleotides (e.g. deletions at a rate of 1 in 100 bases and mismatches and insertions at about 1 in 400 bases) exceeds the error rate obtainable through enzymatic means of replicating an existing nucleic acid (e.g., PCR). Therefore, there is an urgent need for new technology to produce high yield high-fidelity polynucleotides in a cost efficient manner.

SUMMARY

[0006] Aspects of the invention relate to methods, systems and compositions for preparing and/or assembling high fidelity polymers. Also provided herein are devices and methods for processing nucleic acid assembly reactions and assembling nucleic acids. It is an object of this invention to provide practical, economical methods of synthesizing custom polynucleotides. It is a further object to provide methods of producing synthetic polynucleotides that have lower error rates than synthetic polynucleotides made by methods known in the art.

[0007] According to some embodiments, the invention provides a method for producing a target nucleic acid having a predefined sequence. In some embodiments, the method comprises the step of providing a plurality of oligonucleotides, wherein each oligonucleotides comprises (i) an internal sequence identical to a different portion of a sequence of a target nucleic acid, (ii) a 5' sequence flanking the 5' end of the internal sequence and a 3' flanking sequence flanking the 3' end of the internal sequence, each of the flanking sequence comprising a primer recognition site for a primer pair and a restriction enzyme recognition site. The method further comprises, in some embodiments, amplifying at least a subset of the oligonucleotides using the primer pair thereby generating a plurality of amplified oligonucleotides. The plurality of amplified oligonucleotides can then be exposed to a restriction enzyme and ligase in a single pool, wherein the restriction enzyme is capable of recognizing the restriction enzyme recognition site, thereby generating the target nucleic acid.

[0008] In some embodiments, the method comprises subjecting the assembled target nucleic acid to sequence verification. In some embodiments, the amplified double stranded oligonucleotides can comprise a sequence error or msimatch. In some embodiments, the method comprises subjecting the plurality of amplified oligonucleotides to error removal. In some embodiments, the plurality of amplified oligonucleotides can be contacted with a mismatch binding agent. The mismatch binding agent can selectively associate with the double-stranded oligonucleotides comprising a mismatch, resulting in a binding and cleaving action. In some embodiments, the plurality of amplified oligonucleotides can be contacted with a mismatch recognizing agent, for example, a chemical such as lysine, piperidine or the like.

[0009] In some embodiments, the restriction enzyme and the ligase are added to a single pool of amplified oligonucleotides under conditions suitable to promote digestion and ligation thereby generating a mixture comprising the assembled target nucleic acid sequences, and the flanking regions. In some embodiments, each flanking region comprises a common primer recognition site. In some embodiments, the restriction enzyme is a type IIS restriction enzyme. Digestion with the type IIS restriction enzyme can produce a plurality of cohesive end double-stranded construction oligonucleotides and the plurality of cohesive end double stranded construction oligonucleotides can be ligated in a unique linear arrangement.

[0010] In some embodiments, the method comprises amplifying the target nucleic acid using a primer pair capable of recognizing the primer recognition sites at the 5' end and 3' end of the target nucleic acid. In some embodiments, the method comprises sequencing the target nucleic acid to confirm its sequence accuracy, for example, by high throughput sequencing. In some embodiments, the method comprises isolating at least one target nucleic acid having the predefined sequence from a pool of nucleic acid sequences.

[0011] According to some embodiments, the invention provides a method for further processing the isolated nucleic acids. In some embodiments, the method comprises assembling at least two target nucleic acids. The step of assembling can be by hierarchical assembly. In some embodiments, the at least two target nucleic acids are subjected to restriction enzyme digestion and ligation thereby forming a long target nucleic acid construct, for example, at least about 10 kilobases or 100 kilobases in length.

[0012] According to some embodiments, the invention provides a method for producing a target nucleic acid having a predefined sequence in a vector. In some embodiments, a plurality

of oligonucleotides are provided, each oligonucleotide comprising (i) an internal sequence identical to a different portion of a sequence of a target nucleic acid, (ii) a 5' flanking sequence flanking the 5' end of the internal sequence and a 3' flanking sequence flanking the 3' end of the internal sequence, each of the flanking sequence comprising a primer recognition site for a primer pair and a restriction enzyme recognition site for a restriction endonuclease. In some embodiments, at least a subset of the oligonucleotides can be amplified using the primer pair thereby generating a plurality of amplified oligonucleotides. In some embodiments, the plurality of amplified oligonucleotides can be subjected to error removal and/or correction. In some embodiments, a circular vector having a restriction enzyme recognition site for the restriction endonuclease is provided. In some embodiments, the plurality of amplified oligonucleotides and circular vector can be exposed to the restriction enzyme and ligase in a single pool, wherein the restriction enzyme is capable of recognizing the restriction enzyme recognition sites, thereby assembling the target nucleic acid in the vector. In some embodiments, the method further comprises transforming the vector into a host cell and sequence verifying the target nucleic acid sequence.

[0013] According to some embodiments, the invention provides a composition for the assembly of a target nucleic acid having a predefined sequence. In some embodiments, the composition comprises a plurality of oligonucleotides, wherein each oligonucleotide comprises (i) an internal sequence identical to a different portion of a sequence of a target nucleic acid, (ii) 5' flanking sequence flanking the 5' end of the internal sequence and 3' flanking sequence flanking the 3' end of the internal sequence, each of the flanking sequence comprising a primer recognition site for a primer pair and a restriction enzyme recognition site for a restriction endonuclease. In some embodiments, the composition further comprises a restriction endonuclease and/or a ligase. In some embodiments, the composition further comprises a vector comprising a pair of enzyme recognition sites for a restriction endonuclease. In some embodiments, the restriction endonuclease is a type IIS restriction endonuclease.

[0014] In some embodiments, the plurality of oligonucleotides is amplified and/or error corrected.

[0015] In some aspects of the invention, the method of producing a target nucleic acid having a predefined sequence comprises providing a first mixture comprising (i) a restriction enzyme, and (ii) a first pool of oligonucleotides comprising a first oligonucleotide comprising a

sequence identical to the 5' end of the target nucleic acid, a second oligonucleotide comprising a sequence identical to the 3' end of the target nucleic acid; and a plurality of oligonucleotides comprising a sequence identical to a different portion of a sequence of a target nucleic acid, each of the oligonucleotides having an overlapping sequence region corresponding to a sequence region in a next oligonucleotide, the oligonucleotides in the first pool together comprising the target nucleic acid sequence; and exposing the mixture to a ligase, thereby generating the target nucleic acid. The target nucleic acid can then be subjected to sequence verification.

[0016] In some embodiments, the methods of the invention comprise providing a pool of construction oligonucleotides and involve amplification of the oligonucleotides at different stages. The term "construction oligonucleotide" refers to a single stranded oligonucleotide that may be used for assembling nucleic acid molecules that are longer than the construction oligonucleotide itself. Construction oligonucleotides may be single stranded oligonucleotides or double stranded oligonucleotides. In some embodiments, construction oligonucleotides are synthetic oligonucleotides and may be synthesized in parallel on a substrate.

[0017] In some embodiments, the method further comprises prior to providing the first mixture, the step of providing a plurality of construction oligonucleotides, wherein each construction oligonucleotide comprises (i) an internal sequence identical to a different portion of a sequence of a target nucleic acid, (ii) 5' flanking sequence flanking the 5' end of the internal sequence and a 3' flanking sequence flanking the 3' end of the internal sequence, each flanking region comprising a primer recognition site for a primer pair and a restriction enzyme recognition site. In some embodiments, each flanking region can comprise a common primer recognition site. In some embodiments, the plurality of construction oligonucleotides can be amplified. In some embodiments, the oligonucleotides can comprise a sequence error or mismatch. In some embodiments, the plurality of amplified oligonucleotides can be subjected to error removal. For example, the plurality of amplified oligonucleotides can be contacted with a mismatch binding agent, wherein the mismatch binding agent selectively binds and cleaves the double-stranded oligonucleotides comprising a mismatch.

[0018] In some embodiments, the restriction enzyme and the ligase can be added to a single pool of amplified oligonucleotides under conditions suitable to promote digestion and ligation thereby generating a mixture comprising the assembled target nucleic acid sequences, and the flanking regions. In some embodiments, the restriction enzyme can be a type IIS

restriction enzyme and digestion with the type IIS restriction enzyme can produce a plurality of cohesive end double-stranded oligonucleotides and wherein the plurality of cohesive end double stranded oligonucleotides are ligated in a unique linear arrangement.

[0019] In some embodiments, the method further comprises amplifying the target nucleic acid using a primer pair capable of recognizing a primer recognition site at the 5' end of the first oligonucleotide and 3' end of second oligonucleotide.

[0020] In some embodiments, the method further comprises sequencing the target nucleic acid to confirm its sequence accuracy, for example, by high throughput sequencing.

[0021] In some embodiments, the method further comprises isolating at least one target nucleic acid having the predefined sequence from a pool of nucleic acid sequences.

[0022] In some embodiments, the method further comprises processing the target nucleic acids.

[0023] In some embodiments, the method further comprises providing a second mixture comprising (i) a restriction enzyme, and (ii) a second pool of oligonucleotides comprising a first oligonucleotide comprising a sequence identical to the 5' end of the target nucleic acid, a second oligonucleotide comprising a sequence identical to the 3' end of the target nucleic acid; and a plurality of oligonucleotides comprising a sequence identical to a different portion of a sequence of a target nucleic acid, each oligonucleotide having an overlapping sequence region corresponding to a sequence region in a next oligonucleotide, the oligonucleotides in the second pool together comprising the second target nucleic acid. In some embodiments, the second mixture is exposed to a ligase, thereby generating a second target nucleic acid. In some embodiments, the second oligonucleotide of the first pool comprises a restriction endonuclease recognition site for a restriction endonuclease and the first oligonucleotide of the second pool comprises a restriction endonuclease recognition site for the restriction endonuclease.

[0024] In some embodiments, the method further comprises assembling at least two target nucleic acids. In some embodiments, the step of assembling is by hierarchical assembly. In some embodiments, the at least two target nucleic acids are subjected to restriction endonuclease digestion and ligation thereby forming a long target nucleic acid construct. In some embodiments, the long target nucleic acid construct is at least about 10 kilobases in length or at least about 100 kilobases in length.

[0025] In some aspects, the invention relates to a composition for the assembly of a target nucleic acid having a predefined sequence, the composition comprising a plurality of oligonucleotides comprising a first oligonucleotide comprising a sequence identical to the 5' end of the target nucleic acid, a second oligonucleotide comprising a sequence identical to the 3' end of the target nucleic acid; and one or more oligonucleotides comprising a sequence identical to a different portion of a sequence of a target nucleic acid, each of the oligonucleotides having an overlapping sequence region corresponding to a sequence region in a next oligonucleotide, the plurality of oligonucleotides together comprising the target nucleic acid; a plurality of common sequences comprising a primer recognition site for a primer pair and a restriction endonuclease recognition site. In some embodiments, the composition further comprises a restriction endonuclease and/or a ligase. The restriction endonuclease can be a type IIS restriction endonuclease.

[0026] In some embodiments, the plurality of oligonucleotides can be amplified and/or error-corrected. In some embodiments, the composition can further comprise a linearized vector having a 5' compatible with the first oligonucleotide and a 3' end compatible with the second oligonucleotide.

BRIEF DESCRIPTION OF THE FIGURES

[0027] Fig. 1 illustrates an exemplary process for high fidelity nucleic acid assembly according to one embodiment of the invention.

[0028] Fig. 2 illustrates a non-limiting example of assembly method of a polynucleotide having a predefined sequence.

[0029] Fig. 3 illustrates a non-limiting example of assembly method of a polynucleotide having a predefined sequence into a vector.

[0030] Fig. 4 illustrates a non-limiting example of hierarchical assembly method of a polynucleotide having a predefined sequence.

[0031] Fig. 5 illustrates the nucleotide sequence of plasmid pG9-1 with restriction endonuclease recognition sites (underlined).

[0032] Fig. 6 illustrates non-limiting exemplary method of sequence verification.

DETAILED DESCRIPTION OF THE INVENTION

[0033] Aspects of the invention may be useful for optimizing nucleic acid assembly reactions and to reduce the number of incorrectly assembled nucleic acids. The methods and composition of the invention can facilitate the process of obtaining a target sequence having a predefined sequence. Accordingly, the methods and composition of the invention may increase the probability of obtaining a correctly assembled nucleic acid and thereby reduce the cost and time associated with the production of a nucleic acid having a predetermined sequence.

[0034] Aspects of the invention may be used to improve the yield of one or more initial or intermediate assembly reactions. In some embodiments, the methods and compositions of the invention can improve the efficiency of the overall assembly procedure by avoiding the requirement to separate a number of assembly steps, such as for example, enzymatic digestion, purification and ligation steps. Accordingly, some aspects of the invention allows for predictable and/or reliable assembly strategies and can significantly decrease the time and steps needed for gene synthesis and increase the yield and/or accuracy of intermediate product or final nucleic acid products.

[0035] In some aspects of the invention, the assembly process comprises designing and implementing nucleic acid assembly strategies that can accommodate sequence features known or predicted to interfere with one or more assembly steps. For example, the nucleic acid sequence to be synthesized can be analyzed for sequence features, such as repeated sequences, sequences having a significantly high or low GC content, and/or other sequences associated with secondary structures, that can interfere with one or more assembly steps. One of skill in the art will understand that certain sequence features may interfere with multiplex assembly reactions (e.g. polymerase-based extension reactions) and/or promote the formation of unwanted assembly products thereby reducing or preventing the assembly of correct nucleic acid products. In some embodiments, if a plurality of interfering sequence features is identified in a target nucleic acid sequence, a useful strategy may involve separating the interfering sequence features during assembly. For example, a target nucleic acid may be assembled in a process involving a plurality of intermediate fragments or building blocks that are designed to contain only a small number of interfering sequences (e.g., 0, 1, 2, or 3). In some embodiments, each intermediate fragment or building block may contain at most one interfering sequence feature. Accordingly, each intermediate fragment may be assembled efficiently. In some embodiments, the design of the

nucleic acids fragments or building blocks may exclude interfering sequence features from their 5' and/or 3' ends. Accordingly, the interfering sequence features may be excluded from complementary overlapping regions between adjacent starting nucleic acids that are designed for use assembly reaction. This may prevent or reduce interference with sequence-specific hybridization reactions that are important for correct assembly of the nucleic acids. In some embodiments, it may be sufficient to exclude an interfering sequence feature from the immediate 3' and/or 5' end of a building block. For example, an interfering sequence feature may be located at least one nucleotide in from a 3' end and/or 5' end, and preferably 2, 3, 4, 5, or more nucleotides (e.g., 5-10, 10-15, 15-20, or more nucleotides) in from a 3' end and/or 5' end of a building block.

[0036] Aspects of the invention may be used in conjunction with in vitro and/or in vivo nucleic acid assembly procedures.

[0037] Aspects of the methods and compositions provided herein are useful for increasing the accuracy, yield, throughput, and/or cost efficiency of nucleic acid synthesis and assembly reactions. As used herein the terms “nucleic acid”, “polynucleotide”, “oligonucleotide” are used interchangeably and refer to naturally-occurring or synthetic polymeric forms of nucleotides. The oligonucleotides and nucleic acid molecules of the present invention may be formed from naturally occurring nucleotides, for example forming deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) molecules. Alternatively, the naturally occurring oligonucleotides may include structural modifications to alter their properties, such as in peptide nucleic acids (PNA) or in locked nucleic acids (LNA). The solid phase synthesis of oligonucleotides and nucleic acid molecules with naturally occurring or artificial bases is well known in the art. The terms should be understood to include equivalents, analogs of either RNA or DNA made from nucleotide analogs and as applicable to the embodiment being described, single-stranded or double-stranded polynucleotides. Nucleotides useful in the invention include, for example, naturally-occurring nucleotides (for example, ribonucleotides or deoxyribonucleotides), or natural or synthetic modifications of nucleotides, or artificial bases. As used herein, the term monomer refers to a member of a set of small molecules which are and can be joined together to from an oligomer, a polymer or a compound composed of two or more members. The particular ordering of monomers within a polymer is referred to herein as the "sequence" of the polymer. The set of monomers includes but is not limited to example, the set

of common L-amino acids, the set of D-amino acids, the set of synthetic and/or natural amino acids, the set of nucleotides and the set of pentoses and hexoses. Aspects of the invention described herein primarily with regard to the preparation of oligonucleotides, but could readily be applied in the preparation of other polymers such as peptides or polypeptides, polysaccharides, phospholipids, heteropolymers, polyesters, polycarbonates, polyureas, polyamides, polyethyleneimines, polyarylene sulfides, polysiloxanes, polyimides, polyacetates, or any other polymers.

Target nucleic acids

[0038] As used herein, the term "predetermined sequence" means that the sequence of the polymer is known and chosen before synthesis or assembly of the polymer. In particular, aspects of the invention is described herein primarily with regard to the preparation of nucleic acids molecules, the sequence of the oligonucleotide or polynucleotide being known and chosen before the synthesis or assembly of the nucleic acid molecules. In some embodiments of the technology provided herein, immobilized oligonucleotides or polynucleotides are used as a source of material. In various embodiments, the methods described herein use pluralities of oligonucleotides, each sequence being determined based on the sequence of the final polynucleotides constructs to be synthesized. In one embodiment, oligonucleotides are short nucleic acid molecules. For example, oligonucleotides may be from 10 to about 300 nucleotides, from 20 to about 400 nucleotides, from 30 to about 500 nucleotides, from 40 to about 600 nucleotides, or more than about 600 nucleotides long. However, shorter or longer oligonucleotides may be used. Oligonucleotides may be designed to have different length. In some embodiments, the sequence of the polynucleotide construct may be divided up into a plurality of shorter sequences that can be synthesized in parallel and assembled into a single or a plurality of desired polynucleotide constructs using the methods described herein.

[0039] In some embodiments, a target nucleic acid may have a sequence of a naturally occurring gene and/or other naturally occurring nucleic acid (e.g., a naturally occurring coding sequence, regulatory sequence, non-coding sequence, chromosomal structural sequence such as a telomere or centromere sequence, etc., any fragment thereof or any combination of two or more thereof) or a sequence that is not naturally-occurring. In some embodiments, a target nucleic acid may be designed to have a sequence that differs from a natural sequence at one or more positions. In other embodiments, a target nucleic acid may be designed to have an entirely novel

sequence. However, it should be appreciated that target nucleic acids may include one or more naturally occurring sequences, non-naturally occurring sequences, or combinations thereof.

[0040] In some embodiments, methods of assembling libraries containing nucleic acids having predetermined sequence variations are provided herein. Assembly strategies provided herein can be used to generate very large libraries representative of many different nucleic acid sequences of interest. For example, the methods provided herein can be used to assemble libraries having more than 10 different sequence variants. In some embodiments, libraries of nucleic acid are libraries of sequence variants. Sequence variants may be variants of a single naturally-occurring protein encoding sequence. However, in some embodiments, sequence variants may be variants of a plurality of different protein-encoding sequences. Accordingly, one aspect of the invention provided herein relates to the design of assembly strategies for preparing precise high-density nucleic acid libraries. Another aspect of the technology provided herein relates to assembling precise high-density nucleic acid libraries. Aspects of the technology provided herein also provide precise high-density nucleic acid libraries. A high-density nucleic acid library may include more that 100 different sequence variants (e.g., about 10^2 to 10^3; about 10^3 to 10^4; about 10^4 to 10^5; about 10^5 to 10^6; about 10^6 to 10^7; about 10^7 to 10^8; about 10^8 to 10^9; about 10^9 to 10^{10}; about 10^{10} to 10^{11}; about 10^{11} to 10^{12}; about 10^{12} to 10^{13}; about 10^{13} to 10^{14}; about 10^{14} to 10^{15}; or more different sequences) wherein a high percentage of the different sequences are specified sequences as opposed to random sequences (e.g., more than about 50%, more than about 60%, more than about 70%, more than about 75%, more than about 80%, more than about 85%, more than about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%, about 98%, about 99%, or more of the sequences are predetermined sequences of interest).

[0041] In certain embodiments, a target nucleic acid may include a functional sequence (e.g., a protein binding sequence, a regulatory sequence, a sequence encoding a functional protein, etc., or any combination thereof). However, in some embodiments the target nucleic acid may lack a specific functional sequence (e.g., a target nucleic acid may include only non-functional fragments or variants of a protein binding sequence, regulatory sequence, or protein encoding sequence, or any other non-functional naturally-occurring or synthetic sequence, or any non-functional combination thereof). Certain target nucleic acids may include both functional

and non-functional sequences. These and other aspects of target nucleic acids and their uses are described in more detail herein.

[0042] A target nucleic acid may, in some embodiments, be assembled in a single multiplex assembly reaction (e.g., a single oligonucleotide assembly reaction). However, a target nucleic acid may also be assembled from a plurality of nucleic acid fragments, each of which may have been generated in a separate multiplex oligonucleotide assembly reactions. It should be appreciated that one or more nucleic acid fragments generated via multiplex oligonucleotide assembly may, in some embodiments, be combined with one or more nucleic acid molecules obtained from another source (e.g., a restriction fragment, a nucleic acid amplification product, etc.) to form a target nucleic acid. In some embodiments, a target nucleic acid that is assembled in a first reaction may be used as an input nucleic acid fragment for a subsequent assembly reaction to produce a larger target nucleic acid. The terms "multiplex assembly" and "multiplex oligonucleotide assembly reaction" used herein generally refer to assembly reactions involving a plurality of starting nucleic acids (e.g., a plurality of at least partially overlapping nucleic acids) that are assembled to produce a larger final nucleic acid.

Assembly process

[0043] FIG. 1 illustrates a process for assembling a nucleic acid in accordance with one embodiment of the invention. Initially, sequence information is obtained. The sequence information may be the sequence of a predetermined target nucleic acid that is to be assembled. In some embodiments, the sequence may be received in the form of an order from a customer. In some embodiments, the sequence may be received as a nucleic acid sequence (e.g., DNA or RNA). In some embodiments, the sequence may be received as a protein sequence. The sequence may be converted into a DNA sequence. For example, if the sequence obtained is an RNA sequence, the Us may be replaced with Ts to obtain the corresponding DNA sequence. If the sequence obtained is a protein sequence, the protein sequence may be converted into a DNA sequence using appropriate codons for the amino acids.

[0044] In some embodiments, the sequence information may be analyzed to determine an assembly strategy, such as the number and the sequences of the fragments (also referred herein as building blocks, oligonucleotides or intermediate fragments) to be assembled to generate the predefined sequence of the target nucleic acid. In some embodiments, the sequence analysis may involve scanning for the presence of one or more interfering sequence features that are known or

predicted to interfere with oligonucleotide synthesis, amplification or assembly. For example, an interfering sequence structure may be a sequence that has a low GC content (e.g., less than 30% GC, less than 20% GC, less than 10% GC, etc.) over a length of at least 10 bases (e.g., 10-20, 20-50, 50-100, or more than 100 bases), or a sequence that may be forming secondary structures or stem-loop structures. Once passing this filter, the nucleic acid sequence can be divided into smaller pieces, such as oligonucleotide building blocks.

[0045] In some embodiments, after the construct qualification and parsing step, synthetic oligonucleotides for the assembly may be designed (e.g. sequence, size, and number). Synthetic oligonucleotides can be generated using standard DNA synthesis chemistry (e.g. phosphoramidite method). Synthetic oligonucleotides may be synthesized on a solid support, such as for example a microarray, using any appropriate technique known in the art or as described in more detail herein. Oligonucleotides can be eluted from the microarray prior to be subjected to amplification or can be amplified on the microarray. It should be appreciated that different oligonucleotides may be designed to have different lengths.

[0046] In some embodiments, the building blocks oligonucleotides for each target sequence can be amplified. For example, the oligonucleotides can be designed such as having at their 3'end and 5' end a primer binding sequence and the oligonucleotides can be amplified by polymerase chain reaction (PCR) using the appropriate primers pair(s).

[0047] It should be appreciated that synthetic oligonucleotides may have sequence errors. Accordingly, oligonucleotide preparations may be selected or screened to remove error-containing molecules as described in more detail herein. Error containing-oligonucleotides may be double-stranded homoduplexes having the error on both strands (i.e., incorrect complementary nucleotide(s), deletion(s), or addition(s) on both strands). In some embodiments, sequence errors may be removed using a technique that involves denaturing and reannealing the double-stranded nucleic acids. In some embodiments, single strands of nucleic acids that contain complementary errors may be unlikely to reanneal together if nucleic acids containing each individual error are present in the nucleic acid preparation at a lower frequency than nucleic acids having the correct sequence at the same position. Rather, error containing single strands may reanneal with a complementary strand that contains no errors or that contains one or more different errors. As a result, error-containing strands may end up in the form of heteroduplex molecules in the re-annealed reaction product. Nucleic acid strands that are error-free may

reanneal with error-containing strands or with other error-free strands. Reannealed error-free strands form homoduplexes in the reannealed sample. Accordingly, by removing heteroduplex molecules from the re-annealed preparation of oligonucleotides, the amount or frequency of error containing nucleic acids may be reduced. Any suitable method known in the art for removing heteroduplex molecules may be used, including chromatography, electrophoresis, selective binding of heteroduplex molecules, etc. In some embodiments, mismatch binding proteins that selectively (e.g., specifically) bind to heteroduplex nucleic acid molecules may be used. In some embodiments, the mismatch binding protein may be used on double-stranded oligonucleotides or polynucleotides in solution or immobilized onto a support.

[0048] In some embodiments, the oligonucleotides containing errors are removed using a MutS filtration process, for example, using MutS, a MutS homolog, or a combination thereof. In E. coli, the MutS protein, which appears to function as a homodimer, serves as a mismatch recognition factor. In eukaryotes, at least three MutS Homolog (MSH) proteins have been identified; namely, MSH2, MSH3, and MSH6, and they form heterodimers. For example in the yeast, Saccharomyces cerevisiae, the MSH2-MSH6 complex (also known as MutS alpha) recognizes base mismatches and single nucleotide insertion/deletion loops, while the MSH2-MSH3 complex (also known as MutSbeta) recognizes insertions/deletions of up to 12-16 nucleotides, although they exert substantially redundant functions. A mismatch binding protein may be obtained from recombinant or natural sources. A mismatch binding protein may be heat-stable. In some embodiments, a thermostable mismatch binding protein from a thermophilic organism may be used. Examples of thermostable DNA mismatch binding proteins include, but are not limited to: Tth MutS (from Thermus thermophilus), Taq MutS (from Thermus aquaticus), Apy MutS (from Aquifex pyrophilus), Tma MutS (from Thermotoga maritima), homologs thereof any other suitable MutS or any combination of two or more thereof.

[0049] It has been shown that MutS obtained from different species can have different affinity for a specific mismatch or for different mismatch. In some embodiments, a combination of different MutS having different affinities for different mismatch can be used.

[0050] In some embodiments, an enzyme complex using one or more repair proteins can be used. Examples of repair proteins include, but are not limited to, MutS, for mismatch recognition, MutH, for introduction of a nick in the target strand, and MutL, for mediating the

interactions between MutH and MutS, homologs thereof or any combinations thereof. In some embodiments, the mismatch binding protein complex is a MutHLS enzyme complex.

[0051] In some embodiments, a sliding clamp technique may be used for enriching error-free double stranded oligonucleotides. In some embodiments, MutS or homolog thereof can interact with a DNA clamp protein. Examples of DNA clamp proteins include, but are not limited to, the bacterial sliding clamp protein DnaN, encoded by dnaN gene, which can function as a homodimer. In some embodiments, interaction of MutS protein, or homolog thereof, with a clamp protein can increase the effectiveness of MutS in binding mismatches.

[0052] In some embodiments, the oligonucleotides containing errors can be removed using an enzyme from the S1 family of proteins, for example CELI, CELII or a homolog thereof, such as RESI, or a combination thereof. Enzymes from the S1 family of proteins can recognize base mismatches, insertion and deletion loops. In some embodiments, such enzymes can bind preferentially to Holliday junctions after which the recognition site is cleaved, either through only one or both DNA strands. In some embodiments, a thermostable equivalent of a S1 protein may be used.

[0053] In some embodiments, the oligonucleotides containing errors can be removed using a small molecule, chemical or inorganic material that binds to mismatched base sites. At the mismatched site, nucleotide bases are extra-helical and can be susceptible to chemical modification reactions. Materials such permanganate, hydroxylamine, lysine, and or pentaamine ruthenium can be employed in the chemical cleavage method to modify the mismatched thymine and cytosine respectively. The resulting modified DNA can then treated with piperidine to cause a cleavage at the abasic sites. In some embodiments, specificity of cleavage can be monitored using divalent salt.

[0054] In some embodiments, in a next step, the error-corrected oligonucleotides are combined through the sequential removal of common sequences and subsequent ligation into longer, multi-oligonucleotide constructs.

[0055] In some aspects of the invention, the enzymatic digestion common sequence removal step is combined with a ligation step. One of skill in the art will appreciate that the process of the invention allows for a concurrent removal of common sequences and ligation into the target nucleic acid constructs and negate the need of enzymatic removal, bead-based capture and ligation sequential steps. In addition, one of skill in the art will appreciate that the process of

the invention may present a number of advantages over the standard gene assembly process such as:

(1) Increase of the yield efficiency. Using the standard separate enzymatic removal of common sequences, the reaction is stopped after a set time point, with unreacted substrates or undigested oligonucleotides, still present which are the subject of further removal. One of skill in the art will understand that because the ligation reaction creates a desired product which is not a substrate for the enzymatic removal, the combination of the removal and ligation steps has the effect of driving the reaction toward the desired product irreversibly.

(2) Cost efficiency: The methods according to some aspects of the invention are cost efficient since there is no longer a need for the purification steps between the removal of common sequences and the ligation. Because of the elimination of purification steps, aspects of the present method also eliminate the need for biotin-labeled primers. There may be also an associated savings in the form of the reduced lead time for receipt of non-biotinylated primers over their biotin-containing counterparts.

(3) Time efficiency: The time and the number of steps needed for gene synthesis are reduced by removing the purification steps between enzymatic common sequence removal and ligation.

(4) Opportunities to add other sequences easily, without regard for their sizes. Because part of the purification step to remove undesired sequences is based on size, eliminating the purification can remove the size constraint for any additional sequences to be added for the gene synthesis. This can include a one-step ligation into a vector, or addition of common flanking sequences.

(5) The process allows for use of restriction sites in the gene which are used in the gene synthesis process itself. In previous methodologies, these restriction sites could not be used because cut sites would result in small DNA pieces which would be removed in the purification step. Enabling the usage of these restriction sites can allow for recursive (hierarchical) gene synthesis to build longer nucleic acids.

[0056] One of skill in the art would appreciate that after oligonucleotide assembly, the assembly products (e.g. final target nucleic acid or intermediate nucleic acid fragment) may contain sequences containing undesired sequences. The errors may result from sequences errors introduced during oligonucleotide synthesis or during the assembly of oligonucleotides into

longer nucleic acids. In some embodiments, nucleic acids having the correct predefined sequence can be isolated from other nucleic acids sequences (also referred herein as preparative in vitro cloning). In some embodiments, the correct sequence may be isolated by selectively isolating the correct sequence from the other incorrect sequences. For example, nucleic acids having correct sequence can be selectively moved or transferred to a different feature of the support, or to another plate. Alternatively, nucleic acids having an incorrect sequence can be selectively removed from the feature comprising the nucleic acids of interest (see for example, PCT/US2007/011886, which is incorporated by reference herein in its entirety).

[0057] In some embodiments, after oligonucleotide processing and ligation, the assembly constructs or a copy of the assembled constructs can be isolated by clonal isolation. The assembly constructs can be sequence verified using, for example, high throughput sequencing. In some embodiments, sequence determination of the target nucleic acid sequences can be performed using sequencing of individual molecules, such as single molecule sequencing, or sequencing of an amplified population of target nucleic acid sequences, such as polony sequencing. Any suitable methods for sequencing, such as sequencing by hybridization, sequencing by ligation or sequencing by synthesis may be used.

[0058] Some aspects of the invention relate to a gene synthesis platform using methods described herein. In some embodiments, the gene synthesis platform can be combined with a next generation sequencing platform (e.g. sequencing by hybridization, sequencing by synthesis, sequencing by ligation or any other suitable sequencing method).

[0059] In some embodiments, the assembly procedure may include several parallel and/or sequential reaction steps in which a plurality of different nucleic acids or oligonucleotides are synthesized or immobilized, amplified, and are combined in order to be assembled (e.g., by extension or by ligation as described herein) to generate a longer nucleic acid product to be used for further assembly, cloning, or other applications (see PCT application PCT/US09/55267 which is incorporate herein by reference in its entirety).

Oligonucleotides Synthesis

[0060] In some embodiments, the methods and apparatus provided herein use oligonucleotides that are immobilized on a surface or substrate (e.g., support-bound oligonucleotides). As used herein the term "support" and "substrate" are used interchangeably and refers to a porous or non-porous solvent insoluble material on which polymers such as

nucleic acids are synthesized or immobilized. As used herein "porous" means that the material contains pores having substantially uniform diameters (for example in the nm range). Porous materials include paper, synthetic filters etc. In such porous materials, the reaction may take place within the pores. The support can have any one of a number of shapes, such as pin, strip, plate, disk, rod, bends, cylindrical structure, particle, including bead, nanoparticles and the like. The support can have variable widths. The support can be hydrophilic or capable of being rendered hydrophilic and includes inorganic powders such as silica, magnesium sulfate, and alumina; natural polymeric materials, particularly cellulosic materials and materials derived from cellulose, such as fiber containing papers, e.g., filter paper, chromatographic paper, etc.; synthetic or modified naturally occurring polymers, such as nitrocellulose, cellulose acetate, poly (vinyl chloride), polyacrylamide, cross linked dextran, agarose, polyacrylate, polyethylene, polypropylene, poly (4-methylbutene), polystyrene, polymethacrylate, poly(ethylene terephthalate), nylon, poly(vinyl butyrate), polyvinylidene difluoride (PVDF) membrane, glass, controlled pore glass, magnetic controlled pore glass, ceramics, metals, and the like etc.; either used by themselves or in conjunction with other materials. In some embodiments, oligonucleotides are synthesized on an array format. For example, single-stranded oligonucleotides are synthesized in situ on a common support wherein each oligonucleotide is synthesized on a separate or discrete feature (or spot) on the substrate. In preferred embodiments, single stranded oligonucleotides are bound to the surface of the support or feature. As used herein the term "array" refers to an arrangement of discrete features for storing, routing, amplifying and releasing oligonucleotides or complementary oligonucleotides for further reactions. In a preferred embodiment, the support or array is addressable: the support includes two or more discrete addressable features at a particular predetermined location (i.e., an "address") on the support. Therefore, each oligonucleotide molecule of the array is localized to a known and defined location on the support. The sequence of each oligonucleotide can be determined from its position on the support.

[0061] In some embodiments, oligonucleotides are attached, spotted, immobilized, surface-bound, supported or synthesized on the discrete features of the surface or array. Oligonucleotides may be covalently attached to the surface or deposited on the surface. Arrays may be constructed, custom ordered or purchased from a commercial vendor (e.g., Agilent, Affymetrix, Nimblegen). Various methods of construction are well known in the art e.g.,

maskless array synthesizers, light directed methods utilizing masks, flow channel methods, spotting methods etc. In some embodiments, construction and/or selection oligonucleotides may be synthesized on a solid support using maskless array synthesizer (MAS). Maskless array synthesizers are described, for example, in PCT application No. WO 99/42813 and in corresponding U.S. Pat. No. 6,375,903. Other examples are known of maskless instruments which can fabricate a custom DNA microarray in which each of the features in the array has a single-stranded DNA molecule of desired sequence. Other methods for synthesizing construction and/or selection oligonucleotides include, for example, light-directed methods utilizing masks, flow channel methods, spotting methods, pin-based methods, and methods utilizing multiple supports. Light directed methods utilizing masks (e.g., VLSIPS™ methods) for the synthesis of oligonucleotides is described, for example, in U.S. Pat. Nos. 5,143,854, 5,510,270 and 5,527,681. These methods involve activating predefined regions of a solid support and then contacting the support with a preselected monomer solution. Selected regions can be activated by irradiation with a light source through a mask much in the manner of photolithography techniques used in integrated circuit fabrication. Other regions of the support remain inactive because illumination is blocked by the mask and they remain chemically protected. Thus, a light pattern defines which regions of the support react with a given monomer. By repeatedly activating different sets of predefined regions and contacting different monomer solutions with the support, a diverse array of polymers is produced on the support. Other steps, such as washing unreacted monomer solution from the support, can be optionally used. Other applicable methods include mechanical techniques such as those described in U.S. Pat. No. 5,384,261. Additional methods applicable to synthesis of construction and/or selection oligonucleotides on a single support are described, for example, in U.S. Pat. No. 5,384,261. For example, reagents may be delivered to the support by either (1) flowing within a channel defined on predefined regions or (2) "spotting" on predefined regions. Other approaches, as well as combinations of spotting and flowing, may be employed as well. In each instance, certain activated regions of the support are mechanically separated from other regions when the monomer solutions are delivered to the various reaction sites. Flow channel methods involve, for example, microfluidic systems to control synthesis of oligonucleotides on a solid support. For example, diverse polymer sequences may be synthesized at selected regions of a solid support by forming flow channels on a surface of the support through which appropriate reagents

flow or in which appropriate reagents are placed. Spotting methods for preparation of oligonucleotides on a solid support involve delivering reactants in relatively small quantities by directly depositing them in selected regions. In some steps, the entire support surface can be sprayed or otherwise coated with a solution, if it is more efficient to do so. Precisely measured aliquots of monomer solutions may be deposited dropwise by a dispenser that moves from region to region. Pin-based methods for synthesis of oligonucleotides on a solid support are described, for example, in U.S. Pat. No. 5,288,514. Pin-based methods utilize a support having a plurality of pins or other extensions. The pins are each inserted simultaneously into individual reagent containers in a tray. An array of 96 pins is commonly utilized with a 96-container tray, such as a 96-well microtiter dish. Each tray is filled with a particular reagent for coupling in a particular chemical reaction on an individual pin. Accordingly, the trays will often contain different reagents. Since the chemical reactions have been optimized such that each of the reactions can be performed under a relatively similar set of reaction conditions, it becomes possible to conduct multiple chemical coupling steps simultaneously.

[0062] In another embodiment, a plurality of oligonucleotides may be synthesized on multiple supports. One example is a bead based synthesis method which is described, for example, in U.S. Pat. Nos. 5,770,358; 5,639,603; and 5,541,061. For the synthesis of molecules such as oligonucleotides on beads, a large plurality of beads is suspended in a suitable carrier (such as water) in a container. The beads are provided with optional spacer molecules having an active site to which is complexed, optionally, a protecting group. At each step of the synthesis, the beads are divided for coupling into a plurality of containers. After the nascent oligonucleotide chains are deprotected, a different monomer solution is added to each container, so that on all beads in a given container, the same nucleotide addition reaction occurs. The beads are then washed of excess reagents, pooled in a single container, mixed and re-distributed into another plurality of containers in preparation for the next round of synthesis. It should be noted that by virtue of the large number of beads utilized at the outset, there will similarly be a large number of beads randomly dispersed in the container, each having a unique oligonucleotide sequence synthesized on a surface thereof after numerous rounds of randomized addition of bases. An individual bead may be tagged with a sequence which is unique to the double-stranded oligonucleotide thereon, to allow for identification during use.

[0063] Pre-synthesized oligonucleotide and/or polynucleotide sequences may be attached to a support or synthesized in situ using light-directed methods, flow channel and spotting methods, inkjet methods, pin-based methods and bead-based methods set forth in the following references: McGall et al. (1996) Proc. Natl. Acad. Sci. U.S.A. 93:13555; Synthetic DNA Arrays In Genetic Engineering, Vol. 20:111, Plenum Press (1998); Duggan et al. (1999) Nat. Genet. S21:10; Microarrays: Making Them and Using Them In Microarray Bioinformatics, Cambridge University Press, 2003; U.S. Patent Application Publication Nos. 2003/0068633 and 2002/0081582; U.S. Pat. Nos. 6,833,450, 6,830,890, 6,824,866, 6,800,439, 6,375,903 and 5,700,637; and PCT Publication Nos. WO 04/031399, WO 04/031351, WO 04/029586, WO 03/100012, WO 03/066212, WO 03/065038, WO 03/064699, WO 03/064027, WO 03/064026, WO 03/046223, WO 03/040410 and WO 02/24597; the disclosures of which are incorporated herein by reference in their entirety for all purposes. In some embodiments, pre-synthesized oligonucleotides are attached to a support or are synthesized using a spotting methodology wherein monomers solutions are deposited dropwise by a dispenser that moves from region to region (e.g., ink jet). In some embodiments, oligonucleotides are spotted on a support using, for example, a mechanical wave actuated dispenser.

Amplification

[0064] In some embodiments, oligonucleotides may be amplified using an appropriate primer pair with one primer corresponding to each end of the oligonucleotide (e.g., one that is complementary to the 3' end of the oligonucleotide and one that is identical to the 5' end of the oligonucleotide). In some embodiments, an oligonucleotide may be designed to contain a central or internal assembly sequence (corresponding to a target sequence, designed to be incorporated into the final product) flanked by a 5' amplification sequence (e.g., a 5' universal sequence or 5' common amplification sequence) and a 3' amplification sequence (e.g., a 3' universal sequence or 5' common amplification sequence).

[0065] In some embodiments, a synthetic oligonucleotide may include a central assembly sequence flanked by 5' and 3' amplification sequences. The central assembly sequence is designed for incorporation into an assembled nucleic acid. The flanking sequences are designed for amplification and are not intended to be incorporated into the assembled nucleic acid. The flanking amplification sequences may be used as primer sequences to amplify a plurality of different assembly oligonucleotides that share the same amplification sequences but have

different central assembly sequences. In some embodiments, the flanking sequences are removed after amplification to produce an oligonucleotide that contains only the assembly sequence.

[0066] Amplification primers (e.g., between 10 and 50 nucleotides long, between 15 and 45 nucleotides long, about 25 nucleotides long, etc.) corresponding to the flanking amplification sequences may be used to amplify the oligonucleotides (e.g., one primer may be complementary to the 3' amplification sequence and one primer may have the same sequence as the 5' amplification sequence). In some embodiments, a plurality of different oligonucleotides (e.g., about 5, 10, 50, 100, or more) with different central assembly sequences may have identical 5' amplification sequences and identical 3' amplification sequences. These oligonucleotides can all be amplified in the same reaction using the same amplification primers. The amplification sequences may then be removed from the amplified oligonucleotides using any suitable technique to produce oligonucleotides that contain only the assembly sequences. In some embodiments, the amplification sequences are removed by a restriction enzyme as described in more details herein.

[0067] In some embodiments, the oligonucleotides may be amplified while still attached to the support. In some embodiments, the oligonucleotides may be removed or cleaved from the support prior to amplification.

[0068] In some embodiments, the method includes synthesizing a plurality of oligonucleotides or polynucleotides in a chain extension reaction using a first plurality of single stranded oligonucleotides as templates. As noted above, the oligonucleotides may be first synthesized onto a plurality of discrete features of the surface, or may be deposited on the plurality of features of the support. In some embodiments, the oligonucleotides are covalently attached to the support. In some embodiments, the first plurality of oligonucleotides is immobilized to a solid surface. In some embodiments, each feature of the solid surface comprises a high density of oligonucleotides having a different predetermined sequence (e.g., approximately 10^6- 10^8 molecules per feature). The support may comprise at least 100, at least 1,000, at least 10^4, at least 10^5, at least 10^6, at least 10^7, at least 10^8 features. In some embodiments, after amplification, the double-stranded oligonucleotides may be eluted in solution and/or subjected to error reduction and/or assembly to form longer nucleic acid constructs.

Error Reduction

[0069] In some embodiments, each fragment is assembled and fidelity optimized to remove error containing nucleic acids (e.g., using one or more post-assembly fidelity optimization techniques described herein) before being processed to generated cohesive ends. A sequence error may include one or more nucleotide deletions, additions, substitutions (e.g., transversion or transition), inversions, duplications, or any combination of two or more thereof. Oligonucleotide errors may be generated during oligonucleotide synthesis. Different synthetic techniques may be prone to different error profiles and frequencies. In some embodiments, error rates may vary from 1/10 to 1/200 errors per base depending on the synthesis protocol that is used. However, in some embodiments, lower error rates may be achieved. Also, the types of errors may depend on the synthetic techniques that are used. For example, microarray-based oligonucleotide synthesis may result in relatively more deletions than column-based synthetic techniques.

[0070] Some aspects of the invention relate to a polynucleotide assembly process wherein synthetic oligonucleotides are designed and used to assemble polynucleotides into longer polynucleotides constructs. During enzymatic amplification or chain extension reactions, the error in sequence is faithfully replicated. As a result, polynucleotides population synthesized by this method contains both error-free and error-prone sequences. In some embodiments, since synthetic oligonucleotides can contain incorrect sequences due to errors introduced during oligonucleotide synthesis, it can be useful to remove polynucleotide that have incorporated one or more error-containing oligonucleotides during assembly or extension. In some embodiments, one or more assembled polynucleotides may be sequenced to determine whether they contain the predetermined sequence or not. This procedure allows fragments with the correct sequence to be identified. In other embodiments, other techniques may be used to remove error containing nucleic acid fragments. Such nucleic acid fragments can be nascently synthesized oligonucleotides or assembled nucleic acid polymers. It should be appreciated that error containing-nucleic acids can be double-stranded homoduplexes having the error on both strands (i.e., incorrect complementary nucleotide(s), deletion(s), or addition(s) on both strands), because the assembly procedure may involve one or more rounds of polymerase extension (e.g., during assembly or after assembly to amplify the assembled product). During polymerase extension, the input nucleic acid containing an error may serve as a template thereby producing a complementary strand comprising the complementary error. In certain embodiments, a

preparation of double-stranded nucleic acid fragments or duplexes may be suspected to contain a mixture of nucleic acids having the correct predefined sequence as well as nucleic acids containing one or more sequence errors incorporated during assembly. The term "duplex" refers to a nucleic acid molecule that is at least partially double-stranded. A "stable duplex" refers to a duplex that is relatively more likely to remain hybridized to a complementary sequence under a given set of hybridization conditions. In an exemplary embodiment, a stable duplex refers to a duplex that does not contain a basepair mismatch, insertion, or deletion. An "unstable duplex" refers to a duplex that is relatively less likely to remain hybridized to a complementary sequence under a given set of hybridization conditions such as stringent melt. In an exemplary embodiment, an unstable duplex refers to a duplex that contains at least one base-pair mismatch, insertion, or deletion. As used herein the term "stringency" is used in reference to the conditions of temperature, ionic strength, and the presence of other compounds such as organic solvents, under which nucleic acid hybridizations are conducted. Hybridization stringency increases with temperature and/or the solution chemical properties such as the amounts of salts and/or formamide in the hybridization solution during a hybridization process. With "high stringency" conditions, nucleic acid base pairing will occur only between nucleic acid fragments that have a high frequency of complementary base sequences. Stringent conditions may be selected to be about 5°C lower than the thermal melting point (Tm) for a given polynucleotide duplex at a defined ionic strength and pH. The length of the complementary polynucleotide strands and the GC content determine the Tm of the duplex, and thus the hybridization conditions necessary for obtaining a desired specificity of hybridization. The Tm is the temperature (under defined ionic strength and pH) at which 50% of a polynucleotide sequence hybridizes to a perfectly matched complementary strand. In certain cases it may be desirable to increase the stringency of the hybridization conditions to be about equal to the Tm for a particular duplex. Appropriate stringency conditions are known to those skilled in the art or may be determined experimentally by the skilled artisan. See, for example, Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-12.3.6; Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y; S. Agrawal (ed.) Methods in Molecular Biology, volume 20; Tijssen (1993) Laboratory Techniques in biochemistry and molecular biology-hybridization with nucleic acid probes, e.g., part I chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York.

[0071] In some embodiments, sequence errors may be removed using a technique that involves denaturing and reannealing the double-stranded nucleic acids. In some embodiments, single strands of nucleic acids that contain complementary errors may be unlikely to reanneal together if nucleic acids containing each individual error are present in the nucleic acid preparation at a lower frequency than nucleic acids having the correct sequence at the same position. Rather, error containing single strands can reanneal with error-free complementary strand or complementary strands containing one or more different errors or error at different location. As a result, error-containing strands can end up in the form of heteroduplex molecules in the reannealed reaction product. Nucleic acid strands that are error-free may reanneal with error-containing strands or with other error-free strands. Reannealed error-free strands form homoduplexes in the reannealed sample. Accordingly, by removing heteroduplex molecules from the reannealed preparation of nucleic acid fragments, the amount or frequency of error containing nucleic acids can be reduced.

[0072] Heteroduplex formation thus takes place through a process that can be understood as shuffling, by which nucleic acid strands from different populations can be hybridized with one another so that perfect match and mismatch-containing duplexes can be formed. Suitable method for removing heteroduplex molecules include chromatography, electrophoresis, selective binding of heteroduplex molecules that binds preferentially to double stranded DNA having a sequence mismatch between the two strands. The term "mismatch" or "base pair mismatch" indicates a base pair combination that generally does not form in nucleic acids according to Watson and Crick base pairing rules. For example, when dealing with the bases commonly found in DNA, namely adenine, guanine, cytosine and thymidine, base pair mismatches are those base combinations other than the A-T and G-C pairs normally found in DNA. As described herein, a mismatch may be indicated, for example as C/C meaning that a cytosine residue is found opposite another cytosine, as opposed to the proper pairing partner, guanine.

[0073] In some embodiments, oligonucleotide preparations may be selected or screened to remove error-containing molecules as described in more detail herein. In some embodiments, oligonucleotides can be error-corrected using a mismatch-binding agent as described herein.

[0074] In one aspect, the invention relates to a method for producing high fidelity polynucleotides on a solid support. The synthetic polynucleotides are at least about 1, 2, 3, 4, 5, 8, 10, 15, 20, 25, 30, 40, 50, 75, or 100 kilobases (kb), or 1 megabase (mb), or longer. In

exemplary embodiments, a compositions of synthetic polynucleotides contains at least about 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 50%, 60%, 70%, 80%, 90%, 95 % or more, copies that are error free (e.g., having a sequence that does not deviate from a predetermined sequence). The percent of error free copies is based on the number of error free copies in the compositions as compared to the total number of copies of the polynucleotide in the composition that were intended to have the correct, e.g., predefined or predetermined, sequence.

[0075] Some aspects the invention relate to the design of oligonucleotides for high fidelity polynucleotide assembly. Aspects of the invention may be useful to increase the throughput rate of a nucleic acid assembly procedure and/or reduce the number of steps or amounts of reagent used to generate a correctly assembled nucleic acid. In certain embodiments, aspects of the invention may be useful in the context of automated nucleic acid assembly to reduce the time, number of steps, amount of reagents, and other factors required for the assembly of each correct nucleic acid. Accordingly, these and other aspects of the invention may be useful to reduce the cost and time of one or more nucleic acid assembly procedures.

Single-stranded Overhangs

[0076] In some aspects of the invention, nucleic acid fragments being assembled are designed to have overlapping complementary sequences. In some embodiments, the nucleic acid fragments are double-stranded DNA fragments with 3' and/or 5' single-stranded overhangs. These overhangs may be cohesive ends that can anneal to complementary cohesive ends on different nucleic acid fragments. According to aspects of the invention, the presence of complementary sequences (and particularly complementary cohesive ends) on two nucleic acid fragments promotes their covalent assembly. In some embodiments, a plurality of nucleic acid fragments with different overlapping complementary single-stranded cohesive ends can be assembled and their order in the assembled nucleic acid product can be determined by the identity of the cohesive ends on each fragment. For example, the nucleic acid fragments may be designed so that a first nucleic acid has a first cohesive end that is complementary to a first cohesive end of the vector and a second cohesive end that is complementary to a first cohesive end of a second nucleic acid. The second cohesive end of the second nucleic acid may be complementary to a first cohesive end of a third nucleic acid. The second cohesive end of the third nucleic acid may be complementary a first cohesive end of a fourth nucleic acid. And so on

through to the final nucleic acid that has a first cohesive end that may be complementary to a second cohesive end on the penultimate nucleic acid.

[0077] In certain embodiments, the overlapping complementary regions between adjacent nucleic acid fragments are designed (or selected) to be sufficiently different to promote (e.g., thermodynamically favor) assembly of a unique alignment of nucleic acid fragments (e.g., a selected or designed alignment of fragments). It should be appreciated that overlapping regions of different length may be used. In some embodiments, longer cohesive ends may be used when higher numbers of nucleic acid fragments are being assembled. Longer cohesive ends may provide more flexibility to design or select sufficiently distinct sequences to discriminate between correct cohesive end annealing (e.g., involving cohesive ends designed to anneal to each other) and incorrect cohesive end annealing (e.g., between non-complementary cohesive ends).

[0078] In some embodiments, two or more pairs of complementary cohesive ends between different nucleic acid fragments may be designed or selected to have identical or similar sequences in order to promote the assembly of products containing a relatively random arrangement (and/or number) of the fragments that have similar or identical cohesive ends. This may be useful to generate libraries of nucleic acid products with different sequence arrangements and/or different copy numbers of certain internal sequence regions.

[0079] In some embodiments, the second cohesive end of the final nucleic acid may be complementary to a second cohesive end of the vector. According to aspects of the invention, this method may be used to generate a vector containing nucleic acid fragments assembled in a predetermined linear order (e.g., first, second, third, forth, . . . , final). In some embodiments, each of the two terminal nucleic acid fragments (e.g., the terminal fragment at each end of an assembled product) may be designed to have a cohesive end that is complementary to a cohesive end on a vector (e.g., on a linearized vector). These cohesive ends may be identical cohesive ends that can anneal to identical complementary terminal sequences on a linearized vector. However, in some embodiments, the cohesive ends on the terminal fragments are different and the vector contains two different cohesive ends, one at each end of a linearized vector, each complementary to one of the terminal fragment cohesive ends. Accordingly, the vector may be a linearized plasmid that has two cohesive ends, each of which is complementary with one end of the assembled nucleic acid fragments.

[0080] Some aspects of the invention involve double-stranded nucleic acids with single-stranded overhangs. Overhangs may be generated using any suitable technique. In some embodiments, a double-stranded nucleic acid fragment (e.g., a fragment assembled in a multiplex assembly) may be digested with an appropriate restriction enzyme to generate a terminal single-stranded overhang. In some embodiments, fragments that are designed to be adjacent to each other in an assembled product may be digested with the same enzyme to expose complementary overhangs. In some embodiments, overhangs may be generated using a type IIS restriction enzyme. Type IIS restriction enzymes are enzymes that bind to a double stranded nucleic acid at one site, referred to as the recognition site, and make a single double stranded cut outside of the recognition site. The double stranded cut, referred to as the cleavage site, is generally situated 0-20 bases away from the recognition site. The recognition site is generally about 4-7 bp long. All type IIS restriction enzymes exhibit at least partial asymmetric recognition. Asymmetric recognition means that 5'3' recognition sequences are different for each strand of the nucleic acid. The enzyme activity also shows polarity meaning that the cleavage sites are located on only one side of the recognition site. Thus, there is generally only one double stranded cut corresponding to each recognition site. Cleavage generally produces 1-5 nucleotide single-stranded overhangs, with 5' or 3' termini, although some enzymes produce blunt ends. Either cut is useful in the context of the invention, although in some instances those producing single-stranded overhangs are produced. To date, about 80 type IIS enzymes have been identified. Examples include but are not limited to BstF5 I, BtsC I, BsrD I, Bts I, Alw I, Bcc I, BsmA I, Ear I, Mly I (blunt), Ple I, Bmr I, Bsa I, BsmB I, Fau I, Mnl I, Sap I, Bbs I, BciV I, Hph I, Mbo II, BfuA I, BspCN I, BspM I, SfaN I, Hga I, BseR I, Bbv I, Eci I, Fok I, BceA I, BsmF I, BtgZ I, BpuE I, Bsg I, Mme I, BseG I, Bse3D I, BseM I, AcIW I, Alw26 I, Bst6 I, BstMA I, Eam1104 I, Ksp632 I, Pps I, Sch I (blunt), Bfi I, Bso31 I, BspTN I, Eco31 I, Esp3 I, Smu I, Bfu I, Bpi I, BpuA I, BstV2 I, AsuHP I, Acc36 I, Lwe I, Aar I, BseM II, TspDT I, TspGW I, BseX I, BstV1 I, Eco57 I, Eco57M I, Gsu I, and Bcg I. Such enzymes and information regarding their recognition and cleavage sites are available from commercial suppliers such as New England Biolabs, Inc. (Ipswich, Mass., U.S.A.).

[0081] In some embodiments, commercial or engineered restriction enzyme may be used. In some embodiments, Type IIS restriction enzymes can be designed and engineered to produce longer overhang lengths. Designing and engineering restriction enzymes to produce longer

single-stranded overhangs can allow for the joining of a larger number of oligonucleotides together to form longer nucleic acid constructs. For example, BsaI, which produces a 4 nucleotide single-stranded overhang, can be engineered to produce a 5, or 6 or longer single-stranded overhang. Increasing the length of the single-stranded overhang produced by such engineered BsaI can increase the theoretical limit of 17 nucleic acids or oligonucleotides that can be joined.

[0082] In some embodiments, each of a plurality of nucleic acid fragments designed for nucleic acid assembly may have a Type IIS restriction site at each end. The Type IIS restriction sites may be oriented so that the cleavage sites are internal relative to the recognition sequences. As a result, enzyme digestion exposes an internal sequence (e.g., an overhang within an internal sequence) and removes the recognition sequences from the ends. Accordingly, the same Type IIS sites may be used for both ends of all of the nucleic acid fragments being prepared for assembly and/or may be used for linearizing a suitable vector. However, different Type IIS sites also may be used. Two fragments that are designed to be adjacent in an assembled product each may include an identical overlapping terminal sequence and a flanking Type IIS site that is appropriately located to expose complementary overhangs within the overlapping sequence upon restriction enzyme digestion. Accordingly, a plurality of nucleic acid fragments may be generated with different complementary overhangs. The restriction site at each end of a nucleic acid fragment may be located such that digestion with the appropriate Type IIS enzyme removes the restriction site and exposes a single-stranded region that is complementary to a single-stranded region on a nucleic acid fragment that is designed to be adjacent in the assembled nucleic acid product. In some embodiments, one end of each of the two terminal nucleic acid fragments may be designed to have a single-stranded overhang (e.g., after digestion with an appropriate restriction enzyme) that is complementary to a single-stranded overhang of a linearized vector nucleic acid. Accordingly, the resulting nucleic acid fragments and vector may be transformed directly into a host cell. Alternatively, the nucleic acid fragments and vector may be incubated to promote hybridization and annealing of the complementary sequences prior to transformation in the host cell. It should be appreciated that a vector may be prepared using any one of the techniques described herein or any other suitable technique that produces a single-stranded overhang that would be complementary to an end of one of the terminal nucleic acid fragments.

[0083] Enzymatic digestions of DNA with Type II or site-specific restriction enzymes typically generate an overhang of four to six nucleotides. These short cohesive ends may be sufficient for ligating two nucleic acid fragments containing complementary termini. However, when joining multiple nucleic acid fragments together, longer complementary cohesive termini may be preferred to facilitate assembly and to ensure specificity. For example, cohesive ends may be long enough to have sufficiently different sequences to prevent or reduce mispairing between similar cohesive ends. However, their length is preferably not long enough to stabilize mispairs between similar cohesive sequences. In some embodiments, a length of about 9 to about 15 bases may be used. However, any suitable length may be selected for a region that is to be used to generate a cohesive overhang. The importance of specificity may depend on the number of different fragments that are being assembled simultaneously. Also, the appropriate length required to avoid stabilizing mispaired regions may depend on the conditions used for annealing different cohesive ends.

Ligase-based assembly

[0084] Ligase-based assembly techniques may involve one or more suitable ligase enzymes that can catalyze the covalent linking of adjacent 3' and 5' nucleic acid termini (e.g., a 5' phosphate and a 3' hydroxyl of nucleic acid(s) annealed on a complementary template nucleic acid such that the 3' terminus is immediately adjacent to the 5' terminus). Accordingly, a ligase may catalyze a ligation reaction between the 5' phosphate of a first nucleic acid to the 3' hydroxyl of a second nucleic acid if the first and second nucleic acids are annealed next to each other on a template nucleic acid). A ligase may be obtained from recombinant or natural sources. A ligase may be a heat-stable ligase. In some embodiments, a thermostable ligase from a thermophilic organism may be used. Examples of thermostable DNA ligases include, but are not limited to: Tth DNA ligase (from Thermus thermophilus, available from, for example, Eurogentec and GeneCraft); Pfu DNA ligase (a hyperthermophilic ligase from Pyrococcus furiosus); Taq ligase (from Thermus aquaticus), any other suitable heat-stable ligase, or any combination thereof. In some embodiments, one or more lower temperature ligases may be used (e.g., T4 DNA ligase). A lower temperature ligase may be useful for shorter overhangs (e.g., about 3, about 4, about 5, or about 6 base overhangs) that may not be stable at higher temperatures.

[0085] In some embodiments, ligase may be designed and engineered to have a greater degree of specificity so as to minimize unwanted ligation products formed. In some

embodiments, ligase may be used in conjunction with proteins or may be fused with proteins capable of facilitating the interaction of the ligase with nucleic acid molecules and/or of increasing specificity of ligation.

[0086] Non-enzymatic techniques can be used to ligate nucleic acids. For example, a 5'-end (e.g., the 5' phosphate group) and a 3'-end (e.g., the 3' hydroxyl) of one or more nucleic acids may be covalently linked together without using enzymes (e.g., without using a ligase). In some embodiments, non-enzymatic techniques may offer certain advantages over enzyme-based ligations. For example, non-enzymatic techniques may have a high tolerance of non-natural nucleotide analogues in nucleic acid substrates, may be used to ligate short nucleic acid substrates, may be used to ligate RNA substrates, and/or may be cheaper and/or more suited to certain automated (e.g., high throughput) applications. Accordingly, a chemical ligation may be used to form a covalent linkage between a 5' terminus of a first nucleic acid end and a 3' terminus of a second nucleic acid end, wherein the first and second nucleic acid ends may be ends of a single nucleic acid or ends of separate nucleic acids. In one aspect, chemical ligation may involve at least one nucleic acid substrate having a modified end (e.g., a modified 5' and/or 3' terminus) including one or more chemically reactive moieties that facilitate or promote linkage formation. In some embodiments, chemical ligation occurs when one or more nucleic acid termini are brought together in close proximity (e.g., when the termini are brought together due to annealing between complementary nucleic acid sequences). Accordingly, annealing between complementary 3' or 5' overhangs (e.g., overhangs generated by restriction enzyme cleavage of a double-stranded nucleic acid) or between any combination of complementary nucleic acids that results in a 3' terminus being brought into close proximity with a 5' terminus (e.g., the 3' and 5' termini are adjacent to each other when the nucleic acids are annealed to a complementary template nucleic acid) may promote a template-directed chemical ligation. Examples of chemical reactions may include, but are not limited to, condensation, reduction, and/or photo-chemical ligation reactions. It should be appreciated that in some embodiments chemical ligation can be used to produce naturally-occurring phosphodiester internucleotide linkages, non-naturally-occurring phosphamide pyrophosphate internucleotide linkages, and/or other non-naturally-occurring internucleotide linkages.

<u>Concurrent enzymatic removal of common oligonucleotide sequences and ligation of processed oligonucleotides into longer constructs</u>

[0087] FIG. 2 illustrates a method for assembling a nucleic acid in accordance with one embodiment of the invention. In some embodiments, the method comprises concurrent enzymatic removal of common oligonucleotide sequences and ligation of processed oligonucleotide sequences into longer constructs. In some embodiments, the oligonucleotides are amplified by PCR and error corrected as described herein. Amplified oligonucleotides (10), composed of a common priming (amplification) sequence (20) and construct specific payload or internal sequences regions (30) are processed by an appropriate restriction endonuclease (40). In some embodiments, the first and last oligonucleotides contain unique priming sequences (25) for amplification of the target construct. The restriction endonuclease catalyzes the cleavage of the terminal common regions (also referred herein as amplification regions or primer recognition sequences) shared by all of the oligonucleotides (50), leaving internal regions (also referred herein as free payload) with terminal single stranded DNA sequences (60). In some embodiments, the restriction endonuclease is a type IIS restriction endonuclease. These single stranded sequences are designed to instruct the specific interaction of one oligonucleotide with another, allowing the linear arrangement of a number of oligonucleotides into a defined sequence (70). Accordingly, the terminal single stranded DNA sequences can direct the appropriate interaction of oligonucleotides into the correct order, whereby ligase (80) enzyme catalyses the joining of individual oligonucleotides, generating the final target nucleic acid construct (90) or intermediate nucleic acid constructs.

[0088] One of skill in the art will appreciate that if the original common sequence is ligated back together (for example (50) using the terminal sequences complementary to (60)), the presence of the restriction endonuclease can ensure that it may be cut again to generate the free end (60). However, because of the choice of restriction endonuclease, a properly ligated junction (for example between 1' and 2') will not be recognized as a restriction site and will not be undone. The reaction should naturally drive toward the desired product (90).

[0089] In some embodiments, a variant of the process recognizes that the restriction site used for common sequence removal can now be part of the gene to be synthesized. This constraint removal allows for recursive (hierarchical) applications of the gene synthesis method to build longer nucleic acid sequences (as illustrated in FIG. 4). In previous methodologies, where removal and ligation were performed as separate steps, this design was disallowed due to the necessity of a purification step in between the removal and the ligation steps, which was

based partially on size selection. In such methodologies, pieces cut of the desired target sequence could be lost during the purification, resulting in failure to build the desired target sequence. In some embodiments, using the concurrent removal and ligation step of the invention, those cut sequences would be constantly cut and re-ligated, resulting in the presence of some of the target sequence of interest. The amount of the desired sequence may depend, in some embodiments, on the tuning of the relative activities of the restriction enzyme and the ligase.

[0090] As illustrated in FIG. 4, the gene synthesis pieces (390) and (391) can be assembled from oligonucleotide sets (310) and (311). The oligonucleotide sets can be designed with matching restriction endonuclease sites (340) such that the gene synthesis pieces (390) and (391) can be joined using the same concurrent digestion and ligation process (with subsequent amplification). In some embodiments, the second round can have been designed with restriction endonuclease sites (340) using a second restriction enzyme. However, this may be undesirable due to complications of using multiple enzymes in the process. In addition, without the concurrent digestion and ligation, the use of two restrictions enzymes would result in disallowing two restriction enzyme sites from the target sequence, further constraining the genes that can be synthesized.

[0091] Still referring to FIG. 4, the nucleic acid fragment (390) can be amplified using primers (325), and the nucleic acid fragment (391) can be amplified using primers (326). The nucleic acid fragment may then be mixed together and processed in a similar fashion to the previous synthesis step to create the combined nucleic acid fragment (392), where the restriction sites (340) act in a similar manner to the sites (350) in the previous round. The combined target sequence (392) can be amplified using the 5' primer from (325) and the 3' primer from (326).

[0092] In some embodiments, hierarchical assembly strategies may be used in accordance with the methods disclosed herein. One of skill the art will appreciate that the present method can be scalable to multiple nucleic acid fragments, such that the number of nucleic acid fragments in the subsequent round can be similar to the number of nucleic acid fragments in the first round. The hierarchical assembly method can be geometric, allowing very large targets to be constructed in a relatively few number of rounds. For example, a target sequence of 1000 bases (1 kbp) can be constructed from one of the pools (310) or (311). A second round of 10 nucleic acid fragments similar to (390) or (391) would result in a 10 kbp base

target nucleic acid sequence. A third round, using the 10 kbp nucleic acid sequences, would result in a 100 kbp target nucleic acid sequence, derived from the original 100 source pools.

[0093] In some embodiments, a plurality of assembly reactions may be conducted in separate pools. Assembly constructs from the assembly reactions may then be mixed to together to form even longer nucleic acid sequences. In some embodiments, hierarchical assembly may be carried out using restriction endonucleases to form cohesive ends that may be joined together in a desired order. The construction oligonucleotides may be designed and synthesized to contain recognition and cleavage sites for one or more restriction endonucleases at sites that would facilitate joining in a specified order. In some embodiments, one or more Type IIS endonuclease recognition sites may be incorporated into the termini of the construction oligonucleotides to permit cleavage by a Type IIS restriction endonuclease. The order of joining can be determined by hybridization of the complementary cohesive ends.

[0094] In some embodiments, the first pool of oligonucleotides comprises a 3' end oligonucleotide designed to have an additional restriction enzyme recognition site at its 3' end and the second pool of oligonucleotides comprises a 5' end oligonucleotide designed to have an additional restriction enzyme recognition site at its 5' end. In some embodiments, the restriction enzymes are the same. After assembly of the oligonucleotides in each pool, the two subassembly constructs can be subjected to the restriction endonuclease and to ligase in accordance with the methods disclosed herein.

[0095] One of skill in the art would understand that the available assembly space of the synthesis is drastically (geometrically) improved by the aspects of the invention. Previously, to generate a construct of double the sequence size (2n), double the numbers of oligonucleotides were required. For example, to generate a construct (390), double the numbers of oligonucleotides (310) were required, and thus double the numbers of compatible single stranded ends (360) were required. Using the method illustrated in FIG. 4, the junctions for (310) and (311) only have to be compatible with junction (340), thus enabling the assembly of nucleic acids of double the size with only one extra junction used. Therefore, if oligonucleotides (310) and (311) have interfering or incompatible ends, they may still be joined by the process disclosed herein (digestion (340) and ligation) to make target nucleic acid (392), whereas joining would not be possible by solely mixing the oligonucleotide pools (310) and (311) together.

[0096] A variant of the concurrent processing of oligonucleotides and assembly into target constructs and simultaneous entry into a plasmid is illustrated in FIG. 3. Details of the plasmid, pG9-1 (SEQ ID NO. 1) are shown in FIG. 5. The plasmid contains restriction endonuclease recognition sites (underlined text, FIG. 5) that allows a restriction endonuclease (in this case BsaI) to cut the plasmid at two positions, leaving defined single stranded sequences (FIG. 5 - reverse text). Referring to FIG. 3, plasmid (100) (e.g. pG9-1) is introduced into a pool comprising a mixture of oligonucleotides (110) that have been previously amplified and error corrected as described herein. In some embodiments, these oligonucleotide sequences (110) can have common sequences (120) that are recognized by a specific restriction endonuclease (140). In some embodiments, the plasmid (130) can have sequences recognized by the same restriction endonuclease (140). Action of restriction endonuclease (140) upon these sequences results in the removal of the common sequences from the oligonucleotides ((310), (311)) and plasmid (150), exposing single stranded DNA sequences (160). In some embodiments, the restriction enzyme can be a type IIS restriction enzyme. In some embodiments, the single stranded sequences are designed to instruct the specific interaction of one oligonucleotide with another, allowing the arrangement of a number of oligonucleotides into a defined sequence and entry of this ordered sequence of oligonucleotides (170) into the plasmid (100). In some embodiments, ligase (180) enzyme catalyzes the covalent joining of the individual oligonucleotides. The final product is the plasmid (e.g. pG9-1) containing the specified construct derived from joining the oligonucleotides (190). This plasmid (190) may then transformed into a bacteria and sequenced-verified.

[0097] Aspects of the invention relate to the sequence verification of the constructs assembled according to the methods of the invention. Sequence verification of constructs is illustrated in FIG. 6. In this process, a number of constructs (200, C1 to C4) can be generated as shown in FIG. 3 and transformed into bacteria. Bacterial transformants containing plasmid DNA can be selected on solid growth plates (210) using an appropriate antibiotic resistance for selection. After growth, single colonies are picked and pooled, one from each construct plate (220), generating pools of constructs, each pool containing one copy of each construct. In some embodiments, the number of pools can be dependent upon the number of individual constructs that are to be sequenced in order to identify constructs with perfect sequence. As illustrated in FIG. 6, four pools of the four constructs are generated, allowing analysis of four members of

each construct. Plasmid DNA can then be prepared from the pooled material (230). Each pool of plasmid DNA molecules can then prepared for sequencing. This preparation may use one of a variety of methods that cause breakage of DNA into smaller fragments and the attachment of common sequences required for sequencing using, for example, next generation high throughput sequencing. Short pieces of DNA, unique to each of the four pools generated, are contained within these common sequences. These unique pieces of DNA can allow identification of which pool each sequenced construct is derived from. Constructs with the correct sequence can then be recovered by going back to the initial bacterial growth plate and re-growing the corresponding colony containing the plasmid with the wanted construct.

Vectors and Host cells

[0098] Any suitable vector may be used, as the invention is not so limited. For example, a vector may be a plasmid, a bacterial vector, a viral vector, a phage vector, an insect vector, a yeast vector, a mammalian vector, a BAC, a YAC, or any other suitable vector. In some embodiments, a vector may be a vector that replicates in only one type of organism (e.g., bacterial, yeast, insect, mammalian, etc.) or in only one species of organism. Some vectors may have a broad host range. Some vectors may have different functional sequences (e.g., origins or replication, selectable markers, etc.) that are functional in different organisms. These may be used to shuttle the vector (and any nucleic acid fragment(s) that are cloned into the vector) between two different types of organism (e.g., between bacteria and mammals, yeast and mammals, etc.). In some embodiments, the type of vector that is used may be determined by the type of host cell that is chosen.

[0099] It should be appreciated that a vector may encode a detectable marker such as a selectable marker (e.g., antibiotic resistance, etc.) so that transformed cells can be selectively grown and the vector can be isolated and any insert can be characterized to determine whether it contains the desired assembled nucleic acid. The insert may be characterized using any suitable technique (e.g., size analysis, restriction fragment analysis, sequencing, etc.). In some embodiments, the presence of a correctly assembly nucleic acid in a vector may be assayed by determining whether a function predicted to be encoded by the correctly assembled nucleic acid is expressed in the host cell.

[00100] In some embodiments, host cells that harbor a vector containing a nucleic acid insert may be selected for or enriched by using one or more additional detectable or selectable

markers that are only functional if a correct (e.g., designed) terminal nucleic acid fragments is cloned into the vector.

[00101] Accordingly, a host cell should have an appropriate phenotype to allow selection for one or more drug resistance markers encoded on a vector (or to allow detection of one or more detectable markers encoded on a vector). However, any suitable host cell type may be used (e.g., prokaryotic, eukaryotic, bacterial, yeast, insect, mammalian, etc.). For example, host cells may be bacterial cells (e.g., Escherichia coli, Bacillus subtilis, Mycobacterium spp., M. tuberculosis, or other suitable bacterial cells), yeast cells (for example, Saccharomyces spp., Picchia spp., Candida spp., or other suitable yeast species, e.g., S. cerevisiae, C. albicans, S. pombe, etc.), Xenopus cells, mouse cells, monkey cells, human cells, insect cells (e.g., SF9 cells and Drosophila cells), worm cells (e.g., Caenorhabditis spp.), plant cells, or other suitable cells, including for example, transgenic or other recombinant cell lines. In addition, a number of heterologous cell lines may be used, such as Chinese Hamster Ovary cells (CHO).

Applications

[00102] Aspects of the invention may be useful for a range of applications involving the production and/or use of synthetic nucleic acids. As described herein, the invention provides methods for assembling synthetic nucleic acids with increased efficiency. The resulting assembled nucleic acids may be amplified in vitro (e.g., using PCR, LCR, or any suitable amplification technique), amplified in vivo (e.g., via cloning into a suitable vector), isolated and/or purified. An assembled nucleic acid (alone or cloned into a vector) may be transformed into a host cell (e.g., a prokaryotic, eukaryotic, insect, mammalian, or other host cell). In some embodiments, the host cell may be used to propagate the nucleic acid. In certain embodiments, the nucleic acid may be integrated into the genome of the host cell. In some embodiments, the nucleic acid may replace a corresponding nucleic acid region on the genome of the cell (e.g., via homologous recombination). Accordingly, nucleic acids may be used to produce recombinant organisms. In some embodiments, a target nucleic acid may be an entire genome or large fragments of a genome that are used to replace all or part of the genome of a host organism. Recombinant organisms also may be used for a variety of research, industrial, agricultural, and/or medical applications.

[00103] Many of the techniques described herein can be used together, applying combinations of one or more extension-based and/or ligation-based assembly techniques at one

or more points to produce long nucleic acid molecules. For example, concerted assembly may be used to assemble oligonucleotide duplexes and nucleic acid fragments of less than 100 to more than 10,000 base pairs in length (e.g., 100 mers to 500 mers, 500 mers to 1,000 mers, 1,000 mers to 5,000 mers, 5,000 mers to 10,000 mers, 25,000 mers, 50,000 mers, 75,000 mers, 100,000 mers, etc.). In an exemplary embodiment, methods described herein may be used during the assembly of an entire genome (or a large fragment thereof, e.g., about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or more) of an organism (e.g., of a viral, bacterial, yeast, or other prokaryotic or eukaryotic organism), optionally incorporating specific modifications into the sequence at one or more desired locations.

[00104] Nucleic acid molecules generated using methods of the invention can be incorporated into a vector. The vector may be a cloning vector or an expression vector. A vector may comprise an origin of replication and one or more selectable markers (e.g., antibiotic resistant markers, auxotrophic markers, etc.). In some embodiments, the vector may be a viral vector. A viral vector may comprise nucleic acid sequences capable of infecting target cells. Similarly; in some embodiments, a prokaryotic expression vector operably linked to an appropriate promoter system can be used to transform target cells. In other embodiments, a eukaryotic vector operably linked to an appropriate promoter system can be used to transfect target cells or tissues.

[00105] Transcription and/or translation of the constructs described herein may be carried out in vitro (i.e., using cell-free systems) or in vivo (i.e., expressed in cells). In some embodiments, cell lysates may be prepared. In certain embodiments, expressed RNAs or polypeptides may be isolated or purified.

[00106] Aspects of the methods and devices provided herein may include automating one or more acts described herein. In some embodiments, one or more steps of an amplification and/or assembly reaction may be automated using one or more automated sample handling devices (e.g., one or more automated liquid or fluid handling devices). Automated devices and procedures may be used to deliver reaction reagents, including one or more of the following: starting nucleic acids, buffers, enzymes (e.g., one or more ligases and/or polymerases), nucleotides, salts, and any other suitable agents such as stabilizing agents. Automated devices and procedures also may be used to control the reaction conditions. For example, an automated thermal cycler may be used to control reaction temperatures and any temperature cycles that may

be used. In some embodiments, a scanning laser may be automated to provide one or more reaction temperatures or temperature cycles suitable for incubating polynucleotides. Similarly, subsequent analysis of assembled polynucleotide products may be automated. For example, sequencing may be automated using a sequencing device and automated sequencing protocols. Additional steps (e.g., amplification, cloning, etc.) also may be automated using one or more appropriate devices and related protocols. It should be appreciated that one or more of the device or device components described herein may be combined in a system (e.g., a robotic system) or in a micro-environment (e.g., a micro-fluidic reaction chamber). Assembly reaction mixtures (e.g., liquid reaction samples) may be transferred from one component of the system to another using automated devices and procedures (e.g., robotic manipulation and/or transfer of samples and/or sample containers, including automated pipetting devices, micro-systems, etc.). The system and any components thereof may be controlled by a control system.

[00107] Accordingly, method steps and/or aspects of the devices provided herein may be automated using, for example, a computer system (e.g., a computer controlled system). A computer system on which aspects of the technology provided herein can be implemented may include a computer for any type of processing (e.g., sequence analysis and/or automated device control as described herein). However, it should be appreciated that certain processing steps may be provided by one or more of the automated devices that are part of the assembly system. In some embodiments, a computer system may include two or more computers. For example, one computer may be coupled, via a network, to a second computer. One computer may perform sequence analysis. The second computer may control one or more of the automated synthesis and assembly devices in the system. In other aspects, additional computers may be included in the network to control one or more of the analysis or processing acts. Each computer may include a memory and processor. The computers can take any form, as the aspects of the technology provided herein are not limited to being implemented on any particular computer platform. Similarly, the network can take any form, including a private network or a public network (e.g., the Internet). Display devices can be associated with one or more of the devices and computers. Alternatively, or in addition, a display device may be located at a remote site and connected for displaying the output of an analysis in accordance with the technology provided herein. Connections between the different components of the system may be via wire,

optical fiber, wireless transmission, satellite transmission, any other suitable transmission, or any combination of two or more of the above.

[00108] Each of the different aspects, embodiments, or acts of the technology provided herein can be independently automated and implemented in any of numerous ways. For example, each aspect, embodiment, or act can be independently implemented using hardware, software or a combination thereof. When implemented in software, the software code can be executed on any suitable processor or collection of processors, whether provided in a single computer or distributed among multiple computers. It should be appreciated that any component or collection of components that perform the functions described above can be generically considered as one or more controllers that control the above-discussed functions. The one or more controllers can be implemented in numerous ways, such as with dedicated hardware, or with general purpose hardware (e.g., one or more processors) that is programmed using microcode or software to perform the functions recited above.

[00109] In this respect, it should be appreciated that one implementation of the embodiments of the technology provided herein comprises at least one computer-readable medium (e.g., a computer memory, a floppy disk, a compact disk, a tape, etc.) encoded with a computer program (i.e., a plurality of instructions), which, when executed on a processor, performs one or more of the above-discussed functions of the technology provided herein. The computer-readable medium can be transportable such that the program stored thereon can be loaded onto any computer system resource to implement one or more functions of the technology provided herein. In addition, it should be appreciated that the reference to a computer program which, when executed, performs the above-discussed functions, is not limited to an application program running on a host computer. Rather, the term computer program is used herein in a generic sense to reference any type of computer code (e.g., software or microcode) that can be employed to program a processor to implement the above-discussed aspects of the technology provided herein.

[00110] It should be appreciated that in accordance with several embodiments of the technology provided herein wherein processes are stored in a computer readable medium, the computer implemented processes may, during the course of their execution, receive input manually (e.g., from a user).

[00111] Accordingly, overall system-level control of the assembly devices or components described herein may be performed by a system controller which may provide control signals to the associated nucleic acid synthesizers, liquid handling devices, thermal cyclers, sequencing devices, associated robotic components, as well as other suitable systems for performing the desired input/output or other control functions. Thus, the system controller along with any device controllers together form a controller that controls the operation of a nucleic acid assembly system. The controller may include a general purpose data processing system, which can be a general purpose computer, or network of general purpose computers, and other associated devices, including communications devices, modems, and/or other circuitry or components to perform the desired input/output or other functions. The controller can also be implemented, at least in part, as a single special purpose integrated circuit (e.g., ASIC) or an array of ASICs, each having a main or central processor section for overall, system-level control, and separate sections dedicated to performing various different specific computations, functions and other processes under the control of the central processor section. The controller can also be implemented using a plurality of separate dedicated programmable integrated or other electronic circuits or devices, e.g., hard wired electronic or logic circuits such as discrete element circuits or programmable logic devices. The controller can also include any other components or devices, such as user input/output devices (monitors, displays, printers, a keyboard, a user pointing device, touch screen, or other user interface, etc.), data storage devices, drive motors, linkages, valve controllers, robotic devices, vacuum and other pumps, pressure sensors, detectors, power supplies, pulse sources, communication devices or other electronic circuitry or components, and so on. The controller also may control operation of other portions of a system, such as automated client order processing, quality control, packaging, shipping, billing, etc., to perform other suitable functions known in the art but not described in detail herein.

[00112] Various aspects of the present invention may be used alone, in combination, or in a variety of arrangements not specifically discussed in the embodiments described in the foregoing and is therefore not limited in its application to the details and arrangement of components set forth in the foregoing description or illustrated in the drawings. For example, aspects described in one embodiment may be combined in any manner with aspects described in other embodiments.

[00113] Use of ordinal terms such as "first," "second," "third," etc., in the claims to modify a claim element does not by itself connote any priority, precedence, or order of one claim element over another or the temporal order in which acts of a method are performed, but are used merely as labels to distinguish one claim element having a certain name from another element having a same name (but for use of the ordinal term) to distinguish the claim elements.

[00114] Also, the phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting. The use of "including," "comprising," or "having," "containing," "involving," and variations thereof herein, is meant to encompass the items listed thereafter and equivalents thereof as well as additional items.

EQUIVALENTS

[00115] The present invention provides among other things novel methods and devices for high-fidelity gene assembly. While specific embodiments of the subject invention have been discussed, the above specification is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

INCORPORATION BY REFERENCE

[00116] Reference is made to U.S. application 13/986,368, filed April 24, 2013, U.S. application 13/524,164, filed June 15, 2012, and PCT publication PCT/US2009/055267. All publications, patents, patent applications, and sequence database entries mentioned herein are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference.

CLAIMS:

What is claimed is:

1. A method of producing a target nucleic acid having a predefined sequence, the method comprising:

a) providing a first mixture comprising

(i) a first pool of oligonucleotides comprising a first plurality of oligonucleotides comprising a sequence identical to the 5' end of the target nucleic acid, a second plurality of oligonucleotides comprising a sequence identical to the 3' end of the target nucleic acid, and a plurality of oligonucleotides comprising a sequence identical to a different portion of a sequence of a target nucleic acid, each of the oligonucleotides having an overlapping sequence region corresponding to a sequence region in a next oligonucleotide, the oligonucleotides in the first mixture together comprising the target nucleic acid sequence;

(ii) a restriction enzyme, and

b) exposing the first mixture to a ligase, thereby generating the target nucleic acid.

2. The method of claim 1 further comprising subjecting the target nucleic acid to sequence verification.

3. The method of claim 1 further comprising, prior to step (a), providing a plurality of construction oligonucleotides, wherein each construction oligonucleotide comprises (i) an internal sequence identical to a different portion of a sequence of a target nucleic acid, (ii) 5' and 3' flanking sequences flanking the 5'end and the 3' end of the internal sequence, each of the flanking sequence comprising a primer recognition site for a primer pair and a restriction enzyme recognition site.

4. The method of claim 3 further comprising amplifying the plurality of construction oligonucleotides.

5. The method of claim 3 further comprising subjecting the plurality of amplified oligonucleotides to error removal.

6. The method of claim 5 wherein the plurality of amplified oligonucleotides are contacted with a mismatch binding agent, wherein the mismatch binding agent selectively binds and cleaves the double-stranded oligonucleotides comprising a mismatch.

7. The method of claim 4 wherein the restriction enzyme and the ligase are added to a single pool of amplified oligonucleotides under conditions suitable to promote digestion and ligation, thereby generating a mixture comprising the assembled target nucleic acid sequences, and the flanking sequences.

8. The method of claim 1 wherein each flanking sequence comprises a common primer recognition site.

9. The method of claim 1 wherein the restriction enzyme is a Type IIS restriction enzyme.

10. The method of claim 7 wherein digestion with the Type IIS restriction enzyme produces a plurality of cohesive end double-stranded oligonucleotides and wherein the plurality of cohesive end double stranded oligonucleotides are ligated in a unique linear arrangement.

11. The method of claim 1 further comprising amplifying the target nucleic acid using a primer pair capable of recognizing a primer recognition site at the 5' end of the first oligonucleotide and 3' end of second oligonucleotide.

12. The method of claim 1 further comprising sequencing the target nucleic acid to confirm its sequence accuracy.

13. The method of claim 12 wherein the sequencing step by high throughput sequencing.

14. The method of claim 1 further comprising isolating at least one target nucleic acid having the predefined sequence from a pool of nucleic acid sequences.

15. The method of claim 1 further processing the target nucleic acid.

16. The method of claim 1 further comprising

c) providing a second mixture comprising

(i) a second pool of oligonucleotides comprising a first plurality of oligonucleotides comprising a sequence identical to the 5' end of the target nucleic acid, a second plurality of oligonucleotides comprising a sequence identical to the 3' end of the target nucleic acid, and a plurality of oligonucleotides comprising a sequence identical to a different portion of a sequence of a target nucleic acid, each of the oligonucleotides having an overlapping sequence region corresponding to a sequence region in a next oligonucleotide, the oligonucleotides in the second mixture together comprising the second target nucleic acid;

(ii) a restriction enzyme, and

d) exposing the second mixture to a ligase, thereby generating a second target nucleic acid.

17. The method of claim 16 further comprising assembling at least two target nucleic acids.

18. The method of claim 17 wherein the step of assembling is by hierarchical assembly.

19. The method of claim 16 wherein the second plurality of oligonucleotides of the first pool comprises a restriction endonuclease recognition site for a restriction endonuclease and the first plurality of oligonucleotides of the second pool comprises a restriction endonuclease recognition site for the restriction endonuclease.

20. The method of claim 19 wherein the at least two target nucleic acids are subjected to restriction endonuclease digestion and ligation thereby forming a long target nucleic acid construct.

21. The method of claim 20, wherein the long target nucleic acid construct is at least about 10 kilobases in length.

22. The method of claim 20, wherein the long target nucleic acid construct is at least about 100 kilobases in length.

23. A method of producing a target nucleic acid having a predefined sequence, the method comprising:

a) providing a plurality of oligonucleotides, wherein each oligonucleotide comprises (i) an internal sequence identical to a different portion of a sequence of a target nucleic acid, (ii) 5' and 3' flanking sequences flanking the 5' end and the 3' end of the internal sequence, each of the flanking sequence comprising a primer recognition site for a primer pair and a restriction enzyme recognition site for a restriction endonuclease;

b) amplifying at least a subset of the oligonucleotides using the primer pair thereby generating a plurality of amplified oligonucleotides;

c) optionally subjecting the plurality of amplified oligonucleotides to error removal;

d) providing a circular vector having a restriction enzyme recognition site for the restriction endonuclease; and

c) exposing the plurality of amplified oligonucleotides and circular vector to the restriction enzyme and ligase in a single pool, wherein the restriction enzyme is capable of recognizing the restriction enzyme recognition sites, thereby assembling the target nucleic acid in the vector.

24. The method of claim 23 further comprising transforming the vector into a host cell.

25. A composition for the assembly of a target nucleic acid having a predefined sequence comprising:

a) a pool of oligonucleotides comprising a first plurality of oligonucleotides comprising a sequence identical to the 5' end of the target nucleic acid, a second plurality of oligonucleotides comprising a sequence identical to the 3' end of the target nucleic acid, and one or more plurality of oligonucleotides comprising a sequence identical to a different portion of a sequence of a target nucleic acid, each of the oligonucleotides having an overlapping sequence region corresponding to a sequence region in a next oligonucleotide, the oligonucleotides in the pool together comprising the target nucleic acid;

b) a plurality of common sequences comprising a primer recognition site for a primer pair and a restriction endonuclease recognition site;

c) a restriction endonuclease; and

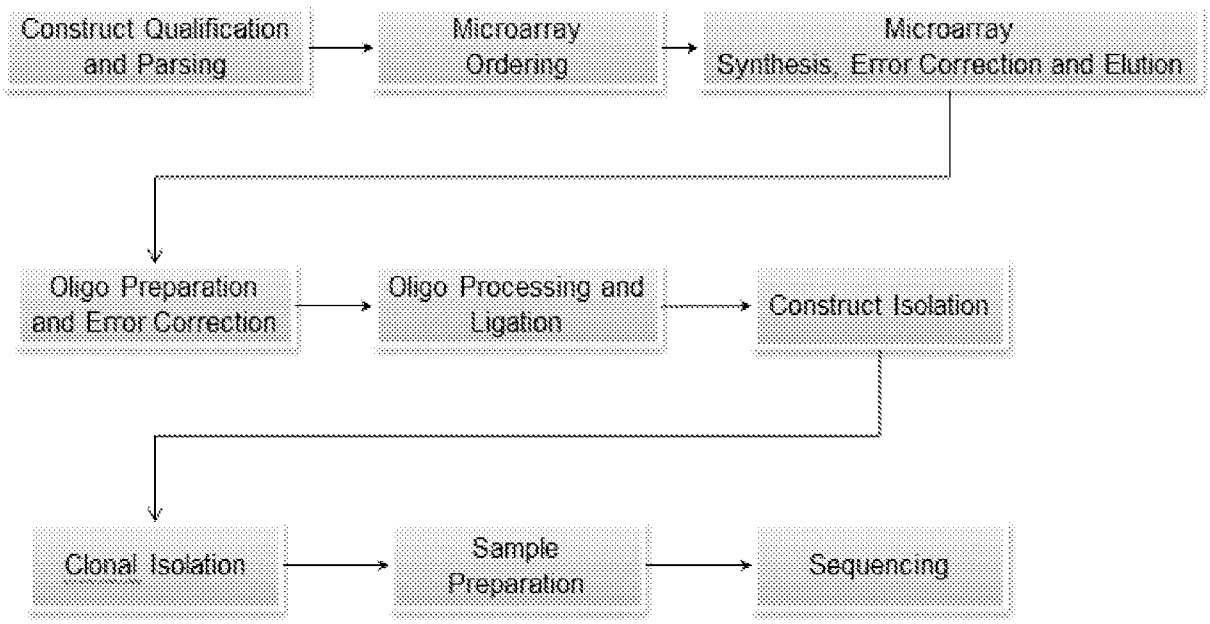
d) a ligase.

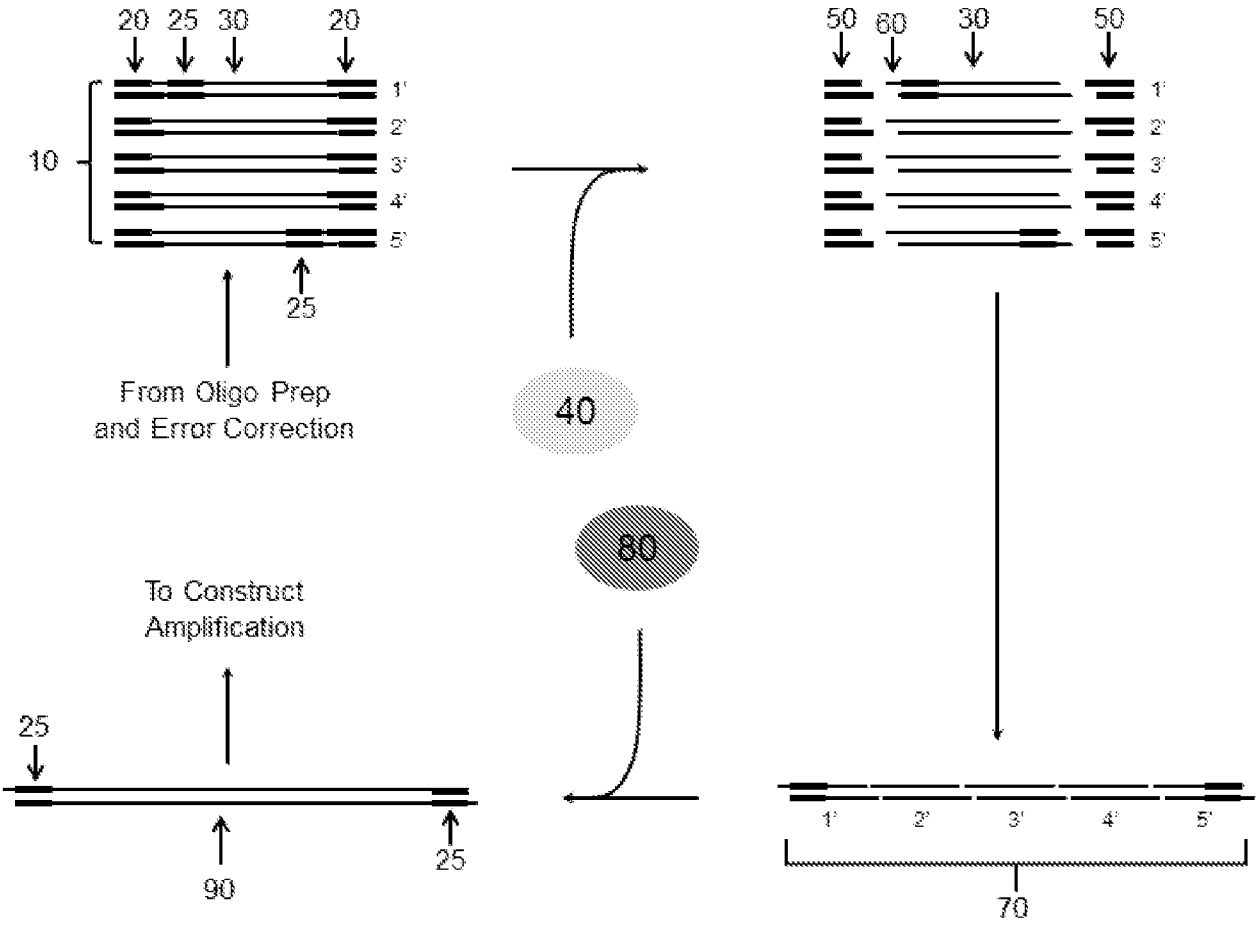
26. The composition of claim 25 wherein the oligonucleotides are amplified.

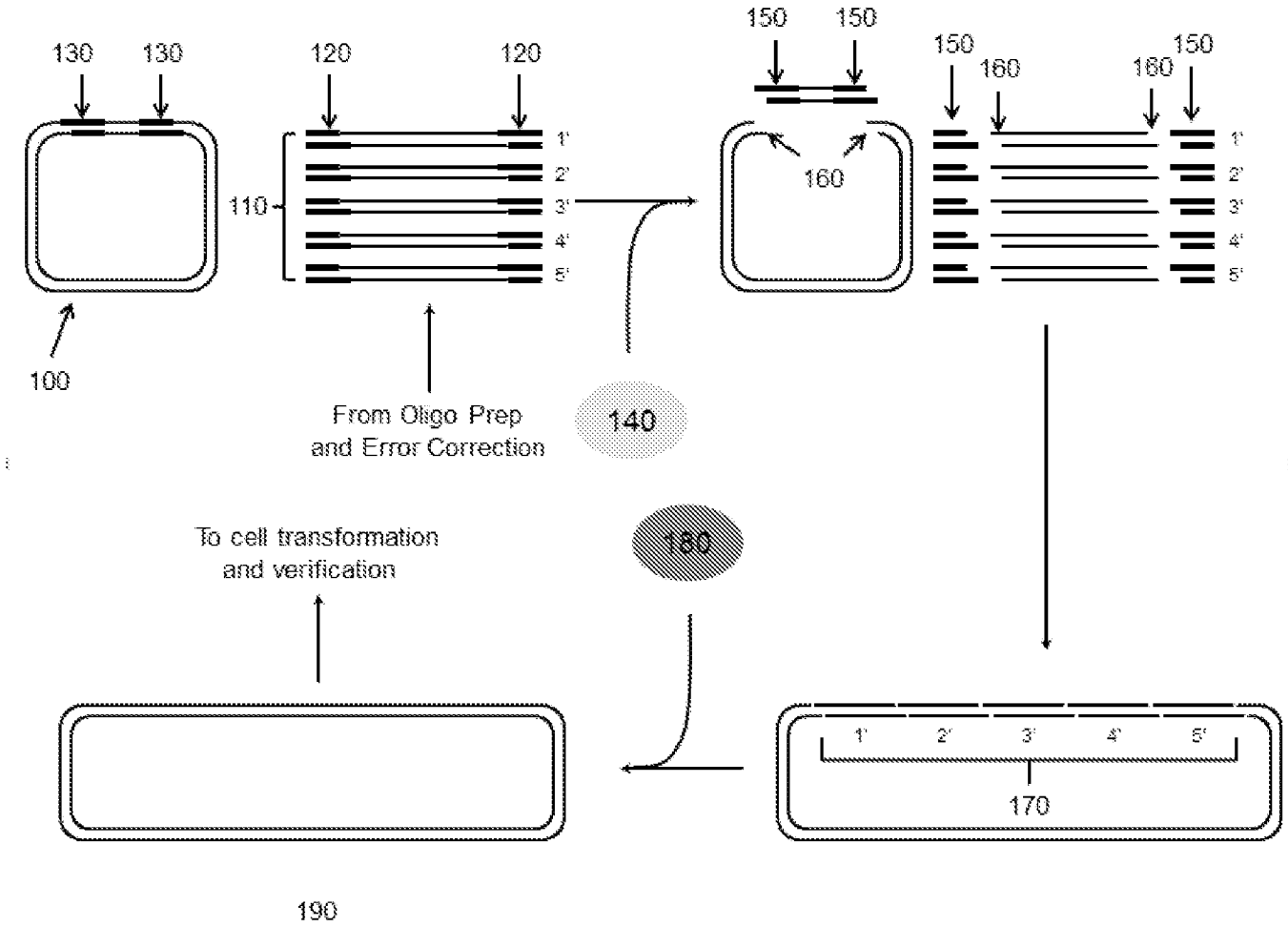
27. The composition of claim 25 wherein the oligonucleotides are error-corrected.

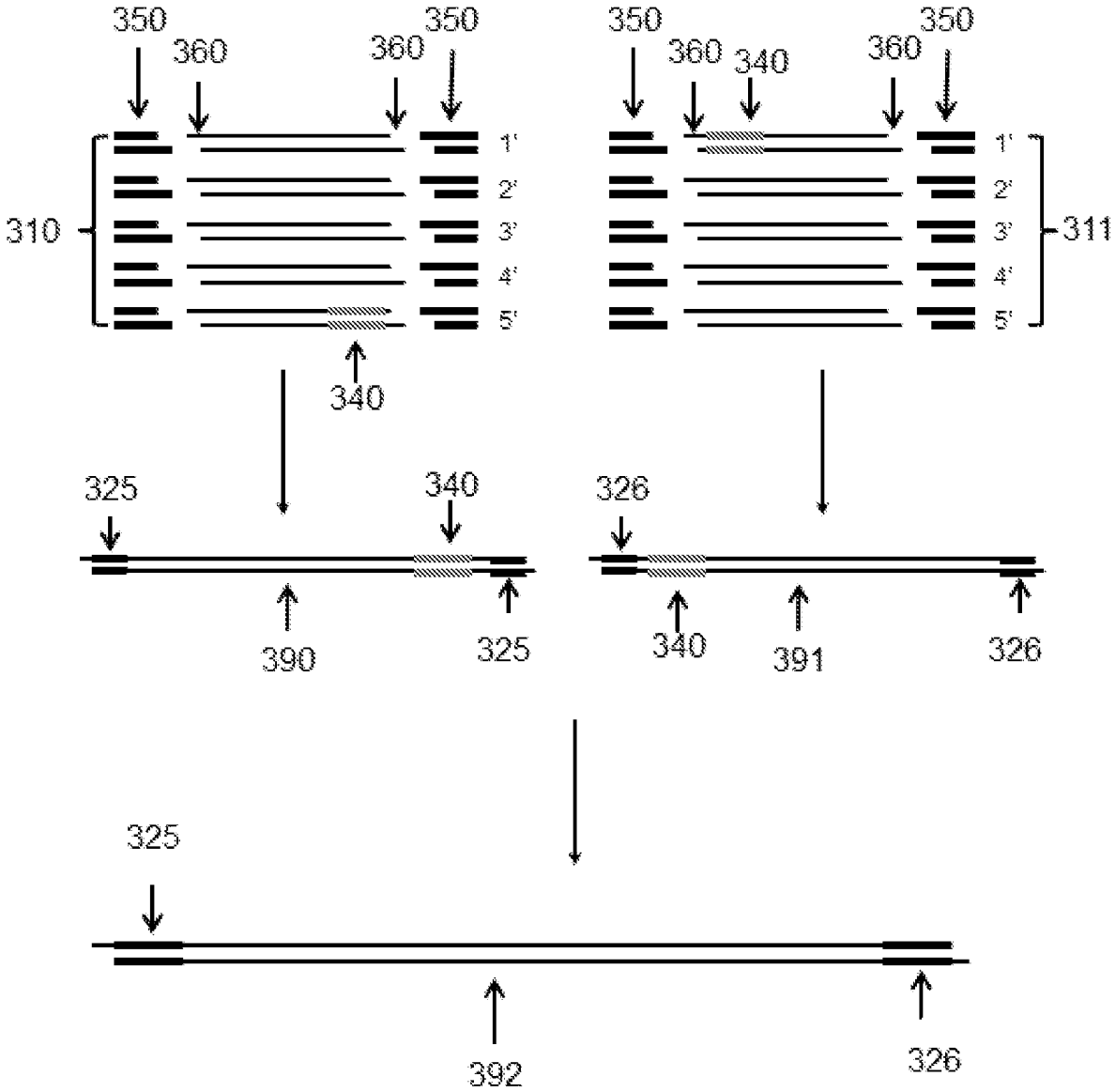
28. The composition of claim 25 further comprising a linearized vector having a 5' end compatible with the first plurality of oligonucleotides and a 3' end compatible with the second plurality of oligonucleotides.

29. The composition of claim 25 wherein the restriction endonuclease is a Type IIS restriction endonuclease.

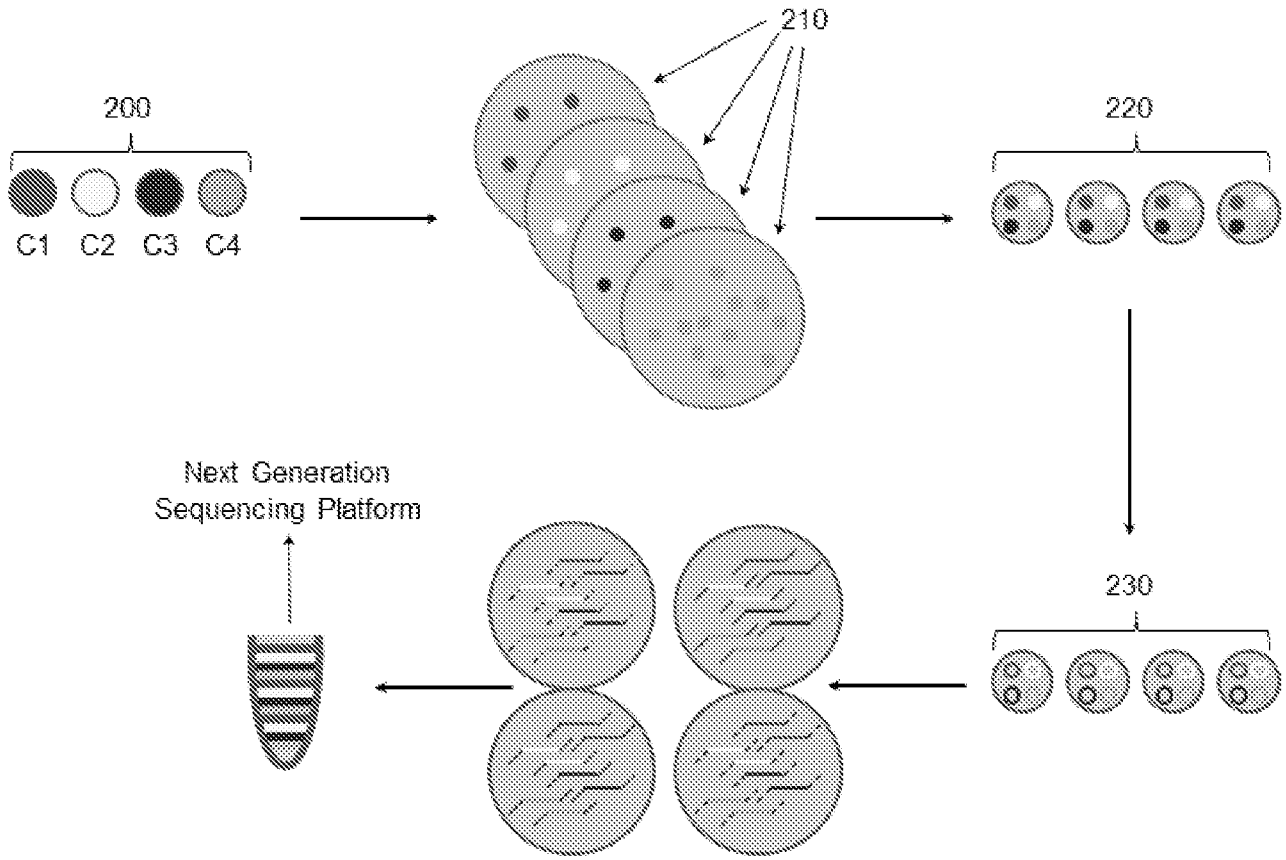

Construct Qualification and Parsing
Microarray Ordering
Microarray Synthesis, Error Correction and Elution
Oligo Preparation and Error Correction
Oligo Processing and Ligation
Construct Isolation
Clonal Isolation
Sample Preparation
Sequencing


20 25 30 20
1'
2'
3'
4'
5'
10
25
From Oligo Prep
and Error Correction
40
50 60 30 50
1'
2'
3'
4'
5'
80
To Construct
Amplification
25
90
25
1' 2' 3' 4' 5'
70


130
130
120
120
110
1'
2'
3'
4'
5'
100
From Oligo Prep
and Error Correction
140
150
150
160
150
160
160
150
To cell transformation
and verification
180
1'
2'
3'
4'
5'
170
190


350
360
360
350
1'
2'
3'
4'
5'
310
340
350
360
340
360
350
1'
2'
3'
4'
5'
311
325
340
390
325
326
340
391
326
325
392
326

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGC
GGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATG
CGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATA
CCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTA
CGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTT
GTAAAACGACGGCCAGTGAATTAGTGTTGAGACCATTCAGCTCCGGTCTCGACACTGAGCTTGGCGTAATCATGGTC
ATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGC
CTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGT
CGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTC
GCTCACTGACTCGCTGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTA
TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGG
CCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTG
GCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACC
CTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTA
TCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGC
CTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACA
GGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAG
AACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAC
AAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGAT
CCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCA
AAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTC
TGACAGTCAGAAGAACTCGTCAAGAAGGCGATAGAAGGCGATGCGCTGCGAATCGGGAGCGGCGATACCGTAAAG
CACGAGGAAGCGGTCAGCCCATTCGCCGCCAAGCTCTTCAGCAATATCACGGGTAGCCAACGCTATGTCCTGATAG
CGGTCCGCCACACCCAGCCGGCCACAGTCGATGAATCCAGAAAAGCGGCCATTTTCCACCATGATATTCGGCAAGC
AGGCATCGCCATGGGTCACGACGAGATCCTCGCCGTCGGGCATGCTCGCCTTGAGCCTGGCGAACAGTTCGGCTG
GCGCGAGCCCCTGATGCTCTTCGTCCAGATCATCCTGATCGACAAGACCGGCTTCCATCCGAGTACGTGCTCGCTC
GATGCGATGTTTCGCTTGGTGGTCGAATGGGCAGGTAGCCGGATCAAGCGTATGCAGCCGCCGCATTGCATCAGCC
ATGATGGATACTTTCTCGGCAGGAGCAAGGTGAGATGACAGGAGATCCTGCCCCGGCACTTCGCCCAATAGCAGCC
AGTCCCTTCCCGCTTCAGTGACAACGTCGAGCACAGCTGCGCAAGGAACGCCCGTCGTGGCCAGCCACGATAGCC
GCGCTGCCTCGTCTTGCAGTTCATTCAGGGCACCGGACAGGTCGGTCTTGACAAAAAGAACCGGGCGCCCCTGCG
CTGACAGCCGGAACACGGCGGCATCAGAGCAGCCGATTGTCTGTTGTGCCCAGTCATAGCCGAATAGCCTCTCCAC
CCAAGCGGCCGGAGAACCTGCGTGCAATCCATCTTGTTCAATCATACTCTTCCTTTTTCAATATTATTGAAGCATTTAT
CAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTT
CCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGG
CCCTTTCGTC


200
C1
C2
C3
C4
210
220
230
Next Generation
Sequencing Platform

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/047370

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C40B 50/06 (2013.01)
USPC - 506/26
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12N 15/09, 15/10, 15/66; C12Q 1/68; C40B 50/06, 50/14 (2013.01)
USPC - 435/91.2; 506/2,16, 26

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

CPC - B01J 19/0046, 2219/00659, 2219/00722; C12N 15/1006, 15/1068 (2013.01)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatBase, Google, PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2011/0172127 A1 (JACOBSON et al) 14 July 2011 (14.07.2011) entire document	1-22, 25-27, 29
Y		23, 24, 28
Y	WO 2011/161413 A2 (CHE et al) 29 December 2011 (29.12.2011) entire document	23, 24, 28
A	WO 2012/064975 A1 (JACOBSON et al) 18 May 2012 (18.05.2012) entire document	1-29
A	US 2010/0028885 A1 (BALASUBRAMANIAN et al) 04 February 2010 (04.02.2010) entire document	1-29

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
05 November 2013	20 NOV 2013
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Blaine R. Copenheaver PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2013/047370

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing filed or furnished:

   a. (means)

   ☐ on paper

   ☒ in electronic form

   b. (time)

   ☒ in the international application as filed

   ☐ together with the international application in electronic form

   ☐ subsequently to this Authority for the purposes of search

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments: