

DE NOVO SYNTHESIZED GENE LIBRARIES

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 61/862445, filed August 5, 2013 and U.S. Provisional Application No. 61/862457, filed August 5, 2013, which applications are incorporated herein by reference.

BACKGROUND OF THE INVENTION

[0002] Highly efficient chemical gene synthesis with high fidelity and low cost has a central role in biotechnology and medicine, and in basic biomedical research.

[0003] De novo gene synthesis is a powerful tool for basic biological research and biotechnology applications. While various methods are known for the synthesis of relatively short fragments in a small scale, these techniques suffer from scalability, automation, speed, accuracy, and cost. There is a need for devices for simple, reproducible, scalable, less error-prone and cost-effective methods that guarantee successful synthesis of desired genes and are amenable to automation.

SUMMARY OF THE DISCLOSURE

[0004] As noted above, there exists a pressing need for methods, devices and systems that can quickly synthesize large gene libraries or relatively longer oligonucleotide fragments efficiently with less error. Similarly, there is also a need for methods that can partition and mix liquid reagents in a microfluidic scale for large numbers of individually addressable reactions in parallel. The present disclosure addresses these needs and provides related advantages as well.

[0004a] In one aspect, the present disclosure provides a method for polynucleotide synthesis, comprising:

- a) providing instructions in a computer readable non-transient medium for synthesis of preselected cDNA sequences; and
- b) synthesizing at least 20,000 polynucleotides encoding for the preselected cDNA sequences, wherein each of the at least 20,000 polynucleotides is at least 100 bases in length, wherein each of the at least 20,000 polynucleotides comprises a first overlap region which is complementary to a second overlap region of another polynucleotide of the at least 20,000 polynucleotides such that a plurality of genes are formed when a subset of the at least 20,000 polynucleotides are assembled, wherein the first overlap region comprises at least 10 bases in length, and wherein the at least 20,000 polynucleotides encode sequences with an

aggregate error rate of less than 1 in 1000 bases compared to the preselected cDNA sequences received in the instructions provided in the computer readable non-transient medium without error correction.

[0004b] In one aspect, the present disclosure provides a method for polynucleotide synthesis, comprising:

- a) providing instructions in a computer readable non-transient medium for synthesis of preselected cDNA sequences; and
- b) synthesizing at least 20,000 polynucleotides encoding for the preselected cDNA sequences, wherein each of the at least 20,000 polynucleotides is at least 100 bases in length, wherein each of the at least 20,000 polynucleotides comprises a first overlap region which is complementary to a second overlap region of another polynucleotide of the at least 20,000 polynucleotides such that a plurality of genes are formed when a subset of the at least 20,000 polynucleotides are assembled, wherein the first overlap region comprises at least 10 bases in length, and wherein at least 89% of the at least 20,000 polynucleotides have no errors compared to the preselected cDNA sequences received in the instructions provided in the computer readable non-transient medium without error correction.

[0004c] In one aspect, the present disclosure provides a method of synthesizing nucleic acids, comprising:

- a) providing predetermined sequences for at least 6,000 preselected nucleic acids;
- b) providing a structure comprising a patterned surface, wherein the patterned surface comprises predetermined regions coated with a first molecule, wherein the first molecule comprises a silane, siloxane, or N-(3-triethoxysilylpropyl)-4-hydroxybutyramide, and wherein each predetermined region coated with the first molecule is surrounded by a region coated with a second molecule, wherein the second molecule binds to the patterned surface and lacks an available reactive group that binds to a nucleoside phosphoramidite;
- c) synthesizing at least 800,000 of non-identical polynucleotides attached to the patterned surface, wherein the at least 800,000 non-identical polynucleotides are each at least 125 bases long and extend from the patterned surface at the predetermined regions coated with the first molecule;
- d) releasing the at least 800,000 non-identical polynucleotides from the patterned surface;
- e) suspending the at least 800,000 non-identical polynucleotides in a solution; and
- f) subjecting the solution comprising the at least 800,000 non-identical polynucleotides to a polymerase chain assembly reaction to assemble the at least 6,000 preselected

nucleic acids, wherein the assembled at least 6,000 preselected nucleic acids encode sequences with an aggregate error rate of less than 1 in 1500 bases compared to the predetermined sequences without correcting errors in the assembled at least 6,000 preselected nucleic acids, and wherein the patterned surface provides for the aggregate error rate of less than 1 in 1500 bases.

[0005] In one aspect, the present disclosure provides a gene library as described herein. The gene library comprises a collection of genes. In some embodiments, the collection comprises at least 100 different preselected synthetic genes that can be of at least 0.5 kb length with an error rate of less than 1 in 3000 bp compared to predetermined sequences comprising the genes. In another aspect, the present disclosure also provides a gene library that comprises a collection of genes. The collection may comprise at least 100 different preselected synthetic genes that can be each of at least 0.5 kb length. At least 90% of the preselected synthetic genes may comprise an error rate of less than 1 in 3000 bp compared to predetermined sequences comprising the genes. Desired predetermined sequences may be supplied by any method, typically by a user, e.g. a user entering data using a computerized system. In various embodiments, synthesized nucleic acids are compared against these predetermined sequences, in some cases by sequencing at least a

chemical species on a substrate; a scanning transport for scanning the substrate adjacent to the print head to selectively deposit the microdroplet at specified sites; a flow cell for treating the substrate on which the microdroplet is deposited by exposing the substrate to one or more selected fluids; an alignment unit for aligning the substrate correctly relative to the print head each time when the substrate is positioned adjacent to the print head for deposition; and not comprising a treating transport for moving the substrate between the print head and the flow cell for treatment in the flow cell, wherein said treating transport and said scanning transport are different elements.

[0064] In yet another aspect, the invention relates to an automated system for synthesizing oligonucleotides on a substrate, said automated system capable of processing one or more substrates, comprising: an inkjet print head for spraying a solution comprising a nucleoside or activated nucleoside on a substrate; a scanning transport for scanning the substrate adjacent to the print head to selectively deposit the nucleoside at specified sites; a flow cell for treating the substrate on which the monomer is deposited by exposing the substrate to one or more selected fluids; an alignment unit for aligning the substrate correctly relative to the print head each time when the substrate is positioned adjacent to the print head for deposition; and not comprising a treating transport for moving the substrate between the print head and the flow cell for treatment in the flow cell, wherein said treating transport and said scanning transport are different elements.

[0065] In yet a further aspect, the disclosure relates to an automated system comprising: an inkjet print head for spraying a microdroplet comprising a chemical species on a substrate; a scanning transport for scanning the substrate adjacent to the print head to selectively deposit the microdroplet at specified sites; a flow cell for treating the substrate on which the microdroplet is deposited by exposing the substrate to one or more selected fluids; and an alignment unit for aligning the substrate correctly relative to the print head each time when the substrate is positioned adjacent to the print head for deposition; and wherein the system does NOT comprise a treating transport for moving the substrate between the print head and the flow cell for treatment in the flow cell.

INCORPORATION BY REFERENCE

[0066] All publications, patents, and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication, patent, or patent application was specifically and individually indicated to be incorporated by reference.

WHAT IS CLAIMED IS:

1. A method for polynucleotide synthesis, comprising:

- a) providing instructions in a computer readable non-transient medium for synthesis of preselected cDNA sequences; and
- b) synthesizing at least 20,000 polynucleotides encoding for the preselected cDNA sequences, wherein each of the at least 20,000 polynucleotides is at least 100 bases in length, wherein each of the at least 20,000 polynucleotides comprises a first overlap region which is complementary to a second overlap region of another polynucleotide of the at least 20,000 polynucleotides such that a plurality of genes are formed when a subset of the at least 20,000 polynucleotides are assembled, wherein the first overlap region comprises at least 10 bases in length, and wherein the at least 20,000 polynucleotides encode sequences with an aggregate error rate of less than 1 in 1000 bases compared to the preselected cDNA sequences received in the instructions provided in the computer readable non-transient medium without error correction.

2. A method for polynucleotide synthesis, comprising:

- a) providing instructions in a computer readable non-transient medium for synthesis of preselected cDNA sequences; and
- b) synthesizing at least 20,000 polynucleotides encoding for the preselected cDNA sequences, wherein each of the at least 20,000 polynucleotides is at least 100 bases in length, wherein each of the at least 20,000 polynucleotides comprises a first overlap region which is complementary to a second overlap region of another polynucleotide of the at least 20,000 polynucleotides such that a plurality of genes are formed when a subset of the at least 20,000 polynucleotides are assembled, wherein the first overlap region comprises at least 10 bases in length, and wherein at least 89% of the at least 20,000 polynucleotides have no errors compared to the preselected cDNA sequences received in the instructions provided in the computer readable non-transient medium without error correction.

3. The method of claim 1 or 2, wherein at least 100,000 polynucleotides are synthesized.

4. The method of claim 1 or 2, wherein at least 750 genes are formed when a subset of the at least 20,000 polynucleotides are assembled.

5. The method of claim 1 or 2, wherein at least 10,000 genes are formed when a subset of the at least 20,000 polynucleotides are assembled.
6. The method of claim 1 or 2, wherein the first overlap region comprises a GC content of 35% to 65%.
7. The method of claim 1 or 2, wherein the first overlap region comprises 10 to 100 bases in length.
8. The method of claim 1 or 2, wherein each polynucleotide is 100 to 300 bases in length.
9. The method of claim 1 or 2, wherein each polynucleotide is 100 to 200 bases in length.
10. The method of claim 1 or 2, wherein each of the genes is at least 0.5 kb in length.
11. The method of claim 1 or 2, wherein each of the genes is at least 3 kb in length.
12. The method of claim 1 or 2, wherein the at least 20,000 polynucleotides are synthesized on a structure.
13. The method of claim 12, wherein the structure is a solid support.
14. The method of claim 1 or 2, further comprising hybridizing each polynucleotide in said subset to at least another polynucleotide in said subset to form the plurality of genes.
15. A method of synthesizing nucleic acids, comprising:
 - a. providing predetermined sequences for at least 6,000 preselected nucleic acids;
 - b. providing a structure comprising a patterned surface, wherein the patterned surface comprises predetermined regions coated with a first molecule, wherein the first molecule comprises a silane, siloxane, or N-(3-triethoxysilylpropyl)-4-hydroxybutyramide, and wherein each predetermined region coated with the first molecule is surrounded by a region coated with a second molecule, wherein the second molecule binds to the patterned surface and lacks an available reactive group that binds to a nucleoside phosphoramidite;

c. synthesizing at least 800,000 of non-identical polynucleotides attached to the patterned surface, wherein the at least 800,000 non-identical polynucleotides are each at least 125 bases long and extend from the patterned surface at the predetermined regions coated with the first molecule;

d. releasing the at least 800,000 non-identical polynucleotides from the patterned surface;

e. suspending the at least 800,000 non-identical polynucleotides in a solution; and

f. subjecting the solution comprising the at least 800,000 non-identical polynucleotides to a polymerase chain assembly reaction to assemble the at least 6,000 preselected nucleic acids, wherein the assembled at least 6,000 preselected nucleic acids encode sequences with an aggregate error rate of less than 1 in 1500 bases compared to the predetermined sequences without correcting errors in the assembled at least 6,000 preselected nucleic acids, and wherein the patterned surface provides for the aggregate error rate of less than 1 in 1500 bases.