

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 of synthesizing nucleic acids, comprising:

- a) providing a predetermined sequences for at least 750 preselected nucleic acids;
- b) providing a structure comprising a surface, wherein the surface comprises a plurality of loci, wherein each locus comprises a first plurality of molecules and is adjacent to a region that comprises a second plurality of molecules, wherein each molecule of the first plurality of molecules binds to the surface and comprises a reactive group capable of binding to a nucleoside phosphoramidite, and wherein each molecule of the second plurality of molecules binds to the surface and lacks the reactive group capable of binding to the nucleoside phosphoramidite;

- c) synthesizing at least 30,000 non-identical polynucleotides extending from the surfaces, wherein the at least 30,000 non-identical polynucleotides are each at least 50 bases long; and
- d) assembling the at least 750 preselected nucleic acids, wherein the at least 750 preselected nucleic acids comprise sequences encoded by the at least 30,000 non-identical polynucleotides, wherein the assembled at least 750 preselected nucleic acids encode sequences with an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences without correcting errors in the at least 750 assembled preselected nucleic acids.

[0004b] In one aspect, the present disclosure provides a method for synthesizing a library of oligonucleotides encoding for predetermined sequences, the method comprising:

- a) providing predetermined sequences for a library of at least 100,000 non-identical oligonucleotides;
- b) providing a structure having a patterned surface comprising a plurality of loci for oligonucleotide extension, wherein each locus comprises a first plurality of molecules and is adjacent to a region that comprises a second plurality of molecules, wherein each molecule of the first plurality of molecules binds to the patterned surface and comprises a reactive group capable of binding to a nucleoside, and wherein each molecule of the second plurality of molecules binds to the patterned surface and lacks the reactive group capable of binding to the nucleoside;
- c) adding a droplet of fluid comprising an extension reaction reagent specific to a locus, wherein the extension reaction reagent comprises a nucleoside phosphoramidite;
- d) allowing sufficient time for an extension reaction step to occur; and
- e) repeating steps (c) and (d) to synthesize the at least 100,000 non-identical oligonucleotides, wherein each non-identical oligonucleotide is at least 50 bases in length and is attached to the patterned surface at a predetermined locus, and wherein the synthesized at least 100,000 non-identical oligonucleotides encode sequences with an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences, and wherein the patterned surface provides for the aggregate error rate of less than 1 in 800 bases.

[0004c] In one aspect, the present disclosure provides an oligonucleotide library, comprising at least 30,000 non-identical oligonucleotides that collectively encode sequences for at least 750 preselected genes, each preselected gene having a predetermined sequence, wherein each of the at least 30,000 non-identical oligonucleotides is synthesized and

comprises at least 30 bases in length, and wherein the at least 30,000 non-identical oligonucleotides encode sequences with an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences for the at least 750 preselected genes.

[0004d] In one aspect, the present disclosure provides a gene library, comprising a collection of at least 750 preselected genes, each gene having a predetermined sequence, wherein each of the at least 750 preselected genes is at least 0.5 kb length in length, wherein each preselected gene is synthesized without correcting errors, and wherein the at least 750 preselected genes have an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences for the at least 750 preselected genes.

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

- a) providing predetermined sequences for at least 750 preselected nucleic acids;
- b) providing a structure comprising a surface;
- c) synthesizing at least 30,000 non-identical polynucleotides extending from the surface wherein the at least 30,000 non-identical polynucleotides are each at least 30 bases long; and
- d) assembling the at least 750 preselected nucleic acids, wherein the at least 750 preselected nucleic acids comprise sequences encoded by the at least 30,000 non-identical polynucleotides, wherein the assembled at least 750 preselected nucleic acids encode sequences with an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences without correcting errors in the at least 750 assembled preselected nucleic acids.

[0004f] In one aspect, the present disclosure provides an oligonucleotide library, comprising at least 30,000 non-identical oligonucleotides that collectively encode sequences for at least 750 preselected genes, each preselected gene having a predetermined sequence, and wherein at least 80% of the non-identical oligonucleotides are error free compared to the predetermined sequences for the at least 750 preselected genes.

[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

WHAT IS CLAIMED:

1. A method of synthesizing nucleic acids, comprising:
 - a) providing predetermined sequences for at least 750 preselected nucleic acids;
 - b) providing a structure comprising a surface, wherein the surface comprises a plurality of loci, wherein each locus comprises a first plurality of molecules and is adjacent to a region that comprises a second plurality of molecules, wherein each molecule of the first plurality of molecules binds to the surface and comprises a reactive group capable of binding to a nucleoside phosphoramidite, and wherein each molecule of the second plurality of molecules binds to the surface and lacks the reactive group capable of binding to the nucleoside phosphoramidite;
 - c) synthesizing at least 30,000 non-identical polynucleotides extending from the surface, wherein the at least 30,000 non-identical polynucleotides are each at least 50 bases long; and
 - d) assembling the at least 750 preselected nucleic acids, wherein the at least 750 preselected nucleic acids comprise sequences encoded by the at least 30,000 non-identical polynucleotides, wherein the assembled at least 750 preselected nucleic acids encode sequences with an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences without correcting errors in the at least 750 assembled preselected nucleic acids.
2. The method of claim 1, wherein the first plurality of molecules comprises a silane, siloxane, or N-(3-triethoxysilylpropyl)-4-hydroxybutyramide.
3. The method of claim 1, wherein the first plurality of molecules comprises 11-acetoxyundecyltriethoxysilane, n-decyltriethoxysilane, (3-aminopropyl)trimethoxysilane, or (3-aminopropyl)triethoxysilane.
4. The method of claim 1, wherein the second plurality of molecules comprises a fluorosilane.
5. The method of claim 4, wherein the fluorosilane is (tridecafluorotetrahydrooctyl)-triethoxysilane.
6. The method of claim 1, wherein the at least 30,000 non-identical polynucleotides attached to the surface are clustered, wherein each cluster is separated from another cluster by a structural barrier on the surface, and wherein each cluster includes a subset of the at least 30,000 non-identical polynucleotides that collectively encodes for a single nucleic acid.
7. The method of claim 6, wherein each cluster comprises 50 to 500 loci.

8. The method of claim 7, wherein each locus is separated from another locus by a structural barrier on the surface.
9. The method of claim 6, wherein each subset of the at least 30,000 non-identical polynucleotides collectively encodes for a single nucleic acid and is located within a channel 1.0 to 1.5 mm in diameter.
10. The method of claim 1, comprising releasing the at least 30,000 non-identical polynucleotides release from the surface after synthesis, wherein the at least 30,000 non-identical polynucleotides are released and clustered, wherein each cluster includes a subset of the at least 30,000 non-identical polynucleotides, and wherein the subset of the at least 30,000 non-identical nucleotides collectively encodes for a single nucleic acid.
11. The method of claim 1, comprising releasing the at least 30,000 non-identical polynucleotides from the surface after synthesis, wherein the release is by gas cleavage.
12. The method of claim 1, wherein steps a) to d) are completed in less than one week.
13. The method of claim 1, wherein steps a) to d) are completed in less than 48 hours.
14. The method of claim 1, wherein the assembled at least 750 preselected nucleic acids are at least 1 kb long on average.
15. The method of claim 1, further comprising treating the assembled at least 750 preselected nucleic acids with an error correction enzyme.
16. 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.

17. A method for synthesizing a library of oligonucleotides encoding for predetermined sequences, the method comprising:

a) providing predetermined sequences for a library of at least 100,000 non-identical oligonucleotides;

b) providing a structure having a patterned surface comprising a plurality of loci for oligonucleotide extension, wherein each locus comprises a first plurality of molecules and is adjacent to a region that comprises a second plurality of molecules, wherein each molecule of the first plurality of molecules binds to the patterned surface and comprises a reactive group capable of binding to a nucleoside, and wherein each molecule of the second plurality of molecules binds to the patterned surface and lacks the reactive group capable of binding to the nucleoside;

c) adding a droplet of fluid comprising an extension reaction reagent specific to a locus, wherein the extension reaction reagent comprises a nucleoside phosphoramidite;

d) allowing sufficient time for an extension reaction step to occur; and

e) repeating steps (c) and (d) to synthesize the at least 100,000 non-identical oligonucleotides, wherein each non-identical oligonucleotide is at least 50 bases in length and is attached to the patterned surface at a predetermined locus, and wherein the synthesized at least 100,000 non-identical oligonucleotides encode sequences with an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences, and wherein the patterned surface provides for the aggregate error rate of less than 1 in 800 bases.

18. The method of claim 17, wherein the first plurality of molecules comprises a silane, siloxane, or N-(3-triethoxysilylpropyl)-4-hydroxybutyramide.

19. The method of claim 17, wherein the first plurality of molecules comprises 11-acetoxyundecyltriethoxysilane, n-decyltriethoxysilane, or (3-aminopropyl)trimethoxysilane.
20. The method of claim 17, wherein the second plurality of molecules comprises a fluorosilane.
21. The method of claim 20, wherein the fluorosilane is (tridecafluorotetrahydrooctyl)-triethoxysilane.
22. The method of claim 17, wherein each locus of the plurality of loci is separated from another locus by a structural barrier of the patterned surface.
23. The method of claim 17, wherein steps (a) to (e) are completed in less than 1 month, 14 days, 7 days, or 2 days.
24. The method of claim 17, wherein each of the at least 100,000 non-identical oligonucleotides comprises a region which is complementary to another of the at least 100,000 non-identical oligonucleotides.
25. The method of claim 17, wherein the at least 100,000 non-identical oligonucleotides encode sequences with an aggregate error rate of less than 1 in 1000 bases compared to the predetermined sequences.
26. The method of claim 17, wherein each non-identical oligonucleotide is at least 100 bases in length.
27. An oligonucleotide library, comprising at least 30,000 non-identical oligonucleotides that collectively encode sequences for at least 750 preselected genes, each preselected gene having a predetermined sequence, wherein each of the at least 30,000 non-identical oligonucleotides is synthesized and comprises at least 30 bases in length, and wherein the at least 30,000 non-identical oligonucleotides encode sequences with an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences for the at least 750 preselected genes.
28. A gene library, comprising a collection of at least 750 preselected genes, each gene having a predetermined sequence, wherein each of the at least 750 preselected genes is at least 0.5 kb length in length, wherein each preselected gene is synthesized without correcting errors, and wherein the at least 750 preselected

genes have an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences for the at least 750 preselected genes.

29. The method of claim 1, comprising releasing the at least 30,000 non-identical polynucleotides from the surface after synthesis, wherein the method comprises no amplification step prior to releasing the at least 30,000 non-identical polynucleotides from the surface.

30. The method of claim 1, wherein the at least 30,000 different oligonucleotides encode sequences with an error rate of less than 1 in 1000 bases compared to predetermined sequences for the at least 30,000 different oligonucleotides.

31. The method of claim 1, wherein the at least 30,000 different oligonucleotides encode sequences with an error rate of less than 1 in 1500 bases compared to predetermined sequences for the at least 30,000 different oligonucleotides.

32. The method of claim 18, wherein the silane is an aminosilane.

33. The method of claim 17, comprising releasing the at least 100,000 non-identical polynucleotides from the surface after synthesis, wherein the method comprises no amplification step prior to releasing the at least 100,000 non-identical polynucleotides from the surface.

34. The oligonucleotide library of claim 27, wherein the at least 30,000 different oligonucleotides encode sequences with an error rate of less than 1 in 1000 bases compared to predetermined sequences for the at least 30,000 different oligonucleotides.

35. The oligonucleotide library of claim 27, wherein each of the at least 30,000 different oligonucleotides comprises a region which is complementary to another of the at least 30,000 different oligonucleotides.

36. The oligonucleotide library of claim 27, wherein the library comprises at least 600,000 different oligonucleotides.

37. The oligonucleotide library of claim 27, wherein each of the at least 30,000 non-identical polynucleotides comprises a first overlap region which is

- complementary to a second overlap region of another non-identical polynucleotide of the at least 30,000 non-identical polynucleotides.
38. The oligonucleotide library of claim 27, wherein each of the polynucleotides varying within a range of lengths are of a different length.
39. The oligonucleotide library of claim 27, wherein the polynucleotides varying within a range of lengths are of at least 10 different lengths.
40. The oligonucleotide library of claim 27, wherein the polynucleotides varying within a range of lengths are of at least 30 different lengths.
41. A method of synthesizing nucleic acids, comprising:
- a) providing predetermined sequences for at least 750 preselected nucleic acids;
 - b) providing a structure comprising a surface;
 - c) synthesizing at least 30,000 non-identical polynucleotides extending from the surface, wherein the at least 30,000 non-identical polynucleotides are each at least 30 bases long; and
 - d) assembling the at least 750 preselected nucleic acids, wherein the at least 750 preselected nucleic acids comprise sequences encoded by the at least 30,000 non-identical polynucleotides, wherein the assembled at least 750 preselected nucleic acids encode sequences with an aggregate error rate of less than 1 in 800 bases compared to the predetermined sequences without correcting errors in the at least 750 assembled preselected nucleic acids.
42. An oligonucleotide library, comprising at least 30,000 non-identical oligonucleotides that collectively encode sequences for at least 750 preselected genes, each preselected gene having a predetermined sequence, and wherein at least 80% of the non-identical oligonucleotides are error free compared to the predetermined sequences for the at least 750 preselected genes.