

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 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

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.

[0004b] 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.

[0004c] 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.

[0004d] 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

WHAT IS CLAIMED IS:

1. 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.
2. 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 1000 bases compared to the predetermined sequences for the at least 750 preselected genes.

3. 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.
4. The oligonucleotide library of claim 2, 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.
5. The oligonucleotide library of claim 2, 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.
6. The oligonucleotide library of claim 2, wherein the library comprises at least 600,000 different oligonucleotides.
7. The oligonucleotide library of claim 2, 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.
8. The oligonucleotide library of claim 2, wherein each of the polynucleotides varying within a range of lengths are of a different length.
9. The oligonucleotide library of claim 2, wherein the polynucleotides varying within a range of lengths are of at least 10 different lengths.
10. The oligonucleotide library of claim 2, wherein the polynucleotides varying within a range of lengths are of at least 30 different lengths.
11. 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.
12. The gene library of claim 3, wherein the at least 750 preselected genes have an aggregate error rate of less than 1 in 1000 bases compared to the predetermined sequences for the at least 750 preselected genes.