BARCODE-BASED NUCLEIC ACID SEQUENCE ASSEMBLY

CROSS-REFERENCE

[0001] This application claims the benefit of U.S. Provisional Application No. 62/865,094, filed June 21, 2019, which application is incorporated herein by reference.

BACKGROUND

[0002] De novo nucleic acid synthesis is a powerful tool for basic biological research and biotechnology applications. While various methods are known for the synthesis of relatively short fragments of nucleic acids on a small scale, these techniques suffer from scalability, automation, speed, accuracy, and cost. Thus, a need remains for efficient methods of variant nucleic acid assembly.

BRIEF SUMMARY

[0002a] In one aspect, the present disclosure refers to a method for nucleic acid assembly, comprising:

(a) providing a first plurality of polynucleotides, comprising up to 2500 bases in length, wherein each polynucleotide of the first plurality of polynucleotides comprises a first terminal region of sequence homology;

(b) providing a second plurality of polynucleotides, comprising up to 2500 bases in length, wherein each polynucleotide of the second plurality of polynucleotides comprises a second terminal region of sequence homology to the first terminal region of sequence homology; and

(c) contacting the first plurality of polynucleotides and the second plurality of polynucleotides with a reaction mixture comprising an exonuclease, a flap endonuclease 1, a polymerase, and a ligase to assemble a library of nucleic acids, wherein at least 80% of the nucleic acids are each present in the library in an amount within 2x of a mean frequency for each of the nucleic acids in the library.

[0002b] In another aspect, the present disclosure refers to a method for nucleic acid assembly, comprising:

(a) de novo synthesizing a first nucleic acid comprising in 5’ to 3’ order: a barcode sequence, a first restriction endonuclease site, a second restriction endonuclease site, and a first hypervariable region sequence;

(b) de novo synthesizing a second nucleic acid comprising in 5’ to 3’ order: a first region of any defined length sequence, a self-cleaving peptide sequence, a first complementary region adjacent to a first variable region sequence, and a first variable region sequence;

(c) contacting the first nucleic acid and the second nucleic to generate a third nucleic acid;

(d) providing a fourth nucleic acid comprising in 5’ to 3’ order: a vector sequence, a second complementary region adjacent to a second variable region sequence, a second variable region sequence, a second hypervariable region sequence, the first restriction endonuclease site, and the barcode sequence;

(e) contacting the third nucleic acid and the fourth nucleic acid with a restriction endonuclease; and

(f) assembling the third nucleic acid and the fourth nucleic acid using a reaction mixture comprising one or more enzymes, comprising an exonuclease, a flap endonuclease 1, a polymerase, and a ligase.

[0002c] In another aspect, the present disclosure refers to a method for nucleic acid assembly, comprising:

(a) de novo synthesizing a first nucleic acid comprising a first variable region sequence;

(b) de novo synthesizing a second nucleic acid comprising a second variable region sequence;

(c) de novo synthesizing a third nucleic acid comprising in 5’ to 3’ order: a first region of fixed variability sequence, a first region of any defined length sequence, a self-cleaving peptide sequence, a first complementary region adjacent to a first variable region sequence, and a second region of fixed variability sequence; and

(d) contacting the first nucleic acid, the second nucleic acid, and the third nucleic acid with a reaction mixture comprising an exonuclease, an endonuclease, a polymerase, and a ligase,

wherein the endonuclease is a flap endonuclease 1.

[0003] Provided herein are methods for nucleic acid assembly, comprising: (a) providing a first plurality of polynucleotides, wherein each polynucleotide of the first plurality of polynucleotides comprises a first terminal region of sequence homology; (b) providing a second plurality of polynucleotides, wherein each polynucleotide of the second plurality of polynucleotides comprises a second terminal region of sequence homology to the first

terminal region of sequence homology; and (c) contacting the first plurality of polynucleotides and the second plurality of polynucleotides with a reaction mixture comprising an exonuclease, an endonuclease, a polymerase, and a ligase to assemble a library of nucleic acids, wherein at least 80% of the nucleic acids are each present in the library in an amount within 2x of a mean frequency for each of the nucleic acids in the library. Further provided herein are methods, wherein the first plurality of polynucleotides comprises up to 100 different sequences. Further provided herein are methods, wherein the second plurality of polynucleotides comprises up to 100 different sequences. Further provided herein are methods, wherein at least 10,000 nucleic acids are assembled. Further provided herein are methods, wherein at least 100,000 nucleic acids are assembled. Further provided herein are methods, wherein each polynucleotide of the first plurality of polynucleotides comprises up to 2500 bases in length. Further provided herein are methods, wherein each polynucleotide of the second plurality of polynucleotides comprises up to 2500 bases in length. Further provided herein are methods, wherein the exonuclease is exonuclease III. Further provided herein are methods, wherein the endonuclease is a flap endonuclease. Further provided