construction, requirement of high volumes of phosphoramidites, an expensive raw material, and production of nanomole amounts of the final product, significantly less than required for downstream steps, and a large number of separate polynucleotides required one 96 well plate to set up the synthesis of one gene.

[0006] Synthesizing of polynucleotides on microarrays provided a significant increase in the throughput of gene synthesis. A large number of polynucleotides could be synthesized on the microarray surface, then cleaved off and pooled together. Each polynucleotide destined for a specific gene contains a unique barcode sequence that enabled that specific subpopulation of polynucleotides to be depooled and assembled into the gene of interest. In this phase of the process, each subpool is transferred into one well in a 96 well plate, increasing throughput to 96 genes. While this is two orders of magnitude higher in throughput than the classical method, it still does not adequately support the design, build, test cycles that require thousands of sequences at one time due to a lack of cost efficiency and slow turnaround times.

BRIEF SUMMARY

[0006a] In one aspect, provided herein is a method of synthesizing a variant nucleic acid library, comprising:

a. providing predetermined sequences encoding for at least 500 polynucleotide sequences, wherein the at least 500 polynucleotide sequences have a preselected codon distribution;
b. de novo synthesizing a plurality of polynucleotides using an array encoding for the at least 500 polynucleotide sequences;
c. assaying an activity for nucleic acids encoded by or proteins translated based on the plurality of polynucleotides; and
d. collecting results from the assay in step (c), wherein the collecting comprises collecting results of predetermined sequences associated with a negative or null result, wherein at least about 70% of a predicted diversity is represented.

[0006b] In one aspect, provided herein is a method of synthesizing a variant nucleic acid library, comprising:

a. providing predetermined sequences encoding for a plurality of polynucleotides, wherein the polynucleotides encode for a plurality of codons having a variant sequence compared to a single reference sequence;
b. selecting a distribution value for codons at a preselected position in the reference sequence;
c. providing machine instructions to randomly generate a set of nucleic acid sequences with a distribution value that aligns with the selected distribution value, wherein the set

of nucleic acid sequences is less than the amount of nucleic acid sequences required to generate a saturating codon variant library;

d. calculating a probability of variants of the codons at the preselected position; and

e. de novo synthesizing the variant nucleic acid library with a preselected distribution when the probability aligns with the selected distribution value using an array, and wherein at least about 70% of a predicted diversity is represented.

[0007] Provided herein are methods of synthesizing a variant nucleic acid library, comprising: (a) providing predetermined sequences encoding for at least 500 polynucleotide sequences, wherein the at least 500 polynucleotide sequences have a preselected codon distribution; (b) synthesizing a plurality of polynucleotides encoding for the at least 500 polynucleotide sequences; (c) assaying an activity for nucleic acids encoded by or proteins translated based on the plurality of polynucleotides; and (d) collecting results from the assay in step (c), wherein the collecting comprises collecting results of predetermined sequences associated with a negative or null result. Further provided herein are methods of synthesizing a variant nucleic acid library, wherein step (d) comprises collecting results for at least 80% of the predetermined sequences. Further provided herein are methods of synthesizing a variant nucleic acid library, wherein step (d) comprises collecting results for at least 90% of the predetermined sequences. Further provided herein are methods of synthesizing a variant nucleic acid library, wherein step (d) comprises collecting results for at least 100% of the predetermined sequences. Further provided herein are methods of synthesizing a variant nucleic acid library, wherein at least about 70% of a predicted diversity is represented. Further provided herein are methods of synthesizing a variant nucleic acid library, wherein at least about 90% of a predicted diversity is represented. Further provided herein are methods of synthesizing a variant nucleic acid library, wherein at least about 95% of a predicted diversity is represented. Further provided herein are methods of synthesizing a variant nucleic acid library, wherein at least 80% of the at least 500 polynucleotide sequences are a correct size. Further provided herein are methods of synthesizing a variant nucleic acid library, wherein at least

CLAIMS

WHAT IS CLAIMED IS:

1. A method of synthesizing a variant nucleic acid library, comprising:
    a. providing predetermined sequences encoding for at least 500 polynucleotide sequences, wherein the at least 500 polynucleotide sequences have a preselected codon distribution;
    b. de novo synthesizing a plurality of polynucleotides using an array encoding for the at least 500 polynucleotide sequences;
    c. assaying an activity for nucleic acids encoded by or proteins translated based on the plurality of polynucleotides; and
    d. collecting results from the assay in step (c), wherein the collecting comprises collecting results of predetermined sequences associated with a negative or null result, wherein at least about 70% of a predicted diversity is represented.
2. The method of claim 1, wherein step (d) comprises collecting results for at least 80% of the predetermined sequences.
3. The method of claim 1, wherein at least about 80% of the at least 500 polynucleotide sequences are each present in the variant nucleic acid library in an amount within 2X of a mean frequency for each of the polynucleotide sequences in the library.
4. The method of claim 1, wherein the activity is cellular activity.
5. The method of claim 4, wherein the cellular activity comprises reproduction, growth, adhesion, death, migration, energy production, oxygen utilization, metabolic activity, cell signaling, response to free radical damage, or any combination thereof.
6. The method of claim 1, wherein the variant nucleic acid library encodes sequences for variant genes or fragments thereof.
7. The method of claim 1, wherein the variant nucleic acid library encodes for at least a portion of an antibody, an enzyme, or a peptide.
8. A method of synthesizing a variant nucleic acid library, comprising:
    a. providing predetermined sequences encoding for a plurality of polynucleotides, wherein the polynucleotides encode for a plurality of codons having a variant sequence compared to a single reference sequence;
    b. selecting a distribution value for codons at a preselected position in the reference sequence;
    c. providing machine instructions to randomly generate a set of nucleic acid sequences with a distribution value that aligns with the selected distribution value, wherein the set

of nucleic acid sequences is less than the amount of nucleic acid sequences required to generate a saturating codon variant library;

d. calculating a probability of variants of the codons at the preselected position; and

e. de novo synthesizing the variant nucleic acid library with a preselected distribution when the probability aligns with the selected distribution value using an array, and wherein at least about 70% of a predicted diversity is represented.

9. The method of claim 8, wherein the variant nucleic acid library when translated encodes for a protein library.

10. The method of claim 8, further comprising performing PCR mutagenesis of a nucleic acid using the variant nucleic acid library as primers for a PCR mutagenesis reaction.

11. The method of claim 8, wherein a codon assignment is used for determining each codon of the plurality of codons having a variant sequence.

12. The method of claim 11, wherein the codon assignment is based on frequency of the codon sequence in an organism.

13. The method of claim 12, wherein the organism is an animal, a plant, a fungus, a protist, an archaeon, a bacterium, or a combination of any of the foregoing.

14. The method of claim 11, wherein the codon assignment is based on a diversity of the codon sequence.