A LIBRARY OF POLYNUCLEOTIDES

FIELD OF THE INVENTION

The invention relates to a library of double stranded (ds) polynucleotide library members of at least 50 bp length comprising a variety of polynucleotide sequences, and a method for synthesizing a double stranded (ds) polynucleotide library using a diverse library of oligonucleotides.

BACKGROUND OF THE INVENTION

Artificial synthesis of polynucleotides is currently achieved through two kinds of methods that are not necessarily exclusive:

The first class of methods for the synthesis of polynucleotides is “chemical synthesis”. This is a process through which single stranded DNA (or RNA) molecules are built by sequentially linking nucleotides, one by one, using phosphoramidite chemistry (Beaucage and Caruthers, 1981). This method allows for building of DNA molecules that have specific, predetermined template sequences of any complexity. Chemical methods are popular due to their inexpensive nature, are easily parallelizable and in some implementations allow for high-throughput production of DNA or RNA in chips (LeProust et al., 2010). The main and utmost disadvantage of these methods is that the yield of the reaction decreases dramatically with the length of the template being synthesized, limiting the size of the molecules, typically, to roughly 200 base-pairs (bp, or bps).

The second class of methods for DNA synthesis are the “assembly methods”, which consist of biochemically joining oligonucleotides and polynucleotides of different sizes and of varying sequences in specific ways in order to obtain a larger molecule that has the desired target sequence. The source of these oligonucleotides is often chemical synthesis, but can also be products of enzymatic digestions of naturally occurring DNAs. These assembly methods are often commercialized under the product name “Gene Synthesis”, a term that is a metonym for the synthesis of large polynucleotide chains (1K-5K bp), but not necessarily of gene-size length. There are several approaches reported in the literature for assembling smaller polynucleotides into the target sequence (Stemmer et al., 1995; Smith et al., 2003; Engler et al., 2008; Gibson et al., 2009; Horspool 2010).

In the past few years “Gibson Assembly” (Gibson et al., 2009) has become a popular method for linking several linear ds DNA fragments (size ranging from about 30 bp up to several Kbp). The method consists of joining many ds DNA fragments that have pairwise overlapping sequence homology. The overlapping homology region between fragments can range between about 15 to 80 bp. No overhangs are necessary, since the enzymatic machinery of the method takes care of producing the overhangs, fill in the gaps and correctly ligate the fragments. This enzymatic machinery makes use of three enzymes: T5 exonuclease, Phusion DNA polymerase and Taq DNA ligase, all in an isothermal reaction. The method is simple and versatile and can produce both linear and circular ds DNA products. The downside of this method is its limitation for automation making it unsuitable for large-scale commercial use.

The common theme in building DNA molecules of thousands of base pairs is to chemically synthesize small fragments of up to few hundred nt or bp and then concatenate these together by cloning, ligation, PCA or Gibson assembly.

Some approaches are suggestive of pre-constructing, possibly through chemical synthesis, a library of oligonucleotides that covers the possible genetic space, or a required subset of it.

Chari and Church propose using synthesized oligonucleotides (200 bases) to produce short DNA fragments and assembly into large DNA segments using in vivo homologous recombination in yeast and *E. coli* (Chari and Church, 2017).

WO 2009/138954 A2 discloses a method for synthesis of larger polynucleotides by solid phase assembly, wherein defined subunits required for assembly of the larger polynucleotide are chemically synthesized according to need.

Pedersen et al. (US2016/0215316A1) propose using a library comprising the space of all possible hexamers (N=4,096 oligos). The six base pair long oligos are then assembled using oligo linkers to form polynucleotides. There are certain limitations pertaining to the concatenation of the oligonucleotides and large-scale DNA synthesis. Because it takes a suitably designed library and manual protocols such as for cloning, employing large volumes of reagents, methods are time consuming. These in turn add substantial costs to the price of synthesis, which increases per bp as the target sequence length increases.

WO2002/081490 discloses an approach utilizing the results of genomic sequence information by computer-directed polynucleotide assembly based upon information

available in databases such as the human genome database. Specifically, it discloses a method of producing a target polynucleotide wherein the target polynucleotide is parsed into a series of contiguous oligonucleotides by a computer program and said target polynucleotide is generated by sequentially adding *de novo* synthesized oligonucleotides to an initiating oligonucleotide in a uni- or bidirectional manner.

WO2004/033619 also discloses an approach utilizing the results of genomic sequence information for computer-directed polynucleotide assembly.

WO99/14318 discloses a method of producing a target polynucleotide using overlapping pairs of oligos having complementary sequences and overhangs. The oligos are annealed sequentially to produce double stranded DNA segments.

WO2019/073072 discloses a method of synthesizing a double stranded polynucleotide having a predefined sequence using a diverse library of short oligonucleotides.

WO2013/017950 discloses a method for the assembly and cloning of a polynucleotide using a method comprising sequential assembly of polynucleotide molecules on a solid support.

WO2012/084923 discloses a library holding fragments of polynucleotides of varying lengths to identify fragments having improved properties.

Although the last few years have seen considerable progress in the techniques for synthesizing DNA, there are still severe restrictions in terms of volume, throughput, purity and, specially, length of DNA.

SUMMARY OF THE INVENTION

It is the object of the present invention to provide an improved method and tools to synthesize a variety double stranded (ds) polynucleotides.

The object is solved by the subject of the present claims, and as further described herein.

According to the invention, there is provided a library of double stranded (ds) polynucleotide library members of at least 12 bp length comprising a variety of polynucleotide core sequences and the same overhangs.

Disclosed herein is a library of double stranded (ds) polynucleotide library members of at least 12 bp length comprising a variety of core sequences and each ds oligo library member further comprising a first overhang sequence and a second

overhang sequence which are the same in each ds oligo library member, wherein the first and second overhang sequences are not complementary to each other polynucleotide core sequences and the same overhangs, wherein the the core sequence is the double stranded part of the library members, and wherein the variation of the core sequences comprises one or more random point mutations.

In some embodiments, the library of double stranded (ds) polynucleotide library members of at least 12 bp length comprising a variety of polynucleotide core sequences and the same overhangs, wherein the variation comprises one or more random point mutations.

Specifically, said overhangs are not identical to each other.

Specifically, said overhangs are not complementary to each other.

Specifically, each of the library members comprises an identical first and an identical second overhang sequence. Specifically, said first and second overhang sequences are not complementary to each other.

According to a specific embodiment, said overhangs are on both, the leading and lagging strand, and wherein each library member comprises

a) the same first overhang sequence which is the 5’ overhang of the leading strand, and the same second overhang sequence which is the 5’ overhang of the lagging strand; or

b) the same first overhang sequence which is the 3’ overhang of the leading strand, and the same second overhang sequence which is the 3’ overhang of the lagging strand;

wherein said first and second overhang sequences are not complementary to each other.

Advantageously, the variety of ds polynucleotides described herein can be pooled and processed together, since they comprise the same overhangs which are not complementary. Thereby, annealing and ligating of the different library members is avoided.

Specifically, the length of each of the library members is at least any one of 12, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290 or 300 bp, up to any one of 320, 350, 380, 400, 500, 600, 700, 800, 900, 1.000 or 2.000 bp or even more.

Specifically, the length of each of the library members is the same, or variable, or may have a length which is identical to the length of a template +/- 5, 10, 15, or 20 bp.

Specifically, said polynucleotide is a DNA molecule.

Specifically, said overhangs are the same within the library.

Specifically, each of said library members has the same 5' or 3' overhang of the leading strand.

Specifically, each of said library members has the same 5' or 3' overhang of the lagging strand.

Specifically, said overhang of the leading strand differs from said overhang of the lagging strand.

According to a specific embodiment, said leading strand comprises a 5' overhang and said lagging strand comprises a 5' overhang.

According to another specific embodiment, said leading strand comprises a 3' overhang and said lagging strand comprises a 3' overhang.

According to yet another specific embodiment, said leading strand comprises a 5' and a 3' overhang. In such case, the lagging strand does not comprises an overhang.

According to another specific embodiment, said lagging strand comprises a 5' and a 3' overhang. In such case, the leading strand does not comprises an overhang.

For the purpose of comparing the sequences of the (or all) library members, specifically, only one template sequence is used.

Specifically, said core sequences of the library members comprise at least one mutation compared to each other, or compared to a template sequence, thereby producing a diversity. Specifically, said mutation is a point mutation, in particular a mutation which differentiates one library member from another one, and/or which differentiates a library member from a template.

Specifically, the number of point mutations within the polynucleotide sequence is limited e.g., wherein the number is 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10, as compared to a template sequence.

Specifically, the 5' and 3' overhangs have a length of at least any one of 4, 5, 6, 7, or 8 nucleotides, preferably up to any one of 15, 14, 13, 12, 11, 10, 9, or 8 nucleotides, preferably 4-8 nucleotides.

Specifically, each of said library members comprises an identical modification selected from the group consisting of phosphorylation, methylation, biotinylation, or linkage to a fluorophore or quencher.

Specifically, each of said library members is immobilized, preferably by binding only one of the 5' and 3' overhangs to a solid carrier.

Specifically, said library members are contained in one library containment, or in a plurality of spatially distinct library containments.

According to a specific embodiment, said library members are comprised in a mixture which is contained in one library containment.

According to a specific embodiment, said library members are provided in an array wherein each of the library members is contained in spatially distinct library containments.

Specifically, each of said library members has a sequence, which is at least any one of 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99% identical to a template. Preferably, each of said library members has sequence identity of at least any one of 30, 31, 32, 33, 34, or 35% compared to a template of the same length as the library member.

Specifically, said template has a predefined length and sequence of interest.

Disclosed herein is the use of a library as defined herein, in a method of synthesizing a variety of target ds polynucleotides each being longer than the library members by assembling said library members with ds oligos that have overhang sequences matching the overhang sequences of the library members, thereby obtaining a library of target ds polynucleotides.

Disclosed herein is a method for producing the library as defined herein, comprising the steps: a) providing a template nucleotide sequence; b) providing a library of single stranded (ss) oligos in separate containments, said ss oligos comprising a variety of sequences comprising one or more point mutations compared to the template; c) selecting pairs of matching ss oligos from said library of ss oligos; d) partially annealing said pairs of matching ss oligos, thereby obtaining a first library of ds oligonucleotides (ds oligos) comprising ds oligo library members, wherein the ds oligo library members comprise a variety of core sequences, and each ds oligo library member

further comprises a first overhang sequence and a second overhang sequence which are the same for each ds oligo library member, wherein the first and second overhang sequences are not complementary to each other; and e) further annealing with ds oligos that have overhangs matching the overhangs of the ds oligo library members of the first library of ds oligos, thereby obtaining a library of ds polynucleotides.

In some embodiments, the method for producing the library described herein, comprising the steps:

a) providing a template nucleotide sequence; and

b) synthesizing a variety of double stranded (ds) polynucleotides of at least 12 bp length comprising a diversity of the core sequence and comprising the same, non-complementary overhangs, wherein each of said ds polynucleotides is at least 30% identical to said template, thereby obtaining a library of ds polynucleotide library members.

In some embodiments, the method for producing the library described herein, comprising the steps:

a) providing a template nucleotide sequence; and

b) synthesizing a variety of double stranded (ds) polynucleotides of at least 12 bp length comprising a diversity of the core sequence and comprising the same (identical in all library members), non-complementary (meaning the overhang of the leading strand is not complementary to the overhang of the lagging strand) overhangs, wherein each of said ds polynucleotides is at least 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99 % identical to said template, thereby obtaining a library of ds polynucleotide library members. Specifically, each of said ds polynucleotides comprises the same overhangs. Preferably, each of said ds polynucleotides comprises

a) the same first overhang sequence is the 5’ overhang of the leading strand, and the same second overhang sequence which is the 5’ overhang of the lagging strand; or

b) the same first overhang sequence is the 3’ overhang of the leading strand, and the same second overhang sequence which is the 3’ overhang of the lagging strand;

wherein said first and second overhang sequences are not complementary to each other. Specific embodiments employ modification, e.g. conjugation, or labelling, or

immobilization of the library members, before enriching or isolating the library members.Specifically, the ds polynucleotides are enriched by polymerase chain reaction (PCR), thereby obtaining copies of the ds polynucleotides comprising said overhang sequences.

In some embodiments, the method of producing a library as defined herein, which is enriched in a predetermined library member which is a ds polynucleotide consisting of a first and a complementary second strand each comprising a polynucleotide core sequence and an overhang, by (i) amplifying the predetermined library member by an enzymatic reaction employing a polymerase to produce amplification products, and:

a) a set of two primer pairs, comprising: i. a first primer pair comprising a forward primer which is complementary to at least the overhang of the leading strand, and a reverse primer, which is complementary to the 3' terminal sequence of the core sequence of the leading strand; and ii. a second primer pair comprising a forward primer which is complementary to the sequence of the overhang of the lagging strand, and a reverse primer, which is complementary to the terminal sequence of the core sequence of the lagging strand; or

b) a set of two primer pairs, comprising: i. a first primer pair comprising a forward primer which is complementary to at least the sequence of the core sequence of the leading strand, and a reverse primer, which is complementary to the overhang of the leading strand; and ii. a second primer pair comprising a forward primer which is complementary to at least the sequence of the core sequence of the lagging strand, and a reverse primer, which is complementary to the overhang of the lagging strand; and

(ii) producing and optionally isolating said amplification products; and (iii) producing the library enriched in said amplification products.

Specifically, said ds polynucleotides are enriched in said library thereby increasing the number of ds polynucleotide molecules that are characterized by one or more of the (distinguishing) features of a library member as further described herein, over the number of ds polynucleotide molecules which do not have such features. In particular, enrichment is by amplification of the library member sequences.

Specifically, library members are enriched by an amplification method e.g. employing an enzymatic rection employing a polymerase, such as polymerase chain reaction (PCR).

According to a specific embodiment, library members are enriched by PCR using primer pairs, in particular at least two different sets of primer pairs.

Specifically, polynucleotide library members described herein are enriched using

a) a first primer pair comprising a forward primer complementary to at least the overhang of the first strand, and a reverse primer complementary to the terminal sequence of the core sequence of the second strand, excluding its overhang; and

b) a second primer pair comprising a forward primer complementary to at least the terminal sequence of the core sequence of the first strand, excluding its overhang, and a reverse primer complementary to at least the overhang of the second strand; and

producing and optionally isolating the amplification products which comprise an overhang on both strands.

According to a specific example, a first and a second primer pair is used in the same amplification reaction, which are e.g. embodiment A or embodiment B.

According to embodiment A:

a) A first primer pair comprises a forward primer which is complementary to at least the overhang of the leading strand, and a reverse primer, which is complementary to the 3' terminal sequence of the core sequence of the leading strand, thereby, excluding the overhang of the lagging strand. In other words, the reverse primer is hybridizing to the part of the polynucleotide sequence which starts with the last nucleotide of the 3' terminus of the leading strand.

b) A second primer pair comprises a forward primer which is complementary to the sequence of the overhang of the lagging strand, however, excluding the overhang of the leading strand, and a reverse primer, which is complementary to the terminal sequence of the core sequence of the lagging strand, thereby, excluding the overhang of the leading strand. In other words, the reverse primer is hybridizing to the part of the polynucleotide sequence which starts with the last nucleotide of the 3' terminus of the lagging strand.

According to embodiment B:

a) A first primer pair comprises a forward primer which is complementary to at least the sequence of the core sequence of the leading strand, thereby, excluding the overhang of the lagging strand, and a reverse primer, which is complementary to the overhang of the leading strand. In other words, the forward primer is hybridizing to the

part of the polynucleotide sequence which starts with the first nucleotide of the leading strand.

b) A second primer pair comprises a forward primer which is complementary to at least the sequence of the core sequence of the lagging strand, thereby, excluding the overhang of the leading strand, and a reverse primer, which is complementary to the overhang of the lagging strand. In other words, the forward primer is hybridizing to the part of the polynucleotide sequence which starts with the first nucleotide of the lagging strand.

By using such first and second primer pair according to either embodiment A or B, a mixture of amplification products is produced, wherein about 20, 21, 22, 23, 24, or 25% of the amplified sequences produce exact copies of the ds polynucleotide sequences which are the amplified library members, which can optionally be separated from the others.

Further provided herein is a method of producing the polynucleotide library described herein, which is enriched in a predetermined library member which is a ds polynucleotide consisting of a first and a complementary second strand each comprising a polynucleotide core sequence and an overhang, by

(i) amplifying the predetermined library member by an enzymatic reaction employing a polymerase to produce amplification products, and:

a) a first primer pair comprising a forward primer complementary to at least the overhang of the first strand, and a reverse primer complementary to the terminal sequence of the core sequence of the second strand, excluding its overhang; and

b) a second primer pair comprising a forward primer complementary to at least the terminal sequence of the core sequence of the first strand, excluding its overhang, and a reverse primer complementary to at least the overhang of the second strand; and

(ii) producing and optionally isolating said amplification products; and

(iii) producing the library enriched in said amplification products.

Specifically, the enzymatic reaction is a polymerase chain reaction (PCR).

According to a specific embodiment of the method of producing a polynucleotide library which is enriched in a predetermined library member, the predetermined library member comprises a tag at the 5'- end of said first and/or second strand, preferably an affinity tag, and wherein each tagged strand is immobilized on a bead via said tag.

Specifically, the predetermined library member comprises a tag at the 3'- end of said first and/or second strand, preferably an affinity tag.

According to a specific embodiment, the variety of ds polynucleotides is synthesized by partially annealing a library of matching single stranded oligonucleotides (ss oligos) thereby obtaining a first library of ds oligonucleotides (ds oligos), each with the same overhangs, and optionally further annealing and ligation with ds oligos that have overhangs matching the overhangs of the first library, thereby obtaining a second library of ds oligos.Specifically, said second library of ds oligos is a library of double stranded (ds) polynucleotide library members as further described herein.

Specifically,

a) the library of ss oligos comprises ss oligos with a length of at least any one of 6, 7, 8, 9, or 10 nt, up to any one of 50, 100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850 or 900 nt or more; and/or

b) the first library of ds oligos comprises ds oligos with a length of at least any one of 6, 7, 8, 9, or 10 bp, up to any one of 50, 100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850 or 900 bp or more; and/or

c) the second library of ds oligos comprises ds oligos with a length of at least any one of 12, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290 or 300 bp, up to any one of 320, 350, 380, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1.000, 1.500, or 2.000 bp or more.

In some embodiments, the use of the library described herein, in a method of synthesizing a variety of target ds polynucleotides each being longer than the library members by assembling said library members with ds oligos that have overhangs matching the overhangs of the library members, thereby obtaining a library of target ds polynucleotides.

In some embodiments, themethod of synthesizing a library described herein, comprising a variety of target ds polynucleotides comprising:

a) providing an oligonucleotide library within an array device, which comprises a diversity of oligonucleotide library members, wherein each of the library members has a different nucleotide sequence and is contained in a separate library containment in an aqueous solution, which diversity includes single stranded oligonucleotides (ss oligos)

and double stranded oligonucleotides (ds oligos) with at least one overhang and covers at least 10.000 pairs of matching oligonucleotides,

b) in a first step, transferring at least a first pair of matching oligonucleotides from said library into a first reaction containment using a liquid handler and assembling the matching oligonucleotides thereby obtaining a first reaction product comprising at least one overhang,

c) in a second and optional further steps, transferring at least a second and optional further pairs of matching oligonucleotides from said library into a second and optional further reaction containments, respectively, using a liquid handler and assembling the matching oligonucleotides thereby obtaining a second and optional further reaction products each comprising at least one overhang, respectively,

d) assembling said first, second and optional further reaction products in a predetermined workflow, thereby producing said target ds polynucleotide with a length of at least 12bp and with an overhang,

wherein the ds polynucleotide library is produced by assembling a variety of either one or more of said first, second or optional further reaction products, which variety comprises a diversity of the core sequence and the same, non-complementary overhangs. Specifically, each of said ds polynucleotides comprises the same, non-complementary overhangs, wherein

a) the same first overhang sequence is the 5’ overhang of the leading strand, and the same second overhang sequence which is the 5’ overhang of the lagging strand; or

b) the same first overhang sequence is the 3’ overhang of the leading strand, and the same second overhang sequence which is the 3’ overhang of the lagging strand;

wherein said first and second overhang sequences are not complementary to each other.

Specifically, said ds polynucleotide library is characterized by the features as further described herein.

Specifically,

a) said ss oligo library members have a length of at least any one of 6, 7, 8, 9, or 10 nt, up to any one of 50, 100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850 or 900 nt or more ; and/or

b) said ds oligo library members have a length of any one of 6, 7, 8, 9, or 10 bp, up to any one of 50, 100, 150, 200, 250, 300, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850 or 900 bp or more; and/or

c) said ds polynucleotide library members have a length of at least any one of 12, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290 or 300 bp, up to any one of 320, 350, 380, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1.000, 1.500, or 2.000bp or more; and/or

d) wherein the overhang is less than 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, or 20 or 10% of the length of the library member, in particular, said 5' and 3' terminal overhang sequences have a length of any one of 4, 5, 6, 7, or 8 nucleotides, preferably up to any one of 15, 14, 13, 12, 11, 10, 9, or 8 nucleotides, preferably 4-8 nucleotides.

Specifically, said polynucleotide sequence is at least any one of 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, or 80% identical to a template. Specifically, said polynucleotide sequence is at least any one of 80, 81, 82, 83, 84, preferably at least 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98 or 99% identical to a template, preferably a template of the target ds polynucleotide.

Specifically, said assembly is by annealing or by a ligation reaction.

According to a specific embodiment, the method described herein comprises method steps for synthesizing a target ds polynucleotide having a predefined sequence, comprising

a) providing an oligonucleotide library within an array device, which comprises a diversity of oligonucleotide library members, wherein each of the library members has a different nucleotide sequence and is contained in a separate library containment in an aqueous solution, which diversity includes single stranded oligonucleotides (ss oligos) and double stranded oligonucleotides (ds oligos) with at least one overhang and covers at least 10.000 pairs of matching oligonucleotides,

b) in a first step, transferring at least a first pair of matching oligonucleotides from said library into a first reaction containment using a liquid handler and assembling the

matching oligonucleotides thereby obtaining a first reaction product comprising at least one overhang,

c) in a second and further steps, transferring at least a second and further pairs of matching oligonucleotides from said library into a second and further reaction containments, respectively, using a liquid handler and assembling the matching oligonucleotides thereby obtaining a second and further reaction products each comprising at least one overhang, respectively,

d) assembling said first, second and further reaction products in a predetermined workflow, thereby producing said target ds polynucleotide with an overhang, optionally followed by a finalization step to prepare blunt ends,

wherein said pairs of matching oligonucleotides and assembly workflow are determined using an algorithm to produce said target ds polynucleotide.

Specifically, a series of different target ds polynucleotides are synthesized using the same oligonucleotide library. Specifically, said different target ds polynucleotides have different sequences and are not fragments of each other.

Specifically, said different target ds polynucleotides have a sequence identity of less than 50%, preferably less than 30%. Specifically, said different target ds polynucleotides have a sequence identity of less than 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22 or 21%. Even more preferably, said different target ds polynucleotides have a sequence identity of less than 20 or 10% to each other, specifically they have a sequence identity of less than 19, 18, 17, 16, 15, 14, 13, 12, 11, 9, 8, 7, 6 or 5%.

Specifically, said target ds polynucleotide is a DNA molecule.

Specifically, one or more amplification steps are performed, e.g. by performing a PCR, preferably of 25 cycles. Specifically, said PCR employs a HiFi thermostable DNA Polymerase (Phusion or Q5) and two oligonucleotides complementary to each of the overhangs of the assembled fragment, and said complementary oligonucleotides including cleavage sites for TypeIIS restriction enzyme (BfuAI). Specifically, the amplified product is contacted with the TypeIIS restriction enzyme, which introduces the original overhangs into the amplified fragments. Specifically, said amplification step is carried out after any one or more of the first, second, third or further assembly steps, wherein the first, second, third or further reaction products, respectively, are amplified.

Specifically, said amplification step is carried out after assembly of the target ds polynucleotide, wherein the target ds polynucleotide is amplified.

Specifically, the predetermined workflow (also referred to as “assembly workflow”) is a hierarchic one, which is specifically characterized as follows:

A hierarchic workflow shall mean parallel or separate production of intermediary assembled matching pairs of polynucleotides which are produced as intermediates, each of the intermediates being assembled in a separate reaction compartment, which intermediates are further assembled to obtain the target polynucleotide or a part thereof. According to a specific example, in a first step, matching pairs of oligonucleotides are combined in parallel and in independent reaction compartments thereby producing in each compartment a polynucleotide that has the combined size of the reagent oligonucleotides and the same overhang length as the reagent oligonucleotides. In a second and subsequent steps, this process is repeated iteratively by using the previous products or other oligonucleotides as reagents thereby producing in each tier a polynucleotide of the combined size of the reagent polynucleotides that maintains the same overhang size. If the step before the last has three compartments, first reacting only two of the compartments carrying matching pairs, and then a further reaction step between this product and the last compartment will produce the target polynucleotide. Alternatively, if the three compartments contain polynucleotides that can form only two matching pairs in total, combining the three compartments the target polynucleotide is produced.

Specifically, the assembly workflow is automated. Specifically, the automated workflow employs microfluidic handlers that are capable of transferring serially or in parallel the full or partial contents of one or several compartments into other pre-specified compartments that may or may not be empty.

Specifically, the assembly workflow is sequence-dependent, meaning that the specific order is determined by the sequence of a template such that when matching pairs are combined at any step in the workflow they result in a larger part of the target ds polynucleotide or finally in the target ds polynucleotide. Specifically, the workflow is determined according to the sequence of a template or the sequence of the target ds polynucleotide.

Specifically, by the method described herein polynucleotides of lengths up to 1.000, 5.000, 10.000 or 100.000 base pairs (bp) or even longer can be produced at a

low price and at a high speed. Specifically, using the ds polynucleotide library and the methods described herein, a variety of target polynucleotides of lengths up to 1.000, 5.000, 10.000 or 100.000 base pairs (bp) or even longer can be produced at a low price and at a high speed.

The method described herein specifically comprises the following components:

A) A pre-built library of oligonucleotides that can be designed to cover the whole genetic sequence space and organizes the oligos in space for an efficient access by a liquid handler or microfluidics device. The access is considered efficient, if the spatial organization of the library diminishes the time needed to access necessary oligonucleotides. Specifically, said access is considered efficient if it diminishes or reduces the total handling time of the library, wherein said total handling time is the time spent handling library members during the synthesis of a target ds polynucleotide. Specifically, said access is further considered efficient if it diminishes the operational costs or diminishes the amount of necessary consumables associated with the access to the oligonucleotides, as compared to other organizations, in particular to spatially randomly placed oligonucleotides or lexicographical ordering. Specifically, the access is considered efficient, if the total handling time of the library is reduced at least by 5, 10, 15, 20, 25 or 50% compared to the total handling time of a randomly or lexicographically organized library.

B) A sequence-specific hierarchical assembly workflow, determined by an algorithm, to produce the long polynucleotide without mismatches.

The oligonucleotide library described herein specifically comprises single stranded (ss) and double stranded (ds) oligonucleotides (oligos), also referred to as oligo library members. These library members are pre-built, provided in storage stable solutions, and located at defined positions within the array device. Oligos of the library are synthesized and stored in the array device until needed.

Specifically, the oligonucleotides are linear polymers of nucleotide monomers and comprise “A” denoting deoxyadenosine, “T” denoting deoxythymidine, “G” denoting deoxyguanosine, and “C” denoting deoxycytidine or besides conventional bases (A, G, C, T) can comprise nucleotide-analogs e.g., inosine and 2’-deoxyinosine and their derivatives (e.g. 7’-deaza-2’-deoxyinosine, 2’-deaza-2’-deoxyinosine),azole- (e.g. benzimidazole, indole, 5-fluoroindole) or nitroazole analogues (e.g. 3-nitropyrrole, 5-nitroindole, 5-nitroimidazole, 4-nitropyrazole, 4-nitrobenzimidazole) and their

derivatives, acyclic sugar analogues (e.g. those drived from hypoxanthine- or indazole derivatives, 3-nitroimidazole, or imidazole-4,5-dicarboxamide), 5'-triphosphates of universal base analogues (e.g. derived from indole derivatives), isocarbostyril and its derivatives (e.g. methylisocarbostyril, 7-propynylisocarbostyril), hydrogen bonding universal base analogues (e.g. pyrrolopyrimidin), and other chemically modified bases (such as diaminopurine, 5-methylcytosine, isoguanine, 5-methyl-isocytosine, K-2'-deoxyribose, P-2'-deoxyribose) or e.g. others modified bases which can have different base-pairing preferences and can pair with more than one natural nucleobase with similar stringency/probability. The monomers are linked by phosphodiester linkage or in certain cases, by peptidyl linkages or by phosphorothioate linkages or by any of the other types of nucleotide linkages.

Specifically, the single stranded DNA oligonucleotide library members of the oligonucleotide library (herein simply referred to as ss oligos) are or comprise natural nucleosides (e.g. adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (e.g., inosine, or 5-methylisocytosine, or 3-nitropyrrole, 5-nitroindole, pyrrolidine, 4-nitroimidazole, 4-nitropyrazole, 4-nitrobenzimidazole, 4-aminobenzimidazole, 5-nitroindazole, 3-nitroimidazole, 5-aminoindole, benzimidazole, 5-fluoroindole, indole, methylisocarbostyril, pyrrolopyrimidine 7-propynylisocarbostryril, 2-aminoadenosine, 2-thiothymidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-amino-adenosine, 7-deaza-adenosine, 7-deaza-guanosine, 8-oxoadenosine, 8-oxoguanosine, 0(6)-methylguanine, and 2-thiocytidine); chemically or biologically modified bases (including methylated bases); intercalated bases; modified sugars (e.g., ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (e.g., phosphorothioates and 5'-N-phosphoramidite linkages).

Specifically, the double stranded DNA oligonucleotide library members (herein simply referred to as ds oligos) are or comprise natural nucleosides (e.g. adenosine, thymidine, guanosine, cytidine, uridine, deoxyadenosine, deoxythymidine, deoxyguanosine, and deoxycytidine); nucleoside analogs (e.g., inosine, or 5-methylisocytosine, or 3-nitropyrrole, 5-nitroindole, pyrrolidine, 4-nitroimidazole, 4-nitropyrazole, 4-nitrobenzimidazole, 4-aminobenzimidazole, 5-nitroindazole, 3-

nitroimidazole, 5-aminoindole, benzimidazole, 5-fluoroindole, indole, methylisocarbostyril, pyrrolopyrimidine 7-propynylisocarbostryril, 2-aminoadenosine, 2-thiothymidine, 3-methyl adenosine, 5-methylcytidine, 2-aminoadenosine, C5-bromouridine, C5-fluorouridine, C5-iodouridine, C5-propynyl-uridine, C5-propynyl-cytidine, C5-methylcytidine, 2-amino-adenosine, 7-deaza-adenosine, 7-deaza-guanosine, 8-oxoadenosine, 8-oxoguanosine, 0(6)-methylguanine, and 2-thiocytidine); chemically or biologically modified bases (including methylated bases); intercalated bases; modified sugars (e.g., ribose, 2'-deoxyribose, arabinose, and hexose); and/or modified phosphate groups (e.g., phosphorothioates and 5'-N-phosphoramidite linkages) and are formed by annealing fully or partially complementary single stranded oligonucleotides.

Specifically, the oligonucleotide library members can be produced by any of the chemical polynucleotide (oligonucleotide) synthesis methods, including the H-phosphonate, phosphodiester, phosphotriester or phosphite triester synthesis methods or any of the massively parallel oligonucleotide synthesis methods, e.g. microarray or microfluidics-based oligonucleotide synthesis (e.g. as described in References (Gao et al. 2001) (LeProust et al. 2010) (Bonde et al. 2014a)).

Specifically, the oligonucleotide library members can be produced by any of the enzymatic polynucleotide (oligonucleotide) synthesis methods, including ssDNA synthesis by DNA polymerase proteins or by reverse transcriptase proteins, which produce hybrid RNA-ssDNA molecules. Specifically, the enzymatic polynucleotide synthesis reaction can occur *in vivo* or *in vitro*.

Specifically, the oligonucleotide library members are produced by synthesizing the oligonucleotide sequence from nucleotide building blocks by any of the polynucleotide synthesis methods, wherein the building blocks are comprised of “A” denoting deoxyadenosine, “T” denoting deoxythymidine, “G” denoting deoxyguanosine, or “C” denoting deoxycytidine or other natural nucleosides (e.g. adenosine, thymidine, guanosine, cytidine, uridine), nucleotide-analogs e.g., inosine and 2’-deoxyinosine and theirs derivatives (e.g. 7’-deaza-2’-deoxyinosine, 2’-deaza-2’-deoxyinosine), azole- (e.g. benzimidazole, indole, 5-fluoroindole) or nitroazole analogues (e.g. 3-nitropyrrol, 5-nitroindol, 5-nitroimidazole, 4-nitropyrazole, 4-nitrobenzimidazole) and their derivatives, acyclic sugar analogues (e.g. those derived from hypoxanthine- or indazole derivatives, 3-nitroimidazole, or imidazole-4,5-dicarboxamide), 5’-triphosphates of universal base

analogues (e.g. derived from indole derivatives), isocarbostyril and its derivatives (e.g. methylisocarbostyril, 7-propynylisocarbostyril), hydrogen bonding universal base analogues (e.g. pyrrolopyrimidine), or any of the other chemically modified bases (such as diaminopurine, 5-methylcytosine, isoguanine, 5-methyl-isocytosine, K-2'-deoxyribose, P-2'-deoxyribose). The building blocks are linked by phosphodiester linkage or peptidyl linkages or by phosphorothioate linkages or by any of the other types of nucleotide linkages.

Preferably, ss oligos have a length of at least 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 nucleotides. In a specific embodiment of the invention, said ss oligos have a length of 6 to 26 nucleotides.

According to specific embodiments, the ss oligos may have a length of at least 50, 60, 70, 80, 90, 100, or even higher, e.g. up to 200 or 400 or more nt.

Specifically the ds polynucleotides described herein can be synthesized by annealing two or more e.g. 2, 4, 6, 8, or 10 ss oligos which are matching and produce the 5' or 3' overhangs.

According to a specific embodiment, the ds polynucleotides described herein can be synthesized by annealing only two ss oligos, each being synthesized by suitable means according to a template.

In such method, the variety of ds polynucleotides can be produced by synthesizing a variety of ss oligos which variety includes ss oligos which differ from others by one or more point mutations compared to a template, but still allow annealing of the ss oligos thereby obtaining a variety the ds polynucleotides.

According to a specific embodiment, the ds polynucleotides described herein, specifically those of the polynucleotide library described herein, can be synthesized by annealing a first set of two ss oligos producing a first ds oligo; and a second set and optionally further sets of two ss oligos producing a second and optionally further ds oligos, respectively, which first, second, and optionally further ds oligos have matching (complementary) overhangs which anneal to a longer ds polynucleotide of the required length.

In such method, the variety of ds polynucleotides can be produced by using a variety of ss oligos in said first, second, and/or further sets of ss oligos, which variety includes ss oligos which differ from others by one or more point mutations compared to

a template, but still allow annealing of the ss oligos and the respective ds oligos thereby obtaining a variety of longer ds polynucleotides.

Specifically, ds oligo library members of the oligonucleotide library described herein have at least one overhang. Specifically, the ds polynucleotide of the polynucleotide library described herein have two overhangs, preferably one on each strand, either on the 5' end or on the 3' end. An overhang is specifically characterized by a reactive (i.e. capable of annealing or hybridizing with another ss oligo or overhang) ss terminal stretch of one or more nucleotides which is part of and/or extending a ds oligo or ds polynucleotide.

The oligonucleotide library described herein may specifically comprise ds oligos with one overhang and a blunt end. A blunt end is specifically characterized by a ds terminal stretch of one or more base pairs which is part of a ds oligo or polynucleotide.

Specifically, ds oligos with overhangs on both ends and no blunt end may be comprised in the oligonucleotide library described herein.

Specifically, ds oligos of the oligonucleorandomtide library described herein may have a length of at least 6 base pairs, and said overhang is not more than half of the respective ds oligo length. Specifically, ds oligos have a length of at least 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 base pairs. Specifically, ds oligos have a length of at least 50, 60, 70, 80, 90, 100, or even higher, e.g. up to 200, 400, 500 or 600 bp or more.

Specifically, if said ds oligo is 6 base pairs long, the overhang is not more than 3 nucleotides long. Specifically, if said ds oligo is 24 base pairs long, the overhang is not more than 12 nucleotides long.

The libraries described herein are specifically comprised of physical oligonucleotides or polynucleotides and synthesized in standardized conditions. Oligonucleotides or polynucleotides can be purified, and may comprise modifications and are ideally kept at a standard concentration and volume in an appropriate buffer and/or excipient, so that they are ready-to-use.

Specifically, any of the following buffer and/or excipients may be used to keep the oligos or polynucleotides in solution: Tris Buffer, T.E. Buffer (Tris-EDTA Buffer) or Nuclease Free Water. Specifically, library members may be kept in Tris Buffer, wherein said Tris Buffer is provided at a concentration of about 10mM (+/- 1mM or 2mM). Specifically, library members may be kept in T.E. Buffer. Specifically, said T.E. Buffer is at least composed of Tris, at a concentration of about 10mM (+/- 1mM or 2mM), and

EDTA, at a concentration of any one of 0,1, 0,2, 0,3, 0,4, 0,5, 0,6, 0,7, 0,8, 0,9 or 1,0mM. Specifically, Nuclease Free Water, is water which has been de-ionized, filtered and autoclaved and is essentially free of contaminating non-specific endonuclease, exonuclease and RNase activity.

Specifically, all library members are kept in a mixture or in a compartmented array device, using the same or different buffer and/or excipients in each case.

A library described herein may comprise thousands of oligos or polynucleotides. Specifically, a library described herein comprises a diversity of library members, wherein each of the library members has a different nucleotide sequence, in particular regarding an oligonucleotide library as described herein, the diversity covers at least 10.000 pairs of matching oligonucleotides. Specifically, a library comprises at least 20.000, 30.000, 40.000, 50.000, 60.000, 70.000, 80.000, 90.000 or 100.000 pairs of matching oligonucleotides. Specifically, a preferred oligo library contains enough pairs of matching oligonucleotides to cover the whole sequence space.

The pairs of matching oligonucleotides described herein refer to single stranded oligonucleotides comprising partially or fully complementary sequences. Said pairs of matching oligos may be present in a library as ss oligos in separate containments or two or more complementary ss oligos may be contained in one containment where they may anneal and form ds oligos. The nucleotide sequences of a pair of matching ss oligos may be complementary in at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides, such that a matching pair can form a new ds polynucleotide molecule by annealing or hybridization of the ss oligo sequences, preferably wherein the ss oligos hybridize in part, thereby obtaining a ds polynucleotide with an overhang, and in partiuclar a 5’ and a 3’ overhang.

An ss oligo may specifically be part of a matching pair consisting of two or three hybridization partners. Specifically, an ss oligo can be used as a first hybridization partner capable of hybridizing with a second hybridization partner, which is another ss oligo or a ds oligo with a complementary overhang.

Specifically, an ss oligo can be used as a first hybridization partner capable of hybridizing with two different ss and/or ds oligos, or two different ds polynucleotides, which are used as second and third hybridization partners. Specifically, the first hybridization partner is a matching ss oligo, wherein a first part of the ss oligo is hybridizing to a second hybridization partner, and a second part of the ss oligo is

hybridizing to a third hybridization partner, thereby obtaining one ds polynucleotide composed of the three hybridization partners without a gap.

A pair of matching ds oligos is specifically characterized by complementary sequences in the respective overhangs of the ds oligos, e.g. wherein the respective overhangs are complementary in at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides, such that the matching pair can form a new ds polynucleotide molecule by hybridization of the overhang sequences.

A library described herein may specifically comprise a diversity of double stranded oligonucleotide or polynucleotide library members, wherein each of the ds library members has a different nucleotide sequence.

Specifically, said diversity of an oligonucleotide library described herein covers at least 100, 500, 1.000, 2.000, 3.000, 4.000, 5.000, 10.000, 20.000, 40.000, 60.000, 80.000, 100.000, 120.000, 140.000, 160.000, 180.000 or 200.000 different ds library members.

Specifically, said diversity of a ds polynucleotide library described herein comprises or consists of at least 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 63, 64, 65, 70, 80, 90, 100, 150, 200, 250, 255, 256, 257, 500, 1.000, 2.000, 3.000, 4.000, 5.000, 10.000, 20.000, 40.000, 60.000, 80.000, 100.000, 120.000, 140.000, 160.000, 180.000 or 200.000 different ds library members.

An oligo library described herein may specifically comprise a diversity of single stranded oligonucleotides library members, wherein each of the ss oligo library members has a different nucleotide sequence. Specifically, said diversity covers at least 100, 500, 1.000, 2.000, 3.000, 4.000, 5.000, 10.000, 20.000, 40.000, 60.000, 80.000, 100.000, 120.000, 140.000, 160.000, 180.000 or 200.000 different ss oligos. Specifically, said ss oligos may be used as linkers, specifically in the assembly of a ds polynucleotide.

Specifically, said diversity means, different library members differ in at least one point mutation, base or base pair. One library member may actually encompass multiple copies of a ss or ds oligonucleotide, or a ds polynucleotide, which copies have the same sequence. Such multiple copies of a library member are specifically contained in only one library containment.

In a specific embodiment of the invention, said diversity covers library members comprising a tag or label e.g., using an affinity ligand such as biotin, e.g. at only one of the 5’ or 3’ terminus, in particular wherein each library members has the same tag or

label at only one of the 5’ or 3’ terminus. This allows separation and purification of the biotinylated library members by reaction or affinity purification recognizing said tag and label, respectively.

In a specific embodiment of the invention, said diversity covers ss oligos and/or ds oligos and/or ds polynucleotides which are phosphorylated. Specific embodiments refer to ss oligos and/or ds oligos and/or and/or ds polynucleotides which are modified by any one or more of phosphorylation, methylation, biotinylation, or linkage to a fluorophore or quencher. Preferably, library members comprise a 5’ phosphorylation.

In a specific embodiment, a library is provided within an array device and library members are contained in separate library containments, each in an aqueous solution. Specifically, said array device is any of a microtiter plate, a microfluidic microplate, a set of capillaries, a microarray or a biochip, preferably a DNA and/or RNA biochip. Said array device may comprise only one, all or any number of the aforementioned containments.

In a further specific embodiment, more than one different library members may be contained in only one library containment. Specifically, said different library members contained in one library containment are ss oligos of such a sequence that they are not capable of annealing to each other. Specifically, said different library members contained in one library containment are ds oligos and/or ds polynucleotides of such a sequence that they are not capable of ligating to the other ds oligos and/or ds polynucleotides contained therein. Specifically, said different library members contained in one library containment are ss oligos and ds oligos (and optionally ds polynucleotides) of such a sequence that they are not capable of annealing to each other.

In a specific embodiment, said separate library containments are spatially arranged in a three-dimensional order, wherein the individual compartments are located within a device at defined coordinates within the x-, y- and z-axes. Specifically, said three-dimensional order comprises at least any one of two, three, four, five, six, seven, eight, nine, ten, fifteen, twenty, thirty, fourty, fifty, sixty or even more stacked library containments, which are at least partially or fully stacked. Preferably, the library containments are placed in different lays, which are laid one above the other in different lays. Specifically, the lays are placed at predefined positions within the three-dimensional order. Preferably, each of said library containments within one lay

comprises a series of library members spatially arranged in a two-dimensional order at predefined positions.

Specifically, the three-dimensional order is predefined by a parameter which primarily serves to shorten synthesis time. Preferably, said parameter is frequency of use, placing those oligos in close proximity to each other which frequently form a matching pair in DNA sequences, e.g. naturally occurring or commonly used in target ds polynucleotides or fragments thereof. Due to the large number of oligos required to build any given sequence, most spatial distributions of oligos in the library would incur into wasted time and resources due to the time needed to scan the library and search for the desired oligos. However, by using a specific distribution of the oligos, there is minimal movement of an automatic device to transfer the pairs of matching oligos into a reaction containment. For example, oligos can be stored in micro-well plates, where the first plate contains the most common matching pairs of oligonucleotides and further plates are arranged in decreasing order until the last plate contains the least frequently used oligos.

In the method described herein, oligonucleotides or ds polynucleotides from the libraries described herein are transferred into a reaction containment using a liquid handler. Specifically, said liquid handler may be a microdroplet handler. Specifically, the liquid handler is automated. Using a liquid handler, a suitable volume of at least any one of 10, 20, 30, 40, 50, 60, 70, 80, 90 100, 200 or 500nL can be transferred, e.g. such that at least any one of 10^9, 10^{10}, 10^{11} or 10^{12} copies of a library member, such as single stranded oligonucleotides, matching pairs of single stranded oligonucleotides and double stranded oligonucleotides or ds polynucleotides, are placed into one reaction containment. Preferably, at least about 10^{11} copies (e.g. 6.06×10^{11} copies) of a specific oligo or ds polynucleotide are placed into one reaction containment to react with another oligo and ds polynucleotide, respectively. Preferably, the volume in which oligos or ds polynucleotides are transferred by the liquid handler is between 10 and 1000nL. More preferably it is between 10 and 500nL and even more preferably it is between 50 and 250nL.

Specifically, a reaction containment is a compartment unit, such as a well, of any one of a microtiter plate, a microfluidic microplate, a set of capillaries, a microarray or a biochip, preferably a DNA and/or RNA biochip. Specifically, reaction containments feature an environment in which one nucleic acid strand bonds to a second nucleic acid strand by complementary strand interactions and hydrogen bonding to produce a double

stranded oligonucleotide. Such conditions include the chemical components and their concentrations (e.g., salts, chelating agents, formamide) of an aqueous or organic solution containing the nucleic acids, and the temperature of the mixture. Other well-known factors, such as the length of incubation time or reaction chamber dimensions may contribute to the environment.

According to a method described herein, oligonucleotides or ds polynucleotides are transferred from a library into a reaction containment and assembled to obtain a reaction product. Specifically, said assembly is by any method of hybridizing ss nucleotide sequences, annealing, and/or ligating, which is an enzymatic and/or chemical reaction. Specifically, said ligation reaction is an enzymatic ligation reaction using ligase, or ribozymes capable of ligation reaction. Preferably T4 DNA ligase, T7 DNA Ligase, T3 DNA Ligase, Taq DNA Ligase, DNA polymerase, or engineered enzymes are used in the ligation reaction. Preferably, the following ligation reaction is used: T4 DNA Ligase, at a concentration of 10 cohesive end units per µL supplemented with 1 mM ATP (Sambrook and Russel, 2014, Chapter 1, Protocol 17).

Specifically, said assembly is directly by hybridizing matching overhangs, or indirectly by hybridizing a suitable ss oligo linker, which ss oligo linker is an ss oligo contained in a library which is selected and transferred from said library to assemble any of said first, second or further reaction products.

Specifically ds polynucleotides as described herein are used as intermediates in a method of synthesizing a longer ds polynucleotides by assembling one or more further assembly molecules selected from ss oligos, ds oligos or ds polynucleotides.

Oligonucleotides or ds polynucleotides are specifically assembled according to a defined workflow. The workflow is specifically designed to avoid mismatches or reaction products which cannot be used for assembly to produce the target ds polynucleotide. If there are partial constructs that can anneal in alternative ways, a runaway, ie. an uncontrolled polymerization reaction, can occur. To avoid combinations of pairs of matching oligonucleotides that would result in unwanted constructs or runaway reactions, pairs of matching oligonucleotides are assembled in a predetermined sequence of assembly steps, ie a specific workflow. Preferably, said specific workflow is not linear but hierarchical, i.e. following an algorithm that provides for intermediate reaction products which are defined non-consecutive parts of the target ds polynucleotide conveniently produced avoiding undesired reaction products to the extent

possible, before such intermediate reaction products are further assembled into further intermediate reaction products or into the target ds polynucleotide sequence.

In a linear workflow, a polynucleotide is assembled in a linear fashion starting at the 3' end of the leading strand, and adding the next oligo to link the 3' end of the leading strand with the 5' end of the next oligo. For example, oligo B is ligated to oligo A, oligo C is ligated to oligo B, oligo D is ligated to oligo C and so forth. This assembly may be achieved simultaneously by adding all oligos to the reaction containment at the same time, or the polynucleotide is extended progressively by successively adding oligos A, B, C, D and so forth to the reaction containment.

A hierarchical workflow may, for example, be necessary when oligo D is capable not only of ligating to oligo C but also to oligo A due to complimentary sequences or overhangs. A linear workflow as described above would result in the unwanted polynucleotide A-D-B-C-D in addition to the desired polynucleotide A-B-C-D. Therefore, the polynucleotide is preferably assembled in a hierarchical workflow. Accordingly, in two separate reaction containments oligos A and B and oligos C and D are ligated, respectively. The ligation reaction will yield the reaction products A-B and C-D which can then be transferred into a third reaction containment, wherein upon ligation the desired polynucleotide A-B-C-D is formed.

Specifically, said workflow is designed using an algorithm. Specifically, said algorithm selects pairs of matching oligonucleotides, polynucleotides and ss oligo linkers, if necessary, and determines the assembly workflow, not by a mere sequence partitioning, but by determining an optimal or near-optimal way to assemble the target ds polynucleotide, avoiding mismatches or undesired reaction products as far as possible. Pairs of matching oligonucleotides and assembly workflow are specifically selected to avoid undesired (incorrect) reactions or reaction products, such as palindromic sequences, runaway reactions and unambiguous assembly. If there are incorrect reaction products besides the correct reaction products, such incorrect reaction products are suitably separated from the correct ones e.g. as follows: using gel electrophoresis to detect oligonucleotides or polynucleotides of a certain size and excising and purifying bands of the gel corresponding to the size of the desired reaction product. Specifically, correct reaction products can be detected by incorporation of tags or labels into the sequence. Specifically, oligos may be captured using biotinylated oligonucleotide adapters capable of hybridizing with the overhang of the oligo wherein,

said adapters are fixed to the substrate and coated with streptavidin. Non-captured incorrect products are eliminated by washing and subsequently, the correct products are released from the adapters by increasing the temperature. Specifically, further separation methods well known in the art may be applied. Specifically, such methods may involve chromatographic or affinity separation methods.

In a specific embodiment of the invention, said target ds polynucleotide has a length of at least 48 nucleotides. Specifically, said target ds polynucleotide has a length of at least any one of 100, 200, 300, 400, 500, 1.000, 10.000, 100.000, 200.000 or 500.000 nucleotides.

Typically, a template is used as a model to synthesize the target ds polynucleotide. Specifically, the nucleotide sequence of said target ds polynucleotide is identical to the nucleotide sequence of a template, and the variety of ds polynucleotides comprised in a library comprises or consists of library members which have a certain sequence identity to such template.

In a specific embodiment, a sequence of interest (SOI) is provided as a single stranded template and/or translated into two single stranded template sequences, based on which the target ds polynucleotide is synthesized. In a certain embodiment, a first template comprises the sequence of the SOI, and a second template comprises the reverse complement to the SOI.

In a further embodiment, said target ds polynucleotide is a proxy ds polynucleotide which has a sequence that is identical to said template, which proxy ds polynucleotide is further modified to obtain a polynucleotide which has a sequence of interest (SOI) which is different from the sequence of the target ds polynucleotide. Typically, the proxy ds polynucleotide is produced as an intermediate product, wherefrom a ds polynucleotide characterized by the SOI can be produced by one or more further steps of mutagenesis.

Specifically, the sequence of said template, according to which said proxy ds polynucleotide is synthesized, is not identical to said SOI. Specifically, the sequence of said template is less than any one of 100, 99, 98, 97, 96, 95, 94, 93, 92, or 91% identical, and/or at least any of 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identical to said SOI.

In a specific embodiment, the terminal nucleotides of the 3’ or 5’ end, or of both ends of the sequence of one strand, or each of the ds strands are removed before partitioning into shorter sequences. Specifically they are removed computationally.

Thereby, a template is produced which is different from the SOI. Specifically, any one of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 or 25 of the terminal nucleotides are removed, of the 3' end, or of the 5' end of the sequence to produce the template. Specifically, said nucleotide(s) are removed to generate overhangs and/or to prepare for finalization of synthesis by producing blunt ends at each of the termini of the target ds polynucleotide.

Specifically, the template is comprised of a single stranded or double stranded sequence. Preferably, said template is single stranded. Specifically, said two single stranded template sequences are annealed generating a double stranded template. Specifically, the sequence of said template is partitioned into shorter sequences, sub-sequences comprising oligonucleotide library members, and positions of said library members in the library are digitally annotated. Specifically, partitioning into sub-sequences depends on the hierarchical workflow and on the library members present in the library.

Specifically, the target ds polynucleotide has blunt ends on both ends.

Specifically, the method provided herein comprises a finalization step.

Specifically, said finalization step serves to add one or more nucleotide(s) which correspond to those previously removed from the 3' end and 5' end, respectively, to prepare the template, aiming to generate blunt ends. Specifically, oligos from the library are selected, which are complementary to the nucleotides at the 3' end and 5' end, respectively, i.e. complementary to the sticky ends of the polynucleotide. Specifically, these oligos are used as primers in a PCR reaction which is prepared to amplify the final product and to add the remaining nucleotides to each strand to synthesize the complete target polynucleotide with blunt ends.

Specifically, said finalization step comprises a purification step of the PCR product employing standard kits, such as the Monarch PCR & DNA clean up kit from New England Biolabs (product nr. T1030), to eliminate remaining oligos, enzymes and reagents, leaving the target ds polynucleotide as a DNA product, ready for downstream applications.

Alternatively, one or both blunt ends of the target ds polynucleotide can be produced by selecting a matching ds oligo with blunt ends or by selecting a ss oligo which is complementary to an overhang, and hybridizing without generating any further overhang thereby producing a blunt end.

Specifically, said nucleotide sequence of a target ds polynucleotide, SOI or template can be of natural or artificial origin.

In order to produce a ds polynucleotide, which has a complicated SOI, in a simpler and thus quicker assembly workflow, a proxy ds polynucleotide with a target sequence less than 100% identical to the SOI may be produced. Said proxy ds polynucleotide produced by the assembly method described herein can then be further modified to produce a ds polynucleotide with a nucleotide sequence 100% identical to the nucleotide sequence of the SOI. Specifically, said proxy ds polynucleotide is further modified by any of directed mutagenesis, endonucleases or exonucleases to obtain a nucleotide sequence identical to the nucleotide sequence of said template.

In a further specific embodiment, the target ds polynucleotide is further modified to produce a derivative thereof, which is any of a ds DNA, ss DNA or RNA molecule.

Specifically, said target ds polynucleotide is modified by site-directed mutagenesis, thereby introducing one or more point mutations which are any of nucleotide insertions, deletions or substitutions.

Specifically, said target ds polynucleotide is modified by enzymatic modification, employing any one or more of methyltransferases, kinases, CRISPR/Cas9, multiplex automated genome engineering (MAGE) using λ-red recombination, conjugative assembly genome engineering (CAGE), the Argonaute protein family (Ago) or a derivative thereof, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), meganucleases, tyrosine/serine site-specific recombinases (Tyr/Ser SSRs), hybridizing molecules, sulfurylases, recombinases, nucleases, DNA polymerases, RNA polymerases or TNases.

In a specific embodiment, said target ds polynucleotide is sequenced to verify the degree of identity with the sequence of a template or a SOI. Any suitable sequencing method may be used, for example any one of SNP genotyping methods, including hybridization-based methods (e.g. molecular beacons, SNP microarrays, restriction fragment length polymorphism, PCR-based methods, including Allele-specific PCR, primer extension-, 5'-nuclease or Oligonucleotide Ligation Assay, Single strand conformation polymorphism, Temperature gradient gel electrophoresis, Denaturing high performance liquid chromatography, High-resolution Melting of the entire amplicon (HRM), SNPlex and surveyor nuclease assay; Sequencing based mutation analysis, including capillary sequencing or high-throughput sequencing of an entire PCR amplicon

of the PTR (amplicon sequencing). Such high-throughput (HT) amplicon sequencing methods include, but are not restricted to polony sequencing, pyrosequencing, Illumina (Solexa) sequencing, SOLiD sequencing, semiconductor sequencing, DNA nanoball sequencing, Heliscope single molecule sequencing, Single molecule real time (SMRT) sequencing, Nanopore DNA sequencing, tunnelling currents DNA sequencing, sequencing by hybridization, sequencing with mass spectrometry, Microfluidic Sanger sequencing, Microscopy-based sequencing, RNAP sequencing.

Specifically, an oligonucleotide library is provided within an array device comprising a diversity of library members, which are single stranded oligonucleotides (ss oligos) and double stranded oligonucleotides (ds oligos) with at least one overhang, wherein each of the library members has a different nucleotide sequence and is contained in a separate library containment in an aqueous solution, which containments are spatially arranged in a three-dimensional order, which diversity covers at least 10.000 pairs of matching oligonucleotides.

Specifically, said library containments are spatially arranged in a three-dimensional order, preferably according to frequency of use, and wherein said three-dimensional order comprises at least any one of two, three, four, five, six, seven, eight, nine, ten, fifteen, twenty, thirty, fourty, fifty, sixty or even more stacked library containments, which are at least partially or fully stacked.

Further described herein is the use of an oligonucleotide library for synthesizing a series of different target double stranded (ds) polynucleotides having a predefined sequence, wherein said different target double stranded (ds) polynucleotides have a sequence identity of less than 50%, preferably less than 30%.

Further provided herein is a method of enriching a ds polynucleotide comprising an overhang on both, the leading and lagging strand, by a polymerase chain reaction (PCR), using

a) a first primer pair comprising a forward primer complementary to at least the overhang of the leading strand, and a reverse primer complementary to the terminal sequence of the lagging strand, excluding the overhang of the lagging strand; and

b) a second primer pair comprising a forward primer complementary to at least the terminal sequence of the leading strand, excluding the overhang of the leading strand, and a reverse primer complementary to at least the overhang of the lagging strand; and

producing and optionally isolating the amplification products which comprise an overhang on both, the leading and lagging strand.

FIGURES

Figure 1A. Source sequences used to construct a library (only fragment of 100 bp is shown), corresponding to four haplotypes of the hyper-variable region II of human mitochondria (Anderson et al., 1981: Gene Bank accession nr.: J01415).

Figure 1B. Scaffold of the design of the oligos that are required to build any possible combination of haplotypes (assuming full heterozygocity of each polymorphic site). Both strands are shown. Each Z-shaped block is an oligo pair where N stands or any of the four bases A, T, G, or C. The number above and below each oligo sequence scaffold indicates the length of the oligo and in parenthesis the number of oligos that are to be present in the library to cover all possible haplotypes at the variable sites. Each ss oligo is to be stored individually in a compartment of the library, except those underlined that are to be stored as annealed pairs.

Figure 2A. Nucleotide sequence of the SOI, called DISCOVER (SEQ ID NO: 1).

Figure 2B. Nucleotide sequence of the 16 oligos constituting SOI DISCOVER.

Figure 2C. Dimer structure of the constituting oligos. Here depicted for D+ and D-, but same structure applies to all other dimers.

Figure 3. Location of oligos in a well plate.

Figure 3A. After annealing, the contents of columns 1 and 3 are transferred to columns 2 and 4, respectively.

Figure 3B. After incubation of the first ligation reaction, the contents of column 2 are transferred to column 4.

Figure 3C. After incubation of the second ligation reaction, the contents of A4 are transferred to well B4 and incubated for the third and last ligation reaction. Well B4 contains the 128 bp target ds polynucleotide.

Figure 4. Acrylamide gel (10%) showing the contents of the process described in Example 2. Lane 1: reactions D+I (well A2 in Fig. 3B). Lane 2: negative control (for ligation) with a 64 bp dsDNA. Lane 3: positive control (for ligation) with a 64 bp dsDNA. Lanes 4 and 5: reactions DI+SC (well A4 in Fig. 3C) in two dilutions. Lanes 6 and 7: two alliquotes of the 128 bp target ds polynucleotide. Lane 8: 50 bp ladder (NEB).

Figure 5A. Partial SOI and its reverse complement (positions 65-100; otherwise as in Fig. 1A). The elements in italic, bold and regular fonts indicate different dimers. The underlined portions highlight the self-complementary overhangs that have to be avoided. Upper sequence SEQ ID NO:18; lower sequence SEQ ID NO:19.

Figure 5B. Partial sequence of the template for producing the proxy ds polynucleotide (positions 65-100). The base pairs with black background indicate the altered sites, which now make the dimer non-self-complementary. (The resulting modified oligos coincide with O-, and V+ of example 2.). Upper sequence is SEQ ID NO:20; lower sequence is SEQ ID NO:21.

Figure 5C. Mutagenizing primers used to modify the proxy ds polynucleotide to produce a ds polynucleotide which has the SOI. The underlined letters indicate the mutagenized bases. Upper sequence is SEQ ID NO:22; lower sequence is SEQ ID NO:23.

Figure 6. Arrangement of the oligos, which were transferred from the library of example 1, on a 96-well plate to prepare them for annealing and hierarchical synthesis.

Figure 7. Agarose gel electrophoresis (2%) showing the results of the hierarchical synthesis process. The top band is the one containing the 608bp product. Left lane is a 600 bp ladder.

Figure 8. Sequences of Example 4.

Figure 9. Sequences of Example 9.

Figure 10. Sequences of Example 6.

Figure 11. Agarose gel electrophoresis (2%) showing the results of the hierarchical synthesis process. The top band is the one containing the 1024 product. Left lane is a 1000 bp ladder.

Figure 12. Schematic of enrichment of polynucleotide library members using PCR primers.

Figure 13. Sequences of Example 8.

Figure 14. shows the results of Sanger sequencing, verifying that 16 variants of SEQ ID NO:218 are present in the library. The displayed sequence is SEQ ID NO:361 (which is the sequence from position 20 to position 80 of SEQ ID NO:218).

DETAILED DESCRIPTION

Specific terms as used throughout the specification have the following meaning.

As used herein, the terms “a”, “an” and “the” are used herein to refer to one or more than one, i.e. to at least one.

The term “**sequence of interest**” or “**SOI**” refers to the desired nucleotide or base pair sequence of the ds polynucleotide which is to be produced by the method provided herein.

The term “**target double stranded (ds) polynucleotide**” refers to a polynucleotide having a predefined sequence, which is produced by the synthesis method provided herein. Specifically, said target double stranded polynucleotide characterized by a sequence which is identical and/or corresponding to a SOI. If the target ds polynucleotide sequence has a sequence which is less than 100% identical to a SOI, the target ds polynucleotide is understood as a proxy ds polynucleotide that can be further modified to produce a ds polynucleotide that has a sequence which is identical and/or corresponding to the SOI.

The term “**proxy double stranded (ds) polynucleotide**” refers to a target double stranded (ds) polynucleotide whose sequence is less than 100% identical and at least 90%, preferably 95% identical to the nucleotide sequence of a SOI. In order to produce a ds polynucleotide having a sequence identical and/or corresponding to the SOI, and which is difficult to synthesize because its sequence may be prone to unambiguous assembly or runaway reactions, a proxy double stranded (ds) polynucleotide may be synthesized first. The sequence of the proxy double stranded polynucleotide is designed to avoid palindromic sequences, runaway reactions and unambiguous assembly and/or to facilitate hierarchical assembly. Specifically, the sequence may be designed computationally. The synthesized proxy ds polynucleotide may then be further modified to produce a ds polynucleotide with a nucleotide sequence identical to the nucleotide sequence of the SOI. Specifically, said proxy ds polynucleotide is further modified by any of directed mutagenesis, endonucleases or exonucleases, and/or enzymatic modification, employing any of methyltransferases, kinases, CRISPR/Cas9, multiplex automated genome engineering (MAGE) using λ-red recombination, conjugative assembly genome engineering (CAGE), the Argonaute protein (Ago) or a derivative thereof, zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), meganucleases, tyrosine/serine site-specific recombinases (Tyr/Ser SSRs), hybridizing molecules, sulfurylases, recombinases, nucleases, DNA polymerases, RNA

polymerases or TNases to obtain a ds polynucleotide which has a sequence that is identical and/or corresponding to the SOI.

The term "**template**" refers to a polynucleotide characterized by a certain sequence, or a polynucleotide sequence, which sequence can be used to synthesize and produce a target ds polynucleotide. If a template is used in a synthesis method provided herein, the so produced target ds polynucleotide has a sequence which is 100% identical to the template.

Specifically, said template is single stranded or double stranded. Such template can be a natural nucleotide sequence or an artificial, computationally designed nucleotide sequence that comprises the desired product. Such template can be identical to a SOI or less than 100% identical to a SOI, preferably less than 95% identical, but at least 80% identical.

Preferably, said template is generated computationally and comprises the sequence of the leading strand of the target ds polynucleotide and the reverse complement of the target polynucleotide, respectively. Typically, two templates are used in the synthesis method described herein, one template for each of the strands of the target ds polynucleotide. When computationally designing a template sequence, compatibility with the experimental strategy used for assembly is preferred.

The term "**single stranded DNA oligonucleotide**", also referred to as "ssDNA oligonucleotide" or simply "ss oligonucleotide" or "ss oligo", shall refer to an oligonucleotide which is a linear polymer of nucleotide monomers. Monomers making up oligonucleotides are capable of specifically binding to a natural polynucleotide by way of a regular pattern of monomer-to-monomer interactions, such as Watson-Crick type of base pairing, base stacking, Hoogsteen or reverse Hoogsteen types of base pairing, wobble base pairing, or the like. ssDNA oligonucleotides described herein typically range in size between 6 and 26, but may be longer. ssDNA oligonucleotides described herein may range in size between 6 and 220 nucleotides, e.g. between 27 and 200 nucleotides. Whenever an oligonucleotide is represented by a sequence of letters (upper or lower case), such as "ATGC," it will be understood that the nucleotides are in 5′→3′ order from left to right and that "A" denotes deoxyadenosine, "T" denotes deoxythymidine, "G" denotes deoxyguanosine, and "C" denotes deoxycytidine. Besides conventional nucleotides (A, G, C, T), modified nucleotides e.g. K-2'-deoxyribose, P-2'-deoxyribose, 2'-deoxyinosine, 2'-deoxyxanthosine or nucleotides with nucleobase analogs may be

used e.g., inosine, or 5-methylisocytosine, or 3-nitropyrrole, 5-nitroindole, pyrrolidine, 4-nitroimidazole, 4-nitropyrazole, 4-nitrobenzimidazole, 4-aminobenzimidazole, 5-nitroindazole, 3-nitroimidazole, 5-aminoindole, benzimidazole, 5-fluoroindole, indole, methylisocarbostyril, pyrrolopyrimidine 7-propynylisocarbostryril. The terminology and atom numbering conventions follow those disclosed in Strachan and Read, Human Molecular Genetics 2 (Wiley-Liss, New York, 1999). Usually oligonucleotides comprise the four natural nucleosides (e.g. deoxyadenosine, deoxycytidine, deoxyguanosine, deoxythymidine for DNA or their ribose counterparts for RNA) linked by phosphodiester or by peptidyl linkages or by phosphorothioate linkages; however, they may also comprise non-natural nucleotide analogs, e.g. including modified bases, sugars, or internucleosidic linkages.

In some embodiments, the single stranded oligonucleotide pools are produced using chemical synthesis methods, e.g. by synthesizing the oligonucleotide sequence from monomer-phosphoramidites, dimer-phosphoramidites (Neuner, Cortese, and Monaci 1998) or trimer-phosphoramidites (Sondek and Shortle 1992), mixture of monomer-phosphoramidites, mixture of dimer-phosphoramidites, mixture of trimer-phosphoramidites or their combination thereof.

In some embodiments, the oligonucleotides are produced and purified from naturally-occurring sources, or synthesized *in vivo*, within the cell undergoing *in vivo* mutagenesis using any of a variety of well-known enzymatic methods e.g. as described in Farzadfard et al. (2014). Specifically, enzymes that synthesize soft-randomized oligonucleotide pools include, but are not limited to low fidelity DNA polymerase proteins or low fidelity reverse transcriptase proteins which incorporate mismatching nucleotides during synthesis with high frequency. Alternatively, mismatching nucleotides are incorporated into the oligos with a higher frequency by the DNA polymerases or reverse transcriptases due to the presence of chemical substances, which are well-known to those skilled in the art.

The term “base pair” or “**bp**”, (used as abbreviation, singular or plural) also “bps” (in plural), refers to any of the pairs of nucleotides connecting the complementary strands of a molecule of DNA or RNA and consisting of a purine linked to a pyrimidine by hydrogen bonds. The pairs are adenine and thymine in DNA, adenine and uracil in RNA, and guanine and cytosine in both DNA and RNA.

The term “**pairs of matching oligonucleotides**” refers to two or more complimentary oligonucleotides. By "complementary" it is meant that the nucleotide sequences of similar regions of two single stranded nucleic acids or overhang parts of one or more ds nucleic acids, have a nucleotide base composition that allow the single-stranded regions to anneal together in a stable, double-stranded hydrogen-bonded region under stringent annealing or amplification conditions, such annealing is also referred to as “hybridization”. When a contiguous sequence of nucleotides of one single-stranded region is able to form a series of "canonical" hydrogen-bonded base pairs with an analogous sequence of nucleotides of the other single-stranded region, such that A is paired with U or T and C is paired with G, the nucleotide sequences are 100% complementary. Besides conventional bases (A, G, C, T), analogs e.g., inosine and 2’-deoxyinosine and their derivatives (e.g. 7’-deaza-2’-deoxyinosine, 2’-deaza-2’-deoxyinosine), azole- (e.g. benzimidazole, indole, 5-fluoroindole) or nitroazole analogues (e.g. 3-nitropyrrol, 5-nitroindol, 5-nitroimidazole, 4-nitropyrazole, 4-nitrobenzimidazole) and their derivatives, acyclic sugar analogues (e.g. those drived from hypoxanthine- or indazole derivatives, 3-nitroimidazole, or imidazole-4,5-dicarboxamide), 5’-triphosphates of universal base analogues (e.g. derived from indole derivatives), isocarbostyril and other hydrophobic analogues, and any of its derivatives (e.g. methylisocarbostyril, 7-propynylisocarbostyril), hydrogen bonding universal base analogues (e.g. pyrrolopyrimidin), and other chemically modified bases (such as diaminopurine, 5-methylcytosine, isoguanine, 5-methyl-isocytosine, K-2’-deoxyribose, P-2’-deoxyribose) can have different base-pairing preferences and can pair with more than one natural nucleobase with similar stringency/probability. In certain cases, the monomers are linked by phosphodiester or by peptidyl linkages or by phosphorothioate linkages.

The term “**double stranded DNA oligonucleotide**”, also referred to as “dsDNA oligonucleotide” or simply “ds oligonucleotide” or “ds oligo”, shall refer to an oligonucleotide which is a linear polymer of nucleotide dimers. Dimers making up oligonucleotides are two complementary nucleotides bound by way of a regular pattern of monomer-to-monomer interactions, such as Watson-Crick type of base pairing, base stacking, Hoogsteen or reverse Hoogsteen types of base pairing, wobble base pairing, or the like. dsDNA oligonucleotides described herein typically range in size between 6 and 26 base pairs (bp), but may be longer. dsDNA oligonucleotides described herein

may range in size between 6 and 200 base pairs, e.g. between 27 and 200 base pairs. Whenever an oligonucleotide is represented by a sequence of letters (upper or lower case), such as “ATGC,” it will be understood that the nucleotides are in 5′→3′ order from left to right and that “A” denotes deoxyadenosine, “T” denotes deoxythymidine, “G” denotes deoxyguanosine, and “C” denotes deoxycytidine. Besides conventional nucleotides (A, G, C, T), modified nucleotides e.g. K-2'-deoxyribose, P-2'-deoxyribose, 2'-deoxyinosine, 2'-deoxyxanthosine or nucleotides with nucleobase analogs may be used e.g., inosine, or 5-methylisocytosine, or 3-nitropyrrole, 5-nitroindole, pyrrolidine, 4-nitroimidazole, 4-nitropyrazole, 4-nitrobenzimidazole, 4-aminobenzimidazole, 5-nitroindazole, 3-nitroimidazole, 5-aminoindole, benzimidazole, 5-fluoroindole, indole, methylisocarbostyril, pyrrolopyrimidine 7-propynylisocarbostryril. The terminology and atom numbering conventions follow those disclosed in Strachan and Read, Human Molecular Genetics 2 (Wiley-Liss, New York, 1999). Usually oligonucleotides comprise the four natural nucleosides (e.g. deoxyadenosine, deoxycytidine, deoxyguanosine, deoxythymidine for DNA or their ribose counterparts for RNA) linked by phosphodiester or by peptidyl linkages or by phosphorothioate linkages; however, they may also comprise non-natural nucleotide analogs, e.g. including modified bases, sugars, or internucleosidic linkages.

The simplest DNA end of a double stranded molecule is called a blunt end. In a blunt-ended molecule, both strands terminate in a base pair. Non-blunt ends are created by various overhangs. The term “**overhang**” as used herein refers to a stretch of unpaired nucleotides at one or both ends of a ds oligo or polynucleotide molecule. These unpaired nucleotides can be in either strand, creating either 3' or 5' overhangs. A double stranded molecule comprising two overhangs is understood as a ds molecule comprising an overhang at each end of the molecule, irrespective of which strand. Thus, the overhangs can be on only one of the strands, i.e. on the same strand on both sides, or on both of the strands, i.e. on both, the sense and anti-sense (leading or lagging) strands of the molecule. The simplest case of an overhang is a single nucleotide. An overhang may comprise or consist of any one of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 nucleotides, or at least any one of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 or 12 nucleotides. An overhang is typically not more than half of a ds oligo length. For example, if said ds oligo is 6 nucleotides long, the overhang is not more than 3 nucleotides long, meaning the overhang can also be 1 or 2 nucleotides long. According to another example, if said ds

oligo is 24 nucleotides long, the overhang is not more than 12 nucleotides long, meaning it can also be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or 11 nucleotides long.

The term "**core sequence**" as used herein, refers to a part of the nucleotide sequence of a double stranded nucleic acid molecule comprising two overhangs, which part is the double stranded part of the nucleic acid molecule, i.e. the sequence minus the overhangs. In other words, in a double stranded nucleic acid molecule comprising two single stranded overhangs, the core sequence is the double stranded segment.

The term "**library**" as used herein regarding an oligonucleotide library, also called "**oligonucleotide library**", shall refer to a collection of library members which are nucleic acid fragments and which comprises at least 10.000 pairs of matching oligonucleotides. Preferably, the library comprises both, single stranded oligonucleotide library members and double stranded oligonucleotide library members. The library members share common features (such as conferred by genomic sequences), but differ in at least one base pair, nucleotide, mutation and/or phenotype. A library typically contains library members which are diverse, besides those that have common features. One particular type of library is a library of randomized mutants of oligonucleotides, generated by random mutagenesis. Another specific example would be a rationally designed (or synthetic) library, e.g. a library which comprises specifically engineered DNA fragments or oligonucleotides. The oligonucleotide library described herein comprises library members suitably composed of oligonucleotides of varying lengths and different sequences, wherein the oligonucleotides may correspond to a certain region of DNA or may even span the entire genetic space. Exemplary, the library provided herein may comprise a diversity of oligonucleotide library members necessary to possibly synthesize any and all naturally occurring polynucleotides of the human chromosomal genome or mitochondrial genome. In a further example said diversity may cover any and all naturally occurring polynucleotides of eukaryotic species other than human, such as e.g. mouse, rat, rabbit, pig, sheep, plants, funghi or yeast. In yet another example said diversity may cover any and all naturally occurring polynucleotides of prokaryotes, such as e.g. achaeans or bacteria.

The library provided herein, specifically comprises at least 10.000 pairs of matching oligonucleotides which are single stranded oligonucleotides, specifically they are ss oligos of varying lengths, comprising partially or fully complementary sequences. Said pairs of matching oligos may be present in the library as ss oligos in separate

containments or two or more complementary ss oligos may be contained in one containment where they may anneal and form ds oligos. The nucleotide sequences of a pair of matching ss oligos may be complementary in at least 1, 2 or 3 nucleotides, preferably at least 4 or more nucleotides, such that a matching pair can form a new ds polynucleotide molecule by hybridization of the ss oligo sequences, preferably wherein the ss oligos hybridize in part, thereby obtaining a ds polynucleotide with an overhang.

The library preferably comprises oligonucleotides which are artificially or chemically synthesized, or chemically modified (e.g. including peptidyl nucleic acids or phosphorothioate bond) oligonucleotides synthesized by suitable methods well-known in the art. The oligonucleotides comprised in the library can also be generated by enzymatic digestion of naturally occurring DNAs. The members of said oligonucleotide library described herein are specifically characterized by different sequences, mutations or nucleobase or nucleotide alterations, e.g. a substitution, or insertion or deletion of one or more subsequent nucleotides. Typically, the library members differ in at least one or more point mutation. Specifically, in some embodiments, the variation covers every possible naturally-occurring nucleobase residue at a certain position. If the mutants are produced by mutagenesis of a parent oligonucleotide, a variety of sequence variations of the parent oligonucleotide is produced.

The term “**library**” as used herein regarding a polynucleotide library, also called “**polynucleotide library**”, shall refer to a collection of double stranded polynucleotide library members comprising a variety of polynucleotide sequences each comprising two overhangs. In other words, the polynucleotide library members are partially double stranded. The library members share common features (such as conferred by genomic sequences), but differ in at least one base pair in the polynucleotide core sequence. Each library member comprises an identical first sequence which is the first overhang sequence and an identical second sequence which is the second overhang sequence. There is no limitation to the first and second sequence, except that they are not complementary to each other, specifically, so as to avoid hybridization of the library members within a polynucleotide library. According to one specific embodiment, the two overhangs are comprised on either the leading or the ladding strand. According to another specific embodiment, one overhang is comprised on the leading strand and one overhang is comprised on the lagging strand.

One particular type of polynucleotide library is a library of randomized mutants of polynucleotides, generated by random mutagenesis. The members of said polynucleotide library described herein are specifically characterized by different sequences, mutations or nucleobase or nucleotide alterations, e.g. a substitution, or insertion or deletion of one or more subsequent nucleotides. Typically, the library members differ in at least one or more point mutation. Specifically, in some embodiments, the variation covers every possible naturally-occurring nucleobase residue at a certain position. If the mutants are produced by mutagenesis of a parent polynucleotide, a variety of sequence variations of the parent polynucleotide is produced.

Another specific example would be a rationally designed (or synthetic) library, e.g. a library which comprises specifically engineered polynucleotides, comprising specifically engineered sequence variations.

The polynucleotide library described herein is typically composed of polynucleotide library members of the same length. The polynucleotide library members may have a length of at least 48, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950 or 1000bp. According to a further specific embodiment, the polynucleotide library members have a length of 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000bp or more.

Specifically, the polynucleotide library members described herein comprise overhangs of about 50% of the library member's polynucleotide sequence, or less than 50%. Preferably the polynucleotide library members described herein comprise overhangs of about 4-8 nucleotides in length.

According to a specific example, the library preferably comprises polynucleotides which are synthesized by assembly of oligonucleotides, specifically, by partially annealing one or more libraries of partially matching ss oligos.

According to a further example, polynucleotide library members described herein are artificially or chemically synthesized, or chemically modified (e.g. including peptidyl nucleic acids or phosphorothioate bond) polynucleotides synthesized by suitable methods well-known in the art. The polynucleotides comprised in the polynucleotide library can also be generated by enzymatic digestion of naturally occurring DNAs.

Exemplary, a target polynucleotide may be synthesized by annealing library members of a polynucleotide library with library members of another polynucleotide library comprising complementary 5' or 3' overhangs.

The diversity of the libraries described herein may further comprise library members which are phosphorylated, methylated, biotinylated or which are linked to fluorophores or quenchers. As described herein, library members may comprise one or more additional phosphoryl groups.

Methylation of library members, the addition of a methyl group to a DNA molecule, preferably to cysteine or adenine, is performed according to suitable DNA methylation methods well-known in the art.

As used herein, biotinylation refers to a method of covalently attaching one or more biotin molecules to a nucleic acid, such as ss oligos or ds oligos. The library members described herein may be biotinylated by suitable methods well-known in the art; preferably it is a method of chemical biotinylation. Oligonucleotides can be readily biotinylated in the course of oligonucleotide synthesis by phosphoramidite methods well-known in the art, which use biotin phosphoramidite.

Members of the libraries described herein may be conjugated to a fluorophore by suitable chemical and enzymatic methods well-known in the art. Exemplary methods used for the fluorescent labeling of nucleic acids may employ a method for enzymatic labeling of DNA with fluorescent dyes e.g., using a Thermo Fisher's ARES DNA labeling kit, which employ a two-step method for enzymatic labeling of DNA with fluorescent dyes. Further exemplary methods may employ a chemical method for labeling nucleic acids without enzymatic incorporation of labeled nucleotides e.g., using a ULYSIS Nucleic Acid Labeling Kit. Further exemplary methods may employ chemical labeling of amine-terminated oligonucleotides to prepare singly labeled fluorescent oligonucleotide conjugates e.g., using an Alexa Fluor Oligonucleotide Amine Labeling Kit. Further exemplary methods may employ DNA arrays/microarrays and other hybridization techniques.

Library members may be linked to one or more quenchers, e.g., substances that absorb excitation energy from a fluorophore, by suitable methods well-known in the art. Examples of quenchers include but are not limited to Dabsyl (dimethylaminoazobenzenesulfonic acid), Black Hole Quenchers, Qxl quenchers, Iowa black FQ, Iowa black RQ and IRDye QC-1.

The term “**point mutation**” or nucleobase alterations as used herein shall refer to a mutation event altering a nucleic acid or amino acid sequence at a certain location, such as by introducing or exchanging single nucleobases or amino acids or introducing gaps. A point mutation or nucleobase alteration may involve a change in one or more single or adjacent or consecutive nucleobases or amino acid residues in a sequence. Specifically, point mutations are introduced into the sequence of a ds polynucleotide or a target polynucleotide in a targeted manner, resulting in a degree of variation compared to a template. In a library comprising a repertoire of mutants covering a limited diversity, the frequency of point mutations in a sequence is limited, such that the mutants share at least a certain sequence identity to a parent (or reference) sequence, which is e.g. at least any of 80%, 90%, 95%, 96%, 97%, 98%, or 99%.

“Percent (%) nucleotide **sequence identity**” with respect to the nucleotide sequences described herein is defined as the percentage of nucleotides in a candidate sequence that are identical with the nucleotides in the specific nucleotide sequence, after aligning the sequence and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity. Those skilled in the art can determine appropriate parameters for measuring alignment, including any algorithms needed to achieve maximal alignment over the full length of the sequences being compared.

The term “**diversity**” as used herein, refers to a degree of versatility characterizing the libraries provided herein. Specifically, said diversity regarding an oligonucleotide library comprises single and double stranded oligos of different lengths and different sequences. For example, the library may comprise all possible sequence variations of 8 nucleobase long ss oligos (herein referred to as octamers), which are 65.536 different ss oligos of 8 nucleobases length, and in addition other ss oligos or ds oligos of different lengths, which are commonly comprised in target sequences and are thus required more often to build any given sequence. Including commonly used single or double stranded oligos into the library’s diversity decreases synthesis cost and increases time efficiency.

Specifically, said diversity may cover an entire genome, for example the human genome. Specifically, said diversity may cover the entire genetic space. Specifically, said diversity may cover a genome or the entire genetic space multiple times in multiple ways. For example by encompassing all possible hexamer, heptamer and/or octamer

sequence combinations. For example, said library may also encompass all or selected 9-mers and 10-mers or of any up to 26-mers.

According to a specific example, the diversity within a pool of oligonucleotides described herein is characterized as follows: the diversity may be determined by the number of mutations within the oligonucleotide sequence. For example, in a single oligonucleotide with a length of 16 nucleotides, the theoretical number of possible single nucleotide changes is 16x3=48 with the four naturally occurring DNA, A, T, G or C nucleotides. For two single nucleotide changes with the four naturally occurring DNA, A, T, G or C nucleotides per oligonucleotide (double mutants) the number of possible sequences is 6.408. For three single nucleotide changes per oligonucleotide (triple mutants) this number is 563.904. For quadruple mutations this number is 36.794.736. These numbers can further increase by incorporating non-natural nucleobases within the oligonucleotide sequence.

Specifically, diversity regarding a polynucleotide library comprises double stranded polynucleotide library members, which differ in at least one nucleobase (base pair) in their sequence, but do not differ in a first and a sequence, which are the 5' and 3' overhangs, respectively. Polynucleotide library members comprise an identical first sequence and an identical second sequence. The first sequence may be identical to the second sequence or not identical. Specifically, the first and second sequence are not complementary such that they would allow annealing or hybridization to each other.

As used herein, the term "**enriched**" regarding a polynucleotide library, refers to the presence of an increased amount of polynucleotides in a polynucleotide library comprising the desired features with reference to a non-enriched polynucleotide library. Specifically, an enriched polynucleotide library comprises about 15, 16, 17, 18, 19, 20, preferably 21, 22, 23, 24 or 25% of **of those polynucleotides** which have the distinguishing features common to all library members (or at least those which need to be purified or separated from others), which is provided in a mixture comprising a reduced amount of those polynucleotides which do not have such distinguishing features. Various methods to enrich a solution in a specific nucleic acid molecule are known to the person skilled in the art. Specifically, the polynucleotide library described herein is enriched by amplifying the specific nucleic acid molecule using polymerase chain rection (PCR) methods.

According to a specific embodiment, the polynucleotide library provided herein may be enriched through targeted amplification of the polynucleotide library members. Specifically, enrichment may be achieved through PCR amplification, wherein two primer sets, each comprising two different primers, are used. The polynucleotide library members are partially double stranded, comprising a double stranded segment, i.e. the core sequence, a 5' or 3' overhang comprised on the 5' or 3' end of the leading strand, respectively and a 3' or 5' overhang comprised on the 3' or 5' end of the lagging strand, respectively. Specifically, each set comprises a primer that is complementary to at least the overhang and a primer that is complementary to the first nucleotides in the double stranded segment of the polynucleotide. Exemplary, the first set of primers comprises a first primer, which is complementary to the first, e.g. 4, 6, 8 or 10, nucleotides of the 5' end of the leading strand including the 5' overhang, and a second primer, which is complementary to the first, e.g. 4, 6, 8 or 10, nucleotides of the leading strand of the double stranded segment of the polynucleotide.

According to a further specific embodiment, the polynucleotide library provided herein may be purified by immobilization of library members on a solid phase using a tag, for example a biotin tag, and enrichment of the immobilized library member using, e.g., PCR amplification. According to a preferred embodiment, two sets of primers, as described above, are used for target specific enrichment and simultaneous elimination of the tag. Specifically, by using a set of primers specific to the 5' end of the leading strand and a set of primers specific to the 5' end of the lagging strand of the polynucleotide that is to be enriched, each comprising a primer that is complementary to at least the overhang and a primer that is complementary to the core sequence of the polynucleotide, the target polynucleotide is amplified without the tag sequence. This has the profound advantage that no additional step is required to remove the tag sequence, e.g. by enzymatic digestion.

The degree of purification is understood as the amount of library members per volume or per total (poly) nucleotide mass. Various methods to determine the degree of purity of a preparation of nucleic acid molecules are known to a person skilled in the art. Specifically, the degree of purity may be determined using gel electrophoresis, next generation sequencing or qPCR.

Exemplary methods for sequencing-based screening of oligonucleotides within a library are the following: SNP genotyping methods, including hybridization-based

methods (e.g. molecular beacons, SNP microarrays, restriction fragment length polymorphism, PCR-based methods, including Allele-specific PCR, primer extension-, 5'-nuclease or Oligonucleotide Ligation Assay, Single strand conformation polymorphism, Temperature gradient gel electrophoresis, Denaturing high performance liquid chromatography, High-resolution Melting of the entire amplicon (HRM), SNPlex and surveyor nuclease assay; Sequencing based mutation analysis, including capillary sequencing or high-throughput sequencing of an entire PCR amplicon of the PTR (amplicon sequencing). Such high-throughput (HT) amplicon sequencing methods include, but are not restricted to polony sequencing, pyrosequencing, Illumina (Solexa) sequencing, SOLiD sequencing, semiconductor sequencing, DNA nanoball sequencing, Heliscope single molecule sequencing, Single molecule real time (SMRT) sequencing, Nanopore DNA sequencing, tunnelling currents DNA sequencing, sequencing by hybridization, sequencing with mass spectrometry, Microfluidic Sanger sequencing, Microscopy-based sequencing, RNAP sequencing.

Each library member may be individually characterized and marked by a selectable marker or a DNA sequence tag or barcode, to facilitate the selection of a library member in the library or the identification of a library member in the library. Alternatively, the genetic mutation may be determined directly by a suitable determination method, e.g. high-throughput sequencing, capillary sequencing or employing specific probes hybridizing with a predefined sequence, to select the corresponding oligonucleotide.

It may be desirable to locate the library members in separate containers, to obtain a library of oligonucleotides in containers. According to a specific embodiment, the library is provided in an array, e.g. a DNA biochip, wherein the array comprises a series of spots on a solid carrier.

The term "**mutagenesis**" as used herein refers to a process of altering the sequence of an oligonucleotide or a polynucleotide. Specifically, site-directed mutagenesis refers to a method for creating a specific mutation in a known nucleotide sequence. This mutation is a specific, targeted change and may comprise single or multiple nucleotide insertions, deletions or substitutions. This task may be performed by restriction enzymes, specifically **endonucleases** and/or **exonucleases**. Endonucleases cleave the phosphodiester bonds in the middle of an oligonucleotide or a polynucleotide,

whereas exonucleases cleave the phosphodiester bonds at the 5’ or 3’ end of an oligonucleotide or a polynucleotide.

The term “**algorithm**” as used herein refers to a self-contained sequence of actions to be performed. An algorithm is an effective method that can be expressed within a finite amount of space and time and in a well-defined formal language for calculating a function. Starting from an initial state and initial input the instructions describe a computation that, when executed, proceeds through a finite number of well-defined successive states, eventually producing "output" and terminating at a final ending state. The transition from one state to the next is necessarily deterministic.

The term “**workflow**” or “**assembly workflow**” refers to the optimal number of oligo subsets and their sequence of assembly into the target ds polynucleotide. In the method provided herein, the sequence of a template may be divided into sub-sequences, corresponding to subsets of oligos, avoiding particular nucleotide synthesis problems, such as palindromic sequences, runaway reactions and unambiguous assembly. In particular, such division into shorter oligonucleotides may be very efficient to shorten the assembly process and to avoid the need of separating unwanted reaction products. Specifically, ligation of subsets of oligos yields intermediate reaction products, also called intermediates, and assembly of intermediate reaction products ultimately yields the target ds polynucleotide. Preferably, additional criteria to those listed above may be used for selecting subsets of oligos. Such additional criteria include, but are not limited to, minimization of the size of the subset of oligos employed in any single ligation reaction (for example to avoid mismatch ligations), minimizing the difference in annealing temperature of members of a subset of oligonucleotide precursors, minimizing the difference in annealing temperatures of the overhangs of different double stranded subunits, whether to employ frame-shifting adaptors or single stranded oligo linkers and whether to minimize the degree of cross-hybridization among the hybrid forming portions of different oligos that make up a subset.

The number of oligos in a subset may vary. Preferably, the size of a subset is in the range of from 1 to 100, or from 2 to 100, more preferably in the range of from 1 to 50, or from 2 to 50 and still more preferably in the range of from 1 to 10, or from 2 to 10.

In a subset, wherein the degree of cross-hybridization has been minimalized, a duplex or triplex consisting of a subunit of the set and the complement of any other subunit of the set contains at least one mismatch. In other words, the sequences of the

oligos of such a subset differ from the sequences of every other oligo of the subset by at least one nucleotide, and more preferably, by at least two oligonucleotides. The number of oligonucleotide tags available in a particular embodiment depends on the number of subunits per tag and on the length of the subunit.

Single stranded oligo linkers having a sequence complimentary to the combined overhangs connect adjacent oligos in the target polynucleotide. Linkers may e.g. comprise 6 bases which connect two adjacent oligonucleotides each with a 3 base long overhang, one on the 3' end and the other on the 5' end, respectively.

In one specific embodiment of the invention the process of determining the assembly workflow is carried out by an algorithm. Candidate divisions of the sequence of the template are systematically examined to find the optimal number and assembly sequence of subsets to divide it into for synthesis in accordance with the method provided herein. Initially the entire template sequence is taken as a single subset, after which smaller and smaller subsets are formed with increasing numbers of candidate oligos in decreasing size until a partitioning is found that fulfills the subset criteria listed above.

The term "**assembly**" or "**assemble**" refers to the formation of an oligonucleotide or polynucleotide by linking and/or hybridizing single stranded and/or double stranded oligos. Specifically, said assembly is by any method of hybridizing ss nucleotide sequences, and/or a ligation reaction which is an enzymatic and/or chemical reaction. Preferably, said assembly is by an in vitro ligation method.

Assembly of the target ds polynucleotide can either be directly by hybridizing matching ss oligos, overhangs of ds oligos, or indirectly by hybridizing one or more suitable ss oligo linkers, wherein a ss oligo linker is contained in the library, and selected and transferred from the library to assemble any of said first, second or further reaction products.

For direct assembly oligonucleotide sequences are joined together by their single stranded oligo parts or overlaps (i.e. the overlapping parts or overhangs), such that the overlaps are included in the continuous sequence only once. Upon aligning two oligonucleotide sequences with an overlap, a continuous sequence is formed which has a length that is the length of both individual oligonucleotides taken together, minus the length of the overlap. Consequently, a continuous sequence is obtained which comprises a segment of each of the aligned oligonucleotides.

For indirect assembly, the target ds polynucleotide or any of said first, second or further reaction products are formed upon aligning ss oligos and joining them through single stranded linkers. For example, two oligonucleotides, each of e.g. 10 bases length, may be joined by an ss oligo linker of e.g. 6 bases length, such that 3 bases of the 3' terminal end of the first oligonucleotide align with the 3 bases of the 5' end of the ss linker and that 3 bases of the 5' end of the second oligonucleotide align with the 3 bases of the 3' end of the ss linker.

The terms "**first, second or further reaction products**" refer to the products of the ligation reactions performed in one or more reaction containments. In a first step, at least a first pair of matching oligonucleotides is transferred from the library into the first reaction containment using a liquid handler and the matching oligonucleotides are assembled in a ligation reaction thereby forming the first reaction product. Specifically, said first, second and further reaction products each comprise at least one overhang. Such overhang of a reaction product allows further assembly with another matching oligonucleotide in the direction of the overhang, e.g. to produce a new reaction product with an overhang, if the matching oligonucleotide included a first part that hybridizes with the overhang of said reaction product, and further included a second part that creates another overhang of the new reaction product. Alternatively, if the matching oligonucleotide only consisted of a part that hybridizes with the overhang over the full length of the overhang, such as to cover all nucleotides of the overhang, a blunt end can be created.

In specific cases, a ds target double stranded (ds) polynucleotide is produced which has a blunt end on one or both termini. Such blunt ends are preferably produced by hybridizing any terminal overhang with a matching ss oligo and/or ds oligo that hybridizes with the full-length of such overhang, without creating a new overhang, thus, producing a blunt end.

In said first step one or multiple pairs of matching oligonucleotides and one or multiple ss oligo linkers are transferred into said first reaction containment using a liquid handler and assembling the matching oligonucleotides thereby obtaining first reaction products. Preferably, the number of matching pairs transferred into said first reaction containment is any one of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or 25, preferably 4 and even more preferably 1, 2, or 3, and the number of ss oligo linkers transferred is any of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or 25, preferably 4 and even more preferably 1, 2 or 3.

In a second and further steps, at least a second and further pairs of matching oligonucleotides are transferred from the library into a second and further reaction containments, respectively, using a liquid handler and assembling the matching oligonucleotides thereby obtaining a second and further reaction products, respectively. In said second step one or multiple pairs of matching oligonucleotides and one or multiple ss oligo linkers are transferred into said second reaction containment. Preferably, the number of matching pairs transferred in said second step is any of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or 25, preferably 4 and even more preferably 1, 2 or 3 and the number of ss oligo linkers transferred is any of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or 25, preferably 4 and even more preferably 1, 2 or 3. In said further step, one or multiple pairs of matching oligonucleotides and one or multiple ss oligo linkers are transferred into said further reaction containment. Preferably, the number of matching pairs transferred in said further step is any of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or 25, preferably 4 and even more preferably 1, 2 or 3 and the number of ss oligo linkers transferred is is any of 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 or 25, preferably 4 and even more preferably 1, 2 or 3.

The number of steps and corresponding reaction products is unlimited. In order to synthesize large target ds polynucleotides steps a series of reaction products may need to be produced for assembly into the target polynucleotide, e.g at least 5, 10, 20, 50, 100, 500, 1.000, 5.000 or more may be necessary.

The terms "**hybridize**," "**hybridization**," "**hybridizing**," "**anneal**," and "**annealing**," as used herein, generally refer to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR, or the enzymatic cleavage of a polynucleotide by a ribozyme.

As used herein, the term “**ligation**” is intended to mean the process during which ends of two nucleic acid fragments are joined together by the formation of a covalent bond, for example a phosphodiester bond under appropriate conditions.

Ligation products, herein also referred to as reaction products, can be formed from both double stranded nucleic acids and single stranded nucleic acids. Double-stranded nucleic acids can be ligated by "sticky end" ligation or "blunt end" ligation. In sticky end ligation, staggered ends comprising terminal overhangs can hybridize to a ligation partner. In blunt end ligation, terminal overhangs are not present and successful ligation depends on transient associations of 5' ends and 3' ends. Blunt end ligations in general are less efficient than sticky end ligations, and various optimizations, such as adjusting concentrations, incubation times, and temperatures, can be applied to improve efficiencies. Single-stranded polynucleotides can also be ligated.

The ligation efficiency between two complementary sequences or sufficiently complementary sequences depends on the operating conditions that are used, and in particular the stringency. The stringency may be understood to denote the degree of homology; the higher the stringency, the higher percent homology between the sequences. The stringency may be defined in particular by the base composition of the two nucleic sequences, and/or by the degree of mismatching between these two nucleic sequences. By varying the conditions, e.g. salt concentration and temperature, a given nucleic acid sequence may be allowed to ligate only with its exact complement (high stringency) or with any somewhat related sequences (low stringency). Increasing the temperature or decreasing the salt concentration may tend to increase the selectivity of a ligation reaction.

The ligation reaction is performed by an enzyme, specifically a DNA ligase enzyme. The DNA ligase catalyzes the formation of covalent phosphodiester linkages, which permanently join the nucleotides together. In addition, T4 DNA ligase can also ligate ssDNA if no dsDNA templates are present, although this is generally a slow reaction. Non-limiting examples of enzymes that can be used for ligation reactions are ATP-dependent double- stranded polynucleotide ligases, NAD+ dependent DNA or RNA ligases, and single-strand polynucleotide ligases. Non-limiting examples of ligases are *Escherichia coli* DNA ligase, *Thermus filiformis* DNA ligase, *Thermus thermophilus* DNA ligase, *Thermus scotoductus* DNA ligase (I and II), CircLigase™ (Epicentre; Madison, WI), T3 DNA ligase, T4 DNA ligase, T4 RNA ligase, T7 DNA ligase, Taq ligase, Ampligase (Epicentre®Technologies Corp.), VanC- type ligase, 9° N DNA Ligase, Tsp DNA ligase, DNA ligase I, DNA ligase III, DNA ligase IV, Sso7-T3 DNA ligase, Sso7-T4 DNA ligase, Sso7-T7 DNA ligase, Sso7-Taq DNA ligase, Sso7-*E.coli* DNA ligase, Sso7-

Amp ligase DNA ligase, and thermostable ligases. Ligase enzymes may be wild-type, mutant isoforms, and genetically engineered variants. Ligation reactions can contain a buffer component, small molecule ligation enhancers, and other reaction components.

Preferably, the T4 DNA ligase is used in the ligation reaction. In the method provided herein the ligation reaction is performed under high-fidelity conditions that block side reactions and minimize mismatches.

Assembly into intermediate reaction products or into the target polynucleotide may be carried out using suitable ligation buffer solutions. The ligation buffer solution is e.g. an aqueous solution, typically in a nuclease free environment, at a pH that ensures the selected ligase will be active; typically, this is a pH of between about 7-9. Preferably, the pH is maintained by Tris-HCl at a concentration of between about 5 mM to 50 mM. The ligation buffer solution may include one or more nuclease inhibitors, usually calcium ion chelators, such as EDTA. Typically, EDTA is included at a concentration of between about 0.1 to 10 mM. The ligation buffer solution includes whatever cofactors are required for the selected ligase to be active. Usually, this is a divalent magnesium ion at a concentration of between about 0.2 mM to 20 mM, typically provided as a chloride salt. For T4 DNA ligase ATP is required as a cofactor. The ligase buffer solution may also include a reducing agent, such as dithiothreitol (DTT) or dithioerythritol (DTE), typically at a concentration of between about 0.1 mM to about 10 mM. Optionally, the ligase buffer may contain agents to reduce nonspecific binding of the oligonucleotides and polynucleotides. Exemplary agents include salmon sperm DNA, herring sperm DNA, serum albumin, Denhardt's solution, and the like. Preferably, ligation conditions are adjusted so that ligation will occur if the first and second oligonucleotides form perfectly matched duplexes with the bases of the contiguous complementary region of the target sequence. However, it is understood that it may be advantageous to permit non-pairing nucleotides on the 5' end of the first oligonucleotide and the 3' end of the second oligonucleotide in some embodiments to aid in detection or to reduce blunt-end ligation. Important parameters in the ligation reaction include temperature, salt concentration, presence or absence and concentration of denaturants such as formamide, concentration of the first and second oligonucleotides and type of ligase employed. Methods of selecting hybridization conditions for the reaction are known to those skilled in the art.

Preferably, ligation occurs under stringent hybridization conditions to ensure that only perfectly matched oligonucleotides hybridize. Typically, stringency is controlled by adjusting the temperature at which hybridization occurs while holding salt concentration at some constant value, e.g. 100 mM NaCl, or the equivalent. Other factors can be relevant, such as the particular sequence of the first and second oligonucleotides, the length of the first and second oligonucleotide and the heat lability of the ligase selected. Preferably, the ligation reaction is carried out at a temperature close to the melting temperature of the hybridized oligonucleotides in the ligation buffer. More preferably, the ligation reaction is carried out at a temperature within 10°C of the melting temperature of the hybridized oligonucleotides in the ligation buffer solution. Most preferably, the ligation reaction is carried out at a temperature in the range of 0 to 5°C below the melting temperature of the hybridized oligonucleotides in the ligation buffer solution.

Ligation may be followed by one or more amplification reactions. In some embodiments, the ligation products, or target polynucleotides are isolated or enriched prior to amplification. Isolation can be achieved by various suitable purification methods including affinity purification and gel electrophoresis. For example, ligation products, or target polynucleotides can be isolated by binding of a selective binding agent immobilized on a support to a tag attached to the capture probe. The support can then be used to separate or isolate the capture probe and any polynucleotide hybridized to the capture probe from the other contents of the sample reaction volume. The isolated polynucleotides can then be used for amplification and further sample preparation steps. In some embodiments, the capture probe is degraded or selectively removed prior to amplification of the circular target polynucleotides. Amplification of reaction products, or target polynucleotides can be achieved by various suitable amplification methods known to those skilled in the art.

The term "**derivative**" refers to an oligonucleotide or a polynucleotide differing from the original oligonucleotide or polynucleotide, but retaining essential properties thereof. Derivatives may e.g. be produced using a ds polynucleotide (e.g. DNA) as a starting material to engineer single stranded DNA, or complementary RNA molecule, to introduce one or more point mutations, or to bind heterologous moieties or tags by chemical and/or enzymatic means.

Generally, derivatives are overall closely similar, and, in many regions, identical to the original oligonucleotide or polynucleotide. As a practical matter, whether any

particular nucleic acid molecule or polypeptide is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Blosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 impaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for.

The library of the present invention may comprise thousands of oligos necessary to cover the whole sequence space. Each of the oligonucleotide library members may be physically placed in a compartment. All compartments may be conveniently provided within one or more parts of a device, which together are provided as “**array device**”. Such array device may be any one or more of a microtiter plate, microfluidic microplate, set of capillaries, microarray or a biochip, preferably a DNA or RNA biochip. Oligos may be conveniently transferred by automated means, e.g. either robotically or via dedicated

fluids using, for example, an automated liquid handler, from such compartments into other compartments herein referred to as reaction compartments, i.e. from one vessel to another. In order to facilitate time efficient assembly of polynucleotides, hierarchies of reactions and respective vessels may be employed corresponding to frequency of use of oligonucleotide library members. The transfer to a new vessel involves the physical movement of a device that picks one or more molecules of an oligo from the respective location, or the pneumatic/hydraulic deposition though microfluidics. Due to the large number of oligos required to theoretically build any given sequence, most spatial distributions of library members in the library would incur into wasted time and resources due to scanning of the library and lengthy travel times of the liquid handler. However, by using a specific distribution of the library members, it can be ensured that that there is minimal movement according to a target sequence. One example is to store into micro-well plates where the first plate comprises the most common pair combinations of oligos, in decreasing order until the last micro-well plate which contains the least-frequently used library members.

Specifically, said separate library containments are **spatially arranged** in a two-dimensional order, wherein the individual compartments are located within a device at defined coordinates within the x- and y-axes. The order is specifically predefined by a parameter which primarily serves to shorten synthesis time. Preferably, said parameter is frequency of use, placing those oligos in close proximity to each other which frequently form a matching pair in DNA sequences, e.g. naturally occurring or commonly used in target ds polynucleotides or fragments thereof. Even more preferably, said separate library containments are spatially arranged in a three-dimensional order, wherein the individual compartments are located within a device at defined coordinates within the x- , y- and z-axes. The order is specifically determined by the frequency of use, placing those oligos in close proximity to each other which frequently form a matching pair in naturally occurring DNA sequences. Specifically, the spatial arrangement of library members may depend on any one of, or a multitude of the following parameter: frequency of use of the oligonucleotides, frequency of occurrence of the oligonucleotides in natural DNA sequences, frequency of occurrence of the oligonucleotides in a set of designed DNA sequences, minimization of handling or access time by the microfluidic device, minimization of operational cost or of amount consumables by the microfluidic device.

In a specific example, said separate library containments are micro-well plates, arranged as stacked plates, optionally barcode labelled, and accessible by an automated microdroplet handler. Library members may be conveniently stored in said stacked micro-well plates, wherein the order and stacking is according to decreasing frequency of use.

As used herein the terms "**liquid handler**", "**automated handler**" or "**microdroplet handler**" refer to any device used in a method of liquid handling, preferably, automated liquid handling, preferably a device as used in sensor-integrated robotic systems. As low-volume dispensing becomes increasingly common in life science, microsyringes have emerged which have a high level of precision with hermetic seals. Some manual or electronic holders are designed to precisely control the piston displacement to ensure the accuracy of the dispensed volume. Besides the syringe, a pipette is another popular tool for liquid handling. The dispensed volume can be at the micro- or sub-microliter level. Multichannel pipettes are recommended for multirouting pipetting at one time. There are both fixed- and adjustable-volume pipettes on the market. The former is more accurate and precise, whereas the latter has a larger scope of applications because the operator can choose different volumes according to need. Besides, high throughput has become critically important in life science research. One of the representative applications is microarray printing. This technology creates an array of biosample spots each at the nanoliter scale to enable the analysis of large numbers of experiments in parallel with only tiny quantities of samples. The process of spotting thousands of biosamples is almost an impossible task with a handheld dispensing tool, making robotic liquid handling an important aspect.

Robotic workstations have multiple advantages over manual liquid handling since robots can work without fatigue, increase the throughput, perform consistently, and ensure accuracy and precision. According to the requirements for the platform with integration and multifunction, there are still more complex systems in which the liquid-handling task is only one part of the function. The generic architecture of liquid-handling may be built up as follows. First, the control center controls a robot that moves between the dispensing part and the washing station of the robotic workstation. The washing station is used to clean the dispensing head for lengthening its life and for ensuring the safety of the sample. Liquid samples are expelled from the dispensing head and deposited on the substrates for further processing. Sensors are incorporated to monitor

the status of the dispensing part such that feedback control can be performed by the control center. Sensors are not always installed on all the workstations but are more and more used to construct the feedback loop for delivering a better performance.

The term “**capillaries**” refers to any of glass capillaries, microfluidic capillaries and autonomous microfluidic capillary systems. Capillary microfluidics are important tools in many different fields. Due to their axisymmetric flow and ability to withstand organic solvents, when compared with their lithographically fabricated polydimethylsiloxane (PDMS) counterparts, glass capillary devices possess advantages for microfluidic applications. In particular, a circular tube is inserted into a square outer flow channel, which greatly simplifies alignment and centering of these devices. These devices can produce small and large droplets, ranging from 10 to multiple hundreds in μm size.

The term “**microtiter plate**” refers to any of well plates, multi-well plates or micro-well plates. These plates are commonly manufactured in a 2:3 rectangular mix with 96, 384, or 1536 wells, although other cavity configurations are available. Some of the other sizes, far less common, available are 6, 24, 3456, and 9600 wells. The wells of the microplate typically hold between tens of nanoliters to several milliliters of liquid.

The term “**microarray**” refers to a supporting material (such as a glass or plastic slide) onto which numerous molecules or fragments usually of DNA or protein are attached in a regular pattern. More specifically, it refers to microscope slides that are printed with thousands of tiny spots in defined positions, wherein said spots are capable of binding DNA or RNA. Such slides are often also referred to as biochips, DNA chips, RNA chips or gene chips. Such microarrays can bind DNA or RNA in a covalent or non-covalent manner and can thus serve as array devices in which oligos are stored in pre-defined locations, ie spots.

“**Microfluidic devices**” enable the manipulation of discrete fluid packets in the form of microdroplets that provide numerous benefits for conducting biological and chemical assays. Among these benefits are a large reduction in the volume of reagent required for assays, the size of sample required, and the size of the equipment itself. Such technology also enhances the speed of biological and chemical assays by reducing the volumes over which processes such as heating, diffusion, and convective mixing occur. Once the droplets are generated, carefully designed droplet operations

allow for the multiplexing of a large number of droplets to enable large-scale complex biological and chemical assays.

The term "**microfluidic microplate**" refers to a combination of microfluidic technology with standard SBS-configured 96-well microplate architecture, in the form of microfluidic microplate technology. A microfluidic microplate allows for the improvement of essential workflows, conservation of samples and reagents, improved reaction kinetics, and the ability to improve the sensitivity of the assay by multiple analyte loading (Kai et al., 2012).

The term "**methyltransferase**" as used herein, can refer to any of DNA methyltransferase, RNA methyltransferase, protein methyltransferase and histone methyltransferase. Methyltransferases can be further subdivided into class I, all of which contain a Rossman fold for binding S-Adenosyl methionine (SAM) and class II methyltransferases, containing a SET domain, which are exemplified by SET domain histone methyltransferases, and class III methyltransferases, which are membrane associated.

The term "**CRISPR/Cas9**" refers to a gene editing method well known to those skilled in the art, as well as modifications thereof. Such modifications include, but are not limited to, fusion of a nuclease-dead Cas9 (dCas9) to cytidine deaminase, enabling site-specific conversion of cytidine to uracil and mutations to the Cas9 protein, which generate versions of the Cas9 protein that only create single-strand DNA cuts (nicks).

The terms "multiplex automated genome engineering" or "**MAGE**" refer to a technique which generally includes introducing multiple nucleic acid sequences into one or more cells such that the entire cell culture approaches a state involving a set of changes to a genome or targeted region. The method can be used to generate one specific configuration of alleles or can be used for combinatorial exploration of designed alleles optionally including additional random, or non-designed, changes.

ssDNA-binding protein mediated recombination, homologous recombination and MAGE-based methods typically include introducing multiple oligonucleotides into a cell including the steps transforming or transfecting cells using transformation medium or transfection medium including oligonucleotides, replacing the transformation medium or transfection medium with growth medium, incubating the cell in the growth medium, and repeating the steps if necessary or desired until multiple nucleic acid mutations have

been introduced into the nucleotide sequence of interest. Increasing the number of cycles of mutagenesis generally increases the diversity of mutations introduced.

MAGE particularly employs a highly efficient lambda phage red recombination system (the λ Red System) which is a process by which the genome of a cell is reprogrammed to perform desired functions via a form of accelerated, directed evolution. The λ Red System includes β, γ, and exo genes, whose products are called Beta, Gam, and Exo, respectively. Gam inhibits the host RecB,C,D exonuclease and the SbcC,D nuclease activities, so that exogenously added linear DNA is not degraded. The Exo protein is a dsDNA-dependent exonuclease that binds to the terminus of each strand while degrading the other strand in a 5′ to 3′ direction. Beta binds to the resulting ssDNA overhangs, ultimately pairing them with a complementary chromosomal DNA target. The λ Red System has been widely utilized for specific gene inactivation in *E. coli*, Salmonella, Citrobacter and Shigella species, and for introducing small biological tags or single genes into these chromosomes.

The term “conjugative assembly genome engineering” or “**CAGE**” refers to a precise method of genome assembly using conjugation to hierarchically combine distinct genotypes from multiple *E. coli* strains into a single chimeric genome. CAGE permits large-scale transfer of specified genomic regions between strains without constraints imposed by in vitro manipulations. Strains are assembled in a pairwise manner by establishing a donor strain that harbors conjugation machinery and a recipient strain that receives DNA from the donor. Within strain pairs, targeted placement of a conjugal origin of transfer and selectable markers in donor and recipient genomes enables the controlled transfer and selection of desired donor-recipient chimeric genomes. By design, selectable markers act as genomic anchor points, and they are recycled in subsequent rounds of hierarchical genome transfer.

“**Ago**” refers to the Argonaute protein which has shown to provide DNA-based DNA interference, where a single-stranded DNA guide could direct Ago-based cleavage of a plasmid DNA target. A key advantage is that, unlike CRISPR–Cas9, there is no requirement of a Protospacer adjacent motif (PAM).

Zinc-finger nucleases (**ZFNs**) and transcription activator-like effector nucleases (**TALENs**) recognize DNA target sites, ranging from 25 to 40 bp in size, in a sequence-specific manner through their DNA-binding domains and generate staggered double strand breaks through the action of FokI nuclease domains on opposite DNA strands.

"**Meganucleases**", also known as homing endonucleases, recognize a specific DNA sequence between 14 and 40 bp upon which they cut and induce a DSB. The efficiencies of meganucleases are reasonably high, and they only require a single custom biopolymer for each target site.

"**Tyrosine/serine site-specific recombinases**" or "Tyr/Ser SSRs", which typically recognize target sequences between 30 and 40 bp in length, were one of the earliest genome-engineering tools to enable homology-directed repair (HDR) in mammalian genomes. Briefly, the target site comprises three parts, a short DNA sequence flanked by two inverted repeats, and recombination can occur between a pair of target sites, where the DNA sequence between the target sites can be deleted, inverted or replaced. Notably, whereas Tyr SSRs utilize a mechanism of strand exchange without creating double strand breaks, Ser SSRs do create double strand breaks, but unlike simpler designer double-strand nucleases, SSRs require concerted cleavage and re-ligation with the donor DNA present.

The foregoing description will be more fully understood with reference to the following examples. Such examples are, however, merely representative of methods of practicing one or more embodiments of the present invention and should not be read as limiting the scope of invention.

EXAMPLES

In the following examples it is described how the library of oligos and the library of polynuceotides is produced, how it is handled and its contents and properties are verified. Furthermore, it is described how a polynucleotide is synthesized according to the method provided herein.

Example 1: Production of the oligonucleotide library

1.1 Determining the spatial structure of the genetic information

A. First, all the sequences of oligonucleotides that are to be included in the library have to be listed. These sequences are pre-computed from an input set of sequences that cover all potentially desired targets. This information can come from a diversity of criteria, such as a subset of possible combinations (e.g. all heptamers, all octamers, etc.), predicted outcome of the digestion of a genome with a set of restriction enzymes or any other computational criterion.

In this example 400 randomly sampled coding genes from the Human reference genome, for a total of ~5 million base pairs, served as basis for the library. Ideally, all reported sequences, not only in the Human genome but, for example in the gene bank, would be taken and processed in the same way as described below and in Fig. 1.

Each reference sequence was systematically partitioned into oligonucleotide dimers of lengths between 8 and 26 bp. Similarly, the reverse-complement of each was computed and it was systematically partitioned into oligonucleotides of lengths between 8 and 26 nt. Next the same process was carried out repeatedly by shifting the sequence first 1, then 2, then 3, up to 15 nucleotides. The library size was fixed to be 536,736 oligos and contains: all possible octamers, weighted by their normalized occurrence with a score between 0 and 1, and longer oligos between 9 and 26 bp, which were prioritized by their incidence in the reference set.

The number of matching pairs in this database is combinatorially high, even if we only count those that overlap by 4 bp. Generally speaking, variant sequences should be processed in a similar way, which increases the multiplicity in a non-linear way. For instance, a window of about 100 bp containing only 16 polymorphic sites adds over 400 oligos and almost 20.000 matching pairs (Fig. 1). The combinatorics imply that when considering more variable sites, the oligo library is populated in a non-polynomial fashion with the number of sequence variants that are considered in its design.

Some of the oligos were conserved across haplotypes and were allocated in the library as paired elements (Fig. 1B). The oligos spanning variable sites (and depending on the extent of the variability) were kept independently as ssDNA elements.

B. The 2-dimensional arrangement of the library was determined by sorting library members according to a preferred criterion. Here, 16-mers were sorted first by sequence shift and second according to their order of first occurrence in the sequence and by alternating conjugate pairs. When alternative oligos occurred for a given position they were sub-sorted according to the frequency of their occurrence. Oligos that are conserved across all input sequences were allocated with their conjugated pair in the same position.

Alternative criteria that reflect both the individual usage of an oligo and also the relative usage of its matching pairs could be lexicographically, length, adjacency of matching pairs, frequency, or any other arbitrary but known way.

C. Next, the first sequence was allocated into a 2-dimensional array corresponding to the position(s) in a 1536 micro-well plate where the actual oligo(s) were to be placed.

D. The subsequent oligos were added until the 1535 remaining wells were all occupied by oligos in an order reflecting the sorting preference of step B.

E. Step C was then repeated with the next 1536 oligos, and so forth until all 33.120 or more oligos were distributed in micro-well plates.

F. The information was stored digitally to keep track of the location of each oligo. At a later step this served two purposes: first, it functioned as a look-up table for easier access to oligos and, secondly, it allowed monitoring of usage and access frequency of every oligo in order to keep track of available volumes.

1.2. Synthesis of the library

Once the sequence was properly structured the actual synthesis of each oligo was carried out. Physically, the library consists of 326 1536-micro-well plates (Corning 1536 well plates, Sigma Aldrich Product Nr. CLS3726-50EA), made of polypropylene (polypropylene is preferred, however any material that minimizes DNA absorption to the surface can be used). Each of the plates was labeled and/or barcoded unambiguously for easy access and for content bookkeeping.

Each produced oligo was located in its predefined plate as determined above. In this example, the oligos were phosphorylated at the 5’ end. Other applications might require treatment with other modifications such as di- or tri-phosphates, biotin, TEG or thiol modifiers, etc. at the 3’, 5’ or both ends, or methylations, etc. Oligos were kept in aqueous solution (nuclease free ddH20 or TRIS 10mM pH 8.0 and 1mM EDTA) at a volume of 10 µL per oligo per micro-well at a concentration of 200 µM (Sambrook and Russell, 2014).

The actual production of the library can be carried out with standard methods of molecular biology by digesting with nucleases naturally occurring DNA, chemically constructed with oligo-synthesizers, etc. followed by separation and purification with HPLC, capillary electrophoresis or other techniques. Because the synthesis and modification of oligonucleotides is standard, it can also be outsourced from many services. According to this example, the library was produced using automatic DNA synthesizers that implement iteratively the chemical reaction of deoxynucleoside

phosphoramidites to covalently bond mononucleotides to a solid-phase-attached polynucleotide (Beaucage and Cartuthers, 1981).

The library was stored at -20°C when not in use for short periods, or -80°C for long term storage.

1.3 Usage of the library

A. The library was thawed by placing the plates at 3°C for at least 60 minutes and then kept on ice or on a cooler plate at a temperature between 3-5°C.

B. Each micro-well plate was vortexed for 30 seconds in an orbital mixer at 2500 rpm and spun down in a centrifuge for 1 min at 900 rpm.

C. Using a low-volume micro-droplet handler (TPP Lab Tech Mosquito X1) 100 nL (recommended range: 50-250 nL) were transferred to a fresh 384 micro-well plate (other capacities such as 96 or 1536, or surface can also be used) that contained 1.8 µL of a solution or solution droplet (recommended range is of 1-5 µL) where the oligos were combined and/or further reacted.

D. In the digital database the used volume of the respective micro-wells was annotated to ensure there was always enough of all required oligos for a further round of usage. Note that some liquid handlers provide accurate and real-time measurement of the used and remaining volumes in each accessed well. This function may aid a more accurate tracking.

E. Once the library had been used, it was returned to storage at -80°C.

1.4 Determining the properties of the library

The main properties defining a library of the present invention are i) defined lengths of the oligonucleotides, ii) single stranded and/ or double-stranded with at least one overhang and iii) a certain number of oligos. The main properties of the library used in this example were i) lengths of oligonucleotides ranging from 8 to 26 nt, ii) presence of single stranded and double stranded oligos with at least one overhang and iii) at least 33.120 oligos are included in the library.

It is desirable to be able to verify that these properties hold for purposes of quality control.

I. Verifying the length of the oligonucleotides.

Using a micro-droplet handler aliquots of 5-10 nL of each micro-well were taken and pooled into a common solution. Alternatively, random aliquots were taken and pooled into 10 different pooled solutions in such a way that each oligo is in only one of the pools. The pool or pools were mixed by vortexing. A small aliquot of a few µL per pooled solution was run through capillary electrophoresis (Kemp, 1998). Alternatively, the samples can be analyzed on a 25% acrylamide gel, and compared with a standard ladder ranging from 6 to 24 bp of ssDNA.

II. Verifying the structure of the oligonucleotides present in the library.

ss oligos, ds oligos and ds oligos with ss overhangs were differentiated by comparing denatured but otherwise untreated samples of a given oligo with a sample treated with an exonuclease, such as E. coli Exonuclease I (e.g. Thermo Scientific Exonuclease I, product nr. EN0581). This enzyme digests ssDNA to mono-nucleotides and di-nucleotides, but leaves dsDNA intact (Lehman and Nussbaum, 1964). Therefore the untreated and treated samples gave one of the following results when inspected through capillary electrophoresis:

- The untreated sample showed a single band within a range of 6-26 nt, and the treated sample showed no bands. This implied that the original sample consisted of ss DNA.
- The untreated sample showed a single band within a range of 6-26 nt, and the treated sample showed the same band. This implied that the original sample consisted of ds DNA (with no overhangs).
- The untreated sample showed two different bands, both within a range of 6-26 nt, and the treated sample showed a single band whose length coincided with the smallest band of the untreated sample. This implied that the original sample consisted of a dimer of DNA that has one overhang. The length of the overhang is the difference of the sizes of the two bands of the untreated sample, and the length of the ds part is that indicated on the treated sample.
- The untreated sample showed a single band, within a range of 6-26 nt, and the treated sample showed a single band whose length was smaller than that of the untreated sample. This implied that the original sample consisted of a dimer of DNA that had two overhangs of equal size. The length of the overhangs is the difference of the sizes of the treated and untreated samples, and the length of the ds part is that indicated by the band of the treated sample.

- The untreated sample showed two bands, both within a range of 6-26 nt, and the treated sample showed a single band whose length is smaller than both of the untreated samples. This implies that the original sample consisted of a dimer of DNA that had two overhangs of different sizes. The lengths of the overhangs are determined by the difference of the sizes of each band relative to the size given by the treated sample, and the length of the ds part is that indicated on the treated sample.

Other analytical techniques, such as HPLC can also reveal in their spectra the composition of an untreated sample, directly indicating the presence of a single species of DNA or of two of them, providing direct evidence of the nature of the oligonucleotides in one well of the library. Also, circular dichroism could be used to distinguish among single and double stranded DNA and even dsDNA with overhangs.

III. *Verifying the number of oligonucleotides and number of matching pairs.* A sample of 50-100 nL of the contents of each micro-well was pooled into a common solution annealed by heating at 95°C for 3 minutes and allowed to cool down at least to room temperature or down to 16°C. The corresponding buffer necessary for ligation was added including necessary cofactors such as Mg+, ATP, etc. Enough ligase (e.g. T4 ligase, NEB, product nr. M0202) to catalyze the reaction (1U per µL of reaction solution) was added. The reaction mix was incubated for an hour at room temperature or overnight at 16°C.

By hypothesis, if there are enough matching pairs, the ligase will covalently link them, resulting in DNA molecules of a range of lengths with random sequences. The distribution of lengths was resolved by using electrophoresis with Agarose 2-4% on TAE. Samples run together with a suitable ladder (on a separate lane; recommended 50 or 100 bp) showed a smear of DNA along the sample lane with no discrete bands. A narrow range of approximately 100-200 bp was isolated by cutting the gel guided by the ladder (Sambrook and Russell, 2014; Ch. 5). Following standard protocols for gel-extraction, the DNA from the excised agarose block was isolated (e.g. Zymoclean gel DNA recovery kit, Zymo research, product nr. D4001T). After purification, the sample was deep-sequenced in order to determine the different sequences in the pool (Bentley et al., 2008).

The following analysis was performed in order to estimate the number of oligos and of matching pairs. If the starting material for the reaction consists of DNAs between 6 and 26 nt, and the sequences are not highly repetitive it could be concluded that, in

average, there are at least 2 x N x 100/26 oligos (N being the number of reported sequences), and up to 2 x N x 200 / 6 oligos and almost as many matching pairs. Further bioinformatic analyses were used to extract the sequences of the oligos, as follows. The first 6 nt of one of the sequences were taken, a search & match for this pattern in the complete sequence pool was performed, and the number of occurrences was annotated. This was repeated for 7 nt, then for 8 nt and so on until 26 nt. By using a statistical T-test, it was determined which number is significantly different from random occurrences. This distinctive pattern was stored in a list of putative oligos and all its occurrences were eliminated from the database. This procedure was repeated with the remaining sequences until only DNA sub-sequences between 6 and 26 nt, that cannot be further partitioned, and which are now added to the list of patterns, were left. The number of identified oligos was called M. Since these oligos were linked to at least one other oligo, it implied that, together with their partial complements in the opposite strand, consecutive oligos were part of matching pairs. Hence there were at least as many matching pairs as number of identified oligos, except for those at the termini. For instance there were on the order of M-N matching pairs. Statistical analysis and bootstrapping simulation was performed to determine whether the identified number can be expected to be a subsample of a larger set of at least 33.120 oligos.

Example 2: Synthesis of a target DNA molecule of 128 bp using the oligonucleotide library of Example 1

In this example it was demonstrated how to synthesize a sequence of 128 bp by means of the method proposed herein. Figure 2A shows the sequence of interest (SEQ ID NO:1), which was termed DISCOVER, and was built from 16 matching pairs (Fig. 2B) that formed 8 ds oligos of 16 nt with 4 nt overhangs on each strand (see Fig. 2C) and 8 complementary sites. Each ds oligo is denoted by the letters D, I, S, C, O, V, E, R and their constituting leading and lagging strands by + and - superscripts, respectively. The oligos were part of the library generated in example 1. It has the following properties: all oligos were phosphorylated at the 5’ end, they were provided at a concentration of 200 µM on nuclease free ddH20 and the used oligos were single-stranded and pure.

A. Preparing the annealing solutions.

In a reaction tube 252 µL on ddH20 with TRIS-HCl (50 mM), MgCl2 (10 mM), DTT (10 mM) and ATP (1 mM) were prepared. The pH was set to 7.5. Some commercial

buffers are ready to mix in H20 such as New England Biolabs' Ligase Reaction Buffer, product nr B0202S, and readily contain the ATP necessary for the ligase activity. Solution was mixed well by vortexing. 28 µL of this solution mix were dispensed into to 8 micro-wells in a 4x2 array. 1 µL of each oligo was transferred to a predefined micro-well of the plate and mixed well by pipetting:

D+ and D- to well A1

I+ and I- to well A2

S+ and S- to well A3

C+ and C- to well A4

O+ and O- to well B1

V+ and V- to well B2

E+ and E- to well B3

R+ and R- to well B4

B. Annealing.

The plate was sealed and incubated in a thermocycler for 5 min at 95°C allowing the matching pairs of ss oligos to anneal. The temperature was then decreased to 16°C with a ramp function that diminished the temperature by 1°C per minute. Once finished the double stranded oligos were kept at 16°C.

C. Preparing the ligation solution.

The ligation solution was prepared on ice by mixing, in the following order, 13.3 µL of nuclease free ddH20, 2 µL of ligase buffer and 4 µL of ATP for a final concentration of 1 mM. The ligation solution was mixed well by vortexing and spun down. 0.7 µL of T4 Ligase (NEB, product nr. M0202) were added for a total of 1 unit per µL of final solution and mixed well by gently pipetting. The solution was kept on ice until needed. 2.5 µL of the ligation solution were transferred to each of the 8 micro-wells containing the ds oligos of B and mixed by pipetting. Afterwards the plate was sealed again.

D. Ligation rounds.

For the first round of ligation the following wells were merged as follows: D+I, S+C, O+V, E+R. This was achieved by transferring the contents of one well into the other (transferring the contents of both wells into a new well is also possible). A scheme was used where the leftmost contents are transferred to the rightmost (Fig. 3A). The ligation reaction mix was incubated at 16°C for at least one hour. This process was repeated by merging the wells DI+SC and OV+ER (Fig. 3B) and again each was

incubated for one hour. For the final ligation round the wells DISC+OVER were merged and incubated for another hour (Fig. 3C). The final volume containing the 128bp product was 140 µL.

E. Purification.

An agarose gel 2% (1 mg agarose in 50 mL TAE supplied with 5 µL of SYBR Safe DNA stain) with a comb of 11 wells was prepared. 4.5 µL of 50 bp ladder (New England Biolabs product nr. N3236 or Invitrogen product nr. 10416014) was added on the first lane and the 140 µL of solution obtained under step D were distributed across the remaining wells. The gel was run at 85 V, 200 mA and 12 Watt for 50 minutes. After the electrophoresis was completed, the gel was placed over a UV trans-illuminator and the bands of the gel that correspond to the 128 bp fragment were excised. Purification of these bands can be performed with commercial Kits for said purpose (e.g. Zymoclean, see previous example), or following any standard protocol for this purpose.

F. Amplification.

To further increase the amount of product, the product obtained under step D was amplified by PCR (Sambrook and Russell, 2014; Chapter 8). The starting 16 nt D- and R+ were used as primers for said amplification. After amplification, the construct was freed from enzymes and primers and separated into two aliquots, one for further use, which was labeled and stored at -20°C and the other one was used to sequence-verify the construct.

Figure 4 depicts an acrylamide gel showing intermediate steps and the final result of this process. In Lanes 6 and 7 the upper band corresponds to the 128 bp target ds polynucleotide. This construct was isolated (from a 2% agarose gel; not shown), purified, amplified and both strands Sanger-sequenced. The resulting sequences were identical to the target and to its reverse complement.

Example 3: Post-processing of target DNA sequences for complex sequences or RNA synthesis.

3.1 Design of proxy ds polynucleotide

In this example a ds polynucleotide was synthesized whose workflow would normally include an ambiguous step such as self-complementary oligo dimer (e.g. Fig. 5A). Since such a self-complementary dimer has to be excluded from the workflow to avoid unwanted runaway reactions, a template sequence was devised by replacing the

self-complementary elements with different bases, in such a way that the resulting assembly workflow was unambiguous. According to this template, a proxy ds polynucleotide was synthesized.

In Fig. 5A the sequence of interest is depicted. The underlined parts indicate those parts of the sequence capable of self-complementation and self-polymerization. In order to avoid these sequences a template sequence (Fig. 5B) was designed which comprises two base pair modifications that span three oligos.

The proxy ds polynucleotide was synthesized with the method presented herein as demonstrated in Example 2. The proxy sequence was chosen to coincide with the oligos O-, and V+ of Example 2, and, consequently, its synthesis proceeded exactly as described above.

Once the proxy ds polynucleotide was synthesized, a ds polynucleotide which has a sequence that is identical to the sequence of interest was produced as follows. The principle of directed mutagenesis was applied, that, upon PCR amplification, replaced the part of the target sequence that was excluded in the synthesized proxy ds polynucleotide with the original target sequence.

After synthesis was completed, and the 128 bp proxy ds polynucleotide was purified, a PCR reaction was prepared. In this reaction mix not only the 3’ end primers but also a pair of “mutagenizing primers” (AttB) were included. These mutagenizing primers had, on either side of the mutagenized element (in this example, the three bases), ten nucleotides that were fully overlapping with the proxy sequences. With these provisions, a standard PCR was performed, to retrieve the ds polynucleotide which has a sequence that is identical to the SOI, (Sambrook and Russell, 2014; Ch. 13) by using, in this example, commercial kits that standardize the reaction conditions and reagents (Taq PCR Kit, New England Biolabs, product nr. E5000S).

3.2 Production of RNAs

RNA molecules with a given target sequence also have to be produced using proxy ds polynucleotides. This was done in two steps. First, the reverse-complement sequence of the RNA sequence of interest (i.e. the DNA sequence) had to be computed. The DNA sequence is the sequence that will be synthesized. Second, a specific promoter sequence was integrated into the template DNA sequence in order to be recognizable by DNA-dependent enzymes that will later transcribe the DNA into the RNA

(Rio, 2011). In this example we used a T7 RNA polymerase I system. The necessary steps are:

A. *Design of DNA template.* For a given RNA sequence of interest, its DNA reverse complement was computed including the T7 RNA pol promoter sequence TAATACGACTCACTATAG (SEQ ID NO: 24) at the 5' end of the reverse complement.

B. *Synthesis of proxy ds polynucleotide.* The proxy ds DNA polynucleotide was synthesized according to the DNA template of step 3.2.A as described in Example 2 (see also Examples 1 and 3.1). After synthesis of the proxy DNA its ends were modified to generate blunt ends. The ss overhangs were blunted by incubation at 25°C for 15 min with one unit per µg of DNA of *E.coli* Polymerase I Large Klenow fragment in the presence of 33uM of each dNTPs and inactivated by adding ETDA 10 mM and heating at 75°C for 20 min (obtained from New England Biolabs, product nr. M0210; Sambrook and Russell, 2014; Ch. 12). Next, the proxy ds polynucleotide was purified and, amplified and purified again: a minimum amount of 1 µg DNA is required for the RNA synthesis reaction described below.

C. *Transcription, post-processing and purification of RNA.* Standard protocols for RNA transcription were followed (for example, the HiScribe T7 ARCA mRNA Kit, New England Biolabs, product nr. E2060, amongst several others) which included the synthesis of the RNA from the proxy DNA. For synthesis of the RNA from the proxy DNA the following protocol was applied:

1-3 ug of DNA were dissolved in a solution composed of 2µL of 2X rNTP Mix, 2µL of T7 RNA Polymerase Mix and 18uL of Nuclease Free Water, followed by incubating at 37°C for 30 min, thereby producing the RNA molecules. The reaction was stopped by adding 2µL of DNAse and incubating 15 min at 37°C to digest the template DNA and then the resulting RNA was purified using spin columns as described in previous examples.

<u>Example 4: Synthesis of a target DNA molecule of 608 bp using the oligonucleotide library of Example 1</u>

In this example it is demonstrated how to synthesize a target ds polynucleotide of 608 bp (SOI is Sequence "Ribbon_test_608", SEQ ID NO:26) using the method provided herein. The oligos were part of the library generated in example 1. Oligos had the same properties as in example 2.

The oligos were prepared in an asymmetric way in the reaction plate in order to obtain partial constructs of different sizes at the fourth ligation. The 608 bp sequence is achieved by completing four ligation rounds to obtain one reaction product of 128 bp, and three of 160 bp, which will then were purified and subject to two more ligation rounds, thereby obtaining each strand of the 608bp target ds polynucleotide.

4.1 Preparing annealing solutions

A master mix of 864 µL of annealing solution was prepared, constituted by 772 µL of ddH20 and 92 µL of T4 ligase buffer. 21.6 µL of this solution mix were dispensed into to 38 micro-wells. 0.7 µL of each oligo (in 150 µM) was transferred to a predefined micro-well of the plate and mixed by pipetting.

Partially complementary ss oligos were derived from the library of Example 1 and placed in specific wells on a 96-well plate as indicated in Fig. 6. For simplicity, the oligos were named according to the position on the plate where they are placed for annealing. As in example 2, the leading and lagging strand are denoted by + and - superscripts respectively; see sequences with SEQ ID NO:27 to 102 in FASTA format. Note that wells in rows E-G, columns 2-7 remained empty on purpose.

4.2 Annealing

Annealing was conducted as in example 2.

4.3 Preparing the ligation solution

The ligation solution was prepared similarly as in example 2 but adjusting the quantities for 80 µL, enough for 38 reactions wells. Namely: 7.2 µL of Nuclease free ddH2O, 8 µL of Ligase buffer, 40 µL of ATP and after vortex mixing, 24.8 µL of T4 ligase, mixed by pipetting.

2 µL of the resulting solution were transferred with a dispenser to each of the 38 reaction wells in B to prepare them for ligation, followed by gentle mixing by using a multichannel pipette.

4.4 First four ligation rounds

For the first round of ligation the complete contents were transferred from wells in rows A and C into rows B and D of columns (1-7) respectively, and from wells E1 and G1 into F1 and H1, respectively. Transfers were done with a multi-channel pipette,

followed by gentle mixing. This scheme is equivalent as in example 2: leftmost contents are transferred to rightmost wells. The plate was sealed and the reaction mix was incubated for at least one hour at 16 °C in a thermocycler. Note that wells E to G from rows 2-7 remained empty.

For the second round of ligation the plate was opened and the complete contents were transferred by pipetting from wells in rows B into row D of columns (1-7), and from well F1 into H1 and mixed. The plate was sealed again and incubated for at least one hour at 16 °C.

For the third round of ligation the plate was opened and the complete contents were transferred by pipetting from wells in row D into row H of columns 1-7 by pipetting followed by mixing. The plate was sealed again and incubated for at least one hour at 16 °C.

For the fourth round of ligation the plate was opened and the complete contents were transferred by pipetting from wells H2, H4 and H6 into wells H3, H5 and H7, respectively, followed by mixing. Note that well H1 was left untouched. The plate was sealed again and incubated for at least one hour at 16 °C.

4.5 Interim purification

Three agarose gels were prepared as in example 2, part E, with a comb of 7 lanes, including the 50 bp ladder. The contents in well H1 from part D was distributed into six lanes of a gel (33 µL on each lane). The contents H3, H5 and H7 were distributed into three lanes each of the other two gel (41 µL on each lane). Gels were ran as indicated in example 2, part E, followed by bands excision as required (128 for the lanes 2-4 of gel 1, and 80 bp for the remaining lanes of gel 1 and of gel 2). Purification was performed as described in example 2, part E, pooling in the same purification column samples containing the same synthons. Each of the 4 samples was eluted with 10 µL of ddH20 (as indicated in the Zymoclean purification kit), warmed at 35 °C to improve elution efficiency. The contents were transferred to a stripe of PCR reaction tubes and labeled from S1 to S4.

A sample of 0.5 µL form S1 and from S4 was taken and diluted in 0.5 µL of ddH2O. These samples were used to estimate the DNA concentration through specrophotometry at 260 nm (nanodrop 2000, Thermo Fisher Scientific), to give 1.52 µg/µL and 1.98 µg/µL

respectively. It was assumed that samples S2 and S3 were on a similar range of molar concentrations.

4.6 Preparing the ligation solution

Samples were placed in ice. To the samples S1 and S4 0.5 µL of ddH20 were added (to compensate the 0.5 µL taken for measurements in part E). Ligations reactions were prepared by adding to each sample 1.14 µL of ligase buffer. 0.3 µL of T4 ligase were added do S1 and S3. Solutions were mixed by pipetting.

4.7 Last two rounds of ligation

For the fifth ligation reaction the complete contents were transferred by pipetting from tubes 1 and 3 into tubes 2 and 4, respectively, followed by mixing. The tubes were closed. The reactions were incubated in a thermocycler at 16 °C for 80 min.

For the last round of ligation reaction the complete contents were transferred by pipetting from tube 2 into tubes 4, followed by mixing. The tubes were closed. The reactions were incubated in a thermocycler at 16 °C for 80 min. This completed the hierarchical synthesis process.

4.8 Final purification

Purification was performed from 2% agarose gel using a comb of 8 lanes. First lane contained 50 bp ladder as in example 2 part E. The complete sample was mixed with 10 µL of purple loading die without SDS and dispensed into a single lane. Gel was run at 100 V, 200 mA, 12 watt for 45 min. Figure 7 shows the resulting gel. The upper band, corresponding to the expected size of 608 bp was excised and purified with Zymo gel extraction kit as in example 2, part E, using 20 µL of ddH20 water warmed to 35°C. Using 0.5 µL if this sample it was estimated spectrophotometrically that the solution contained 10 ng/µL.

4.9 Sequencing

The solution was split into two samples, one of 10 about µL and one of 9.5 µL. To each, a primer (“Primer1” and “Primer2") was added to the solution and sequenced with Sanger methods. Sequencing results in the central reliable region confirmed perfect sequence identity of the target ds polynucleotide with the SOI.

Example 5: Synthesis of a DNA molecule of 10.000 bp using the oligonucleotide library of Example 1

In this example construction of a ds polynucleotide consisting of a sequence of interest of 10.000 bps is demonstrated based on the library design of Example 1 by using oligos of 26 bps that form ds dimers with 4 nucleotide overhangs.

5.1 Sequence processing

A. The reverse complement of the leading strand of the sequence of interest is computed, and in both sequences (leading strand and reverse compliment) the last 4 nucleotides at the 3' ends are removed. This results in two single stranded template sequences, one corresponding to the leading strand of the SOI and the other to the reverse complement of the SOI, minus 4 nucleotides at the 3' ends.

B. The sequences of both ss templates are aligned, resulting in a double stranded template sequence, which is then partitioned into shorter sequences, referred to as oligo subsets or sub-sequences, occurring in the oligos contained in the library and their positions in the library are digitally annotated.

C. A workflow is determined which allows unambiguous assembly of the sub-sequences determined in step B.

5.2 Reaction

All steps below, unless otherwise stated, are carried out at 16°C and all solutions are prepared, and kept, on ice.

A. 700 µL of a solution of 2X ligase buffer in ddH_20 is prepared and 1.8 µL of this master mix solution is dispensed in each well of a 384 microwell-plate

B. 0.1 µL of each of the oligonucleotide library members corresponding to sub-sequences determined in 4.1, step B of the ss template sequence which is the leading strand of the SOI minus 4 nucleotides at the 3' end, is extracted from the library in order of occurrence in the target sequence and dispensed in a micro-well of the 348 microwell plates, starting at well *A1*, *B1*, ..., *P1* and then proceeding to the subsequent column *A2, B2*, etc. until all oligos are dispensed into a well

C. 0.1 µL of each of the oligonucleotide library members corresponding to the sub-sequences determined in 4.1, step B of the ss template sequence which is the

reverse complement of the SOI minus 4 nucleotides at the 3' end, is extracted in reverse sequence order and dispensed to the micro-well plate of step B, starting again at well A1 until all oligos are dispensed into a well. At this point, each micro-well contains two oligos that have 22 complementary bps and overhangs comprised of 4 nucleotides. Taken together, the wells should now contain matching pairs of oligonucleotide library members

D. The micro-well plate is sealed and annealed in a thermocycler starting at 95°C and decreased to 16°C at ramp rate of 1°C per min

E. 800 µL of a master mix ligation reaction solution comprising T4 ligase, at a concentration of 20 cohesive units per µL in ddH20 is prepared and 2 µL of this solution are dispensed into each of the 384 wells of the plate

F. The plate is spun down in a centrifuge by a 1000g pulse

G. The rows that contain solution are enumerated using the following formula: $2^{t-1}k$ where t is the tier number and t=1,2,3,4, and k is the index of the rows with filled wells, $r=k=1,\ldots,16/2^{t-1}$. In this way, in the first tier all rows are enumerated, in the second tier only half, and so on

H. The contents of the wells of each row of odd index are transferred to the wells of the rightmost columns of even index, using a multi-channel micropipette or a liquid handler

I. Right after transferring the contents, the solutions are gently mixed by pipetting directly with the micropipette or handler

J. The reaction is incubated for 60 min allowing the ligation reaction to complete

K. Steps G-J are repeated four more times, until only the last row (*P*) of the micro-well plate is filled, resulting in a total of 24 remaining filled wells

L. The contents of each of the 24 wells (containing 48 µL) are transferred to 24 reaction tubes and prepared for purification in columns following the Monarch PCR & DNA clean up kit from New England Biolabs (product nr. T1030), resulting in 6 µL of purified solution that contain only intermediate reaction products longer than 100 bps

M. The purified solutions are transferred to three fresh strips of 8 PCR tubes and arranged in a 8 row x 3 column fashion

N. 17.5 µL of the solution in step E are taken and 7.5 µL of Ligase buffer (10X) is added for a final concentration of 7X, and 1 µL of this solution is dispensed on each tube

O. The reactions proceed in the same way as in steps H-J 3 more times, resulting in three filled tubes (one on each column of the last row)

P. The contents of column 1 are transferred to column 2, leaving column 3 untouched

Q. The reaction is incubated for 1 hour

R. The contents of column 2 are transferred to column 3

S. The reaction is incubated for 1 hour

T. A 0.8% agarose gel is prepared and the sample loaded together with a 10 kbp ladder. The gel is run at 100 V for 45 minutes

U. The band corresponding to 10K bp is extracted and purified from the gel block using standard protocols and kits (Zymo clean is recommended in this example, see also Example 1).

5.3 Finalization and amplification

A. Two 26 bp long oligos from the library are selected, which are complementary to the last 26 nucleotides at the 3' ends of the SOI, i.e. they also include the 4 nucleotides that were deleted in step A of point 4.1. These two oligos are used as primers in a PCR reaction which is prepared to amplify the final product and to add the remaining 4 bps to each strand to complete the 10.000 bp sequence with blunt ends

B. The PCR product is purified with standard kits as in step L of point 4.2 to eliminate remaining oligos, enzymes and reagents, leaving the final DNA product, i.e. the ds polynucleotide which has a sequence that is identical to the SOI, ready for downstream applications.

<u>Example 6: Production of the polynucleotide library</u>

In this example it is demonstrated how to synthesize a DNA library that comprises a variety of core sequences, from a template and varying at prespecified positions, with the same overhangs.

As template a sequence of 128 bp that contains two variable loci was provided. Fig 10 shows the template sequence (SEQ ID NO:218), indicating in bold two variable sites at positions 29 and 71. Each of the two variables sites will include the four nucleotides A, C, G and T and the final product be a mixed library of the 16 variants of the template.

6.1 Preparing the annealing solutions

The oligos were part of the library generated in example 1. Oligos had the same properties as in example 2 and their sequences are listed in Fig. 10. The annealing solution was prepared as in example 2A but with a total volume of 240 uL, sufficient for total of 14 annealing reactions.

For each variable site 4 pairs of oligos are required, which when annealed, differ only at an internal site and result in the same overhang. These are denoted as 2.1, 2.2, 2.3, 2.4 and 5.1, 5.2, 5.3, 5.4 where the leading strand is labeled by “+” and the lagging strand by “-“ (Fig. 10).

Each of the pairs, as well as the other constituting oligos, are annealed separately as in example 2. Oligos 1, 3, 4, 6, 7 and 8 were transferred to the positions in a microwell plate as indicated in example 2, whereas oligos 2.x+/- placed in wells C1-C4 and oligos 5.x+/- in wells D1-D4. For example, 2.1+ and 2.1- are placed well C1, 2.2+ and 2.2- are placed well C2, etc.

6.2 Annealing

Annealing was conducted as in example 2.

6.3 Preparing the ligation solution

The ligation solution was prepared as in example 2.

6.4 Ligation rounds

Whilst keeping the micro-well plate at 16 °C (alternatively, on ice) the four dimers encoding each of the variable sites of a locus are pooled into a common reaction compartment (reaction tube), A2 in case of oligos 2.x and B1 in case of oligos 5.x. In other words, 5 uL of each of 2.1, 2.2m 2.3 and 2.4 (in wells C1, C2, C3 and C4, respectively) were transferred to well A2, and similarly, 5 uL of each of 5.1, 5.2, 5.3 and 5.4 (in wells D1, D2, D3 and D4, respectively) were transferred to well B1.

The three-tier ligation rounds were conducted exactly as in example 2 to complete the synthesis of a library of 16 different 128 bp polynucleotides that have two variable sites with full polymorphism and the same overhangs.

The 128 bp synthons were pooled together, purified and amplified as described in example 2. Since they have the same overhangs, the polynucleotide variants can be pooled together, without the risk of annealing to one another, thereby significantly simplifying the handling of the library.

To corroborate successful synthesis, the sequences in the sample were verified using Sanger sequencing and next-generation sequencing, showing that the 16 variants are present in the sample as desired (Figure 14).

Example 7: Enrichment of polynucleotides using PCR primers

In this example it is demonstrated how to enrich the yield (the amount) of a desired ligation product (gene fragments) by means of PCR amplification. This enrichment method is also used to enrich polynucleotides of the library of Example 6. Due to the nature of the enzymatic ligation reactions, repeated ligation reaction tiers result in the accumulation of unreacted partial DNA fragments (hereinafter referred to as impurities) which may interfere and further decrease the efficiency of the following ligation steps. Amplification of intermediates synthon products serve to drastically decrease the relative amounts of impurities, thus minimizing their interfering effect in the downstream process.

In order to enable further ligation rounds, this amplification-enrichment purification is implemented by PCR with staggered reannealing steps which result in sticky ends formation.

Staggered reannealing method have been successfully applied for the ligation of PCR products of template DNA to vectors (Ailenberg and Silverman, 1996; Walker et al, 2008) and for site-directed mutagenesis with subsequent ligations into vectors. Our current example of a PCR-based staggered reannealing method is novel in that it is used as a purification or enrichment step in the assembly of a polynucleotide library instead of using it for the insertion of a template DNA into a vector.

The oligos were part of the library generated in example 1. The oligos were prepared in an asymmetric way in the reaction plate in order to obtain partial synthons of different sizes as a result of a third ligation tier. Enrichment of partial synthons (four 128 bp fragments and one of 96 bp) were performed via PCR-based staggered reannealing method allowing to reintroduce sticky ends and proceed with higher tiers of ligation. Thus, the 608 bp target sequence (SEQ ID NO:105) is achieved by completing fourth ligation round of enriched synthons of four 128 bp fragments and one of 96 bp.

7.1 Synthesis of intermediate synthons

Synthesis four synthons of 128 bp and one 96 bp was conducted as described in examples 2 and 4. The oligos were part of the library generated in example 1 and has the following properties: all oligos were phosphorylated at the 5' end, they were provided at a concentration of 150 µM on nuclease free ddH20 and the used oligos were single-stranded and pure.

7.2 Amplification of intermediate synthons

PCR -based method is used for 128 bp Synthon amplification. To generate sticky ends in order to enable further ligation, the method described by Ailenberg and Silverman, 1996 was adopted. For design see Fig.1. Two sets of primers for 2 PCR reactions of each Synthon were used.

The introduction of sticky ends after PCR amplification in Synthon 1, 2, 3, 4, 5 requires following primers:

Syn1_PCR1_FW1	AACGCTACTACTATTAGTAGAATTG	SEQ ID NO:182
Syn1_PCR2_FW2	CTACTACTATTAGTAGAATTG	SEQ ID NO:183
Syn1_PCR1_REV1	TGCGAACGAGTAGATTTAG	SEQ ID NO:184
Syn1_PCR2_REV2	ATTCTGCGAACGAGTAGATTTAG	SEQ ID NO:185
Syn2_PCR1_FW1	GAATTGGGAATCAACTGTTACATGG	SEQ ID NO:186
Syn2_PCR2_FW2	TGGGAATCAACTGTTACATGG	SEQ ID NO:187
Syn2_PCR1_REV1	TAAGAGGTCATTTTTGCGGATGG	SEQ ID NO:188
Syn2_PCR2_REV2	AGGTCATTTTTGCGGATGG	SEQ ID NO:189
Syn3_PCR1_FW1	CTTATCAAAAGGAGCAATTAAAGG	SEQ ID NO:190
Syn3_PCR2_FW2	TCAAAAGGAGCAATTAAAGG	SEQ ID NO:191
Syn3_PCR1_REV1	AAGATTAAGAGGAAGCCCG	SEQ ID NO:192
Syn3_PCR2_REV2	CAAAAAGATTAAGAGGAAGCCCG	SEQ ID NO:193
Syn4_PCR1_FW1	TTTGATGCAATCCGCTTTGCTTCTG	SEQ ID NO:194
Syn4_PCR2_FW2	ATGCAATCCGCTTTGCTTCTG	SEQ ID NO:195
Syn4_PCR1_REV1	TCGTCATAAATATTCCTTG	SEQ ID NO:196
Syn4_PCR2_REV2	GGAATCGTCATAAATATTCATTG	SEQ ID NO:197
Syn5_PCR1_FW1	TTCCGCAGTATTGGACGCTATCCAG	SEQ ID NO:198
Syn5_PCR2_FW2	GCAGTATTGGACGCTATCCAG	SEQ ID NO:199
Syn5_PCR1_REV1	TAAAAACCAAAATAGCGAGAG	SEQ ID NO:200
Syn5_PCR2_REV2	ACGATAAAAACCAAAATAGCGAGAG	SEQ ID NO:201

Overhangs introduced by PCR on 5' end (25% of total PCR product) and on 3' end (another 25% of total PCR product) were the part of the sequence of targeted DNA.

A. PCR reaction.

The PCR 1 and PCR2 reaction mix, each 20ul, comprises ingredients according to PCR Protocol for Phusion® High-Fidelity DNA Polymerase (M0530): 5ul of 5X Phusion GC Buffer, 0.5ul 10mM dNTPs, 1.25ul forward and reverse 10uM primers, 1uL of ligation mixture of 128bp synthon, 0.75ul DMSO, 0.25uL Phusion DNA polymerase and 14.5ul of water.

Thermocycling conditions for PCR of 128 bp, 96 bp Synthons:

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
25-35 Cycles	98°C 55°C 72°C	10 seconds 30 seconds 15 sec (for 128nt)
Final Extension	72°C	7 minutes
Hold	4-10°C	

B. Sticky end PCR production.

Sticky ends were formed by adding 20ul of PCR1 product (for each Synthon) to 20uL of PCR2 product and 40uL (50% of the mix) formamide. After gentle mix, the reaction was heated at 98C for 5min- for DNA denaturation, then incubated at 65C for 5 min to allow reannealing complimentary strands and finally kept for 10 min at 22C. Products were verified on agarose gel (1.5%)

7.3 Purification of PCR products after sticky ends formation (if necessary).

20ul of PCR mix was purified with either SPRI beads according to the manufacture (Biolabs England) protocol (Size Selection Using AMPure XP Beads – NEB Next Fast DNA Library Prep Set for Ion Torrent (E6270) with modifications or using Quick Protocol for Oligonucleotide Cleanup Using the Monarch® PCR & DNA Cleanup Kit (5 µg) (NEB #T1030) with modifications.

After purification amounts of DNA were measured with nanodrop and processed for the ligation steps.

7.4 Ligation of sticky-end containing PCR-products of 128 bp or 96 bp synthons.

To synthesize 608 bp equimolar amounts of intermediate synthons were ligated. The ligation could be set as a “one-pot reaction”, where all five intermediate synthons are mixed (synthon 1 + synthon 2+synthon 3 + synthon 4+synthon 5) and simultaneously ligated or are set as a hierarchical ligation, where in the first round paired ligations are performed between synthon 1 + synthon 2, and between synthon 3 + synthon 4, generating 256 bp fragments and those are, in turn, ligated in a further round and, lastly ligated with Synthon 5.

Master mix for the ligation contained 2uL of 10X T4 ligase Buffer (NEB), 0.5ul 1mM ATP, 1.5 ul T4 DNA ligase (NEB), 1ul ddH20. For paired ligation 7.5 ul of PCR mixture of each synthon was used. For “one-pot reaction” 3ul of PCR mixture each of five synthons was used.

7.5 Final amplification and sequence verification

To verify the 608 bp product, it was amplified from the ligation mix with a pair of primers: Syn1_PCR1_FW1 and Syn5_PCR2_REV2 (sequences provided in example 7.2), with subsequent gel purification as it is in the example 4 (step 4.8). The resulting sample was prepared as it is in the example 4 (step 4.9) and sequenced with Sanger methods.

Example 8: Enrichment of a 1024bp polynucleotide using PCR primers

Using the method of Example 7, a polynucleotide of 1024bp is enriched.The oligos used in this example were part of the library generated in example 1. The oligos were prepared in an asymmetric way in the reaction plate in order to obtain partial synthons of different sizes as a result of a fourth ligation tier. Enrichment of partial synthons (four 256 bp fragments) were performed via PCR-based staggered reannealing method allowing to reintroduce sticky ends and proceed with higher tiers of ligation. Thus, the 1024 bp target sequence (SEQ ID NO:362) is achieved by completing a fifth ligation round of enriched synthons of four 256 bp fragments.

8.1 Synthesis of intermediate synthons

Synthesis of four synthons of 256 bp was conducted as described in examples 2 and 4. The oligos (listed in Fig. 13) were part of the library generated in example 1 and have the following properties: all oligos were phosphorylated at the 5' end, they were provided at a concentration of 150 µM on nuclease free ddH20 and the used oligos were single-stranded and pure.

8.2 Amplification of intermediate synthons

PCR -based method is used for 256 bp Synthon amplification. To generate sticky ends in order to enable further ligation, the method described by Ailenberg and Silverman, 1996 was adopted. For design see Fig.1. Two sets of primers for 2 PCR reactions of each Synthon were used.

The introduction of sticky ends after PCR amplification in Synthon 1, 2, 3, 4 requires the following primers:

Syn1_PCR1_FW1	AACGCTACTACTATTAGTAGAATTG	SEQ ID NO:182
Syn1_PCR2_FW2	CTACTACTATTAGTAGAATTG	SEQ ID NO:183
Syn1_PCR1_REV1	AGGTCATTTTTGCGGATGG	SEQ ID NO:347
Syn1_PCR2_REV2	TAAGAGGTCATTTTTGCGGATGG	SEQ ID NO:348
Syn2_PCR1_FW1	CTTATCAAAAGGAGCAATTAAAGG	SEQ ID NO:349
Syn2_PCR2_FW2	TCAAAAGGAGCAATTAAAGG	SEQ ID NO:350
Syn2_PCR1_REV1	TCGTCATAAATATTCCTTG	SEQ ID NO:351
Syn2_PCR2_REV2	GGAATCGTCATAAATATTCATTG	SEQ ID NO:352
Syn3_PCR1_FW1	TTCCGCAGTATTGGACGCTATCCAG	SEQ ID NO:353
Syn3_PCR2_FW2	GCAGTATTGGACGCTATCCAG	SEQ ID NO:354
Syn3_PCR1_REV1	AATAAAACGAACTAACGG	SEQ ID NO:355
Syn3_PCR2_REV2	CGTTAATAAAACGAACTAACGG	SEQ ID NO:356
Syn4_PCR1_FW1	AACGTAGATTTTTCTTCCCAACG	SEQ ID NO:357
Syn4_PCR2_FW2	TAGATTTTTCTTCCCAACG	SEQ ID NO:358
Syn4_PCR1_REV1	TACAGACCAGGCGCATAG	SEQ ID NO:359
Syn4_PCR2_REV2	GGTGTACAGACCAGGCGCATAG	SEQ ID NO:360

Overhangs introduced by PCR on 5' end (25% of total PCR product) and on 3' end (another 25% of total PCR product) were the part of the sequence of targeted DNA.

A. PCR reaction.

The PCR 1 and PCR2 reaction mix, each 20ul, comprises ingredients according to PCR Protocol for Phusion® High-Fidelity DNA Polymerase (M0530): 5ul of 5X Phusion GC Buffer, 0.5ul 10mM dNTPs, 1.25ul forward and reverse 10uM primers, 1uL of ligation mixture of 256bp synthon, 0.75ul DMSO, 0.25uL Phusion DNA polymerase and 14.5ul of water.

Thermocycling conditions for PCR of 256 bp Synthons:

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
25-35 Cycles	98°C 55°C 72°C	10 seconds 30 seconds 15 sec (for 128nt)
Final Extension	72°C	7 minutes
Hold	4-10°C	

B. Sticky end PCR production.

Sticky ends were formed by adding 20ul of PCR1 product (for each Synthon) to 20uL of PCR2 product and 40uL (50% of the mix) formamide. After gentle mix, the reaction was heated at 98C for 5min- for DNA denaturation, then incubated at 65C for 5 min to allow reannealing complimentary strands and finally kept for 10 min at 22C. Products were verified on agarose gel (1.5%)

8.3 Purification of PCR products after sticky ends formation (if necessary).

20ul of PCR mix was purified with either SPRI beads according to the manufacture (Biolabs England) protocol (Size Selection Using AMPure XP Beads – NEB Next Fast DNA Library Prep Set for Ion Torrent (E6270) with modifications or using Quick Protocol for Oligonucleotide Cleanup Using the Monarch® PCR & DNA Cleanup Kit (5 µg) (NEB #T1030) with modifications.

After purification amounts of DNA were measured with nanodrop and processed for the ligation steps.

8.4 Ligation of sticky-end containing PCR-products of 256 bp synthons.

To synthesize 1024 bp equimolar amounts of intermediate synthons were ligated. The ligation could be set as a “one-pot reaction”, where all four intermediate synthons

are mixed (synthon 1 + synthon 2 + synthon 3 + synthon 4) and simultaneously ligated, or set as a hierarchical ligation, where in the first round paired ligations are performed between synthon 1 + synthon 2, and between synthon 3 + synthon 4, generating 512 bp fragments, and those are, in turn, ligated in a further round.

Master mix for the ligation contained 2uL of 10X T4 ligase Buffer (NEB), 0.5ul 1mM ATP, 1.5 ul T4 DNA ligase (NEB), 1ul ddH20. For paired ligation 8ul of PCR mixture of each synthon was used. For “one-pot reaction” 4ul of PCR mixture each of four synthons was used.

8.5 Final amplification and sequence verification

To verify the 1024 bp product, it was amplified from the ligation mix with a pair of primers: Syn1_PCR1_FW1 and Syn4_PCR2_REV2 (sequences provided in example 7.2), with subsequent gel purification as it is in the example 4 (step 4.8). The resulting sample was prepared as it is in the example 4 (step 4.9) and sequenced with Sanger methods, confirming that the product matched the desired sequence.

Example 9: Method of enrichment of the polynucleotide library, comprising an immobilization step

In this example it is demonstrated how to synthesise in high purity a target ds polynucleotide of 608 bp (SOI is Sequence "Ribbon_M13_608", SEQ ID NO:105) using the method provided herein, through the purification of biotinilated synthons by means of their immobilisation to a solid phase with avidin.

The oligos were part of the library generated in example 1. Oligos had the same properties as in example 2 and their sequences (including biotinilations) are listed in Fig 9. In short, leading oligos placed in row A of the microwell plate (namely A1+, A2+, A3+, A4+ and A5+) were biotinilated and obtained commercially from a CRO that provides them in high purity.

The oligos were prepared in an asymmetric way in the reaction plate in order to obtain partial constructs of different sizes at the fourth ligation. The 608 bp sequence is achieved by completing four ligation rounds to obtain four reaction products of 128 bp, and one of 96 bp. Each of these products contain a 5’ biotin modification. Subsequent immobilisation to avidin-coated magnetic beads allowed to purify the synthons thereby increasing their purity but with an inevitable mass loss, therefore decreasing its yield.

Subsequent enrichment of the purified synthons as explained in example 7 was applied in order to increase the yield. The resulting synthons were subject to two more ligation rounds, thereby obtaining each strand of the 608 bp target ds polynucleotide.

9.1 Assembly of intermediate product synthons

Annealing of the synthons and their four-tier ligation rounds were conducted exactly as in examples 2 and 4. Note that the 5' biotinilated oligos do not interfere with the ligation reactions because these correspond to the 5' overhangs of the product synthons.

Because the ligation reactions are not 100% efficient and because they are somewhat sequence non-specific the reaction solution contains not only the desired products but many other polynucleotides, as revealed by a diversity of bands in acrylamide gels.

Furthermore, next-generation sequencing (outsourced to a CRO) corroborated both that the target synthons were present in the sample and also that incomplete products and incorrect assemblies are present in the samples.

9.2 Purification of intermediate product synthons

To increase the purity of the product synthons these were immobilized to avidin-coated magnetic beads (Invitrogen DynabeadsTM M-270 Streptavidin, Catalog nos. 65305, 65306) by adapting the nucleic acid immobilization protocol described by the provider. Namely, the following steps were applied to each of the five reaction products:

A. Preparation of streptavidin-coated magnetic beads: Wash 50 µL of magnetic beads with 50 µL of 2× B&W buffer (10mMTris-HCl pH 7.5, 1mM EDTA, 2M NaCl, and 0.05% Tween20) three times to remove excess sodium azide bacteriostatic agent and resuspended in 50 µL, 2X B&W buffer. Transfer 20 uL to an empty container (PCR reaction tube) and cool the solution to 4°C. At the same time chill the metal-plate in the fridge.

B. Binding the synthons to the avidin-coated beads: Transfer 20 uL of the solution containing the product synthons (step 9.1) to the tubes containing the beads and incubate for 1h; shake every 10 min.

C. Cleaning the solution from reaction by-products:

- Place the reaction tubes on the cooled metal plate, keeping it on ice.

- Transfer 40uL of supernatant (containing the reaction by-products) by pipetting and store at -12 °C for further analyses, or discard.
- Wash the sample 3 times by adding 40uL of B&W buffer to the container with the beads, re-suspend and discard supernatant.
- Remove tubes form the metal plate.
- Add 40uL of B&W buffer and incubate at 30 °C for 25 min.
- Place the reaction tubes on the metal plate and extract out supernatant. Store it in -15 °C for further analyses or discard.

D. Releasing the product synthons from the avidin-coated beads: Wash the beads with MilliQ water at room temperature 3 times. In a thermocycler, heat up the sample from 30 °C to 70 °C. After reaching thermal equilibrium, leave for 1 sec at 70 °C and remove the sample from the thermocycler.

E. Separate the sample from the beads: Place the reaction tube on the metal plate and transfer the supernatant to new reaction tubes by pipetting.

Result from acrylamide gel electrophoresis (prepared as in example 2) showed that the target synthons increased in purity.

9.3 Amplification of intermediate product synthons

The bead-purified product synthons were further enriched by PCR amplification as in example 7. This step was necessary for two reasons. First, because the product synthons contain 5’ biotinilations their subsequent ligation into longer reaction products is not possible. However, this restriction was overcome by using PCR with appropriate primers that introduce the required overhangs. Second, PCR amplification resulted in an increase in yield, which was required to obtain detectable product amounts in the subsequent ligation tears.

9.4 Last two rounds of ligation, final purification and sequence verification

Further ligation rounds were prepared and accomplished as in example 4 (steps 4.6 and 4.7), with subsequent gel purification as in example 4 (step 4.8). The resulting sample was prepared as in example 4 (step 4.9) and sequenced with Sanger methods.

Example 10: Method of enrichment and purification of the polynucleotide library by template replication, comprising an immobilization step

In this example it is demonstrated how to enrich the yield of intermediate ligation products (gene fragments) by means of template replication with polymerase and replication primers, and with immobilisation to a solid phase with avidin.

10.1 Assembly of intermediate product synthons

Annealing of the synthons and their four-tier ligation rounds was conducted exactly as in examples 2 and 4 to obtain a 256 bp “intermediate synthon” (which comprises the first 256 bp of SEQ ID NO:105,) which would be part of the process to assemble a larger target polynucleotide. The constituting oligos were part of the library generated in example 1 and have the following properties: all oligos were phosphorylated at the 5’ end, with two exceptions. The two oligos forming the 5’ end at both the leading and lagging strand of the 256 bp intermediate synthon contained a 5’ biotin modification to later allow for immobilization of each strand and undergo template replication (explained below). All oligos were provided at a concentration of 150 µM on nuclease free ddH20 and the used oligos were single-stranded and pure.

Subsequent immobilisation to streptavidin-coated magnetic beads allowed to enrich these products with PCR primers and to introduce sticky ends. Note that the 5’ biotinilated oligos do not interfere with the first four rounds of ligation because these modified oligos correspond to the 5’ overhangs of the intermediate synthons.

10.2 Purification and enrichment of intermediate synthons through template replication with immobilization

An enzyme-based (polymerase) method was used for enrichment of the 256 bp intermediate synthon and for the generation of sticky ends, at the same time, to enable further ligation. For this purpose, two sets of primers were used in two separate reactions to introduce overhangs during polymerase enzyme replication at the 5’ end of each the leading- and lagging strand, respectively in each separate reaction. These overhangs were the same for every leading strand and the same for every lagging strand, but the overhang of the leading strands differed from the overhang of the lagging strand. Importantly, the overhang of the leading strands is not complementary to the overhang of the lagging strands.

The two reaction templates were generated by dividing the sample containing the 256 bp intermediate synthon generated under point 10.1 into two equal aliquots, Sample A and Sample B, both at a quantity of 20 uL. The leading and the lagging strand of the 256bp synthon are separately immobilized on streptavidin-coated magnetic beads via the biotin modification at their 5' ends. Specifically, sample A comprises the leading strand immobilized on beads, and sample B comprises the lagging strand immobilized on beads.

In each reaction, the 5' primer contained the overhang of the leading strand (Sample A) and lagging strand (Sample B), while the 3' primer was a complementary sequence of the respective strand from the oligo library. In each reaction one DNA strand (the leading strand in Sample A and the lagging strand in Sample B) was used as a template and thus by each replication cycle a new complementary strand was produced. A schematic of the reaction is provided in Figure 12.

The following steps were applied to both Sample A and Sample B:

A. Preparation of streptavidin-coated magnetic beads: as in example 9.2.

B. Binding the synthons to the avidin-coated beads: as in example 9.2. B.

C. On-bead synthon (template) replication for enrichment: each of the two reaction mixtures, with the total volume of 20ul, comprised ingredients according to example 8.2., but taking into consideration that only FW1 (Sample A) and REV2 (Sample B) primers were used (as indicated in the table). Thermocycler conditions for the replication were as in example 8.2.

D. Cleaning Washes for purification and cleanup: as in example 9.2. C.

E. Separation: as in example 9.2. D and E.

F. Pooling & Reannealing for Target Polynucleotide Recovery: each sample containing either the leading strand (Sample A) and the lagging strand (Sample B) in a pure, enriched form were then pooled and annealed as explained in examples 2.B and 5.2 D.

10.3 Further ligation rounds, final purification and sequence verification

Further ligation rounds were prepared and accomplished as in example 4 (steps 4.6 and 4.7), with subsequent gel purification as in example 4 (step 4.8) or, alternatively with bead-based purification. The resulting sample was prepared as in example 4 (step 4.9)

and sequenced with NGS methods confirming that the sequence of the enriched 256bp synthon corresponded to the template sequence.

REFERENCES

Ailenberg M, Silverman M. (1996) Description of a one step staggered reannealing method for directional cloning of PCR-generated dna using sticky-end ligation without employing restriction enzymes. *IUBMB Life*, 39(4):771-9.

Anderson, S., Bankier, A.T., Barrell, B.G. et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature*, 290:457-465.

Beaucage, S.L. and Caruthers, M.H. (1981) Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* 22:1859–1862.

Bentley, D.R., et al. (65 authors) (2008) Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature*, 456:53–59.

Bonde, M.T., Kosuri, S., Genee, H.J., Sarup-Lytzen, K., Church, G.M., Sommer, M.O.A. and Wang H.H. (2014) Direct Mutagenesis of Thousands of Genomic Targets Using Microarray-Derived Oligonucleotides. *ACS Synthetic Biology* 4(1):17-22.

Chari, R. and Church, G.M. (2017) Beyond editing to writing large genomes. *Nature Reviews Genetics*, In Press.

Engler, C., Kandzia, R. and Marillonnet, S. (2008) A one pot, one step, precision cloning method with high through put capability. *PloS One* 3(11):e3647.

Farzadfard, F. and Timothy, K.L. (2014) Genomically Encoded Analog Memory with Precise in Vivo DNA Writing in Living Cell Populations. *Science* 346(6211): 1256272.

Gao, X., LeProust, E.M., Zhang, H., Srivannavit, O. Gulari, E., Yu, P., Nishiguchi, C., Xiang, Q. and Zhou, X. (2001) A Flexible Light-Directed DNA Chip Synthesis Gated by Deprotection Using Solution Photogenerated Acids. *Nucleic Acids Research* 29(22):4744–50.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison III, C.A. and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5):343-345.

Horspool, D.R., Coope, R.J.N. and Holt, R.A. (2010) Efficient assembly of very short oligonucleotides using T4 DNA Ligase. *BMC Research Notes*, 3:291-299.

Kai, J., Puntambekar A., Santiago N., Lee S.H., Sehy D.W., Moore V., Han J. and Ahn C.H. (2012) A novel microfluidic microplate as the next generation assay platform for enzyme linked immunoassays (ELISA). *Lab Chip*, 12(21):4257-62

Kemp, G. (1998) Capillary electrophoresis: a versatile family of analytical techniques. *Biotechnology and Applied Biochemistry* 27:9-17.

Lehman, I.R. and Nussbaum, A.L. (1964) The deoxyribonucleases of Escherichia coli. V. On the specificity of exonuclease I (phosphodiesterase), *Journal of Biological Chemistry*, 239:2628-2636.

LeProust, E.M., Peck, B.J., Spirin, K., McCuen, H.B., Moore, B., Namsaraev, E., and Caruthers, M.H. (2010) Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process. *Nucleic Acids Research*, 38(8), 2522–2540.

Neuner, P., Cortese, R. and Monaci, P. (1998) Codon-Based Mutagenesis Using Dimer-Phosphoramidites. *Nucleic Acids Research* 26(5):1223–27.

Rio, D.C. (2011). RNA: A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press.

Sambrook, J., and Russell, D. W. (2014). Molecular Cloning. A Laboratory Manual. (3rd ed.). New York: Cold Spring Harbor Laboratory Press.

Smith H.O., Hutchison III, C.A., Pfannkoch, C. and Venter J.C. (2003) Generating a synthetic genome by whole genome assembly: X174 bacteriophage from synthetic oligonucleotides. *Proceedings of the Natural Academy of Sciences of the USA*, 100(26):15440-15445.

Sondek, J., and Shortle, D. (1992). A General Strategy for Random Insertion and Substitution Mutagenesis: Substoichiometric Coupling of Trinucleotide Phosphoramidites. *Proceedings of the National Academy of Sciences* 89(8): 3581–85.

Stemmer, W.P., Crameri, A., Ha, K.D., Brennan, T.M. and Heyneker, H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene*, 1614:49-53.

Walker A, Taylor J, Rowe D, Summers D. (2008) A method for generating sticky-end PCR products which facilitates unidirectional cloning and the one-step assembly of complex DNA constructs. *Plasmid*, 59(3):155-62.

CLAIMS

1. A library of double stranded (ds) polynucleotide library members of at least 12 bp length comprising a variety of core sequences and each ds oligo library member further comprising a first overhang sequence and a second overhang sequence which are the same in each ds oligo library member, wherein the first and second overhang sequences are not complementary to each other, wherein the core sequence is the double stranded part of the library members, and wherein the variation of the core sequences comprises one or more random point mutations.

2. The library of claim 1, wherein the ds oligo library members each comprise a leading and a lagging strand, the leading strand and the lagging strand each having an overhang sequence, and wherein

a) the first overhang sequence is a 5' overhang of the leading strand, and the second overhang sequence is a 5' overhang of the lagging strand; or

b) the first overhang sequence is a 3' overhang of the leading strand, and the second overhang sequence which is a 3' overhang of the lagging strand;

wherein said first and second overhang sequences are not complementary to each other.

3. The library of claim 1 or 2, wherein the first and second overhang sequences have a length of 4-8 nucleotides.

4. The library of any one of claims 1 to 3, wherein each of said library members comprises an identical modification selected from the group consisting of phosphorylation, methylation, biotinylation, or linkage to a fluorophore or quencher.

5. The library of any one of claims 1 to 4, wherein said library members are contained in one library containment, or in a plurality of spatially distinct library containments.

6. The library of any one of claims 1 to 5, wherein each of said library members comprises a sequence which is at least 30% identical to a template.

7. A method for producing the library of any one of claims 1 to 6, comprising the steps:

a) providing a template nucleotide sequence;

b) providing a library of single stranded (ss) oligos in separate containments, said ss oligos comprising a variety of sequences comprising one or more point mutations compared to the template;

c) selecting pairs of matching ss oligos from said library of ss oligos;

d) partially annealing said pairs of matching ss oligos, thereby obtaining a first library of ds oligonucleotides (ds oligos) comprising ds oligo library members, wherein the ds oligo library members comprise a variety of core sequences, and each ds oligo library member further comprises a first overhang sequence and a second overhang sequence which are the same for each ds oligo library member, wherein the first and second overhang sequences are not complementary to each other; and

e) further annealing with ds oligos that have overhangs matching the overhangs of the ds oligo library members of the first library of ds oligos, thereby obtaining a second library of ds polynucleotides.

8. The method of claim 7, wherein the ds polynucleotides are enriched by polymerase chain reaction (PCR).

9. The method of claim 7 or 8, wherein the variety of ds polynucleotides is synthesized by partially annealing a library of matching single stranded oligonucleotides (ss oligos) thereby obtaining a first library of ds oligonucleotides (ds oligos), each with the same first and second overhang sequences.

10. The method of claim 9, wherein

a) the library of ss oligos comprises ss oligos with a length of at least 6 nt; and/or

b) the first library of ds oligos comprises ds oligos with a length of at least 6 bp; and/or

c) the library of ds polynucleotides comprises ds polynucleotides with a length of at least 12 bp.

11. Use of the library of any one of claims 1 to 6, in a method of synthesizing a variety of target ds polynucleotides each being longer than the library members by assembling said library members with ds oligos that have overhang sequences matching the overhang sequences of the library members, thereby obtaining a library of target ds polynucleotides.