

이브러리 컨테인먼트에 함유되고, 다양성은 적어도 하나의 돌출부를 갖는 일본쇄 올리고뉴클레오타이드(ss 올리고) 및 일본쇄 올리고뉴클레오타이드(ds 올리고)를 포함하고 정합 올리고뉴클레오타이드의 적어도 10,000 쌍을 커버하는, 단계,

b) 제1 단계에서, 액체 핸들러를 사용하여 상기 라이브러리로부터 적어도 제1 쌍의 정합 올리고뉴클레오타이드를 제1 반응 컨테인먼트로 전달하고, 정합 올리고뉴클레오타이드를 어셈블리하여, 적어도 하나의 돌출부를 포함하는 제1 반응 생성물을 수득하는 단계,

c) 제2 및 추가 단계에서, 액체 핸들러를 사용하여, 상기 라이브러리로부터 적어도 제2 및 추가 쌍의 정합 올리고뉴클레오타이드를 각각 제2 및 추가 반응 컨테인먼트로 전달하고, 정합 올리고뉴클레오타이드를 어셈블리하여, 각각 적어도 하나의 돌출부를 포함하는 제2 및 추가 반응 생성물을 각각 수득하는 단계,

d) 상기 제1, 제2 및 추가 반응 생성물을 소정 워크플로우로 어셈블리함으로써, 돌출부를 갖는 상기 표적 ds 폴리뉴클레오타이드를 생성한 후, 임의로 평활 말단을 제조하기 위해 최종화 단계를 수행하는 단계를 포함하고,

단계 b) 및 c) 및 어셈블리 워크플로우에 사용된 정합 올리고뉴클레오타이드의 상기 쌍은 상기 표적 ds 폴리뉴클레오타이드를 생성하는 알고리즘을 사용하여 결정되는, 방법을 제공한다.

(52) CPC특허분류

C12N 15/1031 (2013.01)

명세서

청구범위

청구항 1

소정 서열을 갖는 표적 이본쇄(ds) 폴리뉴클레오타이드를 합성하는 방법으로서,

a) 올리고뉴클레오타이드 라이브러리 멤버의 다양성을 포함하는 올리고뉴클레오타이드 라이브러리를 어레이 장치 내에 제공하는 단계로서, 상기 라이브러리 멤버 각각은 상이한 뉴클레오타이드 서열을 갖고 수용액 중의 개별 라이브러리 컨테인먼트(containment)에 함유되고, 다양성은 적어도 하나의 돌출부를 갖는 일본쇄 올리고뉴클레오타이드(ss 올리고) 및 이본쇄 올리고뉴클레오타이드(ds 올리고)를 포함하고 정합 올리고뉴클레오타이드의 적어도 10,000 쌍을 커버하는, 단계,

b) 제1 단계에서, 액체 핸들러를 사용하여 상기 라이브러리로부터 적어도 제1 쌍의 정합 올리고뉴클레오타이드를 제1 반응 컨테인먼트로 전달하고, 정합 올리고뉴클레오타이드를 어셈블리하여, 적어도 하나의 돌출부를 포함하는 제1 반응 생성물을 수득하는 단계,

c) 제2 및 추가 단계에서, 액체 핸들러를 사용하여, 상기 라이브러리로부터 적어도 제2 및 추가 쌍의 정합 올리고뉴클레오타이드를 각각 제2 및 추가 반응 컨테인먼트로 전달하고, 정합 올리고뉴클레오타이드를 어셈블리하여, 각각 적어도 하나의 돌출부를 포함하는 제2 및 추가 반응 생성물을 각각 수득하는 단계,

d) 상기 제1, 제2 및 추가 반응 생성물을 소정 워크플로우로 어셈블리함으로써, 돌출부를 갖는 상기 표적 ds 폴리뉴클레오타이드를 생성한 후, 임의로 평활 말단을 제조하기 위해 최종화 단계를 수행하는 단계를 포함하고,

단계 b) 및 c) 및 어셈블리 워크플로우에 사용된 정합 올리고뉴클레오타이드의 상기 쌍은 상기 표적 ds 폴리뉴클레오타이드를 생성하는 알고리즘을 사용하여 결정되는, 방법.

청구항 2

제1항에 있어서, 일련의 상이한 표적 ds 폴리뉴클레오타이드가 동일한 올리고뉴클레오타이드 라이브러리를 사용하여 합성되는, 방법.

청구항 3

제2항에 있어서, 상기 상이한 표적 ds 폴리뉴클레오타이드가 50% 미만, 바람직하게는 30% 미만의 서열 동일성을 갖는, 방법.

청구항 4

제1항 내지 제3항 중의 어느 한 항에 있어서, 상기 ss 올리고가 6 내지 26 nt의 길이를 갖는, 방법.

청구항 5

제1항 내지 제4항 중의 어느 한 항에 있어서, 상기 ds 올리고가 6 내지 26 bp의 길이를 갖고, 상기 돌출부가 각 ds 올리고 길이의 절반 이하인, 방법.

청구항 6

제1항 내지 제5항 중의 어느 한 항에 있어서, 상기 다양성은, 포스포릴화, 메틸화, 비오틴화, 또는 형광단 또는 켄처에 대한 결합 중의 하나 이상에 의해 변형되는 ss 올리고 및/또는 ds 올리고를 커버하는, 방법.

청구항 7

제1항 내지 제6항 중의 어느 한 항에 있어서, 상기 표적 ds 폴리뉴클레오타이드가 적어도 48 bp의 길이를 갖는, 방법.

청구항 8

제1항 내지 제7항 중의 어느 한 항에 있어서, 상기 어셈블리는, 효소 반응 및/또는 화학 반응인 결찰 반응에 의한 것인, 방법.

청구항 9

제8항에 있어서, 상기 결찰 반응이 리가제, 바람직하게는 T3, T4 또는 T7 DNA 리가제, 또는 폴리머라제 또는 리보자임을 사용하는 효소적 결찰 반응인, 방법.

청구항 10

제1항 내지 제9항 중의 어느 한 항에 있어서, 상기 어셈블리가 정합 돌출부를 직접 하이브리드화함으로써 또는 적합한 ss 올리고 링커를 간접적으로 하이브리드화함으로써 이루어지고, ss 올리고 링커는, 상기 제1, 제2 또는 추가 반응 생성물 중의 어느 것을 어셈블리하기 위해 상기 라이브러리로부터 선택되고 전달되는 상기 라이브러리에 함유된 ss 올리고인, 방법.

청구항 11

제1항 내지 제10항 중의 어느 한 항에 있어서, 상기 표적 ds 폴리뉴클레오타이드의 뉴클레오타이드 서열이 주형과 동일한, 방법.

청구항 12

제1항 내지 제11항 중의 어느 한 항에 있어서, 상기 표적 ds 폴리뉴클레오타이드는, 목적 서열(SOI)을 갖는 상기 폴리뉴클레오타이드를 수득하기 위해 지시된 돌연변이 유발, 엔도뉴클레아제 또는 엑소뉴클레아제 중의 어느 것에 의해 추가로 변형되는, 방법.

청구항 13

제1항 내지 제12항 중의 어느 한 항에 있어서, 상기 표적 ds 폴리뉴클레오타이드는 ds DNA, ss DNA 또는 RNA 분자 중의 어느 하나인 이의 유도체를 생성하기 위해 추가로 변형되고, 바람직하게는 상기 변형 방법은 메틸트랜스페라제, 키나제, CRISPR/Cas9, 다중 자동화 게놈 조작(MAGE), 공액 어셈블리 게놈 조작(CAGE), 아르코나우트 단백질(Ago) 또는 이의 유도체, 징크-핑거 뉴클레아제(ZFN), 전사 활성화인자-유사 이펙터 뉴클레아제(TALEN), 하이브리드화 분자, 설푸릴라제, 리콤포비나제, 뉴클레아제, DNA 폴리머라제, RNA 폴리머라제 또는 TNase 중의 어느 것을 사용하는 효소적 변형 중의 어느 하나인, 방법.

청구항 14

제1항 내지 제13항 중의 어느 한 항에 있어서, 상기 개별 라이브러리 컨테인먼트는 바람직하게는 사용 빈도에 따라 3차원 순서로 공간적으로 배열되고, 상기 3차원 순서는 적어도 부분적으로 적층된 적어도 2개 라이브러리 컨테인먼트를 포함하는, 방법.

청구항 15

제1항 내지 제14항 중의 어느 한 항에 있어서, 상기 어레이 장치는 미세역가 플레이트, 마이크로유체 마이크로 플레이트, 모세관 세트, 마이크로어레이 또는 바이오칩, 바람직하게는 DNA 또는 RNA 바이오칩 중의 어느 것인, 방법.

청구항 16

제1항 내지 제15항 중의 어느 한 항에 있어서, 상기 표적 ds 폴리뉴클레오타이드는 주형 또는 SOI의 서열과의 동일성 정도를 검증하기 위해 서열분석되는, 방법.

청구항 17

적어도 하나의 돌출부를 갖는 일본쇄 올리고뉴클레오타이드(ss 올리고) 및 이본쇄 올리고뉴클레오타이드(ds 올리고)인 라이브러리 멤버의 다양성을 포함하는 어레이 장치 내에 제공된 올리고뉴클레오타이드 라이브러리로서, 상기 라이브러리 멤버 각각은 상이한 뉴클레오타이드 서열을 갖고 수용액 중의 개별 라이브러리 컨테인먼트(containment)에 함유되며, 컨테인먼트는 3차원 순서로 공간적으로 배열되고, 다양성은 정합 올리고뉴클레오타

이드의 적어도 10,000 쌍을 커버하는, 올리고뉴클레오타이드 라이브러리.

청구항 18

제17항에 있어서, 상기 라이브러리 컨테인먼트는 바람직하게는 사용 빈도에 따라 3차원 순서로 공간적으로 배열되고, 상기 3차원 순서는 적어도 부분적으로 적층된 적어도 2개 라이브러리 컨테인먼트를 포함하는, 올리고뉴클레오타이드 라이브러리.

청구항 19

소정 서열을 갖는 일련의 상이한 표적 이분쇄(ds) 폴리뉴클레오타이드를 합성하기 위한 제17항 또는 제18항의 올리고뉴클레오타이드 라이브러리의 용도로서,

상기 상이한 표적 이분쇄(ds) 폴리뉴클레오타이드가 50% 미만, 바람직하게는 30% 미만의 서열 동일성을 갖는, 용도.

발명의 설명

기술 분야

[0001] 본 발명은 올리고뉴클레오타이드의 다양한 라이브러리를 사용하여 소정 서열을 갖는 이분쇄(ds) 폴리뉴클레오타이드를 합성하는 신규한 방법에 관한 것이다.

배경 기술

[0002] 현재, 폴리뉴클레오타이드의 인공 합성은, 반드시 배타적인 것은 아닌 2종류의 방법을 통해 실현되고 있다:

[0003] 폴리뉴클레오타이드의 합성을 위한 방법의 첫번째 부류는 "화학적 합성"이다. 이는, 포스포르아미다이트 화학을 사용하여 뉴클레오타이드를 1개씩 연속적으로 연결함으로써 일본쇄 DNA(또는 RNA) 분자를 구축하는 프로세스이다[참조: Beaucage and Caruthers, 1981]. 이 방법은 임의의 복잡성의 특정한 소정 주형 서열을 갖는 DNA 분자를 구축할 수 있게 한다. 화학적 방법은 이들의 저렴한 성질에 기인하여 인기가 있고, 용이하게 병렬화할 수 있으며, 일부 구현에서는 칩 내에서 DNA 또는 RNA의 고효율 생성을 가능하게 한다[참조: LeProust et al., 2010]. 이들 방법의 주요 및 최대 단점은, 합성되는 주형의 길이에 의해 반응의 수율이 현저히 저하되고, 분자의 크기가 통상 약 200 염기쌍(bp 또는 bps)로 제한된다는 것이다.

[0004] DNA 합성을 위한 방법의 두번째 부류는 "어셈블리 방법"이고, 이는, 목적하는 표적 서열을 갖는 보다 큰 분자를 수득하기 위해, 상이한 크기 및 상이한 서열의 올리고뉴클레오타이드 및 폴리뉴클레오타이드를 특정 방법으로 생화학적으로 결합하는 것으로 이루어진다. 이들 올리고뉴클레오타이드의 공급원은 종종 화학적 합성이지만, 또한 천연 존재 DNA의 효소적 소화의 생성물일 수도 있다. 이들 어셈블리 방법은 종종, 거대 폴리뉴클레오타이드 쇄(1K-5K bp)의 합성을 위한 용어인 제품명 "유전자 합성"으로 상품화되지만, 반드시 유전자-크기의 길이는 아니다. 보다 작은 폴리뉴클레오타이드를 표적 서열과 어셈블리하기 위한 몇몇 접근법이 문헌에 보고되어 있다 [참조: Stemmer et al., 1995; Smith et al., 2003; Engler et al., 2008; Gibson et al., 2009; Horspool 2010].

[0005] 과거 수년간에, "깊은 어셈블리"(Gibson et al., 2009)는 몇몇 선형 ds DNA 단편(크기가 약 30 bp 내지 수 Kbp까지의 범위)을 연결하는 일반적 방법으로 되어 있다. 이 방법은, 쌍 방식으로 중첩 서열 상동성을 갖는 다수의 ds DNA 단편을 결합하는 것으로 이루어진다. 단편 사이의 중첩 상동성 영역은 약 15 내지 80 bp의 범위일 수 있다. 돌출부(overhang)가 필요하지 않은데, 이 방법의 효소적 기구가 돌출부의 생성을 처리하고, 겹을 충전하고, 단편을 정확하게 연결하기 때문이다. 이러한 효소적 기구는 3개 효소: T5 엑소뉴클레아제, 퓨전(Phusion) DNA 폴리머라제 및 Taq DNA 리가제를 모두 등온 반응에 사용하게 한다. 이 방법은 단순하고 범용성이며, 선형 및 환상 ds DNA 생성물을 생성할 수 있다. 이 방법의 결점은 자동화에 제한이 있어, 대규모 상업적 사용에는 적합하지 않다.

[0006] 수천 염기 쌍의 DNA 분자를 구축할 때의 공통 테마는, 최대 수백 nt 또는 bp의 작은 단편을 화학적으로 합성하고, 이어서 이들을 클로닝, 절찰, PCA 또는 깊은 어셈블리에 의해 함께 연결하는 것이다.

[0007] 일부 접근법은, 가능하게는 화학적 합성을 통해, 가능한 유전자 공간을 커버하는 올리고뉴클레오타이드의 라이브러리, 또는 이의 필요한 서브셋을 사전-작성하는 것을 시사한다.

- [0008] 차리 및 처취(Chari and Church)는 합성된 올리고뉴클레오타이드(200 염기)를 사용하여 짧은 DNA 단편을 생성하고 효모 및 이. 콜라이(*E. coli*)에서 생체내 상동성 제조합을 사용하여 큰 DNA 세그먼트로 어셈블리하는 것을 제안한다[참조: Chari and Church, 2017].
- [0009] WO 2009/138954 A2호는 고상 어셈블리에 의해 보다 큰 폴리뉴클레오타이드를 합성하는 방법을 개시하고, 여기서 보다 큰 폴리뉴클레오타이드의 어셈블리에 필요한 정의된 서브유닛은 필요에 따라 화학적으로 합성된다.
- [0010] 피더센 등(Pedersen et al.)은 모든 가능한 육량체의 공간(N=4,096 올리고)을 포함하는 라이브러리를 사용하는 것을 제안한다. 이어서, 6개 염기 쌍의 긴 올리고는 올리고 링커를 사용하여 어셈블리하여 폴리뉴클레오타이드를 형성한다. 올리고뉴클레오타이드의 연결 및 대규모 DNA 합성에 관한 특정한 제한이 있다. 대량의 시약을 사용하는 클로닝 등의 적절하게 설계된 라이브러리 및 수동 프로토콜이 필요하기 때문에, 이 방법은 시간 소비적이다. 이들은 또한 합성 가격에 상당한 비용을 추가하며, 이는 표적 서열 길이가 증가함에 따라 bp당 증가한다.
- [0011] WO2002/081490호는 인간 게놈 데이터베이스 등의 데이터베이스에서 이용가능한 정보에 기초하여 컴퓨터-지향 폴리뉴클레오타이드 어셈블리에 의한 게놈 서열 정보의 결과를 이용하는 접근법을 개시한다. 구체적으로, 표적 폴리뉴클레오타이드를 생성하는 방법을 개시하며, 여기서 표적 폴리뉴클레오타이드는 컴퓨터 프로그램에 의해 일련의 연속 올리고뉴클레오타이드로 해석되고, 상기 표적 폴리뉴클레오타이드는 드 노보 합성된 올리고뉴클레오타이드를 개시 올리고뉴클레오타이드에 일방향 또는 쌍방향 방식으로 연속 부가함으로써 생성된다.
- [0012] WO2004/033619호는 또한 컴퓨터-지향 폴리뉴클레오타이드 어셈블리에 대한 게놈 서열 정보의 결과를 이용하는 접근법을 개시한다.
- [0013] 지난 수년간 DNA 합성 기술에서 상당한 진전이 있었지만, 용적, 처리량, 특히 DNA 길이에 대해서는 여전히 심각한 제한이 있다.

발명의 내용

해결하려는 과제

- [0014] 본 발명의 목적은, 표적 ds 폴리뉴클레오타이드를 합성하는 시간을 단축시키는 것을 목표로 하여, 이분쇄(ds) 폴리뉴클레오타이드를 합성하는 개선된 방법을 제공하는 것이다.
- [0015] 이러한 목적은 본 발명의 주제에 의해 해결된다.

과제의 해결 수단

- [0016] 본 발명에 따르면, 소정 서열을 갖는 표적 이분쇄(ds) 폴리뉴클레오타이드를 합성하는 방법으로서,
- [0017] a) 올리고뉴클레오타이드 라이브러리 멤버의 다양성을 포함하는 올리고뉴클레오타이드 라이브러리를 어레이 장치 내에 제공하는 단계로서, 상기 라이브러리 멤버 각각은 상이한 뉴클레오타이드 서열을 갖고 수용액 중의 개별 라이브러리 컨테인먼트(containment)에 함유되고, 다양성은 적어도 하나의 돌출부를 갖는 일본쇄 올리고뉴클레오타이드(ss 올리고) 및 이분쇄 올리고뉴클레오타이드(ds 올리고)를 포함하고 정합 올리고뉴클레오타이드의 적어도 10,000 쌍을 커버하는, 단계,
- [0018] b) 제1 단계에서, 액체 핸들러를 사용하여 상기 라이브러리로부터 적어도 제1 쌍의 정합 올리고뉴클레오타이드를 제1 반응 컨테인먼트로 전달하고, 정합 올리고뉴클레오타이드를 어셈블리하여, 적어도 하나의 돌출부를 포함하는 제1 반응 생성물을 획득하는 단계,
- [0019] c) 제2 및 추가 단계에서, 액체 핸들러를 사용하여, 상기 라이브러리로부터 적어도 제2 및 추가 쌍의 정합 올리고뉴클레오타이드를 각각 제2 및 추가 반응 컨테인먼트로 전달하고, 정합 올리고뉴클레오타이드를 어셈블리하여, 각각 적어도 하나의 돌출부를 포함하는 제2 및 추가 반응 생성물을 각각 획득하는 단계,
- [0020] d) 상기 제1, 제2 및 추가 반응 생성물을 소정 워크플로우로 어셈블리함으로써, 돌출부를 갖는 상기 표적 ds 폴리뉴클레오타이드를 생성한 후, 임의로 평활 말단을 제조하기 위해 최종화 단계를 수행하는 단계를 포함하고,
- [0021] 단계 b) 및 c) 및 어셈블리 워크플로우에 사용된 정합 올리고뉴클레오타이드의 상기 쌍은 상기 표적 ds 폴리뉴클레오타이드를 생성하는 알고리즘을 사용하여 결정되는, 방법이 제공된다.

- [0022] 구체적으로, 일련의 상이한 표적 ds 폴리뉴클레오타이드는 동일한 올리고뉴클레오타이드 라이브러리를 사용하여 합성된다. 구체적으로, 상기 상이한 표적 ds 폴리뉴클레오타이드는 상이한 서열을 갖고, 서로의 단편이 아니다.
- [0023] 구체적으로, 상기 상이한 표적 ds 폴리뉴클레오타이드는 50% 미만, 바람직하게는 30% 미만의 서열 동일성을 갖는다. 구체적으로, 상기 상이한 표적 ds 폴리뉴클레오타이드는 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 또는 21% 미만의 서열 동일성을 갖는다. 보다 더 바람직하게는, 상기 상이한 표적 ds 폴리뉴클레오타이드는 서로 20% 또는 10% 미만의 서열 동일성을 갖고, 구체적으로 이들은 19, 18, 17, 16, 15, 14, 13, 12, 11, 9, 8, 7, 6 또는 5% 미만의 서열 동일성을 갖는다.
- [0024] 구체적으로, 상기 표적 ds 폴리뉴클레오타이드는 DNA 분자이다.
- [0025] 구체적으로, 예를 들면, 바람직하게는 25 사이클의 PCR을 수행함으로써 하나 이상의 증폭 단계가 수행된다. 구체적으로, 상기 PCR은 HiFi 열안정성 DNA 폴리머라제(Phusion 또는 Q5) 및 조립된 단편의 각각의 돌출부에 상보성인 2개 올리고뉴클레오타이드를 사용하고, 상기 상보성 올리고뉴클레오타이드는 TypeIIS 제한 효소(BfuAI)의 절단 부위를 포함한다. 구체적으로, 증폭된 생성물은 TypeIIS 제한 효소와 접촉되고, 이는 원래의 돌출부를 증폭된 단편에 도입시킨다. 구체적으로, 상기 증폭 단계는, 제1, 제2, 제3 또는 추가 반응 생성물이 각각 증폭되는, 제1, 제2, 제3 또는 추가 어셈블리 단계 중의 어느 하나 이상 후에 수행된다. 구체적으로, 상기 증폭 단계는, 표적 ds 폴리뉴클레오타이드가 증폭되는 표적 ds 폴리뉴클레오타이드의 어셈블리 후에 수행된다.
- [0026] 구체적으로, 소정 워크플로우(또한 "어셈블리 워크플로우"로서 지칭됨)는 다음과 같이 구체적으로 특성화되는 계층 구조이다:
- [0027] 계층적 워크플로우는, 중간체로서 생성되는 폴리뉴클레오타이드의 중간 어셈블리된 정합 쌍의 병행 또는 개별 생성을 의미하고, 각 중간체는 개별 반응 구획에서 어셈블리되고, 중간체는 표적 폴리뉴클레오타이드 또는 이의 일부를 획득하기 위해 추가로 어셈블리된다. 특정 예에 따르면, 제1 단계에서, 올리고뉴클레오타이드의 정합 쌍을 병행하여 독립적 반응 구획에서 조합하고, 이에 의해 시약 올리고뉴클레오타이드의 조합된 크기, 및 시약 올리고뉴클레오타이드와 동일한 돌출부 길이를 갖는 폴리뉴클레오타이드를 각 구획에서 생성한다. 제2 및 후속 단계에서, 시약으로서 이전 생성물 또는 다른 올리고뉴클레오타이드를 사용하여 이 프로세스를 반복적으로 반복하고, 이에 의해 동일한 돌출부 크기를 유지하는 시약 폴리뉴클레오타이드의 조합된 크기의 폴리뉴클레오타이드를 각 계층에서 생성한다. 마지막 이전의 단계가 3개 구획을 갖는 경우, 정합 쌍을 갖는 2개의 구획만을 먼저 반응시킨 다음, 이 생성물과 마지막 구획 사이의 추가 반응 단계는 표적 폴리뉴클레오타이드를 생성할 것이다. 대안적으로, 3개 구획이 총 2개의 정합 쌍만을 형성할 수 있는 폴리뉴클레오타이드를 함유하는 경우, 3개 구획을 조합하면 표적 폴리뉴클레오타이드가 생성된다.
- [0028] 구체적으로, 어셈블리 워크플로우는 자동화된다. 구체적으로, 자동화 워크플로우는, 1개 또는 수개 구획의 전체 또는 일부 내용물을 비어있거나 비어있지 않을 수 있는 다른 사전-지정된 구획으로 직렬 또는 병렬로 전달할 수 있는 마이크로유체 핸들러를 사용한다.
- [0029] 구체적으로, 어셈블리 워크플로우는 서열-의존적이고, 이는, 정합 쌍이 워크플로우의 임의의 단계에서 조합되는 경우, 이들이 표적 ds 폴리뉴클레오타이드의 대부분 또는 최종적으로 표적 ds 폴리뉴클레오타이드를 초래하도록 특정 순서가 주형의 서열에 의해 결정된다는 것을 의미한다. 구체적으로, 워크플로우는 주형의 서열 또는 표적 ds 폴리뉴클레오타이드의 서열에 따라 결정된다.
- [0030] 구체적으로, 본원에 기재된 방법에 의해, 최대 1,000, 5,000, 10,000 또는 100,000 염기 쌍(bp) 이상 길이의 폴리뉴클레오타이드가 저렴한 가격 및 고속으로 생성될 수 있다.
- [0031] 본원에 기재된 방법은 구체적으로 하기 성분을 포함한다: A) 전체 유전자 서열 공간을 커버하도록 설계되고 액체 핸들러 또는 마이크로유체 장치에 의한 효율적 접근을 위해 공간 내의 올리고를 편성하는 올리고뉴클레오타이드의 사전-구축된 라이브러리. 라이브러리의 공간 편성이 필요한 올리고뉴클레오타이드에의 접근에 필요한 시간을 단축하는 경우, 접근은 효율적인 것으로 간주된다. 구체적으로, 상기 접근은, 라이브러리의 총 처리 시간을 감소 또는 저하시키는 경우에 효율적인 것으로 간주되고, 상기 총 처리 시간은 표적 ds 폴리뉴클레오타이드의 합성 중에 라이브러리 멤버를 처리하는데 소비된 시간이다. 구체적으로, 다른 편성, 특히 공간적으로 랜덤하게 배치된 올리고뉴클레오타이드 또는 사전식(lexicographical) 순서와 비교하여, 작업 코스트를 감소시키거나 올리고뉴클레오타이드에의 접근과 연관된 필요한 소모품의 양을 감소시키는 경우, 상기 접근은 추가로 효

올적인 것으로 간주된다. 구체적으로, 라이브러리의 총 처리 시간이, 랜덤으로 또는 사전식 편성된 라이브러리의 총 처리 시간과 비교하여 적어도 5, 10, 15, 20, 25 또는 50% 단축되는 경우, 접근은 효율적인 것으로 간주된다.

[0032] B) 부정합 없이 긴 폴리뉴클레오타이드를 생성하기 위해, 알고리즘에 의해 결정되는 서열-특이적 계층적 어셈블리 워크플로우.

[0033] 본원에 기재된 라이브러리는 구체적으로, 라이브러리 멤버로도 또한 지칭되는, 일본쇄(ss) 및 이본쇄(ds) 올리고뉴클레오타이드(올리고)를 포함한다. 이들 라이브러리 멤버는 사전-구축되고, 저장 안정한 용액에 제공되고, 어레이 장치 내의 정의된 위치에 배치된다. 라이브러리의 올리고를 합성하고, 필요할 때까지 어레이 장치에 저장한다.

[0034] 구체적으로, 올리고뉴클레오타이드는 뉴클레오타이드 단량체의 선형 중합체이고, 테옥시아데노신을 나타내는 "A", 테옥시티미딘을 나타내는 "T", 테옥시구아노신을 나타내는 "G" 및 테옥시시티딘을 나타내는 "C"를 포함하거나, 통상의 염기(A, G, C, T) 이외에, 뉴클레오타이드-유사체, 예를 들면, 이노신 및 2'-테옥시이노신 및 이들의 유도체(예를 들면, 7'-데아자-2'-테옥시이노신, 2'-데아자-2'-테옥시이노신), 아졸-(예를 들면, 벤즈이미다졸, 인돌, 5-플루오로인돌) 또는 니트로아졸 유사체(예를 들면, 3-니트로피롤, 5-니트로인돌, 5-니트로이미다졸, 4-니트로피라졸, 4-니트로벤즈이미다졸) 및 이들의 유도체, 비사이클릭 당 유사체(예를 들면, 하이포크산틴 - 또는 인다졸 유도체, 3-니트로이미다졸, 또는 이미다졸-4,5-디카복사아미드로부터 유래하는 것들), 유니버설 염기 유사체의 5'-트리포스페이트(예를 들면, 인돌 유도체로부터 유래함), 이소카보스티릴 및 이의 유도체(예를 들면, 메틸이소카보스티릴, 7-프로피닐이소카보스티릴), 수소 결합 유니버설 염기 유사체(예를 들면, 피롤로피리미딘), 및 기타 화학적으로 개질된 염기(예컨대, 디아미노푸린, 5-메틸시토신, 이소구아닌, 5-메틸-이소시토신, K-2'-테옥시리보스, P-2'-테옥시리보스) 또는, 예를 들면, 상이한 염기-쌍 우선순위를 가질 수 있고 유사한 엄격성/확률을 갖는 하나 보다 많은 천연 뉴클레오타이드 염기와 쌍을 이룰 수 있는 다른 변형된 염기를 포함할 수 있다. 단량체는 포스포디에스테르 결합에 의해, 또는 특정의 경우에, 펩티드 결합 또는 포스포티오에이트 결합에 의해, 또는 뉴클레오타이드 결합의 임의의 다른 유형에 의해 결합된다.

[0035] 구체적으로, 일본쇄 DNA 올리고뉴클레오타이드 라이브러리 멤버(본원에서 간단히 ss 올리고로서 지칭됨)은 천연 뉴클레오타이드(예를 들면, 아데노신, 티미딘, 구아노신, 시티딘, 우리딘, 테옥시아데노신, 테옥시티미딘, 테옥시구아노신, 및 테옥시시티딘); 뉴클레오타이드 유사체(예를 들면, 이노신, 또는 5-메틸이소시토신, 또는 3-니트로피롤, 5-니트로인돌, 피롤리딘, 4-니트로이미다졸, 4-니트로피라졸, 4-니트로벤즈이미다졸, 4-아미노벤즈이미다졸, 5-니트로인다졸, 3-니트로이미다졸, 5-아미노인돌, 벤즈이미다졸, 5-플루오로인돌, 인돌, 메틸이소카보스티릴, 피롤로피리미딘 7-프로피닐이소카보스티릴, 2-아미노아데노신, 2-티오티미딘, 3-메틸 아데노신, 5-메틸시티딘, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-요오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-시티딘, C5-메틸시티딘, 2-아미노-아데노신, 7-데아자-아데노신, 7-데아자-구아노신, 8-옥소아데노신, 8-옥소구아노신, 0(6)-메틸구아닌, 및 2-티오시티딘); 화학적으로 또는 생물학적으로 변형된 염기(메틸화된 염기 포함); 삽입된 염기; 변형된 당(예를 들면, 리보스, 2'-테옥시리보스, 아라비노즈, 및 핵소즈); 및/또는 변형된 포스페이트 그룹(예를 들면, 포스포티오에이트 및 5'-N-포스포라미다이트 결합)이거나 이를 포함한다.

[0036] 구체적으로, 이본쇄 DNA 올리고뉴클레오타이드 라이브러리 멤버(본원에서 간단히 ds 올리고로서 지칭됨)은 천연 뉴클레오타이드(예를 들면, 아데노신, 티미딘, 구아노신, 시티딘, 우리딘, 테옥시아데노신, 테옥시티미딘, 테옥시구아노신, 및 테옥시시티딘); 뉴클레오타이드 유사체(예를 들면, 이노신, 또는 5-메틸이소시토신, 또는 3-니트로피롤, 5-니트로인돌, 피롤리딘, 4-니트로이미다졸, 4-니트로피라졸, 4-니트로벤즈이미다졸, 4-아미노벤즈이미다졸, 5-니트로인다졸, 3-니트로이미다졸, 5-아미노인돌, 벤즈이미다졸, 5-플루오로인돌, 인돌, 메틸이소카보스티릴, 피롤로피리미딘 7-프로피닐이소카보스티릴, 2-아미노아데노신, 2-티오티미딘, 3-메틸 아데노신, 5-메틸시티딘, 2-아미노아데노신, C5-브로모우리딘, C5-플루오로우리딘, C5-요오도우리딘, C5-프로피닐-우리딘, C5-프로피닐-시티딘, C5-메틸시티딘, 2-아미노-아데노신, 7-데아자-아데노신, 7-데아자-구아노신, 8-옥소아데노신, 8-옥소구아노신, 0(6)-메틸구아닌, 및 2-티오시티딘); 화학적으로 또는 생물학적으로 변형된 염기(메틸화된 염기 포함); 삽입 염기; 변형된 당(예를 들면, 리보스, 2'-테옥시리보스, 아라비노즈, 및 핵소즈); 및/또는 변형된 포스페이트 그룹(예를 들면, 포스포티오에이트 및 5'-N-포스포라미다이트 결합)이거나 이를 포함하고, 상보성 일본쇄 올리고뉴클레오타이드를 완전히 또는 부분적으로 어닐링함으로써 형성된다.

[0037] 구체적으로, 올리고뉴클레오타이드 라이브러리 멤버는, H-포스포네이트, 포스포디에스테르, 포스포트리에스테르 또는 포스파이트 트리에스테르 합성 방법을 포함하는 임의의 화학적 폴리뉴클레오타이드(올리고뉴클레오타이드)

합성 방법 또는 임의의 대량 병렬 올리고뉴클레오타이드 합성 방법, 예를 들면, 마이크로어레이 또는 마이크로 유체-기반 올리고뉴클레오타이드 합성(예를 들면, 문헌[참조: Gao et al. 2001, (LeProust et al. 2010) (Bonde et al. 2014a)]에 기재된 바와 같음)에 의해 생성할 수 있다.

- [0038] 구체적으로, 올리고뉴클레오타이드 라이브러리 멤버는, 하이브리드 RNA-ssDNA 분자를 생성하는, DNA 폴리머라제 단백질 또는 역전사효소 단백질에 의한 ssDNA 합성을 포함하는 임의의 효소적 폴리뉴클레오타이드(올리고뉴클레오타이드) 합성 방법에 의해 생성할 수 있다. 구체적으로, 효소적 폴리뉴클레오타이드 합성 반응은 생체내 또는 시험관내에서 일어날 수 있다.
- [0039] 구체적으로, 올리고뉴클레오타이드 라이브러리 멤버는 임의의 폴리뉴클레오타이드 합성 방법에 의해 뉴클레오타이드 구성 블록으로부터 올리고뉴클레오타이드 서열을 합성함으로써 생성되고, 여기서 구성 블록은 테옥시아데노신을 나타내는 "A", 테옥시티미딘을 나타내는 "T", 테옥시구아노신을 나타내는 "G" 또는 테옥시시티딘을 나타내는 "C", 또는 기타 천연 뉴클레오사이드(예를 들면, 아데노신, 티미딘, 구아노신, 시티딘, 우리딘), 뉴클레오타이드-유사체, 예를 들면, 이노신 및 2'-테옥시이노신 및 이들의 유도체(예를 들면, 7'-데아자-2'-테옥시이노신, 2'-데아자-2'-테옥시이노신), 아졸-(예를 들면, 벤즈이미다졸, 인돌, 5-플루오로인돌) 또는 니트로아졸 유사체(예를 들면, 3-니트로피롤, 5-니트로인돌, 5-니트로이미다졸, 4-니트로피라졸, 4-니트로벤즈이미다졸) 및 이들의 유도체, 비사이클릭 당 유사체(예를 들면, 하이포크산틴- 또는 인다졸 유도체, 3-니트로이미다졸, 또는 이미다졸-4,5-디카복시아미드로부터 유래된 것들), 유니버설 염기 유사체의 5'-트리포스페이트(예를 들면, 인돌 유도체로부터 유래됨), 이소카보스티릴 및 이의 유도체(예를 들면, 메틸이소카보스티릴, 7-프로피닐이소카보스티릴), 수소 결합 유니버설 염기 유사체(예를 들면, 피롤로피리미딘), 또는 임의의 기타 화학적으로 변형된 염기(예컨대, 디아미노푸린, 5-메틸시토신, 이소구아닌, 5-메틸-이소시토신, K-2'-테옥시리보스, P-2'-테옥시리보스)로 이루어져 있다. 구성 블록은 포스포디에스테르 결합 또는 펩티딜 결합에 의해 또는 포스포로티오에이트 결합에 의해 또는 임의의 기타 유형의 뉴클레오타이드 결합에 의해 결합된다.
- [0040] 본 발명의 특정 실시양태에서, 상기 ss 올리고는 6 내지 26 뉴클레오타이드의 길이를 갖는다. 바람직하게는, ss 올리고는 적어도 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15 뉴클레오타이드의 길이를 갖는다. 바람직하게는, ss 올리고는 최대 26, 25, 24, 23, 22, 21, 20, 19, 18, 17 또는 16 뉴클레오타이드의 길이를 갖는다. 추가의 특정 실시양태에서, 상기 ss 올리고는 26 보다 많은 뉴클레오타이드 길이를 갖는다. 바람직하게는, ss 올리고는 100, 90, 80, 70, 60 또는 50 뉴클레오타이드 미만의 길이를 갖는다.
- [0041] 구체적으로, ds 올리고 라이브러리 멤버는 적어도 하나의 돌출부를 갖는다. 돌출부는 구체적으로, ds 올리고 또는 폴리뉴클레오타이드의 일부이고/이거나 이를 신장하는 하나 이상의 뉴클레오타이드의 반응성(즉, 또 다른 ss 올리고 또는 돌출부와 하이브리드화할 수 있음) ss 말단 스트럿치를 특징으로 한다.
- [0042] 라이브러리는 구체적으로 하나의 돌출부 및 평활 말단을 갖는 ds 올리고를 포함한다. 평활 말단은 ds 올리고 또는 폴리뉴클레오타이드의 일부인 하나 이상의 염기 쌍의 ds 말단 스트럿치를 특징으로 한다.
- [0043] 구체적으로, 양 말단에 돌출부를 갖고 평활 말단부를 갖지 않는 ds 올리고가 라이브러리에 포함될 수도 있다.
- [0044] 구체적으로, ds 올리고는 6 내지 26 염기 쌍의 길이를 갖고, 상기 돌출부는 각 ds 올리고 길이의 절반 이하이다. 구체적으로, ds 올리고는 적어도 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15 염기 쌍의 길이를 갖는다. 구체적으로, ds 올리고는 최대 26, 25, 24, 23, 22, 21, 20, 19, 18, 17 또는 16 염기 쌍의 길이를 갖는다. 구체적으로, 상기 ds 올리고가 6 염기 쌍 길이인 경우, 돌출부는 3 뉴클레오타이드 길이 이하이다. 구체적으로, 상기 ds 올리고가 24 염기 쌍 길이인 경우, 돌출부는 12 뉴클레오타이드 길이 이하이다. 또 다른 특정 실시양태에서, 상기 ds 올리고는 26 염기 쌍 초과를 갖는다. 바람직하게는, ds 올리고는 100, 90, 80, 70, 60 또는 50 염기 쌍 미만의 길이를 갖는다.
- [0045] 본원에 기재된 라이브러리는 특히 물리적 올리고뉴클레오타이드에 의해 구성되고, 표준화 조건하에 합성된다. 올리고뉴클레오타이드는 정제되고, 변형을 포함할 수 있고, 이상적으로는 적절한 완충액 및/또는 부형체에 표준 농도 및 용적으로 유지되어, 즉시 사용 가능하다.
- [0046] 구체적으로, 하기 완충액 및/또는 부형제 중의 어느 것을 사용하여 올리고를 용액 중에 유지시킬 수 있다: 트리스 완충액, T.E. 완충액(트리스-EDTA 완충액) 또는 뉴클레아제 비함유 물. 구체적으로, 라이브러리 멤버는 트리스 완충액에 유지시킬 수 있고, 상기 트리스 완충액은 약 10mM(+/- 1mM 또는 2mM)의 농도로 제공된다. 구체적으로, 라이브러리 멤버는 T.E. 완충액에 유지시킬 수 있다. 구체적으로, 상기 T.E. 완충액은 적어도 약 10mM(+/- 1mM 또는 2mM)의 농도의 트리스, 및 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 또는 1.0mM 중의

어느 하나의 농도의 EDTA로 구성된다. 구체적으로, 뉴클레아제 비함유 물은, 탈이온화되고 여과되고 오토클레이빙된 물이고, 본질적으로 오염 비-특이적 엔도뉴클레아제, 엑소뉴클레아제 및 RNase 활성이 없다.

- [0047] 구체적으로, 모든 라이브러리 멤버는 각각의 경우에서 동일하거나 상이한 완충액 및/또는 부형제를 사용하여 구획된 어레이 장치에서 유지된다.
- [0048] 본원에 기재된 라이브러리는 수천의 올리고를 포함할 수 있다. 구체적으로, 본원에 기재된 라이브러리는 올리고뉴클레오타이드 라이브러리 멤버의 다양성을 포함하고, 각각의 라이브러리 멤버는 상이한 뉴클레오타이드 서열을 갖고, 다양성은 적어도 10,000 쌍의 정합 올리고뉴클레오타이드를 커버한다. 구체적으로, 라이브러리는 적어도 20,000, 30,000, 40,000, 50,000, 60,000, 70,000, 80,000, 90,000 또는 100,000 쌍의 정합 올리고뉴클레오타이드를 포함한다. 구체적으로, 라이브러리는 서열 공간 전체를 커버하기 위해 충분한 쌍의 정합 올리고뉴클레오타이드를 함유한다.
- [0049] 본원에 기재된 정합 올리고뉴클레오타이드의 쌍은 부분적 또는 완전히 상보성 서열을 포함하는 일본쇄 올리고뉴클레오타이드를 지칭한다. 정합 올리고의 상기 쌍은 별개의 컨테이너먼트의 ss 올리고로서 라이브러리에 존재할 수 있거나, 2개 이상의 상보성 ss 올리고는 이들이 어닐링하여 ds 올리고를 형성할 수 있는 하나의 컨테이너먼트에 함유될 수 있다. 정합 ss 올리고 쌍의 뉴클레오타이드 서열은 적어도 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 또는 26 뉴클레오타이드에서 상보성일 수 있고, 이에 의해 정합 쌍은 ss 올리고 서열의 하이브리드화(바람직하게는, ss 올리고는 부분적으로 하이브리드화한다)에 의해 신규한 ds 폴리뉴클레오타이드 분자를 형성하여, 돌출부를 갖는 ds 폴리뉴클레오타이드를 수득할 수 있다.
- [0050] ss 올리고는 구체적으로 2 또는 3개 하이브리드화 파트너로 이루어진 정합 쌍의 일부일 수 있다. 구체적으로, ss 올리고는, 또 다른 ss 올리고, 또는 상보성 돌출부를 갖는 ds 올리고인 제2 하이브리드화 파트너와 하이브리드화할 수 있는 제1 하이브리드화 파트너로서 사용될 수 있다.
- [0051] 구체적으로, ss 올리고는, 제2 및 제3 하이브리드화 파트너로서 사용되는 2개의 상이한 ss 및/또는 ds 올리고, 또는 2개의 상이한 ds 폴리뉴클레오타이드와 하이브리드화할 수 있는 제1 하이브리드화 파트너로서 사용될 수 있다. 구체적으로, 제1 하이브리드화 파트너는 정합 ss 올리고이고, ss 올리고의 제1 부분은 제2 하이브리드화 파트너와 하이브리드화하고, ss 올리고의 제2 파트너는 제3 하이브리드화 파트너와 하이브리드화하고, 이에 의해 겹이 없는 3개 하이브리드화 파트너로 구성된 1개의 ds 폴리뉴클레오타이드를 수득한다.
- [0052] 정합 ds 올리고의 쌍은 구체적으로 ds 올리고의 각 돌출부에서 상보성 서열을 특징으로 하고, 예를 들면, 각 돌출부는 적어도 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 또는 26 뉴클레오타이드에서 상보성이고, 따라서 정합 쌍은 돌출부 서열의 하이브리드화에 의해 새로운 ds 폴리뉴클레오타이드 분자를 형성할 수 있다.
- [0053] 본원에 기재된 라이브러리는 구체적으로, 각각의 ds 올리고 라이브러리 멤버가 상이한 뉴클레오타이드 서열을 갖는 일본쇄 올리고뉴클레오타이드 라이브러리 멤버의 다양성을 포함할 수 있다. 구체적으로, 상기 다양성은 적어도 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 또는 200,000개의 상이한 ds 올리고를 커버한다.
- [0054] 본원에 기재된 라이브러리는 구체적으로, 각각의 ss 올리고 라이브러리 멤버가 상이한 뉴클레오타이드 서열을 갖는 일본쇄 올리고뉴클레오타이드 라이브러리 멤버의 다양성을 포함할 수 있다. 구체적으로, 상기 다양성은 적어도 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 또는 200,000개의 상이한 ss 올리고를 커버한다. 구체적으로, 상기 ss 올리고는 특히 ds 폴리뉴클레오타이드의 어셈블리에서 링커로서 사용될 수 있다.
- [0055] 구체적으로, 상기 다양성은 상이한 라이브러리 멤버가 적어도 하나의 염기 또는 염기 쌍에서 상이한 것을 의미한다. 하나의 라이브러리 멤버는 실제로는 동일한 서열의 ss 또는 ds 올리고뉴클레오타이드의 복수 카피를 포함할 수 있다. 라이브러리 멤버의 이러한 복수 카피는 특히 하나의 라이브러리 컨테이너먼트에만 함유되어 있다.
- [0056] 본 발명의 특정 실시양태에서, 상기 다양성은 포스포릴화되는 ss 올리고 및/또는 ds 올리고를 커버한다. 특정 실시양태는 포스포릴화, 메틸화, 비오틴화, 또는 형광단 또는 켄처에의 결합 중의 하나 이상에 의해 변형되는 ss 올리고 또는 ds 올리고를 지칭한다. 따라서, 본원에 기재된 라이브러리는, 비변형된 ss 올리고, 포스포릴화된 ss 올리고, 메틸화된 ss 올리고, 비오틴화된 ss 올리고, 포스포릴화된, 비오틴화된 및 메틸화된 ss 올리고, 비변형된 ds 올리고, 포스포릴화된 ds 올리고, 메틸화된 ds 올리고, 비오틴화된 ds 올리고 및 포스포릴화된, 비오틴화된 및 메틸화된 ds 올리고 중의 어느 것 또는 모두일 수 있는 라이브러리 멤버를 포함한다. 바

람직하게는, 라이브러리 멤버는 5' 포스포릴화를 포함한다. 구체적으로, 본원에 기재된 라이브러리는 형광단 또는 켄처를 포함하는 ss 올리고 및 형광단 또는 켄처를 포함하는 ds 올리고를 추가로 포함한다.

[0057] 본 발명의 특정 실시형태에서, 올리고뉴클레오타이드 라이브러리는 어레이 장치 내에 제공되고, 라이브러리 멤버는 각각 수용액 중의 개별 라이브러리 컨테인먼트 내에 함유된다. 구체적으로, 상기 어레이 장치는 미세역가 플레이트, 마이크로유체 마이크로플레이트, 모세관 세트, 마이크로어레이 또는 바이오칩, 바람직하게는 DNA 및/또는 RNA 바이오칩 중의 어느 것이다. 상기 어레이 장치는 상술된 컨테인먼트 중의 단지 하나, 모든 또는 임의 수를 포함할 수 있다.

[0058] 본 발명의 추가의 특정 실시양태에서, 하나 보다 많은 상이한 라이브러리 멤버는 단지 하나의 라이브러리 컨테인먼트 내에 함유될 수 있다. 구체적으로, 하나의 라이브러리 컨테인먼트에 함유된 상기 상이한 라이브러리 멤버는 이들이 서로 어닐링할 수 없는 이러한 서열의 ss 올리고이다. 구체적으로, 하나의 라이브러리 컨테인먼트에 함유된 상기 상이한 라이브러리 멤버는 이들이 그 속에 함유된 다른 ds 올리고에 결합할 수 없는 이러한 서열의 ds 올리고이다. 구체적으로, 하나의 라이브러리 컨테인먼트에 함유된 상기 상이한 라이브러리 멤버는 이들이 서로 어닐링할 수 없는 이러한 서열의 ss 올리고 및 ds 올리고이다.

[0059] 특정 실시양태에서, 상기 개별 라이브러리 컨테인먼트는 3차원 순서로 공간적으로 배열되어 있고, 개개 구획은 x-, y- 및 z-축 내의 정의된 좌표에서 장치 내에 배치된다. 구체적으로, 상기 3차원 순서는, 적어도 부분적으로 또는 완전히 적층되어 있는, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60 또는 그 이상의 적층된 라이브러리 컨테인먼트 중의 적어도 어느 하나를 포함한다. 바람직하게는, 라이브러리 컨테인먼트는 상이한 레이(lay)에 배치되고, 이는 상이한 레이에 상하로 배치된다. 구체적으로, 레이는 3차원 순서 내의 소정의 위치에 배치된다. 바람직하게는, 하나의 레이 내의 상기 라이브러리 컨테인먼트 각각은 소정 위치에서 2차원 순서로 공간적으로 배열된 일련의 라이브러리 멤버를 포함한다.

[0060] 구체적으로, 3차원 순서는, 주로 합성 시간의 단축에 작용하는 파라미터에 의해 사전-정의된다. 바람직하게는, 상기 파라미터는 사용 빈도이고, 예를 들면, 천연 존재 또는 표적 ds 폴리뉴클레오타이드 또는 이의 단편에 통상 사용되는 DNA 서열 내에 정합 쌍을 빈번히 형성하는 이들 올리고를 서로 근접하여 배치한다. 임의의 제공된 서열을 구축하기 위해 다수의 올리고가 필요하기 때문에, 라이브러리 내의 올리고의 공간 분포 대부분은, 라이브러리를 스캐닝하고 목적하는 올리고를 검색하는데 필요한 시간에 기인하여 시간 및 자연 낭비를 초래할 수 있다. 그러나, 올리고의 특정 분포를 사용함으로써, 정합 올리고의 쌍을 반응 컨테인먼트로 전달하기 위한 자동 장치의 이동이 최소한으로 된다. 예를 들면, 올리고는 마이크로-웰 플레이트에 저장될 수 있고, 제1 플레이트는 올리고뉴클레오타이드의 가장 일반적 정합 쌍을 함유하고, 최후 플레이트가 가장 사용 빈도가 낮은 올리고를 함유할 때까지 추가 플레이트는 감소 순서로 배열된다.

[0061] 본원에 기재된 방법에서, 본원에 기재된 라이브러리로부터의 올리고뉴클레오타이드는 액체 핸들러를 사용하여 반응 컨테인먼트로 전달된다. 구체적으로, 상기 액체 핸들러는 미소액적 핸들러일 수 있다. 구체적으로, 액체 핸들러는 자동화된다. 액체 핸들러를 사용하여 10, 20, 30, 40, 50, 60, 70, 80, 90 100, 200 또는 500nL 중의 적어도 어느 하나의 적합한 용적을 전달할 수 있고, 예를 들면, 일본쇄 올리고뉴클레오타이드, 일본쇄 올리고뉴클레오타이드 및 이분쇄 올리고뉴클레오타이드의 정합 쌍 등의 라이브러리 멤버의 10^9 , 10^{10} , 10^{11} 또는 10^{12} 카피의 적어도 어느 하나는 하나의 반응 컨테인먼트에 넣는다. 바람직하게는, 특정 올리고의 적어도 약 10^{11} 카피(예를 들면, 6.06×10^{11} 카피)를 하나의 반응 컨테인먼트에 넣어 또 다른 올리고와 반응시킨다. 바람직하게는, 올리고가 액체 핸들러에 의해 전달되는 용적은 10 내지 1000 nL이다. 보다 바람직하게는, 이는 10 내지 500 nL 및 보다 더 바람직하게는 50 내지 250 nL이다.

[0062] 구체적으로, 반응 컨테인먼트는 미세역가 플레이트, 마이크로유체 마이크로플레이트, 모세관 세트, 마이크로어레이 또는 바이오칩, 바람직하게는 DNA 및/또는 RNA 바이오칩의 구획 유닛, 예를 들면, 웰이다. 구체적으로, 반응 컨테인먼트는, 상보성 쇠 상호작용 및 수소 결합에 의해 하나의 핵산 쇠가 제2 핵산 쇠에 결합하여 이분쇄 올리고뉴클레오타이드를 생성하는 환경을 특징으로 한다. 이러한 조건은 핵산을 함유하는 수용액 또는 유기 용액의 화학적 성분 및 이들의 농도(예를 들면, 염, 킬레이트제, 포름아미드), 및 혼합물의 온도를 포함한다. 인큐베이션 시간의 길이 또는 반응 챔버 치수 등의 다른 공지된 인자가 환경에 영향을 미칠 수 있다.

[0063] 본원에 제공된 방법에 따르면, 올리고뉴클레오타이드는 라이브러리로부터 반응 컨테인먼트로 전달되고, 어셈블리되어 반응 생성물을 수득한다. 구체적으로, 상기 어셈블리는 ss 뉴클레오타이드 서열을 하이브리드화하는 임의의 방법, 및/또는 효소 반응 및/또는 화학 반응인 결합 반응에 의한 것이다. 구체적으로, 상기 결합 반응은,

결찰 반응을 가능하게 하는 리가제 또는 리보자임을 사용한 효소적 결찰 반응이다. 바람직하게는, T4 DNA 리가제, T7 DNA 리가제, T3 DNA 리가제, Taq DNA 리가제, DNA 폴리머라제, 또는 인공 효소가 결찰 반응에 사용된다. 바람직하게는, 하기 결찰 반응이 사용된다: 1mM ATP가 보충된 μL 당 10 응집 말단 유닛의 농도에서 T4 DNA 리가제(Sambrook and Russel, 2014, Chapter 1, Protocol 17).

[0064] 구체적으로, 상기 어셈블리는 정합 돌출부를 하이브리드화하여 직접, 또는 적합한 ss 올리고 링커를 하이브리드화하여 간접적으로 수행되고, ss 올리고 링커는 상기 제1, 제2 또는 추가 반응 생성물을 어셈블리하기 위해 상기 라이브러리로부터 선택되고 전달되는 상기 라이브러리에 함유된 ss 올리고이다.

[0065] 올리고뉴클레오타이드는 정의된 워크플로우에 따라 구체적으로 어셈블리된다. 워크플로우는, 표적 ds 폴리뉴클레오타이드를 생성하기 위해 어셈블리에 사용될 수 없는 부정합 또는 반응 생성물을 회피하도록 특별히 설계된다. 별도의 방법에서 어닐링할 수 있는 부분 작제물이 있는 경우, 폭주, 즉 비조절된 중합 반응이 발생할 수 있다. 바람직하지 않은 작제물 또는 폭주 반응을 초래할 수 있는 정합 올리고뉴클레오타이드 쌍의 조합을 회피하기 위해, 정합 올리고뉴클레오타이드 쌍은 어셈블리 단계, 즉 특정 워크플로우의 소정 서열에서 어셈블리된다. 바람직하게는, 상기 특정 워크플로우는 선형이 아니고 계층적이며, 즉 중간 반응 생성물이 추가의 중간 반응 생성물 또는 표적 ds 폴리뉴클레오타이드 서열에 추가로 어셈블리되기 전에, 바람직하지 않은 반응 생성물을 가능한 회피하여 편리하게 생성된 표적 ds 폴리뉴클레오타이드의 불연속 부분으로서 정의되는 중간 반응 생성물을 제공하는 알고리즘에 따른다.

[0066] 선형 워크플로우에서, 폴리뉴클레오타이드는 리딩 쇄의 3' 말단에서 개시하여 선형 방식으로 어셈블리되고, 후속 올리고를 부가하여 리딩 쇄의 3' 말단을 후속 올리고의 5' 말단에 연결시킨다. 예를 들면, 올리고 B는 올리고 A에 연결되고, 올리고 C는 올리고 B에 연결되고, 올리고 D는 올리고 C 등에 연결된다. 이러한 어셈블리는, 모든 올리고를 동시에 반응 컨테인먼트에 부가함으로써 동시에 달성될 수 있거나, 폴리뉴클레오타이드는 올리고 A, B, C, D 등을 반응 컨테인먼트에 연속 부가함으로써 서서히 확장된다.

[0067] 계층적 워크플로우는, 예를 들면, 올리고 D가 올리고 C 뿐만 아니라 상보성 서열 또는 돌출부에 기인하여 올리고 A에도 결찰되는 경우에 필요할 수 있다. 상기 기재된 선형 워크플로우는 목적하는 폴리뉴클레오타이드 A-B-C-D에 추가하여 바람직하지 않은 폴리뉴클레오타이드 A-D-B-C-D를 초래할 것이다. 따라서, 폴리뉴클레오타이드는 바람직하게는 계층적 워크플로우에서 어셈블리된다. 따라서, 2개의 별개 반응 컨테인먼트에서 올리고 A와 B 및 올리고 C와 D가 각각 결찰된다. 결찰 반응은 반응 생성물 A-B 및 C-D를 생성하고, 이는 이어서 제3 반응 컨테인먼트로 전달될 수 있으며, 결찰시에 목적하는 폴리뉴클레오타이드 A-B-C-D가 형성된다.

[0068] 구체적으로, 상기 워크플로우는 알고리즘을 사용하여 결정된다. 구체적으로, 상기 알고리즘은, 필요한 경우, 정합 올리고뉴클레오타이드 및 ss 올리고 링커의 쌍을 선택하고, 단순한 서열 분할에 의한 것이 아니라, 표적 ds 폴리뉴클레오타이드를 어셈블리하기 위한 최적 또는 거의 최적 방식을 결정함으로써 어셈블리 워크플로우를 결정하여, 부정합 또는 바람직하지 않은 반응 생성물을 가능한 한 회피한다. 정합 올리고뉴클레오타이드 쌍 및 어셈블리 워크플로우는 구체적으로 바람직하지 않은 (정확하지 않은) 반응 또는 반응 생성물, 예컨대, 회문 서열, 폭주 반응 및 모호하지 않은 어셈블리를 회피하도록 선택된다. 정확한 반응 생성물 이외에 부정확한 반응 생성물이 있는 경우, 이러한 부정확한 반응 생성물은, 예를 들면, 다음과 같이 정확한 것으로부터 적합하게 분리된다: 겔 전기영동을 사용하여 특정 크기의 올리고뉴클레오타이드 또는 폴리뉴클레오타이드를 검출하고, 바람직하지 않은 반응 생성물의 크기에 상응하는 겔의 밴드를 소화하고 정제하는 것. 구체적으로, 정확한 반응 생성물은 태그 또는 표지를 서열에 도입함으로써 검출할 수 있다. 구체적으로, 올리고는 올리고의 돌출부와 하이브리드화할 수 있는 비오틴화 올리고뉴클레오타이드 어댑터를 사용하여 포획할 수 있고, 상기 어댑터는 기질에 고정되고 스트렙트아비딘으로 코팅된다. 비-포획된 부정확한 생성물은 세척에 의해 제거하고, 후속적으로, 정확한 생성물은 온도를 증가시킴으로써 어댑터로부터 방출시킨다. 구체적으로, 당해 기술분야에 공지된 추가의 분리 방법을 적용할 수 있다. 구체적으로, 이러한 방법은 크로마토그래피 또는 친화성 분리 방법을 포함할 수 있다.

[0069] 본 발명의 특정 실시양태에서, 상기 표적 ds 폴리뉴클레오타이드는 적어도 48 뉴클레오타이드의 길이를 갖는다. 구체적으로, 상기 표적 ds 폴리뉴클레오타이드는 100, 200, 300, 400, 500, 1,000, 10,000, 100,000, 200,000 또는 500,000 뉴클레오타이드 중의 적어도 어느 하나의 길이를 갖는다.

[0070] 전형적으로, 주형은 표적 ds 폴리뉴클레오타이드를 합성하기 위한 모델로서 사용된다. 구체적으로, 상기 표적 ds 폴리뉴클레오타이드의 뉴클레오타이드 서열은 주형의 뉴클레오타이드 서열과 동일하다.

- [0071] 특정 실시양태에서, 목적 서열(SOI)은 일본쇄 주형으로서 제공되고/되거나 2개의 일본쇄 주형 서열로 번역되며, 이에 기초하여 표적 ds 폴리뉴클레오타이드가 합성된다. 특정 실시양태에서, 제1 주형은 SOI의 서열을 포함하고, 제2 주형은 SOI에 대한 역 상보체를 포함한다.
- [0072] 본 발명의 추가의 실시양태에서, 상기 표적 ds 폴리뉴클레오타이드는 상기 주형과 동일한 서열을 갖는 프록시 ds 폴리뉴클레오타이드이고, 프록시 ds 폴리뉴클레오타이드는, 표적 ds 폴리뉴클레오타이드의 서열과 상이한 목적 서열(SOI)을 갖는 폴리뉴클레오타이드를 수득하기 위해 추가로 변형된다. 전형적으로, 프록시 ds 폴리뉴클레오타이드는 중간 생성물로서 생성되고, 이로부터 SOI를 특징으로 하는 ds 폴리뉴클레오타이드가 돌연변이유발의 하나 이상의 추가 단계에 의해 생성될 수 있다.
- [0073] 구체적으로, 이에 따라 상기 프록시 ds 폴리뉴클레오타이드가 합성되는 상기 주형의 서열은 상기 SOI와 동일하지 않다. 구체적으로, 상기 주형의 서열은 상기 SOI에 대해 100, 99, 98, 97, 96, 95, 94, 93, 92, 또는 91% 중의 어느 하나 미만으로 동일하고/하거나 90, 91, 92, 93, 94, 95, 96, 97, 98, 또는 99% 중의 어느 것으로 동일하다.
- [0074] 본원에 제공된 방법의 특정 실시양태에서, 한 쇠의 서열의 3' 또는 5' 말단 또는 양 말단의 말단 뉴클레오타이드, 또는 각 ds 쇠는 보다 짧은 서열로 분배하기 전에 제거된다. 구체적으로, 이들은 계산에 의해 제거된다. 이에 의해, SOI와 상이한 주형이 생성된다. 구체적으로, 말단 뉴클레오타이드의 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 또는 25 중의 어느 하나가 서열의 3' 말단 또는 5' 말단으로부터 제거되어 주형을 생성한다. 구체적으로, 상기 뉴클레오타이드(들)는 표적 ds 폴리뉴클레오타이드의 각 말단에서 평활 말단을 생성함으로써 돌출부를 생성하고/하거나 합성의 최종화를 준비한다.
- [0075] 구체적으로, 주형은 일본쇄 또는 이본쇄 서열로 이루어진다. 바람직하게는, 상기 주형은 일본쇄이다. 구체적으로, 상기 2개 일본쇄 주형 서열을 정렬시켜 이본쇄 주형을 생성한다. 구체적으로, 상기 주형의 서열은 보다 짧은 서열, 올리고뉴클레오타이드 라이브러리 멤버를 포함하는 서브-서열로 분할되고, 라이브러리 중의 상기 라이브러리 멤버의 위치는 디지털로 주석이 달린다. 구체적으로, 서브-서열로의 분할은 계층적 워크플로우 및 라이브러리에 존재하는 라이브러리 멤버에 의존한다.
- [0076] 구체적으로, 표적 ds 폴리뉴클레오타이드는 양 말단에 평활 말단을 갖는다.
- [0077] 구체적으로, 본원에 제공된 방법은 최종화 단계를 포함한다.
- [0078] 구체적으로, 상기 최종화 단계는, 각각 3' 말단 및 5' 말단으로부터 사전에 제거된 것들에 상응하는 하나 이상의 뉴클레오타이드(들)를 부가하여, 평활 말단을 생성하는 것을 목적으로 하는 주형을 제조하는 역할을 한다. 구체적으로, 라이브러리로부터의 올리고가 선택되고, 이는 각각 3' 말단 및 5' 말단의 뉴클레오타이드에 상보성이고, 즉 폴리뉴클레오타이드의 점착 말단에 상보성이다. 구체적으로, 이들 올리고는, 최종 생성물을 증폭시키고 나머지 뉴클레오타이드를 각 쇠에 부가하여 평활 말단의 완전한 표적 폴리뉴클레오타이드를 합성하기 위해 준비되는 PCR 반응에서 프라이머로서 사용된다.
- [0079] 구체적으로, 상기 최종화 단계는 표준 키트, 예컨대, 뉴 잉글랜드 바이오랩(New England Biolabs)(product nr. T1030)의 모나크(Monarch) PCR & DNA 클린 업 키트를 사용하여, 나머지 올리고, 효소 및 시약을 제거하고 하류 적용에 적합한 DNA 생성물로서 표적 ds 폴리뉴클레오타이드를 잔류시키는 PCR 생성물의 정제 단계를 포함한다.
- [0080] 또는, 표적 ds 폴리뉴클레오타이드의 한쪽 또는 양 평활 말단은 평활 말단을 갖는 정합 ds 올리고를 선택하거나 돌출부에 상보성인 ss 올리고를 선택하고 임의의 추가 돌출부를 생성하지 않고서 하이브리드화하여 평활 말단을 생성함으로써 생성된다.
- [0081] 구체적으로, 표적 ds 폴리뉴클레오타이드, SOI 또는 주형의 상기 뉴클레오타이드 서열은 천연 또는 인공 기원의 것일 수 있다.
- [0082] 복잡한 SOI를 갖는 ds 폴리뉴클레오타이드를 보다 간단하고 따라서 보다 신속하게 어셈블리 워크플로우에서 생성하기 위해, SOI와 100% 미만 동일한 표적 서열을 갖는 프록시 ds 폴리뉴클레오타이드를 생성할 수 있다. 이어서, 본원에 기재된 어셈블리 방법에 의해 생성된 상기 프록시 ds 폴리뉴클레오타이드는 SOI의 뉴클레오타이드 서열과 100% 동일한 뉴클레오타이드 서열을 갖는 ds 폴리뉴클레오타이드를 생성하기 위해 추가로 변형시킬 수 있다. 구체적으로, 상기 프록시 ds 폴리뉴클레오타이드는 상기 주형의 뉴클레오타이드 서열과 동일한 뉴클레오타이드 서열을 수득하기 위해 지시된 돌연변이유발, 엔도뉴클레아제 또는 엑소뉴클레아제 중의 어느 하나에 의해 추가로 변형된다.

- [0083] 본 발명의 추가의 특정 실시양태에서, 표적 ds 폴리뉴클레오타이드는, ds DNA, ss DNA 또는 RNA 분자 중의 어느 하나인 이의 유도체를 생성하기 위해 추가로 변형된다.
- [0084] 구체적으로, 상기 표적 ds 폴리뉴클레오타이드는 부위-지시된 돌연변이유발에 의해 변형되고, 이에 의해 뉴클레오타이드 삽입, 결실 또는 치환 중의 어느 하나인 하나 이상의 점 돌연변이를 도입한다.
- [0085] 구체적으로, 상기 표적 ds 폴리뉴클레오타이드는, 메틸트랜스페라제, 키나제, CRISPR/Cas9, λ-레드 재조합을 사용한 다중 자동화 게놈 조작(MAGE), 공액 어셈블리 게놈 조작(CAGE), 아르코나우트 단백질 계열(Ago) 또는 이의 유도체, 징크-핑거 뉴클레아제(ZFN), 전사 활성화인자-유사 이펙터 뉴클레아제(TALEN), 메가뉴클레아제, 티로신/세린 부위-특이적 리코비나제(Tyr/Ser SSRs), 하이브리드화 분자, 설푸릴라제, 리코비나제, 뉴클레아제, DNA 폴리머라제, RNA 폴리머라제 또는 TNase 중의 어느 하나를 사용한 효소적 변형에 의해 변형된다.
- [0086] 본 발명의 특정 실시양태에서, 상기 표적 ds 폴리뉴클레오타이드는 주형 또는 SOI의 서열과의 동일성 정도를 검증하기 위해 서열분석된다. 임의의 적합한 서열분석 방법, 예를 들면, 하이브리드화-기반 방법(예를 들면, 분자 비콘, SNP 마이크로어레이, 제한 단편 길이 다형성)을 포함하는 SNP 유전자형 방법, 대립유전자-특이적 PCR, 프라이머 신장-, 5'-뉴클레아제 또는 올리고뉴클레오타이드 결찰 분석, 일본쇄 형태 다형성, 온도 구배 젤 전기영동, 변성 고성능 액체 크로마토그래피, 전체 앰플리콘의 고-해상도 용해(HRM), SNPlex 및 서베이어 뉴클레아제 검정을 포함하는 PCR-기반 방법; PTR의 전체 PCR 앰플리콘의 모세관 서열분석 또는 고효율 서열분석(앰플리콘 서열분석)을 포함하는 서열분석 기반 돌연변이 분석 중의 어느 하나를 사용할 수 있다. 이러한 고-효율(HT) 앰플리콘 서열분석 방법은 플로니 서열분석, 피로서열분석, 일루미나(Solexa) 서열분석, SOLiD 서열분석, 반도체 서열분석, DNA 나노볼 서열분석, 헬리스크프 단일 분자 서열분석, 단일 분자 실시간(SMRT) 서열분석, 나노포어 DNA 서열분석, 터널링 전류 DNA 서열분석, 하이브리드화에 의한 서열분석, 질량 분석에 의한 서열분석, 마이크로유체 생거 서열분석, 현미경-기반 서열분석, RNAP 서열분석을 포함하지만, 이로써 한정되지 않는다.
- [0087] 본원에 제공된 발명에 따르면, 올리고뉴클레오타이드 라이브러리는 라이브러리 멤버의 다양성을 포함하는 어레이 장치 내에 제공되고, 이는 적어도 하나의 돌출부를 갖는 일본쇄 올리고뉴클레오타이드(ss 올리고) 및 이본쇄 올리고뉴클레오타이드(ds 올리고)이며, 각각의 라이브러리 멤버는 상이한 뉴클레오타이드 서열을 갖고 수용액 중의 개별 라이브러리 컨테인먼트에 함유되고, 컨테인먼트는 3차원 순서로 공간적으로 배열되고, 다양성은 정합 올리고뉴클레오타이드의 적어도 10,000 쌍을 커버한다.
- [0088] 구체적으로, 상기 라이브러리 컨테인먼트는, 바람직하게는 사용 빈도에 따라, 3차원 순서로 공간적으로 배열되고, 상기 3차원 순서는, 적어도 부분적으로 또는 완전히 적층되어 있는, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30, 40, 50, 60 또는 그 이상의 적층된 라이브러리 컨테인먼트 중의 적어도 어느 하나를 포함한다.
- [0089] 추가로, 본 발명은 소정 서열을 갖는 일련의 상이한 표적 이본쇄(ds) 폴리뉴클레오타이드를 합성하기 위한 본원에 기재된 올리고뉴클레오타이드 라이브러리의 용도를 제공하고, 상기 상이한 표적 이본쇄(ds) 폴리뉴클레오타이드는 50% 미만, 바람직하게는 30% 미만의 서열 동일성을 갖는다.

도면의 간단한 설명

- [0090] 도 1A. 인간 미토콘드리아의 초가변 영역 II의 4개의 일배체(haplotype)에 상응하는 라이브러리의 작체에 사용된 공급원 서열(100 bp의 단편만이 제시되어 있다)(Anderson et al., 1981: Gene Bank accession nr.: J01415).
- 도 1B. 일배체형의 임의의 가능한 조합을 구축하는데 필요한 올리고의 설계의 스케폴드(각 다형성 부위의 완전한 헤테로접합성을 상징). 양 쇄가 제시되어 있다. 각 Z형 블록은 N이 있는 올리고 쌍 또는 4개 염기 A, T, G 또는 C 중의 어느 것이다. 각 올리고의 각 올리고 서열 스케폴드 상하의 숫자는 올리고의 길이를 나타내고, 괄호 안에는 가변 부위에서 가능한 모든 일배체를 커버하기 위해 라이브러리에 존재해야 하는 올리고의 숫자를 나타낸다. 각각의 ss 올리고는, 어닐링된 쌍으로서 저장되는 밀줄친 것들을 제외하고는, 라이브러리의 구획에 개별적으로 저장되어야 한다.
- 도 2A. DISCOVER로 불러오는 SOI의 뉴클레오타이드 서열(서열번호 1).
- 도 2B. SOI DISCOVER를 구성하는 16 올리고의 뉴클레오타이드 서열.
- 도 2C. 구성 올리고의 이량체 구조. 여기서는 D+ 및 D-에 대해 나타내지만, 동일한 구조는 모든 다른 이량체에 적용된다.

도 3. 웰 플레이트 내의 올리고의 위치.

도 3A. 어닐링 후, 컬럼 1 및 3의 내용은 각각 컬럼 2 및 4로 전달된다.

도 3B. 제1 결합 반응의 인큐베이션 후, 컬럼 2의 내용은 컬럼 4로 전달된다.

도 3C. 제2 결합 반응의 인큐베이션 후, A4의 내용은 웰 B4로 전달되고, 제3 및 최종 결합 반응을 위해 인큐베이션한다. 웰 B4는 128 bp 표적 ds 폴리뉴클레오타이드를 함유한다.

도 4. 실시예 2에 기재된 프로세스의 내용을 나타내는 아크릴아미드 겔(10%). 레인 1: 반응물 D+I(도 2B의 웰 A2). 레인 2: 64 bp dsDNA를 사용한 네가티브 대조군(결찰용). 레인 3: 64 bp dsDNA를 사용한 포지티브 대조군(결찰용). 레인 4 및 5: 2개 회색역에서의 반응물 DI+SC(도 2C의 웰 A4). 레인 6 및 7: 표적 ds 폴리뉴클레오타이드. 레인 8: 50 bp 래더(NEB).

도 5A. 부분적 SOI 및 이의 역 상보체(위치 65-100; 그 이외는 도 1A와 동일). 이탤릭체, 볼드체 및 통상의 폰트의 요소는 상이한 이량체를 나타낸다. 밑줄친 부분은 회피되어야 하는 자가-상보성 돌출부를 강조한다. 상부 서열 서열번호 18; 하부 서열 서열번호 19.

도 5B. 프록시 ds 폴리뉴클레오타이드(위치 65-100)를 생성하기 위한 주형의 부분 서열. 흑색 배경의 염기 쌍은 변경된 부위를 나타내고, 이는 이제 이량체를 비-자가-상보성으로 되게 한다. (수득되는 변형된 올리고는 실시예 2의 0- 및 V+와 일치한다.). 상부 서열 서열번호 20; 하부 서열 서열번호 21.

도 5C. SOI를 갖는 ds 올리고뉴클레오타이드를 생성하기 위해 프록시 ds 폴리뉴클레오타이드를 변형시키는데 사용된 돌연변이유발 프라이머. 밑줄친 문자는 돌연변이유발된 염기를 나타낸다. 상부 서열 서열번호 22; 하부 서열 서열번호 23.

도 6. 어닐링 및 계층적 합성용으로 이들을 제조하기 위해 96-웰 플레이트 상에서, 실시예 1의 라이브러리로부터 전달된 올리고의 배열.

도 7. 계층적 합성 프로세스의 결과를 나타내는 아가로즈 겔 전기영동(2%). 상부 밴드는 608 생성물을 함유하는 것이다. 좌측 레인은 50 bp 래더이다.

도 8. 실시예 4의 서열.

발명을 실시하기 위한 구체적인 내용

- [0091] 본 명세서를 통해 사용되는 특정 용어는 하기 의미를 갖는다.
- [0092] 본원에 사용되는 바와 같이, 용어 "a", "an" 및 "the"는 하나 또는 하나 이상, 즉 적어도 하나를 지칭하기 위해 본원에서 사용된다.
- [0093] 용어 "목적 서열" 또는 "SOI"는, 본원에 제공된 방법에 의해 생성되는 ds 폴리뉴클레오타이드의 목적 뉴클레오타이드 또는 염기 쌍 서열을 지칭한다.
- [0094] 용어 "표적 이본쇄(ds) 폴리뉴클레오타이드"는, 본원에 제공된 합성 방법에 의해 생성되는 소정 서열을 갖는 폴리뉴클레오타이드를 지칭한다. 구체적으로, 상기 표적 이본쇄 폴리뉴클레오타이드는 SOI와 동일하고/하거나 이에 상응하는 서열을 특징으로 한다. 표적 ds 폴리뉴클레오타이드 서열이 SOI와 100% 미만 동일한 서열을 갖는 경우, 표적 ds 폴리뉴클레오타이드는, SOI와 동일하고/하거나 이에 상응하는 서열을 갖는 ds 폴리뉴클레오타이드를 생성하도록 추가로 변형시킬 수 있는 프록시 ds 폴리뉴클레오타이드로서 이해된다.
- [0095] 용어 "프록시 이본쇄(ds) 폴리뉴클레오타이드"는, 이의 서열이 SOI의 뉴클레오타이드 서열과 100% 미만 동일한 및 적어도 90%, 바람직하게는 95% 동일한 표적 이본쇄(ds) 폴리뉴클레오타이드를 지칭한다. SOI와 동일한 및/또는 이에 상응하는 서열을 갖고 이의 서열이 명확한 어셈블리 또는 폭주 반응을 일으키기 쉽기 때문에 합성하기 곤란한 ds 폴리뉴클레오타이드를 생성하기 위해, 프록시 이본쇄(ds) 폴리뉴클레오타이드를 먼저 합성할 수 있다. 프록시 이본쇄 폴리뉴클레오타이드의 서열은 회문 서열, 폭주 반응 및 명확한 어셈블리를 회피하고/하거나 계층적 어셈블리를 용이하게 하도록 설계된다. 구체적으로, 서열은 계산적으로 설계될 수 있다. 이어서, 합성된 프록시 ds 폴리뉴클레오타이드는 SOI의 뉴클레오타이드 서열과 동일한 뉴클레오타이드 서열을 갖는 ds 폴리뉴클레오타이드를 생성하기 위해 추가로 변형시킬 수 있다. 구체적으로, 상기 프록시 ds 폴리뉴클레오타이드는 임의의 지시된 돌연변이유발, 엔도뉴클레아제 또는 엑소뉴클레아제 및/또는 효소적 변형에 의해, 임의의 메틸트랜스페라제, 키나제, CRISPR/Cas9, λ-레드 재조합을 사용한 다중 자동화 게놈 조작(MAGE), 공액 어셈블

리 게놈 조작(CAGE), 아르고나우트 단백질(Ago) 또는 이의 유도체, 징크-핑거 뉴클레아제(ZFN), 전사 활성화인자-유사 이펙터 뉴클레아제(TALEN), 메가뉴클레아제, 티로신/세린 부위-특이적 리콤비나제(Tyr/Ser SSR), 하이브리드화 분자, 설푸릴라제, 리콤비나제, 뉴클레아제, DNA 폴리머라제, RNA 폴리머라제 또는 TNases를 사용하여 추가로 변형시켜, SOI와 동일하고/하거나 이에 상응하는 서열을 갖는 ds 폴리뉴클레오타이드를 수득한다.

[0096] 용어 "주형"은 특정 서열을 특징으로 하는 폴리뉴클레오타이드 또는 폴리뉴클레오타이드 서열을 지칭하고, 이 서열을 사용하여 표적 ds 폴리뉴클레오타이드를 합성하고 생성할 수 있다. 주형이 본원에 제공된 합성 방법에 사용되는 경우, 이렇게 생성된 표적 ds 폴리뉴클레오타이드는 주형과 100% 동일한 서열을 갖는다.

[0097] 구체적으로, 상기 주형은 일본쇄 또는 이본쇄이다. 이러한 주형은 천연 뉴클레오타이드 서열 또는 목적하는 생성물을 포함하는 인공적, 계산에 의해 설계된 뉴클레오타이드 서열일 수 있다. 이러한 주형은 SOI와 동일할 수 있거나 SOI와 100% 미만 동일, 바람직하게는 95% 미만 동일, 그러나 적어도 80% 동일할 수 있다.

[0098] 바람직하게는, 상기 주형은 계산에 의해 생성되고, 표적 ds 폴리뉴클레오타이드의 리딩 쇄의 서열 및 표적 폴리뉴클레오타이드의 역 상보체를 각각 포함한다. 전형적으로, 2개 주형이 본원에 기재된 합성 방법에 사용되고, 표적 ds 폴리뉴클레오타이드의 각각의 쇄에 대해 하나의 주형이 사용된다. 주형 서열을 계산적으로 설계하는 경우, 어셈블리에 사용되는 실험 전략과의 호환성이 우선된다.

[0099] "ssDNA 올리고뉴클레오타이드" 또는 단순한 "ss 올리고뉴클레오타이드" 또는 "ss 올리고"로서도 지칭되는 용어 "일본쇄 DNA 올리고뉴클레오타이드"는 뉴클레오타이드 단량체의 선형 중합체인 올리고뉴클레오타이드를 지칭한다. 올리고뉴클레오타이드를 구성하는 단량체는, 왓슨-크릭형의 염기쌍 형성, 염기 스택킹, 호그스틴형 또는 역 호그스틴형의 염기쌍 형성, 워블 염기쌍 형성 등의 단량체-대-단량체 상호작용의 규칙적 패턴에 의해 천연 폴리뉴클레오타이드에 특이적으로 결합할 수 있다. 본원에 기재된 ssDNA 올리고뉴클레오타이드는 전형적으로 크기가 6 내지 26의 범위이지만, 이보다 더 길 수도 있다. 본원에 기재된 ssDNA 올리고뉴클레오타이드는 크기가 6 내지 220 뉴클레오타이드, 예를 들면, 27 내지 200 뉴클레오타이드의 범위일 수 있다. 올리고뉴클레오타이드가 "ATGC" 등의 문자의 서열(대문자 또는 소문자)로 나타내어지는 경우에는 항상, 뉴클레오타이드가 좌측으로부터 우측으로 5'→3'의 순서로 존재하고, "A"는 데옥시아데노신을 나타내고, "T"는 데옥시티미딘을 나타내고, "G"는 데옥시구아노신을 나타내고, "C"는 데옥시시티딘을 나타내는 것으로 이해된다. 통상의 뉴클레오타이드(A, G, C, T) 이외에, 변형된 뉴클레오타이드, 예를 들면, K-2'-데옥시리보스, P-2'-데옥시리보스, 2'-데옥시이노신, 2'-데옥시크산토신, 또는 뉴클레오염기 유사체를 갖는 뉴클레오타이드, 예를 들면, 이노신, 또는 5-메틸이소시토신, 또는 3-니트로피롤, 5-니트로인돌, 피롤리딘, 4-니트로이미다졸, 4-니트로피라졸, 4-니트로벤즈이미다졸, 4-아미노벤즈이미다졸, 5-니트로인다졸, 3-니트로이미다졸, 5-아미노인돌, 벤즈이미다졸, 5-플루오로인돌, 인돌, 메틸이소카보스티릴, 피롤로피리미딘 7-프로피닐이소카보스티릴이 사용될 수 있다. 용어 및 원자 번호 넘버링 규칙은 문헌[참조: Strachan and Read, Human Molecular Genetics 2 (Wiley-Liss, New York, 1999)]에 개시된 것들에 따른다. 통상, 올리고뉴클레오타이드는 포스포디에스테르 또는 펩티드 결합 또는 포스포로티오에이트 결합에 의해 결합된 4개 천연 뉴클레오사이드(예를 들면, DNA의 데옥시아데노신, 데옥시시티딘, 데옥시구아노신, 데옥시티미딘 또는 RNA의 이들의 리보스 대응물)을 포함하지만, 그러나 이들은 또한, 예를 들면, 변형된 염기, 당 또는 뉴클레오사이드간 결합을 포함하여 비-천연 뉴클레오타이드 유사체를 포함할 수 있다.

[0100] 일부 실시양태에서, 일본쇄 올리고뉴클레오타이드 풀은 화학 합성 방법을 사용하여, 예를 들면, 단량체-포스포르아미다이트, 이량체-포스포르아미다이트(Neuner, Cortese, and Monaci 1998) 또는 삼량체-포스포르아미다이트(Sondek and Shortle 1992), 단량체-포스포르아미다이트의 혼합물, 이량체-포스포르아미다이트의 혼합물, 삼량체-포스포르아미다이트의 혼합물 또는 이들의 조합으로부터 올리고뉴클레오타이드 서열을 합성함으로써 생성된다.

[0101] 일부 실시양태에서, 올리고뉴클레오타이드는 다양한 공지된 효소적 방법 중의 어느 하나를 사용하여, 예를 들면, 문헌[참조: Farzadfard et al. (2014)]에 기재된 바와 같이, 생체내 돌연변이유발을 받은 세포 내에서 천연-존재 공급원으로부터 생성 및 정제되거나 생체내에서 합성된다. 구체적으로, 소프트-랜덤화 올리고뉴클레오타이드 풀을 합성하는 효소는, 합성 동안 부정합 뉴클레오타이드를 높은 빈도로 도입하는 저신뢰성의 DNA 폴리머라제 단백질 또는 저신뢰성의 역전사효소 단백질을 포함하지만, 이들로 한정되지 않는다. 또는, 부정합 뉴클레오타이드는 당업자에게 공지된 화학 물질의 존재에 기인하여, DNA 폴리머라제 또는 역전사효소에 의해 보다 높은 빈도로 올리고에 도입된다.

[0102] 용어 "염기 쌍" 또는 "bp", 약어, 단수 또는 복수로 사용) 또는 "bps"(복수)는 DNA 또는 RNA 분자의 상보성 쇄

를 연결하고 수소 결합에 의해 피리미딘에 결합된 퓨린으로 이루어진 뉴클레오타이드의 임의의 쌍을 지칭한다. 쌍은 DNA의 아데닌 및 티민, RNA의 아데닌 및 우라실, 및 DNA 및 RNA 둘 다의 구아닌 및 시토신이다.

[0103] 용어 "**정합 올리고뉴클레오타이드의 쌍**"은 2개 이상의 상보성 올리고뉴클레오타이드를 지칭한다. "상보성"이란, 2개의 일본쇄 핵산 또는 하나 이상의 ds 핵산의 돌출부 부분의 유사한 영역의 뉴클레오타이드 서열이, 일본쇄 영역을 엄격한 어닐링 또는 증폭 조건하에 안정한 이본쇄 수소-결합된 영역에서 함께 어닐링하는 뉴클레오타이드 염기 조성을 갖는 것을 의미하고, 이러한 어닐링은 또한 "하이브리드화"로서 지칭된다. 하나의 일본쇄의 뉴클레오타이드의 연속 서열이 다른 일본쇄 영역의 뉴클레오타이드의 유사한 서열과의 일련의 "표준적" 수소-결합된 염기 쌍을 형성할 수 있는 경우, A는 U 또는 T와 쌍형성하고, C는 G와 쌍형성하고, 뉴클레오타이드 서열은 100% 상보성이다. 통상의 염기(A, G, C, T) 이외에, 유사체, 예를 들면, 이노신 및 2'-데옥시이노신 및 이들의 유도체(예를 들면, 7'-데아자-2'-데옥시이노신, 2'-데아자-2'-데옥시이노신), 아졸-(예를 들면, 벤즈이미다졸, 인돌, 5-플루오로인돌) 또는 니트로아졸 유사체(예를 들면, 3-니트로피롤, 5-니트로인돌, 5-니트로이미다졸, 4-니트로피라졸, 4-니트로벤즈이미다졸) 및 이들의 유도체, 비사이클릭 당 유사체(예를 들면, 하이포크산틴- 또는 인다졸 유도체, 3-니트로이미다졸, 또는 이미다졸-4,5-디카복사아미드로부터 유래된 것들), 유니버설 염기 유사체의 5'-트리포스페이트(예를 들면, 인돌 유도체로부터 유래됨), 이소카보스티릴 및 기타 수소성 유사체, 및 이의 임의의 유도체(예를 들면, 메틸이소카보스티릴, 7-프로피닐이소카보스티릴), 수소 결합 유니버설 염기 유사체(예를 들면, 피롤로피리미딘), 및 기타 화학적으로 변형된 염기(예컨대, 디아미노퓨린, 5-메틸시토신, 이소구아닌, 5-메틸-이소시토신, K-2'-데옥시리보스, P-2'-데옥시리보스)는 상이한 염기-쌍 형성 선호도를 가질 수 있고, 유사한 엄격성/확률로 하나 이상의 천연 뉴클레오타이드와 쌍형성할 수 있다. 특정의 경우에, 단량체는 포스포디에스테르 또는 펩티딜 결합 또는 포스포로티오에이트 결합에 의해 결합된다.

[0104] "dsDNA 올리고뉴클레오타이드" 또는 간단히 "ds 올리고뉴클레오타이드" 또는 "ds 올리고"로도 지칭되는 용어 "**이본쇄 DNA 올리고뉴클레오타이드**"는 뉴클레오타이드 이량체의 선형 중합체인 올리고뉴클레오타이드를 지칭한다. 올리고뉴클레오타이드를 구성하는 이량체는 왓슨-크릭형의 염기쌍 형성, 염기 스택킹, 호그스틴형 또는 역 호그스틴형의 염기쌍 형성, 워블 염기쌍 형성 등의 단량체-대-단량체 상호작용의 규칙적 패턴에 의해 결합된 2개 상보성 뉴클레오타이드이다. 본원에 기재된 dsDNA 올리고뉴클레오타이드는 전형적으로 크기가 6 내지 26 염기 쌍(bp) 범위이지만, 보다 길 수도 있다. 본원에 기재된 dsDNA 올리고뉴클레오타이드는 크기가 6 내지 200 염기 쌍, 예를 들면, 27 내지 200 염기 쌍의 범위일 수 있다. 올리고뉴클레오타이드가 문자의 서열(대문자 또는 소문자), 예컨대, "ATGC"에 의해 나타내어지는 경우에는 항상, 뉴클레오타이드는 좌측으로부터 5'→3'의 순서이고, "A"는 데옥시아데노신을 나타내고, "T"는 데옥시티미딘을 나타내고, "G"는 데옥시구아노신을 나타내고, "C"는 데옥시시티딘을 나타내는 것으로 이해된다. 통상의 뉴클레오타이드(A, G, C, T) 이외에, 변형된 뉴클레오타이드, 예를 들면, K-2'-데옥시리보스, P-2'-데옥시리보스, 2'-데옥시이노신, 2'-데옥시크산토신 또는 뉴클레오타이드 유사체를 갖는 뉴클레오타이드, 예를 들면, 이노신, 또는 5-메틸이소시토신, 또는 3-니트로피롤, 5-니트로인돌, 피롤리딘, 4-니트로이미다졸, 4-니트로피라졸, 4-니트로벤즈이미다졸, 4-아미노벤즈이미다졸, 5-니트로인다졸, 3-니트로이미다졸, 5-아미노인돌, 벤즈이미다졸, 5-플루오로인돌, 인돌, 메틸이소카보스티릴, 피롤로피리미딘 7-프로피닐이소카보스티릴이 사용될 수 있다. 용어 및 원자 넘버링 규칙은 문헌[참조: Strachan and Read, Human Molecular Genetics 2 (Wiley-Liss, New York, 1999)]에 기재된 것들에 따른다. 통상적으로, 올리고뉴클레오타이드는 포스포디에스테르 또는 펩티딜 결합 또는 포스포로티오에이트 결합에 의해 결합된 4개 천연 뉴클레오타이드(예를 들면, DNA의 데옥시아데노신, 데옥시시티딘, 데옥시구아노신, 데옥시티미딘 또는 RNA의 이들의 리보스 대응물)를 포함하지만, 그러나 이들은 또한, 예를 들면, 변형된 염기, 당, 또는 뉴클레오타이드간 결합을 포함하여, 비-천연 뉴클레오타이드 유사체를 포함할 수 있다.

[0105] 이본쇄 분자의 가장 단순한 DNA 말단은 평활 말단으로 불리운다. 평활-말단 분자에서, 양 쇠는 염기 쌍에서 종결된다. 비-평활 말단은 다양한 돌출부에 의해 생성된다. 본원에 사용된 용어 "**돌출부**"는 ds 올리고 또는 폴리뉴클레오타이드 분자의 한 말단 또는 양 말단에서 쌍을 형성하지 않은 뉴클레오타이드의 스트렛치를 지칭한다. 이들 쌍을 형성하지 않은 뉴클레오타이드는 어느 하나의 쇠에 존재하여 어느 하나의 3' 또는 5' 돌출부를 생성한다. 돌출부의 가장 단순한 경우는 단일 뉴클레오타이드이다. 돌출부는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 또는 12 뉴클레오타이드 중의 어느 하나 또는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 또는 12 뉴클레오타이드 중의 적어도 어느 하나를 포함하거나 이들로 이루어질 수 있다. 돌출부는 전형적으로 ds 올리고 길이의 절반 이하이다. 예를 들면, 상기 ds 올리고가 6 뉴클레오타이드 길이인 경우, 돌출부는 3 이하의 뉴클레오타이드 길이이고, 이는 돌출부가 또한 1 또는 2 뉴클레오타이드 길이일 수 있음을 의미한다. 또 다른 예에 따르면, 상기 ds 올리고가 24 뉴클레오타이드 길이인 경우, 돌출부는 12 이하의 뉴클레오타이드 길이이고, 이는 또한 1,

2, 3, 4, 5, 6, 7, 8, 9, 10 또는 11 뉴클레오타이드 길이일 수 있음을 의미한다.

- [0106] 본원에 사용된 용어 "**라이브러리**"는 핵산 분자(예를 들면, 올리고뉴클레오타이드 라이브러리)이고 정합 올리고뉴클레오타이드의 적어도 10,000 쌍을 포함하는 라이브러리 멤버의 집합을 지칭한다. 라이브러리 멤버는 일본쇄 올리고뉴클레오타이드 또는 이본쇄 올리고뉴클레오타이드일 수 있다. 라이브러리 멤버는 공통 기능(예컨대, 게놈 서열에 의해 부여됨)을 공유하지만, 적어도 하나의 염기 쌍, 뉴클레오타이드, 돌연변이 및/또는 표현형이 상이하다. 라이브러리는 전형적으로, 공통 기능을 갖는 것들 이외에, 다양한 라이브러리 멤버를 함유한다. 한 가지 특정 유형의 라이브러리는 랜덤 돌연변이유발에 의해 생성된 올리고뉴클레오타이드의 랜덤화 돌연변이체의 라이브러리이다. 또 다른 특정 예는 합리적으로 설계된(또는 합성) 라이브러리, 예를 들면, 특별히 조작된 DNA 단편 또는 올리고뉴클레오타이드를 포함하는 라이브러리일 것이다. 본원에 기재된 라이브러리는 다양한 길이 및 상이한 서열의 올리고뉴클레오타이드로 적합하게 구성된 라이브러리 멤버를 포함하고, 올리고뉴클레오타이드는 DNA의 특정 영역에 상응할 수 있거나, 전체 유전자 공간으로 신장할 수도 있다. 예시적으로, 라이브러리는 인간 염색체 게놈 또는 미토콘드리아 게놈의 임의의 및 모든 천연 존재 폴리뉴클레오타이드를 가능한 합성하는데 필요한 다양한 올리고를 포함할 수 있다. 특정 예에서, 상기 다양성은 인간 이외의 다른 진핵생물 중, 예를 들면, 마우스, 랫트, 래빗, 돼지, 양, 식물, 곰팡이(funghi) 또는 효모의 임의의 및 모든 천연 존재 폴리뉴클레오타이드를 커버할 수 있다. 또 다른 예에서, 상기 다양성은 원핵생물, 예컨대, 아카에안 또는 세균의 임의의 및 모든 천연 존재 폴리뉴클레오타이드를 커버할 수 있다.
- [0107] 본원에 제공된 라이브러리는 특히, 일본쇄 올리고뉴클레오타이드인 정합 올리고뉴클레오타이드의 적어도 10,000 쌍을 포함하고, 구체적으로 이들은 부분 또는 완전 상보성 서열을 포함하는 상이한 길이의 ss 올리고이다. 정합 올리고의 상기 쌍은 별개의 컨테이너먼트 내의 ss 올리고로서 라이브러리에 존재할 수 있거나, 2개 이상의 상보성 ss 올리고는 이들이 어닐링하고 ds 올리고를 형성할 수 있는 하나의 컨테이너먼트에 함유될 수 있다. 정합 ss 올리고 쌍의 뉴클레오타이드 서열은 적어도 1, 2 또는 3 뉴클레오타이드, 바람직하게는 적어도 4 이상의 뉴클레오타이드에서 상보성일 수 있고, 정합 쌍은 ss 올리고 서열의 하이브리드화에 의해 새로운 ds 폴리뉴클레오타이드 분자를 형성할 수 있고, 바람직하게는 ss 올리고는 부분적으로 하이브리드화하여 돌출부를 갖는 ds 폴리뉴클레오타이드를 수득한다.
- [0108] 라이브러리는 바람직하게는, 인공적 또는 화학적으로 합성된 올리고뉴클레오타이드, 또는 당해 기술분야에 공지된 적합한 방법에 의해 합성된 화학적으로 변형된(예를 들면, 펩티딜 핵산 또는 포스포포티오에이트 결합을 포함) 올리고뉴클레오타이드를 포함한다. 라이브러리에 포함된 올리고뉴클레오타이드는 또한 천연 존재 DNA의 효소적 소화에 의해 생성될 수 있다. 본원에 기재된 상기 올리고뉴클레오타이드 라이브러리의 멤버는 특히 상이한 서열, 돌연변이 또는 뉴클레오타이드 염기 또는 뉴클레오타이드 변화, 예를 들면, 하나 이상의 후속 뉴클레오타이드의 치환, 또는 삽입 또는 결실을 특징으로 한다. 전형적으로, 라이브러리 멤버는 적어도 하나 이상의 점 돌연변이가 상이하다. 구체적으로, 일부 실시양태에서, 변이는 특정 위치에서 모든 가능한 천연-존재 뉴클레오타이드 염기쌍을 커버한다. 돌연변이가 친계 올리고뉴클레오타이드의 돌연변이유발에 의해 생성되는 경우, 친계 올리고뉴클레오타이드의 다양한 서열 변이체가 생성된다.
- [0109] 본원에 기재된 라이브러리의 다양성은 포스포릴화, 메틸화, 비오틴화되거나 형광단 또는 쉐퍼에 연결되는 라이브러리 멤버를 추가로 포함할 수 있다. 본원에 기재된 바와 같이, 라이브러리 멤버는 하나 이상의 추가의 포스포릴 그룹을 포함할 수 있다.
- [0110] 라이브러리 멤버의 메틸화, DNA 분자, 바람직하게는 시스테인 또는 아데닌에 대한 메틸 그룹의 부가는 당해 기술분야에 공지된 적합한 DNA 메틸화 방법에 따라 수행된다.
- [0111] 본원에 사용된 바와 같이, 비오틴화는 핵산, 예컨대, ss 올리고 또는 ds 올리고에 하나 이상의 비오틴 분자를 공유 결합시키는 방법을 지칭한다. 본원에 기재된 라이브러리 멤버는 당해 기술분야에 공지된 적합한 방법에 의해 비오틴화될 수 있고; 바람직하게는 이는 화학적 비오틴화의 방법이다. 올리고뉴클레오타이드는, 비오틴 포스포르아미다이트를 사용하는, 당해 기술분야에 공지된 포스포르아미다이트 방법에 의해 올리고뉴클레오타이드 합성의 과정에서 용이하게 비오틴화될 수 있다.
- [0112] 본원에 기재된 라이브러리의 멤버는, 당해 기술분야에 공지된 적합한 화학적 및 효소적 방법에 의해 형광단에 접합시킬 수 있다. 핵산의 형광 표지화에 사용된 예시적 방법은, 예를 들면, 썬모 피셔 ARES DNA 표지화 키트를 사용하는, 형광 염료로 DNA를 효소적 표지하는 방법을 사용할 수 있고, 이는 형광 염료에 의한 DNA의 효소적 표지화를 위한 2단계 방법을 사용한다. 추가의 예시적 방법은, 예를 들면, ULYSIS 핵산 표지화 키트를 사용하여, 표지된 뉴클레오타이드의 효소적 도입 없이 핵산을 표지하는 화학적 방법을 사용할 수 있다. 추가의 예시

적 방법은, 예를 들면, 알렉사 플루오르 올리고뉴클레오타이드 아민 표지화 키트를 사용하여, 단일 표지된 형광 올리고뉴클레오타이드를 제조하기 위해 아민-중결된 올리고뉴클레오타이드의 화학적 표지화를 이용할 수 있다. 추가의 예시적 방법은 DNA 어레이/마이크로어레이 및 기타 하이브리드화 기술을 사용할 수 있다.

[0113] 라이브러리 멤버는, 당해 기술분야에 공지된 적합한 방법에 의해, 하나 이상의 켄처, 예를 들면, 형광단으로부터 여기 에너지를 흡수하는 물질에 연결시킬 수 있다. 켄처의 예는 다브실 (디메틸아미노아조벤젠설포산), 블랙 홀 켄처, Qx1 켄처, Iowa 블랙 FQ, Iowa 블랙 RQ 및 IRDye QC-1을 포함하지만, 이들로 한정되지 않는다.

[0114] 본원에 사용된 용어 "점 돌연변이" 또는 뉴클레옴기 변화는, 예컨대, 단일 뉴클레옴기 또는 아미노산을 도입 또는 교환하거나 갭을 도입함으로써 특정 위치에서 핵산 또는 아미노산 서열을 변화시키는 돌연변이 사상을 지칭한다. 점 돌연변이 또는 뉴클레옴기 변화는 서열 중의 하나 이상의 단일 또는 인접 또는 연속 뉴클레옴기 또는 아미노산 잔기의 변화를 포함할 수 있다. 한정된 다양성을 커버하는 돌연변이체의 레퍼토리를 포함하는 라이브러리에서, 점 돌연변이의 빈도는, 돌연변이체가 친계(또는 참조) 서열과 적어도 특정 서열 동일성을 공유하도록 한정되고, 이는, 예를 들면, 80%, 90%, 95%, 96%, 97%, 98%, 또는 99% 중의 적어도 어느 하나이다.

[0115] 본원에 기재된 뉴클레오타이드 서열과 관련하여 "퍼센트(%) 뉴클레오타이드 서열 동일성"은, 최대 퍼센트 서열 동일성을 달성하기 위해, 필요에 따라, 서열을 정렬시키고 갭을 도입시킨 후, 특정 뉴클레오타이드 서열의 뉴클레오타이드와 동일한 후보 서열의 뉴클레오타이드의 퍼센트로서 정의되고, 서열 동일성의 일부로서 임의의 보존적 치환을 고려하지 않는다. 당해 기술분야의 숙련가는, 비교되는 서열의 전체 길이에 걸쳐 최대 정렬을 달성하는데 필요한 임의의 알고리즘을 포함하는, 정렬을 측정하기 위한 적절한 파라미터를 결정할 수 있다.

[0116] 본원에 사용된 용어 "다양성"은, 본원에 제공된 라이브러리를 특성화하는 정도의 다양성을 지칭한다. 구체적으로, 상기 다양성은 상이한 길이 및 상이한 서열의 일본쇄 및 이본쇄 올리고를 포함한다. 예를 들면, 라이브러리는 8 뉴클레옴기 길이의 ss 올리고(본원에서 옥타머로서 지칭됨)의 모든 가능한 서열 변이를 포함할 수 있고, 이는 8 뉴클레옴기 길이의 65,536 상이한 ss 올리고 및 또한 상이한 길이의 다른 ss 올리고 또는 ds 올리고이며, 이는 통상 표적 서열에 포함되고 따라서 임의의 소정 서열을 구축하기 위해 보다 빈번히 요구된다. 통상 사용되는 단일 또는 이본쇄 올리고를 라이브러리 다양성에 포함시키는 것은 합성 비용을 감소시키고 시간 효율을 증가시킨다.

[0117] 구체적으로, 상기 다양성은, 예를 들면, 인간 게놈 등의 전체 게놈을 커버할 수 있다. 구체적으로, 상기 다양성은 게놈 또는 전체 유전자 공간을 복수의 방식으로 복수회 커버할 수 있다. 예를 들면, 모든 가능한 핵사머, 헵타머 및/또는 옥타머 서열 조합을 포함시키는 것에 의해. 예를 들면, 상기 라이브러리는 또한 모든 또는 선택된 9-머 및 10-머 또는 임의의 최대 26-머를 포함할 수 있다.

[0118] 특정의 예에 따르면, 본원에 기재된 올리고뉴클레오타이드의 풀 내의 다양성은 하기와 같이 특성화된다: 다양성은 올리고뉴클레오타이드 서열 내의 돌연변이의 수에 의해 결정될 수 있다. 예를 들면, 16 뉴클레오타이드의 길이를 갖는 단일 올리고뉴클레오타이드에서, 가능한 단일 뉴클레오타이드 변화의 이론적 수는 4개 천연 발생 DNA, A, T, G 또는 C 뉴클레오타이드에서 $16 \times 3 = 48$ 이다. 올리고뉴클레오타이드(이중 돌연변이체)당 4개 천연 존재 DNA, A, T, G 또는 C 뉴클레오타이드에서 2개의 단일 뉴클레오타이드 변화의 경우, 가능한 서열의 수는 6,408이다. 올리고뉴클레오타이드(삼중 돌연변이체)당 3개 단일 뉴클레오타이드 변화의 경우, 이 수는 563,904이다. 사중 돌연변이의 경우, 이 수는 36,794,736이다. 이들 수는 올리고뉴클레오타이드 서열 내에 비-천연 뉴클레옴기를 도입함으로써 추가로 증가시킬 수 있다.

[0119] 라이브러리 내의 올리고뉴클레오타이드의 서열분석-기반 스크리닝을 위한 예시적 방법은 다음과 같다: 하이브리드화-기반 방법(예를 들면, 분자 비콘, SNP 마이크로어레이, 제한 단편 길이 다형성)을 포함하는 SNP 유전자형 방법, 대립유전자-특이적 PCR, 프라이머 신장-, 5'-뉴클레아제 또는 올리고뉴클레오타이드 결찰 검정, 일본쇄 형태 다형성, 온도 구배 겔 전기영동, 변성 고성능 액체 크로마토그래피, 전체 앰플리콘의 고해상도 용해(HRM), SNPlex 및 측량 뉴클레아제 검정을 포함하는 PCR-기반 방법; PTR의 전체 PCR 앰플리콘의 모세관 서열분석 또는 고효율 서열분석(앰플리콘 서열분석)을 포함하는 서열분석 기반 돌연변이 분석. 이러한 고효율(HT) 앰플리콘 서열분석 방법은 폴리니 서열분석, 피로서열분석, 일루미나(Solexa) 서열분석, SOLiD 서열분석, 반도체 서열분석, DNA 나노볼 서열분석, 헬리스코프 단일 분자 서열분석, 단일 분자 실시간(SMRT) 서열분석, 나노포어 DNA 서열분석, 터널링 전류 DNA 서열분석, 하이브리드화에 의한 서열분석, 질량 분석에 의한 서열분석, 마이크로유체 생성 서열분석, 현미경-기반 서열분석, RNAP 서열분석을 포함하지만, 이들로 한정되지 않는다.

- [0120] 각 라이브러리 멤버는, 라이브러리 내의 라이브러리 멤버의 선택 또는 라이브러리 내의 라이브러리 멤버의 동정을 용이하게 하기 위해, 선택가능한 마커 또는 DNA 서열 태그 또는 바코드에 의해 개별적으로 특징지어지고 마킹될 수 있다. 또는, 유전자 돌연변이는, 적합한 결정 방법, 예를 들면, 고효율 서열분석, 모세관 서열분석, 또는 소정의 서열과 하이브리드화하는 특정 프로브의 사용에 의해 직접 결정하여, 상응하는 올리고뉴클레오타이드를 선택할 수 있다.
- [0121] 용기 내의 올리고뉴클레오타이드의 라이브러리를 수득하기 위해, 라이브러리 멤버를 개별 용기 내에 배치하는 것이 바람직할 수 있다. 특정 실시양태에 따르면, 라이브러리는 어레이, 예를 들면, DNA 바이오칩에 제공되고, 여기서 어레이는 고체 담체 상에 일련의 스폿을 포함한다.
- [0122] 본원에 사용된 용어 "**돌연변이유발**"은, 올리고뉴클레오타이드 또는 폴리뉴클레오타이드의 서열을 변화시키는 프로세스를 지칭한다. 구체적으로, 부위-지시된 돌연변이유발은 공지된 뉴클레오타이드 서열에 특정 돌연변이를 생성하는 방법을 지칭한다. 이 돌연변이는 특이적 표적화 변화이고, 단일 또는 복수 뉴클레오타이드 삽입, 결실 또는 치환을 포함할 수 있다. 이 작업은, 제한 효소, 구체적으로는 **엔도뉴클레아제** 및/또는 **엑소뉴클레아제**에 의해 수행할 수 있다. 엔도뉴클레아제는 올리고뉴클레오타이드 또는 폴리뉴클레오타이드의 중앙에서 포스포디에스테르 결합을 절단하는 반면, 엑소뉴클레아제는 올리고뉴클레오타이드 또는 폴리뉴클레오타이드의 5' 또는 3' 말단에서 포스포디에스테르 결합을 절단한다.
- [0123] 본원에 사용된 용어 "**알고리즘**"은, 실행되는 작동의 자가-완결형 서열을 지칭한다. 알고리즘은, 한정된 양의 공간과 시간 내에서, 관수를 계산하기 위한 명확하게 정의된 형식 언어로 표현될 수 있는 효과적인 방법이다. 초기 상태 및 초기 입력으로부터 출발하여, 명령은 실행시에 명확하게 정의된 연속 상태의 유한 수를 통해 진행하고, 최종적으로 "출력"을 생성하고, 최종적 완료 상태에서 종료하는 계산을 기재한다. 한 상태에서 다음 상태로의 전이는 필연적으로 확정적이다.
- [0124] 용어 "**워크플로우**" 또는 "**어셈블리 워크플로우**"는, 최적 수의 올리고 서브세트 및 이들의 표적 ds 폴리뉴클레오타이드로의 어셈블리 서열을 지칭한다. 본원에 제공된 방법에서, 주형의 서열은 올리고의 서브세트에 상응하는 서브-서열로 분할되어, 특정 뉴클레오타이드 합성 문제, 예컨대, 팔린드롬 서열, 폭주 반응 및 명확한 어셈블리를 회피할 수 있다. 특히, 보다 짧은 올리고뉴클레오타이드로의 이러한 분할은 어셈블리 프로세스를 단축시키고 바람직하지 않은 반응 생성물을 분리할 필요를 회피하기 위해 매우 효율적일 수 있다. 구체적으로, 올리고의 서브세트의 절찰은, 중간체로도 불리우는 중간 반응 생성물을 생성하고, 중간 반응 생성물의 어셈블리는 최종적으로 표적 ds 폴리뉴클레오타이드를 생성한다. 바람직하게는, 올리고의 서브세트를 선택하기 위해, 상기 수록된 것들에 대한 추가의 기준을 사용할 수 있다. 이러한 추가 기준은, 임의의 단일 절찰 반응에 사용되는 올리고의 서브세트의 크기를 최소화하고(예를 들면, 부정합 절찰을 회피하기 위해), 올리고뉴클레오타이드 전구체의 서브세트의 멤버의 어닐링 온도의 차이를 최소화하고, 상이한 이분쇄 서브유닛의 돌출부의 어닐링 온도의 차이를 최소화하고, 프레임-쉬프트 어댑터 또는 일본쇄 올리고 링커를 사용할지의 여부, 및 서브세트를 구성하는 상이한 올리고의 하이브리드 형성 부분 사이에서 교차-하이브리드화의 정도를 최소화하는지의 여부를 포함하지만, 이들로 한정되지 않는다.
- [0125] 서브세트 내의 올리고의 수는 상이할 수 있다. 바람직하게는, 서브세트의 크기는 1 내지 100, 또는 2 내지 100의 범위, 보다 바람직하게는 1 내지 50, 또는 2 내지 50의 범위 및 여전히 보다 바람직하게는 1 내지 10 또는 2 내지 10의 범위이다.
- [0126] 교차-하이브리드화의 정도가 최소화되어 있는 서브세트에서, 세트의 서브유닛 및 세트의 임의의 다른 서브유닛의 상보체로 이루어진 이중쇄 및 삼중쇄는 적어도 하나의 부정합을 함유한다. 달리 말하면, 이러한 서브세트의 올리고의 서열은 서브세트의 모든 다른 올리고의 서열과 적어도 하나의 뉴클레오타이드, 및 보다 바람직하게는 적어도 2개의 올리고뉴클레오타이드가 상이하다. 특정 실시양태에서 이용가능한 올리고뉴클레오타이드 태그의 수는 태그당 서브유닛의 수 및 서브유닛의 길이에 의존한다.
- [0127] 조합된 돌출부에 상보성인 서열을 갖는 일본쇄 올리고 링커는 표적 폴리뉴클레오타이드의 인접 올리고를 연결한다. 링커는, 예를 들면, 인접하는 2개 올리고뉴클레오타이드를 각각 3' 말단 및 5' 말단 상에 각각 3 염기 길이의 돌출부로 연결하는 6 염기를 포함할 수 있다.
- [0128] 본 발명의 한 가지 특정 실시양태에서, 어셈블리 워크플로우를 결정하는 프로세스는 알고리즘에 의해 수행된다. 주형의 서열의 후보 분할을 계통적으로 조사하여, 본원에 제공된 방법에 따라 합성을 위해 분할하는 서브세트의 최적 수 및 어셈블리 서열을 발견한다. 먼저, 전체 주형 서열을 단일 서브세트로서 취하고, 이어서 상기 수록

된 서브세트 기준을 만족하는 분할이 발견될 때까지, 크기 감소시에 후보 올리고의 수를 증가시키면서, 보다 작아지는 서브세트가 형성된다.

- [0129] 용어 "**어셈블리**" 또는 "**어셈블**"은 일본쇄 및/또는 이본쇄 올리고를 연결 및/또는 하이브리드화시킴으로써 올리고뉴클레오타이드 또는 폴리뉴클레오타이드의 형성을 지칭한다. 구체적으로, 상기 어셈블리는 ss 뉴클레오타이드 서열을 하이브리드화하는 임의의 방법 및/또는 효소 반응 및/또는 화학 반응인 결합 반응에 의한다. 바람직하게는, 상기 어셈블리는 시험관내 결합 방법에 의한다.
- [0130] 표적 ds 폴리뉴클레오타이드의 어셈블리는 정합 ss 올리고, ds 올리고의 돌출부를 하이브리드화함으로써 직접 수행할 수 있거나, 하나 이상의 적합한 ss 올리고 링커를 하이브리드화함으로써 간접적으로 수행할 수 있고, 여기서 ss 올리고 링커는 라이브러리에 함유되고, 상기 제1, 제2 또는 추가의 반응 생성물 중의 어느 하나를 어셈블리하기 위해 라이브러리로부터 선택 및 전달된다.
- [0131] 직접 어셈블리의 경우, 올리고뉴클레오타이드 서열은 이들의 일본쇄 올리고 부분 또는 중첩부(즉, 중첩하는 부분 또는 돌출부)에 의해 함께 결합되고, 중첩부는 연속 서열에 단지 1회만 포함된다. 2개 올리고뉴클레오타이드 서열을 중첩하여 정렬시키면, 함께 취한 양방 개개 올리고뉴클레오타이드의 길이로부터 중첩 길이를 뺀 길이를 갖는 연속 서열이 형성된다. 결과적으로, 각각의 정렬된 올리고뉴클레오타이드의 세그먼트를 포함하는 연속 서열이 수득된다.
- [0132] 간접 어셈블리의 경우, 표적 ds 폴리뉴클레오타이드 또는 상기 제1, 제2 또는 추가 반응 생성물 중의 어느 하나는 ss 올리고를 정렬시켜 이들을 일본쇄 링커를 통해 결합시키면 형성된다. 예를 들면, 2개 올리고뉴클레오타이드, 예를 들면, 각각 10 염기 길이는, 예를 들면, 6 염기 길이의 ss 올리고 링커에 의해 결합되고, 제1 올리고뉴클레오타이드의 3' 말단의 3 염기는 ss 링커의 5' 말단의 3 염기로 정렬하고, 제2 올리고뉴클레오타이드의 5' 말단의 3 염기는 ss 링커의 3' 말단의 3 염기와 정렬한다.
- [0133] 용어 "**제1, 제2 또는 추가 반응 생성물**"은, 하나 이상의 반응 컨테이너먼트에서 수행된 결합 반응의 생성물을 지칭한다. 제1 단계에서, 적어도 정합 올리고뉴클레오타이드의 제1 쌍을, 액체 핸들러를 사용하여 라이브러리로부터 제1 반응 컨테이너먼트로 전달하고, 정합 올리고뉴클레오타이드를 결합 반응으로 어셈블리하여 제1 반응 생성물을 형성한다. 구체적으로, 상기 제1, 제2 및 추가 반응 생성물은 각각 적어도 하나의 돌출부를 포함한다. 반응 생성물의 이러한 돌출부는, 예를 들면, 정합 올리고뉴클레오타이드가 상기 반응 생성물의 돌출부와 하이브리드화하는 제1 부분을 포함하고 새로운 반응 생성물의 또 다른 돌출부를 생성하는 제2 부분을 추가로 포함하는 경우, 돌출부를 갖는 새로운 반응 생성물을 생성하기 위해, 돌출부의 방향에서 또 다른 정합 올리고뉴클레오타이드의 추가 어셈블리를 가능하게 한다. 또는, 예컨대, 돌출부의 모든 뉴클레오타이드를 커버하기 위해, 정합 올리고뉴클레오타이드가 돌출부의 전체 길이에 걸쳐 돌출부와 하이브리드화하는 부분으로만 이루어진 경우, 평활 말단을 생성할 수 있다.
- [0134] 특정의 경우에, 한 말단 또는 양 말단에 평활 말단을 갖는 ds 표적 이본쇄(ds) 폴리뉴클레오타이드가 생성된다. 이러한 평활 말단은 바람직하게는, 임의의 말단 돌출부를, 새로운 돌출부를 생성하지 않으면서, 이러한 돌출부의 전장과 하이브리드화하여 평활 말단을 생성하는, 정합 ss 올리고 및/또는 ds 올리고와 하이브리드화함으로써 생성된다.
- [0135] 상기 제1 단계에서, 액체 핸들러를 사용하고 정합 올리고뉴클레오타이드를 어셈블리하여 제1 반응 생성물을 수득함으로써 정합 올리고뉴클레오타이드의 하나 또는 복수 쌍 및 하나 또는 복수의 ss 올리고 링커를 상기 제1 반응 컨테이너먼트로 전달한다. 바람직하게는, 상기 제1 반응 컨테이너먼트로 전달된 정합 쌍의 수는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 또는 25 중의 어느 하나, 바람직하게는 4 및 보다 더 바람직하게는 1, 2, 또는 3이고, 전달된 ss 올리고 링커의 수는 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 또는 25 중의 어느 하나, 바람직하게는 4 및 보다 더 바람직하게는 1, 2 또는 3이다.
- [0136] 제2 및 추가 단계에서, 액체 핸들러를 사용하고 정합 올리고뉴클레오타이드를 어셈블리하여 각각 제2 및 추가 반응 생성물을 수득함으로써 정합 올리고뉴클레오타이드의 적어도 제2 및 추가 쌍을 라이브러리로부터 각각 제2 및 추가 반응 컨테이너먼트로 전달한다. 상기 제2 단계에서, 정합 올리고뉴클레오타이드의 하나 또는 복수의 쌍 및 하나 또는 복수의 올리고 링커는 상기 제2 반응 컨테이너먼트로 전달된다. 바람직하게는, 상기 제2 단계에서 전달된 정합 쌍의 수는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 또는 25 중의 어느 하나, 바람직하게는 4 및 보다 더 바람직하게는 1, 2 또는 3이고, 전달된 ss 올리고 링커의 수는 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 또는 25 중의 어느 하나, 바람직하게는 4 및 보다 더 바람직하게는 1, 2 또는 3이다. 상기 추가 단계에

서, 정합 올리고뉴클레오타이드의 하나 또는 복수의 쌍 및 하나 또는 복수의 ss 올리고 링커를 상기 추가 반응 컨테이너먼트로 전달한다. 바람직하게는, 상기 추가 단계에서 전달된 정합 쌍의 수는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 또는 25 중의 어느 하나, 바람직하게는 4 및 보다 더 바람직하게는 1, 2 또는 3이고, 전달된 ss 올리고 링커의 수는 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 또는 25 중의 어느 하나, 바람직하게는 4 및 보다 더 바람직하게는 1, 2 또는 3이다.

[0137] 단계의 수 및 상응하는 반응 생성물은 무제한이다. 거대 표적 ds 폴리뉴클레오타이드를 합성하기 위해, 표적 폴리뉴클레오타이드로 어셈블리하기 위해, 일련의 반응 생성물을 생성할 필요가 있다. 예를 들면, 적어도 5, 10, 20, 50, 100, 500, 1,000, 5,000 이상이 필요할 수 있다.

[0138] 본원에 사용된 용어 "하이브리드화하다", "하이브리드화", "하이브리드화하는", "어닐링하다" 및 "어닐링"은 일반적으로, 하나 이상의 폴리뉴클레오타이드가 반응하여, 뉴클레오타이드 잔기의 염기 사이의 수소 결합을 통해 안정화되는 복합체를 형성하는 반응을 지칭한다. 수소 결합은 왓슨 크릭 염기 쌍형성, 호그스테인 결합, 또는 임의의 기타 서열 특이적 방식으로 발생할 수 있다. 복합체는, 이중쇄 구조를 형성하는 2개의 쇠, 다중쇄 복합체를 형성하는 3개 이상의 쇠, 단일 자가-하이브리드화 쇠, 또는 이들의 임의의 조합을 포함할 수 있다. 하이브리드화 반응은 보다 광범위한 프로세스, 예컨대, PCR의 개시, 또는 리소자임에 의한 폴리뉴클레오타이드의 효소적 절단에서 단계를 구성할 수 있다.

[0139] 본원에 사용된 바와 같이, "결찰"은, 적절한 조건하에서 이본쇄를 형성하기 위해, 2개 핵산 서열이 분자간 화학적 결합(예를 들면, 수소 결합)으로 서로 어닐링하는 프로세스를 의미하는 것이다.

[0140] 본원에서 반응 생성물로서도 지칭되는 결찰 생성물은 이본쇄 핵산 및 일본쇄 핵산 둘 다로부터 형성될 수 있다. 이본쇄 핵산은 "점착 말단" 결찰 또는 "평활 말단" 결찰에 의해 결찰될 수 있다. 점착 말단 결찰에서, 말단 돌출부를 포함하는 엇갈림(staggered) 말단은 결찰 파트너에 하이브리드화할 수 있다. 평활 말단 결찰에서, 말단 돌출부는 존재하지 않고, 성공적 결찰은 5' 말단 및 3' 말단의 일시적 연관에 의존한다. 평활 말단 결찰은 일반적으로 점착 말단 결찰보다 덜 효율적이고, 다양한 최적화, 예컨대, 농도, 인큐베이션 시간 및 온도의 조절을 적용하여 효율을 개선시킬 수 있다. 일본쇄 폴리뉴클레오타이드도 또한 결찰시킬 수 있다.

[0141] 2개의 상보성 서열 또는 충분히 상보성 서열 사이의 결찰 효율은 사용된 조작 조건 및 특히 엄격성에 의존한다. 엄격성은 상동성의 정도를 나타내는 것으로 이해될 수 있고; 엄격성이 높으면, 서열 사이의 상동성 비율이 높다. 엄격성은 특히 2개 핵산 서열의 염기 조성에 의해 및/또는 이들 2개 핵산 서열 사이의 부정합의 정도에 의해 정의될 수 있다. 조건, 예를 들면, 염 농도 및 온도를 변화시킴으로써, 소정 핵산 서열은 이의 정확한 상보체(높은 엄격성) 또는 다소 관련된 서열(낮은 엄격성)으로만 결찰할 수 있다. 온도를 증가시키거나 염 농도를 저하시키는 것은 결찰 반응의 선택성을 증가시키는 경향이 있다.

[0142] 결찰 반응은 효소, 구체적으로 DNA 리가제 효소에 의해 수행된다. DNA 리가제는 공유 포스포디에스테르 결합의 형성을 촉매하고, 이는 뉴클레오타이드를 함께 영구적으로 결합한다. 추가로, dsDNA 주형이 존재하지 않는 경우, T4 DNA 리가제도 결찰시킬 수 있지만, 이는 일반적으로 느린 반응이다. 결찰 반응에 사용될 수 있는 효소의 비-제한적 예는 ATP-의존성 이본쇄 폴리뉴클레오타이드 리가제, NAD⁺ 의존성 DNA 또는 RNA 리가제, 일본쇄 폴리뉴클레오타이드 리가제이다. 리가제의 비-제한적 예는 에스케리키아 콜라이(*Escherichia coli*) DNA 리가제, 썬무스 필리포르미스(*Thermus filiformis*) DNA 리가제, 썬무스 썬모필루스(*Thermus thermophilus*) DNA 리가제, 썬무스 스코토독투스(*Thermus scotoductus*) DNA 리가제(I 및 II), CircLigase™(Epicentre; Madison, WI), T3 DNA 리가제, T4 DNA 리가제, T4 RNA 리가제, T7 DNA 리가제, Taq 리가제, Ampligase(Epicentre® Technologies Corp.), VanC-형 리가제, 9° N DNA 리가제, Tsp DNA 리가제, DNA 리가제 I, DNA 리가제 III, DNA 리가제 IV, Sso7-T3 DNA 리가제, Sso7-T4 DNA 리가제, Sso7-T7 DNA 리가제, Sso7-Taq DNA 리가제, Sso7-이. 콜라이(*E. coli*) DNA 리가제, Sso7-Amp 리가제 DNA 리가제, 및 열안정한 리가제이다. 리가제 효소는 야생형, 돌연변이체 이소형 및 유전자 조작된 변이체일 수 있다. 결찰 반응은 완충액 성분, 소분자 결찰 인핸서 및 기타 반응 성분을 함유할 수 있다.

[0143] 바람직하게는, T4 DNA 리가제를 결찰 반응에 사용한다. 본원에 제공된 방법에서, 결찰 반응은, 부 반응을 차단하고 부정합을 최소화하는 고신뢰성 조건하에서 수행된다.

[0144] 중간 반응 생성물 또는 표적 폴리뉴클레오타이드로의 어셈블리는 적합한 결찰 완충 용액을 사용하여 수행할 수 있다. 결찰 완충 용액은, 예를 들면, 전형적으로 뉴클레아제 비함유 환경하에, 선택된 리가제가 활성적인 것을 보장하는 pH에서 수용액이고; 전형적으로, 이는 약 7 내지 9의 pH이다. 바람직하게는, pH는 약 5 mM 내지 50

mM의 농도에서 트리스-HCl에 의해 유지된다. 결찰 완충 용액은 하나 이상의 뉴클레아제 억제제, 통상 칼슘 이온 길레이트제, 예컨대, EDTA를 포함할 수 있다. 전형적으로, EDTA는 약 0.1 내지 10 mM의 농도로 포함된다. 결찰 완충 용액은 선택된 리가제가 활성적하도록 하는데 요구되는 보조인자를 포함한다. 통상적으로, 이는 약 0.2 mM 내지 20 mM의 농도의 2가 마그네슘 이온이고, 통상은 염화물 염으로서 제공된다. T4 DNA 리가제의 경우, 보조인자로서 ATP가 요구된다. 리가제 완충 용액은 또한 환원제, 예컨대, 디티오트레이톨(DTT) 또는 디티오에리트리톨(DTE)를 통상 약 0.1 mM 내지 약 10 mM의 농도로 포함할 수 있다. 임의로, 리가제 완충액은 올리고뉴클레오타이드 및 폴리뉴클레오타이드의 비특이적 결합을 감소시키기 위한 제제를 함유할 수 있다. 예시적 제제는 연어 정자 DNA, 청어 정자 DNA, 혈청 알부민, 덴하르트 용액 등을 포함한다. 바람직하게는, 제1 및 제2 올리고뉴클레오타이드가 표적 서열의 인접한 상보성 영역의 염기와 완전히 정합된 이본쇄를 형성하는 경우에 결찰이 발생하도록 결찰 조건을 조정한다. 그러나, 일부 실시양태에서, 제1 올리고뉴클레오타이드의 5' 말단 및 제2 올리고뉴클레오타이드의 3' 말단에 비-쌍형성 뉴클레오타이드를 허용하여, 검출을 보조하거나 평활 말단 결찰을 감소시키는 것이 유리할 수 있는 것으로 이해된다. 결찰 반응의 중요한 파라미터는 온도, 염 농도, 변성제, 예컨대, 포름아미드의 존재 또는 부재 및 농도, 제1 및 제2 올리고뉴클레오타이드의 농도 및 사용된 리가제의 종류를 포함한다. 반응을 위한 하이브리드화 조건을 선택하는 방법은 당업자에게 공지되어 있다.

[0145] 바람직하게는, 결찰은 엄격한 하이브리드화 조건하에 발생하고, 완전히 정합된 올리고뉴클레오타이드만이 하이브리드화하는 것을 보장한다. 전형적으로, 엄격성은 염 농도를 어느 정도 일정한 값, 예를 들면, 100 mM NaCl 또는 등가의 것으로 유지하면서 하이브리드화가 발생하는 온도를 조정함으로써 조절한다. 다른 인자, 예컨대, 제1 및 제2 올리고뉴클레오타이드의 특정 서열, 제1 및 제2 올리고뉴클레오타이드의 길이 및 선택된 리가제의 열 불안정성이 관련될 수 있다. 바람직하게는, 결찰 반응은, 결찰 완충액 중의 하이브리드화된 올리고뉴클레오타이드의 용해 온도에 근접한 온도에서 수행한다. 보다 바람직하게는, 결찰 반응은 결찰 완충 용액 중의 하이브리드화된 올리고뉴클레오타이드의 용해 온도의 10°C 이내의 온도에서 수행한다. 가장 바람직하게는, 결찰 반응은 결찰 완충 용액 중의 하이브리드화된 올리고뉴클레오타이드의 용해 온도보다 0 내지 5°C 낮은 범위의 온도에서 수행한다.

[0146] 결찰 후에, 하나 이상의 증폭 반응을 수행할 수 있다. 일부 실시양태에서, 결찰 생성물 또는 표적 폴리뉴클레오타이드는 증폭 전에 단리 또는 농후화된다. 단리는, 친화성 정제 및 겔 전기영동을 포함하는 다양한 적합한 정제 방법에 의해 달성할 수 있다. 예를 들면, 결찰 생성물 또는 표적 폴리뉴클레오타이드는 지지체 상에 고정된 선택적 결합제를 포획 프로브에 부착된 태그에 결합시킴으로써 단리할 수 있다. 이어서, 지지체를 사용하여, 포획 프로브 및, 포획 프로브에 하이브리드화된 임의의 폴리뉴클레오타이드를 샘플 반응 용액의 다른 내용물로부터 분리 또는 단리할 수 있다. 이어서, 단리된 폴리뉴클레오타이드는, 증폭 및 추가의 샘플 제조 단계에 사용할 수 있다. 일부 실시양태에서, 포획 프로브는 환상 표적 폴리뉴클레오타이드의 증폭 전에 분해 또는 선택적으로 제거된다. 반응 생성물 또는 표적 폴리뉴클레오타이드의 증폭은 당업자에게 공지된 다양한 적합한 증폭 방법에 의해 달성할 수 있다.

[0147] 용어 "**유도체**"는, 본래 올리고뉴클레오타이드 또는 폴리뉴클레오타이드와는 상이하지만 이의 본질적 특성을 보유하는 올리고뉴클레오타이드 또는 폴리뉴클레오타이드를 지칭한다. 유도체는, 예를 들면, 일본쇄 DNA 또는 상보성 RNA 분자를 조작하기 위해, 하나 이상의 점 돌연변이를 도입하기 위해 또는 화학적 및/또는 효소적 수단에 의해 이중성 모이어티 또는 태그를 결합시키기 위해 출발 물질로서 ds 폴리뉴클레오타이드(예: DNA)를 사용하여 생성할 수 있다.

[0148] 일반적으로, 유도체는 전체적으로 매우 유사하고, 다수의 영역에서 본래 올리고뉴클레오타이드 또는 폴리뉴클레오타이드와 동일하다. 실제 문제로서, 임의의 특정 핵산 분자 또는 폴리펩타이드가 본 발명의 뉴클레오타이드 서열과 적어도 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일한지의 여부는 공지된 컴퓨터 프로그램을 사용하여 통상적으로 결정할 수 있다. 쿼리 서열(본 발명의 서열) 및 대상 서열(또한 글로벌 서열 정렬로서 지칭됨) 사이의 전체 최고 정합을 결정하는 바람직한 방법은 브루트래그 등(Brutlag et al.)(Comp. App. Biosci. (1990) 6:237-245)의 알고리즘에 기초하여 FASTDB 컴퓨터 프로그램을 사용하여 결정할 수 있다. 서열 정렬에서, 쿼리 및 대상 서열은 둘 다 DNA 서열이다. RNA 서열은 U를 T로 전환함으로써 비교할 수 있다. 상기 글로벌 서열 정렬의 결과는 동일성 퍼센트이다. 대상 서열이, 내부 결실 때문이 아니라 5' 또는 3' 결실 때문에 쿼리 서열보다 짧은 경우, 수동으로 결과를 수정할 필요가 있다. 이는, FASTDB 프로그램이 동일성 퍼센트를 계산할 때에 대상 서열의 5' 및 3' 절단을 고려하지 않기 때문이다. 예를 들면, 90 염기 대상 서열을 100 염기 쿼리 서열로 정렬하여, 동일성 퍼센트를 결정한다. 결실은 대상 서열의 5' 말단에서 발생하고, 따라서

FASTDB 정렬은 5' 말단의 최초 10 염기의 정합/정렬을 나타내지 않는다. 손상된 10 염기는 서열의 10%(정합되지 않은 5' 및 3' 말단에서 염기의 수/쿼리 서열 중의 염기의 총 수)를 나타내고, 따라서 FASTDB 프로그램에 의해 계산된 동일성 스코어 퍼센트로부터 10%를 뺀다. 나머지 90 염기가 완전히 정합된 경우, 최종 동일성 퍼센트는 90%일 것이다. 또 다른 예에서, 90 염기 대상 서열은 100 염기 쿼리 서열과 비교된다. 이때의 결실은 내부 결실이고, 따라서 대상 서열의 5' 또는 3' 상에, 쿼리와 정합/정렬되지 않은 염기는 없다. 이 경우에, FASTDB에 의해 계산된 동일성 퍼센트는 수동으로 수정되지 않는다. 이 경우에도, 쿼리 서열과 정합/정렬되지 않은 대상 서열의 염기 5' 및 3'만이 수동으로 수정된다.

[0149] 본 발명의 라이브러리는, 전체 서열 공간을 커버하는데 필요한 수천의 올리고를 포함할 수 있다. 각각의 올리고뉴클레오타이드 라이브러리 멤버는 구획에 물리적으로 배치될 수 있다. 모든 구획은, 함께 "어레이 장치"로서 제공되는, 장치의 하나 이상의 부분 내에 편리하게 제공될 수 있다. 이러한 어레이 장치는 미세역가 플레이트, 마이크로유체 마이크로플레이트, 모세관 세트, 마이크로어레이 또는 바이오칩, 바람직하게는 DNA 또는 RNA 바이오칩 중의 어느 하나 이상일 수 있다. 올리고는, 자동화 수단, 예를 들면, 이러한 구획으로부터 본원에서 반응 구획으로 불리우는 다른 구획으로, 즉 한 용기로부터 별개의 용기로 로봇에 의해 또는, 예를 들면, 자동화 액체 핸들러를 사용하여 전용 유체를 통해 편리하게 수송될 수 있다. 폴리뉴클레오타이드의 시간 효율적 어셈블리를 용이하게 하기 위해, 올리고뉴클레오타이드 라이브러리 멤버의 사용 빈도에 상응하여 반응 계층 및 각각의 용기를 사용할 수 있다. 새로운 용기의 전달은 올리고의 하나 이상의 분자를 각각의 장소로부터 취득하는 장치의 물리적 이동, 또는 마이크로유체를 통한 공기압/수압 침착을 포함한다. 소정 서열을 이론적으로 구축하는데 필요한 올리고의 수가 많기 때문에, 라이브러리 내의 라이브러리 멤버의 대부분의 공간 분포는 라이브러리의 스캐닝 및 액체 핸들러의 긴 이동 시간에 기인하여 낭비된 시간 및 자원으로 발생할 것이다. 그러나, 라이브러리 멤버의 특정 분포를 사용함으로써, 표적 서열에 따라 최소 이동을 확실하게 수행할 수 있다. 하나의 예는, 최초 플레이트가 올리고의 가장 일반적 쌍 조합을 포함하는 마이크로-웰 플레이트에, 가장 사용 빈도가 낮은 라이브러리 멤버를 함유하는 최종 마이크로-웰 플레이트까지 내림 차순으로 저장하는 것이다.

[0150] 구체적으로, 상기 개별 라이브러리 컨테이너먼트는 2차원 순서로 **공간적으로 배열**되고, 개별 구획은 x- 및 y-축 내의 정의된 좌표로 장치 내에 배치된다. 순서는, 주로 합성 시간을 단축시키기 위해 작용하는 파라미터에 의해 특별히 사전에 정의된다. 바람직하게는, 상기 파라미터는 사용 빈도이고, 이들 올리고를 서로 근접하여 배치하고, 이는, 예를 들면, 천연에 존재하거나 표적 ds 폴리뉴클레오타이드 또는 이의 단편에 통상적으로 사용되는 DNA 서열에서 정합 쌍을 빈번하게 형성한다. 보다 바람직하게는, 상기 개별 라이브러리 컨테이너먼트는 3차원 순서로 공간적으로 배열되고, 개개 구획은 x-, y- 및 z-축 내에 정의된 좌표로 장치 내에 배치된다. 순서는 사용 빈도에 의해 명확하게 결정되고, 천연에 존재하는 DNA 서열에서 빈번하게 정합 쌍을 형성하는 이들 올리고를 서로 근접하여 배치한다. 구체적으로, 라이브러리 멤버의 공간 배열은 하기 파라미터 중의 어느 하나 또는 복수에 의존한다: 올리고뉴클레오타이드의 사용 빈도, 천연 DNA 서열에서 올리고뉴클레오타이드의 발생 빈도, 설계된 DNA 서열 세트에서 올리고뉴클레오타이드의 발생 빈도, 마이크로유체 장치에 의한 핸들링 또는 접근 시간의 최소화, 마이크로유체 장치에 의한 조작 비용 또는 소비품 양의 최소화.

[0151] 특정 예에 있어서, 상기 개별 라이브러리 컨테이너먼트는, 적층된 플레이트로서 배열되고 임의로 바코드 표지되고 자동화 마이크로액적 핸들러에 의해 접근가능한 마이크로-웰 플레이트이다. 라이브러리 멤버는 편리하게 상기 적층된 마이크로-웰 플레이트에 저장될 수 있고, 순서 및 적층은 사용 빈도의 감소에 따른다.

[0152] 본원에 사용된 바와 같이, 용어 "**액체 핸들러**", "**자동화 핸들러**" 또는 "**마이크로액적 핸들러**"는 액체 핸들링, 바람직하게는 자동화 액체 핸들링의 방법에 사용된 임의의 장치, 바람직하게는 센서-통합 로봇 시스템에 사용된 장치를 지칭한다. 저-용적 분배는 생명 과학에서 점차 일반적으로 되기 때문에, 기밀 밀봉으로 높은 수준의 정밀도를 갖는 마이크로시린지가 등장했다. 일부 수동 또는 전자 홀더는 피스톤 변위를 정확하게 조절하여 분배된 용적의 정밀도를 보장하도록 설계된다. 시린지 외에, 피펫은 액체 핸들링을 위한 또 다른 일반적 도구이다. 분배된 용적은, 마이크로- 또는 서브-마이크로리터 수준으로 할 수 있다. 멀티채널 피펫은 한번에 다중-라우팅 피펫팅을 위해 권장된다. 시장에는 고정- 및 가변-용적 피펫 둘 다가 있다. 전자는 보다 정확하고 정밀한 반면, 오퍼레이터가 필요에 따라 상이한 용적을 선택할 수 있기 때문에 후자의 적용 범위는 보다 크다. 그 외에, 고효율은 생명 과학 연구에서 매우 중요해지고 있다. 대표적 적용 중의 하나는 마이크로어레이 프린팅이다. 이 기술은, 나노리터 규모에서 각각 바이오샘플 스팟의 어레이를 생성하여, 소량의 샘플만으로 다수의 실험을 병행하여 분석할 수 있게 한다. 수천의 바이오샘플을 스포팅하는 프로세스는 핸들 분배 도구로는 거의 불가능한 작업이고, 로봇 액체 핸들링을 중요한 측면으로 되게 한다.

[0153] 로봇은 피로 없이 동작할 수 있고 처리량을 증가시키고 일정하게 실행하고 정확도 및 정밀도를 보장할 수 있기

때문에, 로봇 워크스테이션은 수동 액체 핸들링보다 다수의 이점을 갖는다. 통합 및 다기능을 갖는 플랫폼의 요건에 따르면, 액체 핸들링 작업이 기능의 일부에 불과한, 여전히 보다 복잡한 시스템이다. 액체-핸들링의 일반적 구조는 다음과 같이 구축할 수 있다. 먼저, 조절 센터는 로봇 워크스테이션의 분배 부분과 세척 스테이션 사이를 이동하는 로봇을 제어한다. 세척 스테이션은, 분배 헤드를 세정하여 이의 수명을 연장시키고 샘플의 안전성을 보장하기 위해 사용된다. 액체 샘플은 분배 헤드로부터 배출되고, 기관 상에 침착되어 추가로 처리된다. 센서는, 조절 센터에 의해 피드백 조절이 수행될 수 있도록, 분배 부분의 상태를 모니터링하기 위해 도입된다. 센서는 모든 워크스테이션에 항상 설치되는 것은 아니고, 보다 양호한 성능을 제공하는 피드백 루프를 구성하기 위해 점점 더 사용된다.

[0154] 용어 "**모세관**"은, 유리 모세관, 마이크로유체 모세관 및 자율 마이크로유체 모세관 시스템 중의 어느 하나를 지칭한다. 모세관 마이크로유체는 다수의 상이한 분야에서 중요한 도구이다. 이들의 축대칭 유동 및 유기 용매에 내성인 능력에 기인하여, 이들의 리소그래피 제조된 폴리디메틸실록산(PDMS) 대응물과 비교하여, 유리 모세관 장치는 마이크로유체 적용을 위한 이점을 보유한다. 특히, 원형 튜브는 정방형 외측 유동 채널에 삽입되고, 이는 이들 장치의 정렬 및 센터링을 크게 간소화시킨다. 이들 장치는 μm 크기로 10 내지 수백의 범위인 작은 액적 및 큰 액적을 생성할 수 있다.

[0155] 용어 "**미세역가 플레이트**"는 웰 플레이트, 다중-웰 플레이트 또는 마이크로-웰 플레이트 중의 어느 하나를 지칭한다. 이들 플레이트는 통상 96, 384 또는 1536 웰을 갖는 2:3 장방형 혼합으로 통상 제조되지만, 다른 캐비티 구성도 이용가능하다. 이용가능한 보다 덜 통상적인 다른 크기의 일부는 6, 24, 3456, 및 9600 웰이다. 마이크로플레이트의 웰은 전형적으로 수십 나노리터 내지 수 밀리리터의 액체를 보유한다.

[0156] 용어 "**마이크로어레이**"는 통상 DNA 또는 단백질의 다수의 분자 또는 단편이 규칙적 패턴으로 부착되어 있는 지지 재료(예컨대, 유리 또는 플라스틱 슬라이드)를 지칭한다. 보다 구체적으로, 이는 정의된 위치에 수천의 작은 스폿이 프린팅되어 있는 현미경 슬라이드를 지칭하고, 여기서 상기 스폿은 DNA 또는 RNA에 결합할 수 있다. 이러한 슬라이드는 바이오칩, DNA 칩, RNA 칩 또는 유전자 칩으로서 또한 종종 지칭된다. 이러한 마이크로어레이는 공유 또는 비-공유 방식으로 DNA 또는 RNA에 결합할 수 있고, 따라서 올리고가 소정 장소, 즉 스폿에 저장되는 어레이 장치로서 기능할 수 있다.

[0157] "**마이크로유체 장치**"는, 생물학적 및 화학적 검정을 실시하기 위한 다수의 이점을 제공하는 마이크로액적의 형태로 개별 유체 패킷의 조작을 가능하게 한다. 이들 이점에는, 검정에 필요한 시약의 용적, 요구된 샘플의 크기 및 장비 자체의 크기의 대폭적 감소이다. 이러한 기술은 또한 프로세스, 예컨대, 가열, 확산 및 대류 혼합이 발생하는 용적을 감소시킴으로써 생물학적 및 화학적 검정의 속도를 증강시킨다. 액적이 생성되면, 신중하게 설계된 액적 조작은, 다수의 액적의 다중화를 허용하여, 대규모의 복잡한 생물학적 및 화학적 검정을 가능하게 한다.

[0158] 용어 "**마이크로유체 마이크로플레이트**"는 마이크로유체 기술과, 마이크로유체 마이크로플레이트 기술의 형태로, 표준 SBS-구성된 96-웰 마이크로플레이트 구조의 조합을 지칭한다. 마이크로유체 마이크로플레이트는, 중요한 워크플로우의 개선, 샘플 및 시약의 보존, 개선된 반응 속도, 및 다중 분석물 로딩에 의한 검정의 감도를 개선시키는 능력의 개선을 가능하게 한다[참조: Kai et al., 2012].

[0159] 본원에서 사용된 용어 "**메틸트랜스페라제**"는 DNA 메틸트랜스페라제, RNA 메틸트랜스페라제, 단백질 메틸트랜스페라제 및 히스톤 메틸트랜스페라제 중의 어느 하나를 지칭할 수 있다. 메틸트랜스페라제는 추가로, 모두 S-아데노실 메티오닌(SAM)을 결합하기 위한 로스만 폴드를 함유하는 부류 I, 및 SET 도메인 히스톤 메틸트랜스페라제에 의해 예시되는, SET 도메인을 함유하는 부류 II 메틸트랜스페라제, 및 막 연관되는 부류 III 트랜스페라제로 세분될 수 있다.

[0160] 용어 "**CRISPR/Cas9**"는, 당업자에게 공지된 유전자 편집 방법, 뿐만 아니라 이의 변형을 지칭한다. 이러한 변형은 뉴클레아제-소멸 Cas9(dCas9)의 시티딘 데아미나제로의 융합, 시티딘의 우라실로의 부위-특이적 전환, 일본쇄 DNA 절단물(닉)만을 생성하는 Cas9 단백질의 버전을 생성하는, Cas9 단백질로의 돌연변이를 포함하지만, 이들로 한정되지 않는다.

[0161] 용어 "**다중 자동화 게놈 조작**" 또는 "**MAGE**"는, 전체 세포 배양이 게놈 또는 표적화 영역으로의 일련의 변화를 포함하는 상태에 근접하도록, 하나 이상의 세포로 복수의 핵산 서열을 도입하는 것을 일반적으로 포함하는 기술을 지칭한다. 이 방법은, 대립유전자의 하나의 특정 구성을 생성하기 위해 사용될 수 있거나, 추가의 랜덤 또는 비-설계된 변화를 임의로 포함하는 설계된 대립유전자의 조합 탐색을 위해 사용될 수 있다.

- [0162] ssDNA-결합 단백질 매개된 재조합, 상동성 재조합 및 MAGE-기반 방법은 전형적으로, 올리고뉴클레오타이드를 포함하는 형질전환 배지 또는 형질감염 배지를 사용하여 세포를 형질전환 또는 형질감염시키는 단계, 형질전환 배지 또는 형질감염 배지를 성장 배지로 교환하는 단계, 세포를 성장 배지에서 인큐베이팅하는 단계 및, 필요하거나 바람직한 경우, 복수의 핵산 돌연변이가 목적 뉴클레오타이드 서열 내로 도입될 때까지 상기 단계들을 반복하는 단계를 포함하는 단계를 포함하여, 복수의 올리고뉴클레오타이드를 세포에 도입하는 것을 포함한다. 돌연변이유발 사이클의 수를 증가시키는 것은 일반적으로 도입된 돌연변이의 다양성을 증가시킨다.
- [0163] MAGE는 특히, 세포의 게놈이 가속된, 지시된 진화의 형태를 통해 목적하는 기능을 수행하도록 재프로그래밍되는 프로세스인, 매우 효율적인 람다 파지 레드 재조합 시스템(λ 레드 시스템)을 사용한다. λ 레드 시스템은 β , γ 및 엑소 유전자를 포함하고, 이들 생성물은 각각 Beta, Gam 및 Exo로 불리운다. Gam은 숙주 RecB,C,D 엑소 뉴클레아제 및 SbcC,D 뉴클레아제 활성을 억제하고, 따라서 외인적으로 부가된 선형 DNA는 분해되지 않는다. Exo 단백질은, 각 쇠의 말단에 결합하고 다른 쇠를 5'으로부터 3' 방향으로 분해하는 dsDNA-의존성 엑소뉴클레아제이다. Beta는 수득되는 ssDNA 돌출부에 결합하고, 최종적으로 이들을 상보성 염색체 DNA 표적과 쌍형성한다. λ 레드 시스템은, 이. 콜라이(*E. coli*), 살모넬라, 시트로박터 및 시겔라 종에서 특정 유전자 불활성화, 및 이들 염색체로의 작은 생물학적 태그 또는 단일 유전자의 도입에 광범위하게 이용되고 있다.
- [0164] 용어 "공액 어셈블리 게놈 조작" 또는 "CAGE"는 공액을 사용하여, 복수의 이. 콜라이의 상이한 유전자형을 단일 키메라 게놈으로 계층적으로 조합하는 게놈 어셈블리의 정확한 방법을 지칭한다. CAGE는 시험관내 조작에 의해 부여된 제약 없이 균주간의 특정 게놈 영역의 대규모 전달을 가능하게 한다. 접합 기능을 구비하는 공여체 균주와 공여체로부터 DNA를 수용하는 수용체 균주를 확립함으로써 균주는 쌍 방식으로 어셈블리된다. 균주 쌍 내에서, 공여체 및 수용체 게놈에서 전달 및 선택가능한 마커의 공액 기점의 표적화된 배치는 목적하는 공여체-수용체 키메라 게놈의 조절된 전달 및 선택을 가능하게 한다. 설계상, 선택가능한 마커는 게놈 앵커 포인트로서 작용하고, 이들은 계층적 게놈 전달의 후속 라운드에서 재이용된다.
- [0165] "Ago"는 DNA-기반 DNA 간섭을 제공하는 것을 나타내는 아르고나우트 단백질을 지칭하고, 여기서 일본쇄 DNA 가이드는 플라스미드 DNA 표적의 Ago-기반 절단을 지시할 수 있다. 주요 이점은, CRISPR-Cas9와는 달리, 프로토스페이서 인접 모티프(PAM)의 필요가 없다는 것이다.
- [0166] 징크-핑거 뉴클레아제(ZFN) 및 전사 활성화인자-유사 이펙터 뉴클레아제(TALEN)은, 25 내지 40 bp 크기 범위의 DNA 표적 부위를 이들의 DNA-결합 도메인을 통해 서열-특이적 방식으로 인식하고, 반대측 DNA 쇠에 대한 FokI 뉴클레아제 도메인의 작용을 통해 엇갈림 이분쇄 절단을 생성한다.
- [0167] 호밍 엔도뉴클레아제로도 또한 공지된 "메가뉴클레아제"는, 이들이 전달하고 DSB를 유도하는, 14 내지 40 bp의 특정 DNA 서열을 인식한다. 메가뉴클레아제의 효율은 상당히 높고, 이들은 각 표적 부위에 대해 단일 맞춤형 바이오플리머만을 필요로 한다.
- [0168] 전형적으로 30 내지 40 bp 길이의 표적 서열을 인식하는, "티로신/세린 부위-특이적 리콤비나제" 또는 "Tyr/Ser SSR"은 포유동물 게놈에서 상동성-지시된 수복(HDR)을 가능하게 하는 초기 게놈-조작 도구의 하나이다. 요약하면, 표적 부위는 3개 부분, 2개 역방향 반복체에 의해 인접하는 짧은 DNA 서열을 포함하고, 재조합은 표적 부위의 쌍 사이에서 발생할 수 있고, 표적 부위 사이의 DNA 서열은 결실, 역전 또는 치환될 수 있다. 특히, Tyr SSR은 이분쇄 절단을 생성하지 않으면서 쇠 교환의 메커니즘을 이용하고, Ser SSR은 이분쇄 절단을 생성하지만, 보다 단순한 설계자 이분쇄 뉴클레아제와는 달리, SSR은 협조하는 절단 및 존재하는 공여체 DNA와의 재-결찰을 필요로 한다.
- [0169] 상기 기재는, 하기 실시예를 참조하여 보다 완전히 이해될 것이다. 그러나, 이러한 실시예는 본 발명의 하나 이상의 실시양태를 실시하는 방법의 단순한 대표이고, 본 발명의 범위를 한정하는 것으로 읽히는 것은 아니다.
- [0170] 실시예
- [0171] 하기 실시예에서는, 올리고의 라이브러리가 어떻게 생성되고 어떻게 처리되고 이의 내용 및 특성이 검증되는지에 대해 기재된다. 추가로, 본원에 제공된 방법에 따라 폴리뉴클레오타이드가 어떻게 합성되는지가 기재되어 있다.
- [0172] 실시예 1: 라이브러리의 생성
- [0173] **1.1 유전자 정보의 공간 구조의 결정**
- [0174] A. 먼저, 라이브러리에 포함되어야 하는 올리고뉴클레오타이드의 모든 서열을 수록해야 한다. 이들 서열은 잠

제적으로 바람직한 모든 표적을 커버하는 서열의 입력 세트로부터 사전-계산된다. 이 정보는 기준의 다양성, 예컨대, 가능 조합의 서브세트(예를 들면, 모든 헵타머, 모든 옥타머 등), 제한 효소의 세트 또는 임의의 기타 계산 기준에 의한 게놈의 소화의 예측 결과로부터 취득할 수 있다.

- [0175] 본 실시예에서, 16569 염기 쌍을 갖는 인간 미토콘드리아 게놈(Gene Bank accession nr. J01415; Anderson et al., 1981)가 라이브러리의 기초로서 기능했다. 이상적으로는, 모든 보고된 서열은 하기 및 도 1에 기재된 것과 동일한 방법으로 취득 및 처리된다.
- [0176] 참조 게놈 서열은 8 내지 26 bp 길이의 올리고뉴클레오타이드 이량체로 분할되었다. 유사하게는, 역-상보체를 계산하고, 8 내지 26 nt 길이의 올리고뉴클레오타이드로 분할했다. 이것에 의해, 4 nt 돌출부를 갖는 ds 구조를 형성할 수 있는 합계 2,070 올리고가 수득된다(도 1 참조). 이어서, 서열을 먼저 1, 이어서 2, 이어서 3, 최대 15 뉴클레오타이드에 쉬프트하여 동일한 프로세스를 반복적으로 수행했다. 그 결과, 라이브러리에는 $2070 \times 16 = 33,120$ 올리고가 함유되어 있었다.
- [0177] 이들 데이터베이스에는, 4 bp씩 증첩하는 것들을 계산하는 경우, 적어도 16544개의 정합 쌍이 있다. 일반적으로 말하면, 변이체 서열은 동일한 방식으로 처리되어야 하고, 이는 비선형 방식으로 다중도를 증가시킨다. 예를 들면, 16 다형 부위만을 함유하는 약 100 bp의 윈도우는 400 초과 올리고 및 거의 20,000 정합 쌍을 부가한다(도 1). 조합론은, 보다 많은 가변 부위를 고려하는 경우, 올리고 라이브러리가 비-다항식으로, 이의 설계에서 고려되는 서열 변이체의 수로 입력되는 것을 의미한다.
- [0178] 올리고의 몇몇은 일배체 전체에 보존되어 있고, 쌍형성 요소로서 라이브러리에 할당되어 있다(도 1B). 가변 부위로 스패닝하는 올리고(및 가변성의 범위에 따라)는 ssDNA 요소로서 독립적으로 유지되었다.
- [0179] B. 라이브러리의 2-차원 배열은, 바람직한 기준에 따라 라이브러리 멤버를 선별함으로써 결정했다. 여기서, 16-머를 먼저 서열 시프트에 의해 선별하고, 서열에서의 최초 발생 순서에 따라 공액 쌍을 교대하여 두번째 선별했다. 대체 올리고가 소정 위치에서 발생한 경우, 이들은 발생 빈도에 따라 서브-선별했다. 모든 입력 서열에 걸쳐 보존된 올리고는, 동일한 위치에서 공액 쌍이 할당되었다.
- [0180] 올리고의 개개 용도 및 또한 이의 정합 쌍의 상대적 용도 둘 다를 반영하는 대체 기준은 사전순, 길이, 정합 쌍의 인접성, 빈도 또는 기타 임의의 공지된 방법일 수 있다.
- [0181] C. 이어서, 제1 서열을, 실제 올리고(들)이 배치되는 1536 마이크로-웰 플레이트 내의 위치(들)에 상응하는 2차원 어레이로 할당했다.
- [0182] D. 나머지 1535 웰이 모두, 단계 B의 선별 우선도를 반영하는 순서로 올리고에 의해 점유될 때까지 후속 올리고를 첨가했다.
- [0183] E. 이어서, 후속 1536 올리고로 단계 C를 반복하고, 모든 33,120개 이상의 올리고가 마이크로-웰 플레이트에 분배될 때까지 반복했다.
- [0184] F. 정보는, 각 올리고의 장소를 추적하기 위해 디지털로 저장했다. 후속 단계에서, 이는 2개 목적으로 사용했다; 첫째, 이는 올리고로의 접근을 용이하게 하기 위한 룩-업(look-up) 테이블로서 기능하고, 둘째, 이는 이용 가능한 용적을 추적하기 위해 모든 올리고의 용도 및 접근 빈도의 모니터링을 가능하게 했다.
- [0185] **1.2. 라이브러리의 합성**
- [0186] 서열이 적절하게 구조화되면, 각 올리고의 실제 합성을 수행했다. 물리적으로, 인간 미토콘드리아 게놈을 작제하기 위해 사용된 라이브러리는, 폴리프로필렌으로 제조된 27개의 1536 마이크로-웰 플레이트(Corning 1536 웰 플레이트, Sigma Aldrich Product Nr. CLS3726-50EA)로 구성된다(폴리프로필렌이 바람직하지만, 표면에 대한 DNA 흡수를 최소화하는 임의의 물질이 사용될 수 있다. 각각의 플레이트는, 용이한 접근 및 내용 복기평을 위해 명확하게 표지 및/또는 바코드화했다.
- [0187] 생성된 각 올리고는, 상기 결정된 바와 같이 소정 플레이트에 배치했다. 이 예에서, 올리고는 5' 말단에서 포스포릴화시킨다. 다른 적용은 3', 5' 또는 양 말단에서 다른 변형, 예컨대, 디- 또는 트리-포스페이트, 비오틴, TEG 또는 티올 변형체 등, 또는 메틸화 등을 사용한 처리를 필요로 할 수 있다. 올리고는 수용액(뉴클레아제 비함유 ddH₂O 또는 TRIS 10mM pH 8.0 및 1mM EDTA)에 마이크로-웰당 올리고당 10 μ l의 용적으로 200 μ M의 농도로 유지했다(Sambrook and Russell, 2014).
- [0188] 라이브러리의 실제 생성은, 천연 존재 DNA를 뉴클레아제로 소화시키고, 올리고-합성기 등으로 화학적으로 작제

한 다음, HPLC, 모세관 전기영동 또는 기타 기술로 분리 및 정제함으로써 분자 생물학의 표준 방법으로 수행할 수 있다. 올리고뉴클레오타이드의 합성 및 변형은 표준이기 때문에, 또한 다수의 서비스로부터 외부위탁할 수 있다. 이러한 예에 따르면, 라이브러리는, 모노뉴클레오타이드를 고상-부착된 폴리뉴클레오타이드에 공유 결합하는 데옥시뉴클레오사이드 포스포르아미다이트의 화학 반응을 반복적하는 실행하는 자동 DNA 합성기를 사용하여 생성했다(Beaucage and Cartuthers, 1981).

[0189] 라이브러리는, 단기간 동안에 사용하지 않는 경우에 -20°C에서, 및 장기간 저장을 위해 -80°C에서 저장했다.

[0190] **1.3 라이브러리의 용도**

[0191] A. 라이브러리는, 플레이트를 3°C에서 적어도 60분 동안 방치함으로써 해동시키고, 아이스에 또는 쿨러 플레이트 상에 3 내지 5°C의 온도에서 유지시켰다.

[0192] B. 각 마이크로-웰 플레이트를 오비탈 혼합기에서 2500 rpm으로 30초 동안 와동시키고, 원심분리기로 900 rpm에서 1분 동안 스핀 다운했다.

[0193] C. 소량의 마이크로-액적 핸들러(TPP Lab Tech Mosquito X1)를 사용하여, 100 nL(권장 범위: 50 내지 250 nL)을 새로운 384 마이크로-웰 플레이트(다른 용량; 예컨대, 96 또는 1536, 또는 표면이 또한 사용될 수 있음)로 전달시키고, 이는 올리고가 조합 및/또는 추가로 반응되는 1.8 μL의 용액 또는 용액 액적(권장 범위는 1 내지 5 μL이다)을 함유했다.

[0194] D. 디지털 데이터베이스에서, 각 마이크로-웰의 사용된 용적에 주석을 달아, 추가 라운드의 사용을 위해 필요한 모든 올리고가 항상 충분한 것을 보장했다. 일부 액체 핸들러는, 각각의 접근된 웰에서 사용된 및 나머지 용적의 정확한 및 실시간 측정을 제공하는 것에 유의한다. 이러한 기능은 보다 정확한 추적을 보조할 수 있다.

[0195] E. 라이브러리를 사용한 경우, -80°C의 저장으로 반송했다.

[0196] **1.4 라이브러리의 특성의 결정**

[0197] 본 발명의 라이브러리를 정의하는 주요 특성은 1) 올리고뉴클레오타이드의 정의된 길이, ii) 적어도 하나의 돌출부를 갖는 일본쇄 및/또는 이본쇄 및 iii) 특정 수의 올리고이다. 본 예에서 사용된 라이브러리의 주요 특성은 i) 8 내지 25 nt 범위의 올리고뉴클레오타이드의 길이, ii) 적어도 하나의 돌출부를 갖는 일본쇄 및 이본쇄 올리고의 존재, 및 iii) 적어도 33,120 올리고가 라이브러리에 포함되는 것이다.

[0198] 이들 특성이 품질 관리의 목적으로 유지되는 것을 검증하는 것이 바람직하다.

[0199] *I. 뉴클레오타이드의 길이 확인*

[0200] 마이크로-액적 핸들러를 사용하여, 각 마이크로-웰의 5 내지 10 nL의 분취량을 취득하고, 공통 용액으로 풀링했다. 또는, 랜덤 분취량을 취하고, 각각의 올리고가 하나의 풀에만 존재하도록 하는 방식으로 10개의 상이한 풀링 용액으로 풀링했다. 풀 또는 풀들은 와동에 의해 혼합했다. 풀링된 용액당 수 μL의 작은 분취량을 모세관 전기영동을 통해 작동시켰다(Kemp, 1998). 또는, 샘플은, 25% 아크릴아미드 겔 상에서 분석하고, 6 내지 24 bp의 ssDNA의 표준 래더와 비교했다.

[0201] *II. 라이브러리에 존재하는 올리고뉴클레오타이드의 구조 확인.*

[0202] ss 올리고, ds 올리고, 및 ss 돌출부를 갖는 ds 올리고는, 소정 올리고의 변성된, 그러나 무처리된 샘플을 엑소뉴클레아제, 예컨대, 이. 콜라이 엑소뉴클레아제 I으로 처리한 샘플과 비교함으로써 구별했다(예를 들면, Thermo Scientific Exonuclease I, product nr. EN0581). 이 효소는 ssDNA를 모노-뉴클레오타이드 및 디-뉴클레오타이드로 소화시키지만, dsDNA는 그대로 잔류시킨다(Lehman and Nussbaum, 1964). 따라서, 무처리된 및 처리된 샘플은, 모세관 전기영동으로 검사하는 경우, 하기 결과 중의 하나를 제공했다:

[0203] - 무처리된 샘플은 6 내지 26 nt 범위 내에서 단일 밴드를 나타내고, 처리된 샘플은 밴드를 나타내지 않았다. 이는, 본래 샘플이 ss DNA로 구성되어 있음을 시사한다.

[0204] - 무처리된 샘플은 6 내지 26 nt 범위 내에서 단일 밴드를 나타내고, 처리된 샘플은 동일한 밴드를 나타냈다. 이는 본래 샘플이 ds DNA(돌출부 없음)로 구성되어 있음을 시사한다.

[0205] - 무처리된 샘플은 6 내지 26 nt 범위 내에서 둘다 2개의 상이한 밴드를 나타내고, 처리된 샘플은 단일 밴드를 나타내고, 이의 길이는 무처리 샘플의 최소 밴드와 일치했다. 이는 본래 샘플이 하나의 돌출부를 갖는 DNA의 이량체로 구성되어 있음을 시사한다. 돌출부의 길이는, 무처리 샘플의 2개 밴드의 크기의 차이이고, ds 부분의

길이는 처리된 샘플에 나타난 것이다.

- [0206] - 무처리된 샘플은 6 내지 26 nt 범위 내에서 단일 밴드를 나타내고, 처리된 샘플은 길이가 무처리 샘플의 것보다 짧은 단일 밴드를 나타냈다. 이는, 본래 샘플이 동일한 크기의 2개 돌출부를 갖는 DNA의 이량체로 구성되어 있음을 시사한다. 돌출부의 길이는 처리된 및 무처리된 샘플의 크기의 차이이고, ds 부분의 길이는 처리된 샘플의 밴드에 의해 나타난 것이다.
- [0207] - 무처리된 샘플은 6 내지 26 nt 범위 내에서 둘다 2개 밴드를 나타내고, 처리된 샘플은 길이가 무처리 샘플의 둘보다 짧은 단일 밴드를 나타냈다. 이는, 본래 샘플이 상이한 크기의 2개 돌출부를 갖는 DNA의 이량체로 구성되어 있음을 시사한다. 돌출부의 길이는 처리된 샘플에 의해 제공된 크기와 비교하여 각 밴드의 크기의 차이에 의해 결정되고, ds 부분의 길이는 처리된 샘플 상에 나타난 것이다.
- [0208] 다른 분석 기술, 예컨대, HPLC는 또한 무처리 샘플의 조성을 이들의 스펙트럼에서 나타내고, 이는 단일 종의 DNA 또는 2종의 DNA의 존재를 직접 나타내고, 라이브러리의 1개 웰에서 올리고뉴클레오타이드의 성질의 직접적 증거를 제공한다. 또한, 원형 이색성을 사용하여, 일본쇄 및 이본쇄 DNA, 추가로 돌출부를 갖는 dsDNA를 구별할 수 있다.
- [0209] *III. 올리고뉴클레오타이드의 수 및 정합 쌍의 수의 확인.*
- [0210] 각 마이크로-웰의 내용물의 50 내지 100 nL의 샘플을, 95°C에서 3분 동안 가열하여 어닐링한 일반 용액에 풀링하고, 적어도 실온 또는 16°C까지 냉각시켰다. 결합에 필요한 상응하는 완충액을, 필요한 보조인자, 예컨대, Mg²⁺, ATP 등을 포함하여 첨가했다. 반응을 촉매하기에 충분한 리가제(예를 들면, T4 리가제, NEB, product nr. M0202)(반응 용액 μ L당 1U)를 첨가했다. 반응 혼합물을 1시간 동안 실온에서 또는 밤새 16°C에서 인큐베이션했다.
- [0211] 가설에서는, 정합 쌍이 충분한 경우, 리가제는 이들을 공유 결합하고, 랜덤 서열을 갖는 길이 범위의 DNA 분자를 생성한다. 길이의 분포는, TAE에서 아가로스 2 내지 4%를 사용한 전기영동을 사용함으로써 해결했다. 적합한 래더와 함께 샘플을 실행하면(별개 레인에서; 50 또는 100 bp를 권장), 불연속 밴드 없이 샘플 레인을 따라 DNA의 스메어가 나타났다. 래더에 의해 안내된 겔을 절단함으로써 대략 100 내지 200 bp의 좁은 범위가 단리된다[참조: Sambrook and Russell, 2014; Ch. 5]. 겔-추출의 표준 프로토콜에 따라, 절제된 아가로스로부터 DNA를 단리했다(예를 들면, Zymoclean 겔 DNA 회수 키트, Zymo research, product nr. D4001T). 정제 후, 풀 내의 상이한 서열을 결정하기 위해, 샘플을 딥-서열분석했다(Bentley et al., 2008).
- [0212] 후속 분석은, 올리고 및 정합 쌍의 수를 평가하기 위해 수행했다. 반응의 출발 물질이 6 내지 26 nt의 DNA로 구성되고 서열의 반복성이 매우 높지 않은 경우, 평균하여, 적어도 $2 \times N \times 100 / 26$ 올리고(N은 보고된 서열의 수이다) 및 최대 $2 \times N \times 200 / 6$ 올리고 및 거의 동등한 다수의 정합 쌍이 존재한다. 추가로, 다음과 같이, 생물정보 분석을 사용하여, 올리고의 서열을 추출했다. 서열 중 하나의 최초 6 nt를 취하고, 완전한 서열 풀에서 이 패턴의 검색 및 정합을 수행하고, 발생 수에 주석을 달았다. 이것을 7 nt, 이어서 8 nt 등, 26 nt까지 반복했다. 통계적 T-시험을 사용하여, 랜덤 발생과 유의적으로 상이한 수를 결정했다. 이 독특한 패턴은 추정 올리고의 목록에 저장하고, 모든 이의 발생은 데이터베이스로부터 제거했다. 추가로 분할되지 않고 이제 패턴의 목록에 추가되는 6 내지 26 nt 사이의 DNA 서브-서열만이 잔류할 때까지 이 절차를 나머지 서열로 반복한다. 동정된 올리고의 수는 M으로 불리운다. 이들 올리고는 적어도 하나의 다른 올리고와 결합되어 있기 때문에, 반대 측 쇄의 부분적 상보체와 함께, 연속 올리고는 정합 쌍의 일부인 것으로 시사된다. 따라서, 말단에 있는 것을 제외하여, 적어도 동정된 올리고의 수와 동일한 수의 정합 쌍이 존재했다. 예를 들면, M-N 정합 쌍의 순서가 존재했다. 통계 분석 및 부트스트랩 시뮬레이션을 수행하여, 동정된 수가 적어도 33,120 올리고의 거대 세트의 서브샘플인 것으로 예상할 수 있는지를 결정했다.
- [0213] 실시예 2: 128 bp의 표적 DNA 분자의 합성
- [0214] 본 실시예에서는, 본원에 제안된 방법에 의해 128 bp의 서열을 합성하는 방법을 입증했다. 도 2A는, DISCOVER로서 불리우고, 각 쇄에 4 nt 돌출부(도 2C 참조)와 8개 상보성 부위를 갖는 16 nt의 8 ds 올리고를 형성한 16개 정합 쌍(도 2B)로부터 구성된 목적 서열을 나타낸다. 각 ds 올리고는 문자 D, I, S, C, O, V, E, R 및 이들의 구성 리딩 및 래깅 쇄를 각각 + 및 - 위치자로 나타낸다. 올리고는, 실시예 1에서 생성된 라이브러리의 일부이다. 이는 하기 특성을 갖는다: 모든 올리고는 5' 말단에서 포스포릴화되었고, 뉴클레아제 비함유 ddH₂O에 200 μ M의 농도로 제공되고, 사용된 올리고는 일본쇄로 순수했다.

- [0215] *A. 어닐링 용액의 제조*
- [0216] 반응 튜브에서, TRIS-HCl(50 mM), MgCl₂(10 mM), DTT(10 mM) 및 ATP(1 mM)을 갖는 ddH₂O 상의 252 μL를 제조했다. pH는 7.5로 설정했다. 일부 상업적 완충액, 예컨대, 뉴 잉글랜드 바이오랩스(New England Biolabs) 리가제 반응 완충액(product nr B0202S)는 H₂O에서 혼합할 수 있고, 리가제 활성화에 필요한 ATP를 용이하게 함유한다. 용액을 와동에 의해 양호하게 혼합했다. 이 용액 혼합물의 28 μL를, 4×2 어레이의 8개 마이크로-웰에 분배했다. 각 올리고의 1 μL를 플레이트의 소정 마이크로-웰로 전달하고, 피펫팅에 의해 양호하게 혼합했다:
- [0217] 웰 A1에 D+ 및 D-
 [0218] 웰 A2에 I+ 및 I-
 [0219] 웰 A3에 S+ 및 S-
 [0220] 웰 A4에 C+ 및 C-
 [0221] 웰 B1에 O+ 및 O-
 [0222] 웰 B2에 V+ 및 V-
 [0223] 웰 B3에 E+ 및 E-
 [0224] 웰 B4에 R+ 및 R-
 [0225] *B. 어닐링*
- [0226] 플레이트를 밀봉하고, 써모사이클러에서 95°C에서 5분 동안 인큐베이팅하여 ss 올리고의 정합 쌍을 어닐링시켰다. 이어서, 온도를 1분당 1°C씩 저하시키는 램프 기능을 사용하여, 온도를 16°C까지 저하시켰다. 종료되면, 이분쇄 올리고를 16°C에서 유지했다.
- [0227] *C. 결찰 용액의 제조*
- [0228] 결찰 용액은, 아이스 상에서, 하기 순서, 13.3 μL의 뉴클레아제 비함유 ddH₂O, 2 μL의 리가제 완충액 및 4 μL의 ATP를 최종 농도 1 mM로 혼합함으로써 제조했다. 결찰 용액은 와동 및 스핀 다운에 의해 양호하게 혼합했다. 0.7 μL의 T4 리가제(NEB, product nr. M0202)를 첨가하여, 최종 용액 1 μL당 총 1 유닛으로 하고, 온화하게 피펫팅하여 양호하게 혼합했다. 필요할 때까지 용액을 아이스 상에 유지했다. 2.5 μL의 결찰 용액을, B의 ds 올리고를 함유하는 8개 마이크로-웰의 각각에 전달하고, 피펫팅에 의해 혼합했다. 그 후, 플레이트를 다시 밀봉했다.
- [0229] *D. 결찰 라운드*
- [0230] 결찰의 최초 라운드의 경우, 하기 웰을 다음과 같이 병합했다: D+I, S+C, O+V, E+R. 이는, 한 웰의 내용물을 다른 웰로 전달함으로써 달성했다(양 웰의 내용물을 새로운 웰로 전달하는 것도 가능하다). 가장 좌측 내용물이 가장 우측으로 전달되는 도식이 사용되었다(도 3A). 결찰 반응 혼합물을 16°C에서 적어도 1시간 동안 인큐베이팅했다. 이 프로세스는, 웰 DI+SC 및 OV+ER을 병합함으로써 반복하고(도 3B), 다시 각각을 1시간 동안 인큐베이팅했다. 최종 결찰 라운드의 경우, 웰 DISC+OVER을 병합하고, 추가로 1시간 동안 인큐베이팅했다(도 3C). 128 bp 생성물을 함유하는 최종 용적은 140 μL였다.
- [0231] *E. 정제*
- [0232] 11 웰의 콤보를 갖는 2% 아가로스 겔(5 μL의 SYBR Safe DNA 염료가 공급된 50 mL TAE 중의 1 mg 아가로스)을 제조했다. 4.5 μL의 50 bp 래더(New England Biolabs product nr. N3236 또는 Invitrogen product nr. 10416014)를 최초 레인에 첨가하고, 단계 D에서 수득된 140 μL의 용액을 나머지 웰에 분배했다. 겔을 85 V, 200 mA 및 12 와트에서 50분 동안 실행했다. 전기영동을 완료한 후, 겔을 UV 트랜스-일루미네이터 상에 위치시키고, 128 bp 단편에 상응하는 겔의 밴드를 절제했다. 이들 밴드의 정제는, 상기 목적(예를 들면, Zymoclean, 상기 실시예 참조)을 위해 시판 키트를 사용하여, 또는 이 목적을 위한 임의의 표준 프로토콜에 따라 수행할 수 있다.
- [0233] *F. 증폭*
- [0234] 생성물의 양을 추가로 증가시키기 위해, 단계 D에서 수득된 생성물을 PCR에 의해 증폭시켰다(Sambrook and

Russell, 2014; Chapter 8). 출발 16 nt D- 및 R+를 상기 증폭을 위한 프라이머로서 사용했다. 증폭 후, 작제물은 효소 및 프라이머를 함유하지 않고, 2개 분취량으로 분리하고, 하나는 추가 사용을 위해 표지하고 -20°C에서 저장하고, 다른 하나는 작제물의 서열 검증에 사용했다.

[0235] 도 4는, 중간 단계 및 이 프로세스의 최종 결과를 나타내는 아크릴아미드 겔을 나타낸다. 레인 6 및 7에서, 상부 밴드는 128 bp 표적 ds 폴리뉴클레오타이드에 상응한다. 이 작제물을 단리하고(2% 아가로스 겔로부터; 제시하지 않음), 정제하고, 증폭시키고, 양 쇄를 생거-서열분석했다. 수득되는 서열은 표적 및 이의 역 상보체와 동일했다.

[0236] 실시예 3: 복합체 서열 또는 RNA 합성을 위한 표적 DNA 서열의 후처리.

[0237] **3.1 프록시 ds 폴리뉴클레오타이드의 설계**

[0238] 본 실시예에서는, 이의 워크플로우가 통상 모호한 단계를 포함하는 ds 폴리뉴클레오타이드, 예컨대, 자가-상보성 올리고 이량체를 합성했다(예를 들면, 도 5A). 이러한 자가-상보성 이량체는, 불필요한 폭주 반응을 회피하기 위해 워크플로우로부터 제외될 필요가 있기 때문에, 수득된 어셈블리 워크플로우가 명확해지도록 하는 방식으로, 자가-상보성 요소를 상이한 염기로 치환함으로써 주형 서열을 고안했다. 이러한 주형에 따라, 프록시 ds 폴리뉴클레오타이드를 합성했다.

[0239] 도 5A에는, 목적 서열이 제시되어 있다. 밑줄친 부분은, 자가-상보성 및 자가-중합이 가능한 서열의 부분을 나타낸다. 이들 서열을 회피하기 위해, 3개 올리고에 스페닝하는 2개 염기 쌍 변형을 포함하는 주형 서열(도 5B)를 설계했다.

[0240] 프록시 ds 폴리뉴클레오타이드는, 실시예 2에서 입증된 바와 같이 본원에 제시된 방법으로 합성했다. 프록시 서열은, 실시예 2의 올리고 0- 및 V+와 일치하도록 선택했고, 그 결과, 이의 합성은 상기 기재된 바와 같이 정확하게 진행했다.

[0241] 프록시 ds 폴리뉴클레오타이드가 합성되면, 목적 서열과 동일한 서열을 갖는 ds 폴리뉴클레오타이드를 다음과 같이 생성했다. 지시된 돌연변이유발의 원리를 적용하고, PCR 증폭시, 합성된 프록시 ds 폴리뉴클레오타이드에서 제외된 표적 서열의 일부를 본래 표적 서열로 치환했다.

[0242] 합성이 완료되고 128 bp 프록시 ds 폴리뉴클레오타이드를 정제한 후, PCR 반응을 준비했다. 이 반응 혼합물에서는, 3' 말단 프라이머 뿐만 아니라 "돌연변이유발 프라이머"(AttB)의 쌍이 포함되었다. 이들 돌연변이유발 프라이머는, 돌연변이유발 요소(본 실시예에서, 3개 염기)의 양 측면에, 프록시 서열과 완전히 중첩하는 10개 뉴클레오타이드를 갖는다. 이들 규정에 의해, 표준 PCR을 수행하여, 본 실시예에서는 반응 조건 및 시약을 표준화하는 상업용 키트(Taq PCR Kit, New England Biolabs, product nr. E5000S)를 사용하여, SOI(Sambrook and Russell, 2014; Ch. 13)과 동일한 서열을 갖는 ds 폴리뉴클레오타이드를 검색했다.

[0243] **3.2 RNA의 생성**

[0244] 소정 표적 서열을 갖는 RNA 분자는 또한, 프록시 ds 폴리뉴클레오타이드를 사용하여 생성해야 한다. 이는 2개 단계로 수행했다. 먼저, 목적 RNA 서열의 역-상보성 서열(즉, DNA 서열)을 계산해야 했다. DNA 서열은 합성되는 서열이다. 둘째, 특이적 프로모터 서열을 주형 DNA 서열에 통합시켜, 이후에 DNA를 RNA로 전사하는 DNA-의 존성 효소에 의해 인식할 수 있게 한다(Rio, 2011). 본 실시예에서는, T7 RNA 폴리머라제 I 시스템을 사용했다. 필요한 단계는 다음과 같다:

[0245] A. DNA 주형의 설계. 소정 목적 RNA 서열의 경우, 역 상보체의 5' 말단에 T7 DNA pol 프로모터 서열 TAATACGACTCACTATAG(서열번호 24)을 포함하는 이의 DNA 역 상보체를 계산했다.

[0246] B. 프록시 ds 폴리뉴클레오타이드의 합성. 프록시 ds DNA 폴리뉴클레오타이드는 실시예 2에 기재된 바와 같이 단계 3.2.A의 DNA 주형에 따라 합성했다(또한 실시예 1 및 3.1 참조). 프록시 DNA의 합성 후, 이의 말단을 변형시켜 평활 말단을 생성했다. ss 돌출부는, 33 μM의 각 dNTP의 존재하에, 이. 콜라이 폴리머라제 I 거대 클레노브 단편의 μg당 1 유닛을 25°C에서 15분 동안 인큐베이션하여 평활화시키고, EDTA 10 mM을 첨가하고 75°C에서 20분 동안 가열하여 불활성화시켰다(New England Biolabs로부터 수득함, product nr. M0210; Sambrook and Russell, 2014; Ch. 12). 이어서, 프록시 ds 폴리뉴클레오타이드를 정제하고, 증폭시키고 다시 정제했다: 하기 기재된 RNA 합성 반응에는 1 μg DNA의 최소량이 필요하다.

[0247] C. RNA의 전사, 후처리 및 정제. 프록시 DNA로부터의 RNA의 합성을 포함하는, RNA 전사의 표준 프로토콜(여러

가지 중에서, 예를 들면, HiScribe T7 ARCA mRNA 키트, New England Biolabs, product nr. E2060)을 따랐다. 프록시 DNA로부터 RNA의 합성에는 하기 프로토콜이 적용되었다:

- [0248] 1 내지 3 μg 의 DNA를, 2 μL 의 $2\times$ rNTP Mix, 2 μL 의 T7 RNA 폴리머라제 Mix, 및 18 μL 의 뉴클레아제 비함유 물로 구성된 용액에 용해시키고, 37°C에서 30분 동안 인큐베이팅하여, RNA 분자를 생성했다. 2 μL 의 DNase를 첨가하고 37°C에서 15분간 인큐베이팅하여 주형 DNA를 소화시킴으로써 반응을 중지시키고, 이전 실시예에 기재된 바와 같이, 스핀 컬럼을 사용하여 수득되는 RNA를 정제했다.
- [0249] 실시예 4: 608 bp의 표적 DNA 분자의 합성
- [0250] 본 실시예에서는, 본원에 제공된 방법을 사용하여, 608 bp의 표적 ds 폴리뉴클레오타이드(SOI는 서열 "Ribbon_test_608"임, 서열번호 26)을 합성하는 방법을 입증한다. 올리고는 실시예 1에서 생성된 라이브러리의 일부이다. 올리고는 실시예 2와 동일한 특성을 갖는다.
- [0251] 올리고는, 제4 결찰에서 상이한 크기의 부분적 작제물을 수득하기 위해, 반응 플레이트에서 비대칭적 방법으로 제조했다. 608 bp 서열은, 4개 결찰 라운드를 완료하여 128 bp의 하나의 반응 생성물 및 160 bp의 3개 생성물을 수득하고, 이어서 정제하고 2회 이상의 결찰 라운드에 제공하고, 이에 의해 608 bp 표적 ds 폴리뉴클레오타이드의 각 쇄를 수득한다.
- [0252] **4.1 어닐링 용액의 제조**
- [0253] 772 μL 의 ddH₂O 및 92 μL 의 T4 리가제 완충액으로 구성된 864 μL 의 어닐링 용액의 마스터 혼합물을 준비했다. 이 용액 혼합물의 21.6 μL 를 38 마이크로-웰에 분주했다. 0.7 μL 의 각 올리고(150 μM)를 플레이트의 소정 마이크로-웰로 전달하고, 피펫팅에 의해 혼합했다.
- [0254] 부분적 상보성 ss 올리고는, 실시예 1의 라이브러리로부터 유래하고, 도 6에 제시된 바와 같이 96-웰 플레이트의 특정 웰에 배치했다. 단순화하기 위해, 올리고는, 어닐링을 위해 배치되는 플레이트 상의 위치에 따라 명명했다. 실시예 2와 같이, 리딩 및 래깅 쇄는 각각 + 및 - 위치자로 표시되고; 서열번호 27의 서열을 FASTA 포맷으로 참조한다. 열 E 내지 G, 컬럼 2 내지 7의 웰은 의도적으로 비어 두었음에 주의한다.
- [0255] **4.2 어닐링**
- [0256] 어닐링은 실시예 2와 같이 수행했다.
- [0257] **4.3 결찰 용액의 준비**
- [0258] 결찰 용액은, 실시예 2와 유사하게 준비했지만, 38개 반응 웰에 충분한 양을 80 μL 로 조정했다. 즉, 7.2 μL 의 뉴클레아제 비함유 ddH₂O, 8 μL 의 리가제 완충액, 40 μL 의 ATP 및 와동 혼합후, 24.8 μL 의 T4 리가제를 피펫팅에 의해 혼합했다.
- [0259] 수득되는 용액 2 μL 를 분배기에서 B의 38개 반응 웰의 각각에 전달하여 결찰을 준비하고, 이어서 다중채널 피펫을 사용하여 온화하게 혼합했다.
- [0260] **4.4 제1 4회 결찰 라운드**
- [0261] 결찰의 제1 라운드의 경우, 각각 열 A 및 C의 웰로부터 컬럼(1-7)의 열 B 및 D로, 및 각각 웰 E1 및 G1로부터 F1 및 H1로 완전한 내용물을 전달했다. 전달은 다중채널 피펫으로 수행하고, 이어서 온화한 혼합을 수행했다. 이 도식은 실시예 2와 동일하다: 가장 좌측 내용물이 가장 우측 웰로 전달된다. 플레이트를 밀봉하고, 반응 혼합물을 적어도 1시간 동안 16°C에서 써모사이클러에서 인큐베이팅한다. 열 2 내지 7로부터 웰 E 내지 G는 의도적으로 비워둠에 주의한다.
- [0262] 결찰의 제2 라운드의 경우, 플레이트를 개방하고, 열 B의 웰로부터 컬럼(1-7)의 열 D로, 및 웰 F1으로부터 H1으로 완전한 내용물을 전달하고, 혼합했다. 플레이트를 다시 밀봉하고, 적어도 1시간 동안 16°C에서 인큐베이팅한다.
- [0263] 결찰의 제3 라운드의 경우, 플레이트를 개방하고, 피펫팅에 의해 열 D의 웰로부터 컬럼 1-7의 열 H로 피펫팅함으로써 완전한 내용물을 전달하고, 이어서 혼합했다. 플레이트를 다시 밀봉하고, 적어도 1시간 동안 16°C에서 인큐베이팅한다.
- [0264] 결찰의 제4 라운드의 경우, 플레이트를 개방하고, 각각 웰 H2, H4 및 H6으로부터 웰 H3, H5 및 H7로 피펫팅하여 완전한 내용물을 전달하고, 이어서 혼합한다. 웰 H1은 그 상태로 잔류함에 주의한다. 플레이트를 다시 밀봉하

고, 적어도 1시간 동안 16°C에서 인큐베이팅한다.

[0265] **4.5 중간 정제**

[0266] 3개 아가로스 겔은 실시예 2, 파트 E와 동일하게, 50 bp 래더를 포함하는 7 레인의 콤프로 제조했다. 파트 D의 웰 H1 중의 내용물을 겔의 6 레인으로 분배했다(각 레인에 33 μ L). 내용물 H3, H5 및 H7은, 다른 2개 웰 각각의 3개 레인에 분배했다(각 레인에 41 μ L). 실시예 2, 파트 E에 지시된 바와 같이 겔을 실행하고, 이어서 필요에 따라 밴드를 절제했다(겔 1의 레인 2 내지 4에 대해 128, 및 겔 1 및 겔 2의 나머지 레인에 대해 80 bp). 정제는 실시예 2, 파트 E에 기재된 바와 같이 수행하고, 동일한 신포를 함유하는 동일한 정제 컬럼 샘플을 풀링했다. 4개 샘플은 각각 10 μ L의 ddH2O로 용출시키고(Zymoclean 정제 키트에 제시된 바와 같음), 35°C로 가온하여 용출 효율을 개선시켰다. 내용물을 PCR 반응 튜브의 스트립으로 전달하고, S1 내지 S4로 표지했다.

[0267] S1 및 S4로부터 0.5 μ L 형태의 샘플을 취하고, 0.5 μ L의 ddH2O에서 희석시켰다. 이들 샘플을 사용하여, 260 nm에서 분광광도법(nanodrop 2000, Thermo Fisher Scientific)에 의해 DNA 농도를 평가하여, 각각 1.52 μ g/ μ L 및 1.98 μ g/ μ L를 수득했다. 샘플 S2 및 S3은 유사한 범위의 분자 농도에 존재하는 것으로 가정했다.

[0268] **4.6 결찰 용액의 준비**

[0269] 샘플을 아이스에 위치시켰다. 샘플 S1 및 S4에 0.5 μ L의 ddH2O를 첨가했다(파트 E에서 측정을 위해 취한 0.5 μ L를 보정하기 위해). 결찰 반응물은 각 샘플에 1.14 μ L의 리가제 완충액을 첨가하여 제조했다. 0.3 μ L의 T4 리가제를 S1 및 S3에 첨가했다. 용액을 피펫팅에 의해 혼합했다.

[0270] **4.7 결찰의 최종 2회 라운드**

[0271] 제5 결찰 반응의 경우, 각각 튜브 1 및 3으로부터 튜브 2 및 4로 피펫팅하여 완전한 내용물을 전달하고, 이어서 혼합했다. 튜브를 밀폐했다. 반응물을 씨모사이클러에서 16°C로 80분 동안 인큐베이팅했다.

[0272] 결찰 반응의 최종 라운드의 경우, 튜브 2로부터 튜브 4로 피펫팅하여 완전한 내용물을 전달하고, 이어서 혼합했다. 튜브를 밀폐했다. 반응물을 씨모사이클러에서 16°C로 80분 동안 인큐베이팅했다. 이는 계층적 합성 프로세스를 완성했다.

[0273] **4.8 최종 정제**

[0274] 8 레인의 콤프로 사용하여 2% 아가로스 겔로부터 정제를 수행했다. 제1 레인은 실시예 2의 파트 E와 같이 50 bp 래더를 함유했다. 완전한 샘플을 SDS의 부재하에 10 μ L의 자색 로딩과 혼합하고, 단일 레인에 분배했다. 겔을 100 V, 200 mA, 12 watt에서 45분 동안 실행했다. 도 7은 수득되는 결과를 나타낸다. 608 bp의 예상된 크기에 상응하는 상부 밴드를 절제하고, 20 μ L의 ddH2O 물을 사용하여, 실시예 2, 파트 E와 같이 Zymo 겔 추출 키트로 정제하고, 35°C로 가온했다. 이 샘플의 경우, 0.5 μ L를 사용하면, 용액이 10 ng/ μ L를 함유하는 것이 분광광도적으로 평가되었다.

[0275] **4.9 서열분석**

[0276] 용액을 2개 샘플, 약 10 μ L의 하나 및 9.5 μ L의 하나로 분할했다. 각각에 대해, 프라이머("Primer1" 및 "Primer2")를 용액에 첨가하고, 생거 방법으로 서열분석했다. 중앙의 신뢰성 영역의 서열분석 결과는 SOI와 표적 ds 폴리뉴클레오타이드의 완전한 서열 동일성을 확인했다.

[0277] 실시예 5: 10,000 bp의 DNA 분자의 합성

[0278] 본 실시예에서, 10,000 bp의 목적 서열로 이루어진 ds 폴리뉴클레오타이드의 작제는, 4 뉴클레오타이드 돌출부를 갖는 ds 이량체를 형성하는 26 bp의 올리고를 사용함으로써 실시예 1의 라이브러리 설계에 기초하여 입증된다.

[0279] **5.1. 서열 처리**

[0280] A. 목적 서열의 리딩 쇠의 역 상보체를 계산하고, 양쪽 서열(리딩 쇠 및 역 상보체)에서 3' 말단의 최종 4 뉴클레오타이드를 제거한다. 이는, 하나가 SOI의 리딩 쇠에 상응하고 다른 하나가 SOI의 역 상보체에 상응하고 3' 말단의 4 뉴클레오타이드를 제외한 2개 일본쇄 주형 서열을 생성한다.

[0281] B. 양 ss 주형의 서열을 정렬하여 이본쇄 주형 서열을 생성하고, 이어서 라이브러리에 함유된 올리고에서 발생 하는, 올리고 서브세트 또는 서브-서열로서 지칭되는 보다 짧은 서열로 분할하고, 라이브러리 내의 이들의 위치

는 디지털로 주석을 단다.

- [0282] C. 단계 B에서 결정된 서브-서열의 명확한 어셈블리를 가능하게 하는 워크플로우를 결정한다.
- [0283] **5.2. 반응**
- [0284] 하기의 모든 단계는, 달리 언급하지 않는 한, 16°C에서 수행하고, 모든 용액을 준비하고, 아이스에서 보관한다.
- [0285] A. ddH₂O 중의 2× 리가제 완충액의 용액 700 μL를 준비하고, 마스터 혼합 용액 1.8 μL를 384 마이크로웰-플레이트의 각 웰에 분주한다.
- [0286] B. 3' 말단에서 4 뉴클레오타이드를 뺀 SOI의 리딩 쇠인 ss 주형 서열의 단계 B의 4.1에서 결정된 서브-서열에 상응하는 올리고뉴클레오타이드 라이브러리 멤버 0.1 μL 각각을 표적 서열에서의 발생 순서로 라이브러리로부터 추출하고, 348 마이크로웰 플레이트의 마이크로-웰에 분주하고, 모든 올리고가 웰에 분주될 때까지, 웰 A1, B1, ..., P1에서 개시한 다음 후속 컬럼 A2, B2 등으로 진행한다.
- [0287] C. 3' 말단에서 4 뉴클레오타이드를 뺀 SOI의 역 상보체인 ss 주형 서열의 단계 B의 4.1에서 결정된 서브-서열에 상응하는 올리고뉴클레오타이드 라이브러리 멤버 0.1 μL 각각을 역 서열 순서로 추출하고, 단계 B의 마이크로-웰 플레이트에 분주하고, 모든 올리고가 웰에 분주될 때까지 웰 A1에서 다시 개시한다. 이 시점에서, 각 마이크로-웰은, 4 뉴클레오타이드로 구성된 22 상보성 bs 및 돌출부를 갖는 2개 올리고를 함유한다. 이와 함께, 웰은 이제 올리고뉴클레오타이드 라이브러리 멤버의 정합 쌍을 함유해야 한다.
- [0288] D. 마이크로-웰 플레이트를 밀봉하고, 95°C에서 개시하는 써모사이클러에서 어닐링하고, 1°C/분의 램프 속도에서 16°C까지 저하시킨다.
- [0289] E. ddH₂O 중의 μL당 20 응집 단위의 농도로 T4 리가제를 포함하는 마스터 혼합물 결찰 반응 용액 800 μL를 제조하고, 이 용액 2 μL를 플레이트의 각 384 웰에 분주한다.
- [0290] F. 플레이트를 원심분리기에서 1000g 펄스로 스핀 다운한다.
- [0291] G. 용액을 함유하는 열은 하기 식을 사용하여 열거한다: $2^{t-1}k$ (여기서, t는 층 번호이고, t=1,2,3,4이고, k는 충전된 웰에서 열의 인덱스이고, r=k=1, ..., 16/2^{t-1}이다). 이러한 방식으로, 제1 층에서는 모든 열을 열거하고, 제2 층에서는 절반만 열거한다.
- [0292] H. 홀수 인덱스의 각 열의 웰의 내용물은, 다중-채널 마이크로피펫 또는 액체 핸들러를 사용하여, 짝수 인덱스의 가장 우측 컬럼의 웰에 전달한다.
- [0293] I. 내용물을 전달한 직후, 용액은 마이크로피펫 또는 핸들러로 직접 피펫팅하여 온화하게 혼합한다.
- [0294] J. 반응을 60분 동안 인큐베이팅하여 결찰 반응을 완성시킨다.
- [0295] K. 마이크로-웰 플레이트의 최종 열(P)만이 충전될 때까지 단계 G-J를 추가로 4회 반복하고, 총 24개 나머지 충전된 웰을 생성한다.
- [0296] L. 24 웰(48 μL 함유)의 각각의 내용물을 24 반응 튜브로 전달하고, 뉴 잉글랜드 바이오랩(product nr. T1030)으로부터 Monarch PCR & DNA 클린 업 키트에 따라 컬럼에서 정제하기 위해 준비하여, 100 bp를 초과하는 중간 반응 생성물만을 함유하는 정제된 용액 6 μL를 수득한다.
- [0297] M. 정제된 용액을 8 PCR 튜브의 3개 신선한 스트립으로 전달하고, 8 열 × 3 컬럼 방법으로 배열한다.
- [0298] N. 단계 E에서 17.5 μL의 용액을 취하고, 7.5 μL의 리가제 완충액(10×)을 가하여 7×의 최종 농도로 하고, 1 μL의 이 용액을 각 튜브에 분주한다.
- [0299] O. 반응은 단계 H-J와 동일한 방식으로 추가로 3회 진행시켜, 3개의 충전된 튜브를 수득한다(최후 열의 각 컬럼 상에 1개).
- [0300] P. 컬럼 1의 내용물을 컬럼 2로 전달하고, 컬럼 3은 그대로 유지한다.
- [0301] Q. 반응물을 1시간 동안 인큐베이팅한다.
- [0302] R. 컬럼 2의 내용물을 컬럼 3으로 전달한다.
- [0303] S. 반응물을 1시간 동안 인큐베이팅한다.

- [0304] T. 0.8% 아가로스 겔을 준비하고, 샘플을 10 kbp 래더와 함께 로딩한다. 겔은 100 V에서 45분 동안 실행한다.
- [0305] U. 10K bp에 상응하는 밴드를 추출하고, 표준 프로토콜 및 키트를 사용하여 겔 블록으로부터 정제한다(Zymo 클린이 본 실시예에서 권장된다, 또한 실시예 1 참조).
- [0306] **5.3. 최종화 및 증폭**
- [0307] A. 라이브러리로부터 2개의 26 bp 길이의 올리고를 선택하고, 이는 SOI의 3' 말단에서 최후 26 뉴클레오타이드에 상보성이고, 즉 이들은 또한 포인트 4.1의 단계 A에서 결실된 4 뉴클레오타이드를 포함한다. 이들 2개 올리고는, 최종 생성물을 증폭시키고 나머지 4 bp를 각 쇄에 부가하여 평할 말단을 갖는 10,000 bp 서열을 완성시키기 위해 준비되는 PCR 반응에서 프라이머로서 사용된다.
- [0308] B. 포인트 4.2의 단계 L과 같이 PCR 생성물을 표준 키트로 정제하여, 나머지 올리고, 효소 및 시약을 제거하고, 최종 DNA 생성물, 즉 하류 적용을 위해 준비된, SOI와 동일한 서열을 갖는 ds 폴리뉴클레오타이드를 잔류시킨다.
- [0309] 참조문헌
- [0310] Anderson, S., Bankier, A.T., Barrell, B.G. et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature*, 290:457-465.
- [0311] Beaucage, S.L. and Caruthers, M.H. (1981) Deoxynucleoside phosphoramidites—a new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahedron Letters* 22:1859-1862.
- [0312] Bentley, D.R., et al. (65 authors) (2008) Accurate Whole Human Genome Sequencing using Reversible Terminator Chemistry. *Nature*, 456:53-59.
- [0313] 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.
- [0314] Chari, R. and Church, G.M. (2017) Beyond editing to writing large genomes. *Nature Reviews Genetics*, In Press.
- [0315] 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.
- [0316] 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.
- [0317] 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.
- [0318] 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.
- [0319] 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.
- [0320] 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
- [0321] Kemp, G. (1998) Capillary electrophoresis: a versatile family of analytical techniques. *Biotechnology and Applied Biochemistry* 27:9-17.
- [0322] 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.
- [0323] 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

도면2

A)

> DISCOVER 서열

서열번호 1

TTTCTTCTTTTCGAGTTTCTATATCCGTCGCTGGTCTGAACGGAAAAATCATCGACAATCTATGCCTACCGTTGGC
TGCTCATGCGTCCITOCGACAATCCCATGTTCTGCTCGCATCCGTTTCTCTGC

B)

> D+

서열번호 2

TTTCTTCTTTTCGAGTT

> D-

서열번호 3

TAGAAACTCGAAAGAA

> I+

서열번호 4

TCTATATCCGTCGCTG

> I-

서열번호 5

AGACCAGCGACGGATA

> S+

서열번호 6

GTCTGAACGGAAAAAT

> S-

서열번호 7

GATGATTTTTCGTTTC

> C+

서열번호 8

CATCGCACAACTATG

> C-

서열번호 9

TAGGCATAGATTGTGC

> O+

C)

> D+

서열번호 2

TTTCTTCTTTTCGAGTT----

> D-

서열번호 25

----AAGAAAGCTCAAGAT

서열번호 10

CCTACCGTTGGCTGCT

> O-

서열번호 11

CATGAGCAGCCAACGG

> V+

서열번호 12

CATGCGTCCITTCGAC

> V-

서열번호 13

GATTGTGGGAAGGACG

> E+

서열번호 14

AATCCCATGTTCTGCT

> E-

서열번호 15

TGCGAGACGAACATGG

> R+

서열번호 16

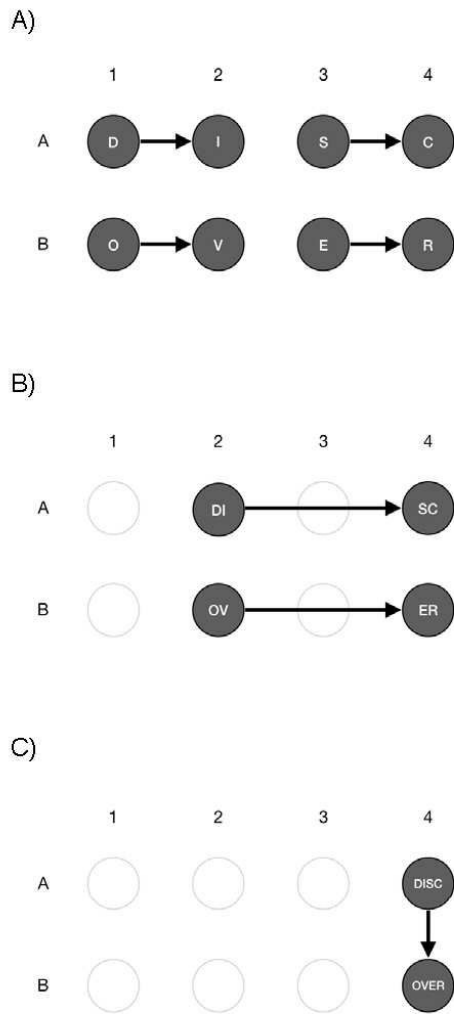
CGCATCCGTTTCTCTGC

> R-

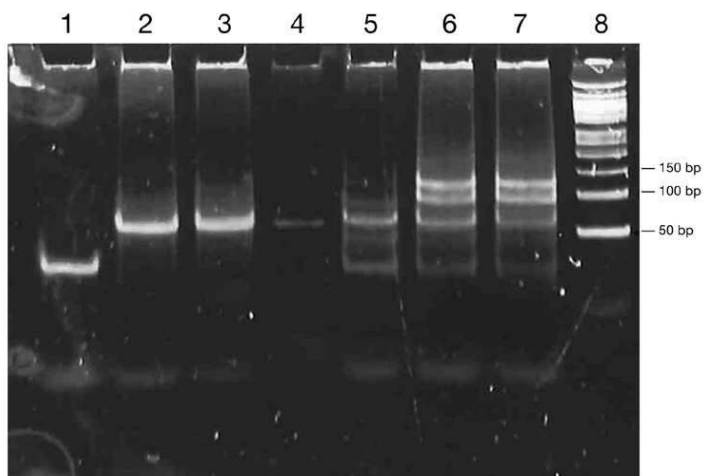
서열번호 17

GCGTGCAGGAAACGGA

도면3



도면4



도면5

A)
 CCTACCGTTGGCTGCTAATCCGTCCTTCCGACAATC...
 ...GGCAACCGACGATTAGGCAGGAAGGCTGTTAG...

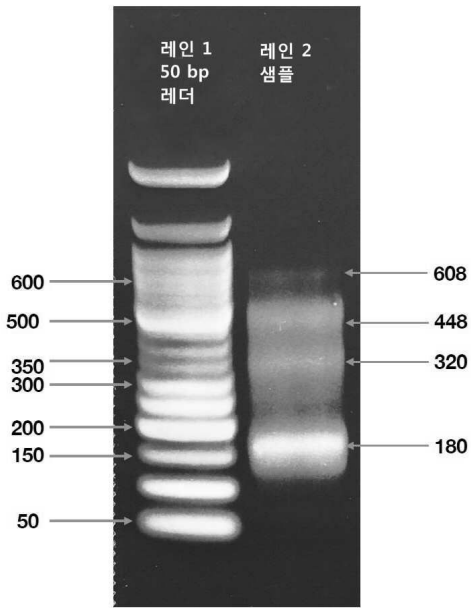
B)
 CCTACCGTTGGCTGCTCATGCGTCCTTCCGACAATC...
 ...GGCAACCGACGAGTACGCAGGAAGGCTGTTAG...

C)
 GTTGGCTGCTAATCCGTCCTTCCG
 CGGAAGGACGCATGAGCAGCCAAC

도면6

| | A | B | C | D | E | F | G | H |
|----|------------|------------|------------|------------|------------|------------|------------|------------|
| 1 | A1+ A1- | B1+ B1- | C1+ C1- | D1+ D1- | E1+ E1- | F1+ F1- | G1+ G1- | H1+ H1- |
| 2 | A2+ A2- | B2+ B2- | C2+ C2- | D2+ D2- | X | X | X | H2+ H2- |
| 3 | A3+ A3- | B3+ B3- | C3+ C3- | D3+ D3- | X | X | X | H3+ H3- |
| 4 | A4+ A4- | B4+ B4- | C4+ C4- | D4+ D4- | X | X | X | H4+ H4- |
| 5 | A5+ A5- | B5+ B5- | C5+ C5- | D5+ D5- | X | X | X | H5+ H5- |
| 6 | A6+ A6- | B6+ B6- | C6+ C6- | D6+ D6- | X | X | X | H6+ H6- |
| 7 | A7+ A7- | B7+ B7- | C7+ C7- | D7+ D7- | X | X | X | H7+ H7- |
| 8 | X | X | X | X | X | X | X | X |
| 9 | X | X | X | X | X | X | X | X |
| 10 | X | X | X | X | X | X | X | X |
| 11 | X | X | X | X | X | X | X | X |
| 12 | X | X | X | X | X | X | X | X |

도면7



도면8a

> Ribbon_test_608 (서열번호 26)

```

TTTCTTCTTCGAGTTTCTATATCCGTCGCTGGTCTGAACGGAAAAATCATCGCACAACTATGCCTACC
GTTGGCTGCTCATGCGTCCTCCGACAATCCCATGTTTCGTCTCGCATCCGTTTCCCTGCACGCACCCCCC
CTGTACTTTGGAAAGCGGCCATCTTAACACTCTCCCACTTTTAAATGCGTCGAAGCCCTGGGCATCTG
GTTTCCACTAGCCTAGTCGGGTGTTGGATACGCCGAGAGTTATGGTGTAGCTGTTGCGCGAACCACGG
GTGGAAATGCTGACCGATTTCAAAATAGTTCTCAGGAAGCCGATGGCAGTTACGGCTTGCGACTCGGGG
CACCGTGAGCCTCTTCTCCCTCTAGAAGTCGAAGCAAGGGACACTATCCTAAATGCCATGGACTAGCGCG
CGCGAAATCGATGCACTCCTTATTAATGTGATCTGCGCAAGTGTTCAGCCATCGGTCATTTTGC GTTGA
TATTCGGTTCTTTGATTTGCGTGCCATGCTTATAACAGGACACTTATTGTGCCCCAGCTTTCTCATGCA
AGTGGGTTTTTCTCTAGCTACTGTGGTGTGCGCTCATCAATACTCCAG
    
```

도면 8b

| | |
|-------------------|------------------|
| > A1+ (서열번호 27) | ACGCACCCCCCTGT |
| TTTCTTCTTTTCGAGTT | |
| > A1- (서열번호 28) | > A2- (서열번호 44) |
| TAGAAACTCGAAAGAA | AAGTACAGGGGGGGT |
| > B1+ (서열번호 29) | > B2+ (서열번호 45) |
| TCTATATCCGTCGCTG | ACTTTGGAAAGCGGCC |
| > B1- (서열번호 30) | > B2- (서열번호 46) |
| AGACCAGCGACGGATA | AGATGGCCGCTTTCCA |
| > C1+ (서열번호 31) | > C2+ (서열번호 47) |
| GTCTGAACGGAAAAAT | ATCTTAACACTCTCCC |
| > C1- (서열번호 32) | > C2- (서열번호 48) |
| GATGATTTTTCCGTTT | AGTTGGGAGAGTGTTA |
| > D1+ (서열번호 33) | > D2+ (서열번호 49) |
| CATCGACAATCTATG | AACTTTTAAATGCGT |
| > D1- (서열번호 34) | > D2- (서열번호 50) |
| TAGGCATAGATTGTGC | ITCGACGCATTTAAAA |
| > E1+ (서열번호 35) | > H2+ (서열번호 51) |
| CCTACCGTTGGCTGCT | CGAAGCCCTGGGCATC |
| > E1- (서열번호 36) | > H2- (서열번호 52) |
| CATGAGCAGCCACCG | ACCRAGATGCCAGGGC |
| > F1+ (서열번호 37) | > A3+ (서열번호 53) |
| CATGCGTCCTTCGAC | TGGTTTCCACTAGCCT |
| > F1- (서열번호 38) | > A3- (서열번호 54) |
| GATTGTCGGAAGGACG | GACTAGGCTAGTGGA |
| > G1+ (서열번호 39) | > B3+ (서열번호 55) |
| AATCCCATGTTTCGTCT | AGTCGGGTGTTGGATA |
| > G1- (서열번호 40) | > B3- (서열번호 56) |
| TGOGAGACGAACATGG | GGCGTATCCAACACC |
| > H1+ (서열번호 41) | > C3+ (서열번호 57) |
| CGCATCCGTTTCCTGC | CGCGAGAGTTATGGT |
| > H1- (서열번호 42) | > C3- (서열번호 58) |
| CGGTGCAGGAAACGGA | CTACACCATAACTCTC |
| > A2+ (서열번호 43) | > D3+ (서열번호 59) |
| | GTAGCTGTGTCCGCGA |

도면8c

- > D3- (서열번호 60)
CGGTTCGCGCACACAG

- > H3+ (서열번호 61)
ACCGACGGGTGGAAAT
- > H3- (서열번호 62)
AGCAATTTCCACCCGT
- > A4+ (서열번호 63)
TGCTGACCGATTTCAC
- > A4- (서열번호 64)
TATTTGAAAATCGGTC
- > B4+ (서열번호 65)
AATAGTTCTCAGGAAG
- > B4- (서열번호 66)
TCGGCTTCCTGAGAAC
- > C4+ (서열번호 67)
CCGATGGCAGTTACGG
- > C4- (서열번호 68)
CAAGCCGTAAGTCCCA
- > D4+ (서열번호 69)
CTTGCGACTCGGGGCA
- > D4- (서열번호 70)
ACGGTGCCCCGAGTCG
- > H4+ (서열번호 71)
CCGTGAGCCTCTTCTC
- > H4- (서열번호 72)
GAGGGAGAAGAGGCTC
- > A5+ (서열번호 73)
CCTCTAGAAGTCGAAG
- > A5- (서열번호 74)
CTTGCTTCGACTTCTA
- > B5+ (서열번호 75)
CAAGGGACACTATCCT

- > B5- (서열번호 76)
ATTTAGGATAGTGTCC
- > C5+ (서열번호 77)
AAATGCCATGGACTAG
- > C5- (서열번호 78)
CGCGCTAGTCCATGGC

- > D5+ (서열번호 79)
CGCGCGCGAAAATCGAT
- > D5- (서열번호 80)
GTGCATCGATTTTCGCG
- > H5+ (서열번호 81)
GCACTCCTTATTAATG
- > H5- (서열번호 82)
ATCACATTAATAAGGA
- > A6+ (서열번호 83)
TGATCTGCGCAAGTTG
- > A6- (서열번호 84)
TGAACAACCTTGCGCAG
- > B6+ (서열번호 85)
TTCAGCCATCGGTCAT
- > B6- (서열번호 86)
CAAAATGACCGATGGC
- > C6+ (서열번호 87)
TTTGCGTTGATATTCG
- > C6- (서열번호 88)
GAACCGAATATCAACG
- > D6+ (서열번호 89)
GTTCCTTGATTTGCGT
- > D6- (서열번호 90)
TGGCACGCAAAATCAAA
- > H6+ (서열번호 91)
GCCATGCTTATAACAG
- > H6- (서열번호 92)

도면8d

```

TGTCCTGTTATAAGCA
> A7+ (서열번호 93)
GACACTTATTGTGCCC
> A7- (서열번호 94)
GCTGGGGCACAATAAG
> B7+ (서열번호 95)
CAGCTCTTCTCATGCA
> B7- (서열번호 96)
CACTTGCATGAGAAGA
> C7+ (서열번호 97)
AGTGCGGTTTTCTCTA

> C7- (서열번호 98)
TAGCTAGAGAAAACCG
> D7+ (서열번호 99)
GCTACTGTGGTGTGCG
> D7- (서열번호 100)
TGAGCGCACACCACAG
> H7+ (서열번호 101)
CTCATCAATACTCCAG
> H7- (서열번호 102)
TTATCTGGAGTATTGA

> Primer1 (서열번호 103)
TTTCTTCTTTTCGAGTTTCTATATCCGTCGCTG
> Primer2 (서열번호 104)
TTATCTGGAGTATTGATGAGCGCAC

```

서열 목록

SEQUENCE LISTING

```

<110> Ribbon Biolabs GmbH
<120> A NOVEL METHOD FOR SYNTHESIS OF POLYNUCLEOTIDES USING A DIVERSE LIBRARY OF OLIGONUCLEOTIDES
<130> RB001P
<160> 104
<170> BiSSAP 1.3
<210> 1
<211> 128
<212> DNA
<213> Artificial Sequence
<220>
<223> Polynucleotide
<400> 1
tttctttttt cgagtttcta tatccgtcgc tggcttgaac ggaaaaatca tcgcacaatc    60
tatgcctacc gttggctgct catgcgtcct tccgacaate ccatgttcgt ctgcgatcgg    120
tttctctgc                                     128

<210> 2
<211> 16

```

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 2
 tttcttcttt cgagtt 16
 <210> 3
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 3
 tagaaactcg aaagaa 16
 <210> 4
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 4
 tctatatccg tcgctg 16

 <210> 5
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 5
 agaccagcga cggata 16
 <210> 6
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Oligonucleotide
 <400> 6
 gtctgaacgg aaaaat 16
 <210> 7
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 7
 gatgatTTTT cgttc 16
 <210> 8
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 8
 catgcacaa tctatg 16
 <210> 9
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 9
 taggcataga ttgtgc 16
 <210> 10
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide

<400> 10
 cctaccgttg gctgct 16

<210> 11
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide

<400> 11
 catgagcagc caacgg 16

<210> 12
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide

<400> 12
 catgctcct tccgac 16

<210> 13
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide

<400> 13
 gattgtcgga aggacg 16

<210> 14
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide

<400> 14

aatcccatgt tcgtct 16

<210> 15

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide

<400> 15

tgcgagacga acatgg 16

<210> 16

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide

<400> 16

cgcacccggtt tcctgc 16

<210> 17

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide

<400> 17

gcgtgcagga aacgga 16

<210> 18

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Polynucleotide

<400> 18

cctaccgttg gctgctaac cgtccttccg acaatc 36

<210> 19

<211> 32
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Polynucleotide
 <400> 19
 ggcaaccgac gattaggcag gaaggctggt ag 32

<210> 20
 <211> 36
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Polynucleotide
 <400> 20
 cctaccgttg gctgctcatg cgtccttccg acaatc 36

<210> 21
 <211> 32
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Polynucleotide
 <400> 21
 ggcaaccgac gactacgcag gaaggctggt ag 32

<210> 22
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Polynucleotide
 <400> 22
 gttggtgct aatccgtcct tccg 24

<210> 23
 <211> 24

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> Polynucleotide
 <400> 23
 cggaaggacg catgagcagc caac 24
 <210> 24
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Polynucleotide
 <400> 24
 taatacgact cactatag 18
 <210> 25
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 25
 aagaaagctc aaagat 16

 <210> 26
 <211> 608
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Polynucleotide
 <400> 26
 tttcttcttt cgagtttcta tatccgtcgc tggctgaac ggaaaaatca tgcacaatc 60
 tatgcctacc gttggctgct catgcgtcct tccgacaate ccatgttcgt ctgcgatccg 120
 tttctgcac gcaccccccc ctgtactttg gaaagcggcc atcttaacac tctccaact 180
 ttttaaatgc gtcgaagccc tgggcatctg gtttcacta gcctagtcgg gtgttgata 240
 cgccgagagt tatggtgtag ctgtgtgcgc gaaccgacgg gtggaattg ctgaccgatt 300

ttcaaatagt ttcaggaag ccgatggcag ttacggcttg cgactcgggg caccgtgagc 360

ctcttctccc tctagaagtc gaagcaaggg aactatcct aaatgccatg gactagcgcg 420

cgcgaaatcg atgcactcct tattaatgtg atctgcgcaa gttgttcagc catcggtcat 480

tttgcgttga tattcggttc tttgatttgc gtgccatgct tataacagga cacttattgt 540

gccccagctc ttctcatgca agtgcggttt tctctageta ctgtggtgtg cgctcatcaa 600

tactccag 608

<210> 27

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide

<400> 27

tttcttcttt cgagtt 16

<210> 28

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide

<400> 28

tagaaactcg aaagaa 16

<210> 29

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> Oligonucleotide

<400> 29

tctatatccg tcgctg 16

<210> 30

<211> 16

<212> DNA

<213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 30
 agaccagcga cggata 16

<210> 31
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 31
 gtctgaacgg aaaaat 16

<210> 32
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 32
 gatgattttt ccgttc 16

<210> 33
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 33
 catcgcaaa tctatg 16

<210> 34
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Oligonucleotide
 <400> 34
 taggcataga ttgtgc 16
 <210> 35
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 35
 cctaccgttg gctgct 16
 <210> 36
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 36
 catgagcagc caacgg 16
 <210> 37
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 37
 catgctcct tccgac 16
 <210> 38
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide

<400> 38
gattgtcgga aggacg 16

<210> 39
<211> 16
<212> DNA
<213> Artificial Sequence
<220>
<223> Oligonucleotide

<400> 39
aatcccatgt tcgtct 16

<210> 40
<211> 16
<212> DNA
<213> Artificial Sequence
<220>
<223> Oligonucleotide

<400> 40
tgcgagacga acatgg 16

<210> 41
<211> 16
<212> DNA
<213> Artificial Sequence
<220>
<223> Oligonucleotide

<400> 41
cgcatccggtt tcctgc 16

<210> 42
<211> 16
<212> DNA
<213> Artificial Sequence
<220>
<223> Oligonucleotide

<400> 42

gcgtgcagga aacgga 16

<210> 43
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 43

acgcaccccc ccctgt 16

<210> 44
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 44

aagtacaggg ggggt 16

<210> 45
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 45

actttgaaa gcgcc 16

<210> 46
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 46

agatggccgc tttcca 16

<210> 47
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 47
 atcttaacac tctccc 16
 <210> 48
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 48
 agttgggaga gtgta 16

 <210> 49
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 49
 aactttttaa atgcgt 16
 <210> 50
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 50
 ttcgacgcat ttaaaa 16
 <210> 51
 <211> 16

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 51
 cgaagccctg ggcac 16

<210> 52
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 52
 accagatgcc cagggc 16

<210> 53
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 53
 tggtttcac tagcct 16

<210> 54
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 54
 gactaggcta gtggaa 16

<210> 55
 <211> 16
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 55
 agtcgggtgt tggata 16
 <210> 56
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 56
 ggcgtatcca acaccc 16
 <210> 57
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 57
 cgccgagagt tatggt 16

 <210> 58
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 58
 ctacaccata actctc 16
 <210> 59
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> Oligonucleotide
 <400> 59
 gtagctgtgt gcgcga 16
 <210> 60
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 60
 cggttcgcgc acacag 16

 <210> 61
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 61
 accgacgggt gaaat 16
 <210> 62
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 62
 agcaatttc accgt 16
 <210> 63
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 63

| | |
|---------------------------|----|
| tgctgaccga ttttca | 16 |
| <210> 64 | |
| <211> 16 | |
| <212> DNA | |
| <213> Artificial Sequence | |
| <220> | |
| <223> Oligonucleotide | |
| <400> 64 | |
| tatttgaaaa tcggtc | 16 |
| <210> 65 | |
| <211> 16 | |
| <212> DNA | |
| <213> Artificial Sequence | |
| <220> | |
| <223> Oligonucleotide | |
| <400> 65 | |
| aatagtcttc aggaag | 16 |
| <210> 66 | |
| <211> 16 | |
| <212> DNA | |
| <213> Artificial Sequence | |
| <220> | |
| <223> Oligonucleotide | |
| <400> 66 | |
| tcggcttctt gagaac | 16 |
| <210> 67 | |
| <211> 16 | |
| <212> DNA | |
| <213> Artificial Sequence | |
| <220> | |
| <223> Oligonucleotide | |
| <400> 67 | |
| ccgatggcag ttacgg | 16 |

<210> 68
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 68
 caagccgtaa ctgcca 16
 <210> 69
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 69
 cttgcgactc ggggca 16

 <210> 70
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 70
 acggtgcccc gagtcg 16
 <210> 71
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 71
 ccgtgagcct cttctc 16
 <210> 72
 <211> 16

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 72
 gagggagaag aggctc 16

<210> 73
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 73
 cctctagaag tcgaag 16

<210> 74
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 74
 cttgcttcga cttcta 16

<210> 75
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 75
 caaggacac tatcct 16

<210> 76
 <211> 16
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 76
 atttaggata ggtgcc 16
 <210> 77
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 77
 aaatgccatg gactag 16
 <210> 78
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 78
 cgcgctagtc catggc 16
 <210> 79
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 79
 cgcgcgcaa atcgat 16
 <210> 80
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> Oligonucleotide
 <400> 80
 gtgcatcgat ttcgcg 16
 <210> 81
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 81
 gcactcctta ttaatg 16

 <210> 82
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 82
 atcacattaa taagga 16
 <210> 83
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 83
 tgatctgcg aagttg 16
 <210> 84
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 84

tgaacaactt gcgcag 16

<210> 85
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 85

ttcagccatc ggtcat 16

<210> 86
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 86

caaatgacc gatggc 16

<210> 87
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 87

tttgcgttga tattcg 16

<210> 88
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 88

gaaccgaata tcaacg 16

<210> 89
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 89
 gttccttgat ttgcgt 16
 <210> 90
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 90
 tggcagcaa atcaaa 16
 <210> 91
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 91
 gccatgctta taacag 16
 <210> 92
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 92
 tgtcctgtta taagca 16
 <210> 93
 <211> 16

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 93
 gacacttatt gtgcc 16

<210> 94
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 94
 gctggggcac aataag 16

<210> 95
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 95
 cagctcttct catgca 16

<210> 96
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 96
 cacttgcag agaaga 16

<210> 97
 <211> 16
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 97
 agtgcggttt tctcta 16
 <210> 98
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 98
 tagctagaga aaaccg 16
 <210> 99
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 99
 gctactgtgg tgtgcg 16

 <210> 100
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 100
 tgagcgacaca ccacag 16
 <210> 101
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> Oligonucleotide
 <400> 101
 ctcatcaata ctccag 16
 <210> 102
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Oligonucleotide
 <400> 102
 ttatctggag tattga 16

 <210> 103
 <211> 32
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Primer
 <400> 103
 tttcttcttt cgagtttcta tatccgtcgc tg 32
 <210> 104
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Primer
 <400> 104
 ttatctggag tattgatgag cgcac 25