

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7586880号
(P7586880)

(45)発行日 令和6年11月19日(2024.11.19)

(24)登録日 令和6年11月11日(2024.11.11)

(51)国際特許分類

F I

C 4 0 B 40/06 (2006.01)

C 4 0 B 40/06

G 1 6 B 50/00 (2019.01)

G 1 6 B 50/00

C 1 2 N 15/09 (2006.01)

C 1 2 N 15/09

1 1 0

C 1 2 N 15/09

Z

請求項の数 27 (全62頁)

(21)出願番号	特願2022-182278(P2022-182278)	(73)特許権者	519173978
(22)出願日	令和4年11月15日(2022.11.15)		カタログ テクノロジーズ, インコーポ
(62)分割の表示	特願2019-547250(P2019-547250)		レイテッド
原出願日	平成29年11月16日(2017.11.16)		アメリカ合衆国 マサチューセッツ 0 2
(65)公開番号	特開2023-29836(P2023-29836A)		1 2 9, チャールズタウン, メイン
(43)公開日	令和5年3月7日(2023.3.7)	(74)代理人	ストリート 5 2 9, スイート 1 2 7
審査請求日	令和4年11月15日(2022.11.15)		100079108
(31)優先権主張番号	62/457,074	(74)代理人	弁理士 稲葉 良幸
(32)優先日	平成29年2月9日(2017.2.9)	(74)代理人	100109346
(33)優先権主張国・地域又は機関	米国(US)	(74)代理人	弁理士 大貫 敏史
(31)優先権主張番号	62/423,058	(74)代理人	100117189
(32)優先日	平成28年11月16日(2016.11.16)	(74)代理人	弁理士 江口 昭彦
(33)優先権主張国・地域又は機関		(74)代理人	100134120
	最終頁に続く	(72)発明者	弁理士 内藤 和彦
			ナサニエル ロケット
			最終頁に続く

(54)【発明の名称】 核酸ベースのデータ記憶

(57)【特許請求の範囲】

【請求項 1】

情報を核酸配列に書き込むための方法であって、前記方法は、

(a) 複数の構成要素を構築することであって、前記複数の構成要素の各個々の構成要素は、核酸配列を含む、ことと、

(b) 前記複数の構成要素のうちの2つ以上の構成要素と一緒に繰り返しライゲーションし、これにより、複数の識別子を生成することであって、前記複数の識別子の各識別子は、2つ以上の構成要素を含み、前記複数の識別子の個々の識別子は、個々のシンボルに対応し、前記シンボルは、前記情報の少なくとも一部に対応する、ことと、

(c) 前記複数の識別子の少なくともサブセットを含む識別子ライブラリを選択的にキャプチャまたは増幅することであって、前記複数の識別子の前記サブセットの各識別子は、前記識別子ライブラリを識別するバーコードを含む、ことと

を含む、方法。

【請求項 2】

前記個々のシンボルは、1つ以上の可能なシンボル値のうちの1つである、請求項 1 に記載の方法。

【請求項 3】

前記個々のシンボルは、2つの可能なシンボル値のうちの1つである、請求項 1 に記載の方法。

【請求項 4】

1つのシンボル値は、前記識別子ライブラリ内の別個の識別子の不在によって表される、請求項1に記載の方法。

【請求項5】

前記2つの可能なシンボル値は、0および1のビット値であり、0の前記ビット値をもつ前記個々のシンボルは、前記識別子ライブラリ内の別個の識別子の不在によって表され、1の前記ビット値をもつ前記個々のシンボルは、前記識別子ライブラリ内の前記別個の識別子の存在によって表され、逆も同様である、請求項3に記載の方法。

【請求項6】

前記識別子ライブラリ内の前記個々の識別子を生成することは、2つ以上の層からの前記2つ以上の構成要素をライゲーションすることを含み、前記2つ以上の層の各層は、構成要素の別個のセットを含む、請求項1に記載の方法。

10

【請求項7】

前記識別子ライブラリからの前記個々の識別子は、前記2つ以上の層の各層からの1つの構成要素を含む、請求項6に記載の方法。

【請求項8】

前記2つ以上の構成要素は、固定順序で組み立てられる、請求項7に記載の方法。

【請求項9】

前記2つ以上の構成要素は、任意の順序で組み立てられる、請求項7に記載の方法。

【請求項10】

前記2つ以上の構成要素は、前記2つ以上の層のうちの異なる層からの2つの構成要素間に設置された1つ以上のパーティション構成要素と組み立てられる、請求項7に記載の方法。

20

【請求項11】

前記個々の識別子が、前記2つ以上の層のサブセットの各層からの1つの構成要素を含む、請求項6に記載の方法。

【請求項12】

前記個々の識別子が、前記2つ以上の層の各々からの少なくとも1つの構成要素を含む、請求項6に記載の方法。

【請求項13】

前記2つ以上の構成要素が、粘着末端ライゲーション、バイオブリックアセンブリ、ゴールド Dengate アセンブリ、ギブソンアセンブリ、リガーゼサイクリング反応、または鋳型指向ライゲーションを使用してライゲーションされる、請求項1に記載の方法。

30

【請求項14】

前記識別子ライブラリ内の前記個々の識別子を生成することは、少なくとも1つのリコンビナーゼまたはヌクレアーゼを含む試薬と親識別子を接触させることにより、前記親識別子内の少なくとも1つの構成要素またはその一部を削除、交換、または挿入することを含む、請求項1に記載の方法。

【請求項15】

前記親識別子が、ヌクレアーゼ特異的標的部位、リコンビナーゼ認識部位、または別個のスペーサ配列によって隣接された複数の構成要素を含む、請求項14に記載の方法。

40

【請求項16】

前記ヌクレアーゼは、Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Cas9、Transcription activator-like effector nucleases (TALENs)、ジンクフィンガーヌクレアーゼ、およびそれらの任意の組み合わせからなる群から選択される、請求項14に記載の方法。

【請求項17】

前記核酸配列は、前記情報のメタデータを記憶するか、または前記情報を隠蔽する、請求項1に記載の方法。

【請求項18】

50

前記メタデータは、前記情報の源、前記情報の意図されたレシピエント、前記情報の元のフォーマット、前記情報を符号化するために使用される機器類および方法、前記情報を前記識別子ライブラリに書き込む日付および時間、前記情報に対してなされる修正、ならびに / または他の情報への参照に対応する二次情報を含む、請求項 17 に記載の方法。

【請求項 19】

2 つ以上の識別子ライブラリが組み合わせられる、請求項 1 に記載の方法。

【請求項 20】

前記サブセットは、それぞれが同一の値を有するシンボルに対応する複数の識別子を含む、請求項 1 に記載の方法。

【請求項 21】

前記複数の識別子は、読取り動作、書込み動作、アクセス動作、コピー動作、および削除動作のために選択される、請求項 1 に記載の方法。

【請求項 22】

前記個々のシンボルの誤りチェックまたは誤り検出をさらに含む、請求項 1 に記載の方法。

【請求項 23】

前記誤りチェックまたは誤り検出は、チェックサムを含む、請求項 22 に記載の方法。

【請求項 24】

前記情報は、前記識別子ライブラリの 1 つ以上のコピーを作成することによってコピーされる、請求項 1 に記載の方法。

【請求項 25】

前記情報は、識別子の標的サブセットを前記識別子ライブラリから隔離することによってクエリされる、請求項 1 に記載の方法。

【請求項 26】

前記情報は、前記識別子ライブラリから複数の識別子の前記核酸配列を決定することによって読み取られる、請求項 1 に記載の方法。

【請求項 27】

前記バーコードは、前記情報のメタデータを記憶するか、前記情報を隠蔽する、請求項 1 に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

相互参照

本出願は、2016 年 11 月 16 日に提出された米国仮特許出願番号第 62 / 423, 058 号、2017 年 2 月 9 日に提出された米国仮特許出願番号第 62 / 457, 074 号、および 2017 年 3 月 2 日に提出された米国仮特許出願番号第 62 / 466, 304 号の利益を主張しており、これら仮特許出願の各々は、参考として本明細書中に全体が援用される。

【背景技術】

【0002】

背景

核酸デジタルデータ記憶は、長期間にわたって情報を符号化し (encoding) 記憶するための安定した手法であり、データは、磁気テープまたはハードドライブ記憶システムよりも高い密度で記憶される。加えて、低温乾燥条件で記憶された核酸分子内に記憶されたデジタルデータは、60, 000 年後またはそれよりも長い間にわたって、取り出し可能である。

【0003】

核酸分子内に記憶されたデジタルデータにアクセスするために、核酸分子は、配列決定され得る。したがって、核酸デジタルデータ記憶は、頻繁にアクセスされないが長期間にわたって記憶またはアーカイブされることになる大量の情報を有することがあるデータを

10

20

30

40

50

記憶するための理想的な方法であり得る。

【 0 0 0 4 】

現在の方法は、配列内の塩基対塩基関係がデジタル情報（たとえば、バイナリコード）に直接的に翻訳するように、デジタル情報（たとえば、バイナリコード）を塩基単位核酸配列に符号化することに依拠する。デジタル的に符号化された情報のビットストリームまたはバイトに読み込まれ得る塩基単位配列内に記憶されたデジタルデータの配列決定は、誤りが起こりやすく、新規塩基単位核酸合成のコストは高価になり得るので符号化するのはコストが高い場合がある。核酸デジタルデータ記憶を実行する新しい方法の機会は、コストがかからず商業的に実施しやすい、データを符号化し取り出すための手法を提供し得る。

10

【 発明の概要 】

【 課題を解決するための手段 】

【 0 0 0 5 】

要旨

一意の核酸配列をもつビットストリーム内の各ビットロケーションを指定することと、プール内での対応する一意の核酸配列の存在または不在によってそのロケーションにおけるビット値を指定することを含む、プール内の一意の核酸配列の存在または不在においてビット値情報を符号化することによって、塩基単位合成なしに核酸（たとえば、デオキシリボ核酸、DNA）分子内にデジタル情報を符号化するための方法およびシステム。しかし、より一般に、核酸配列の一意のサブセットによってバイトストリーム内の一意のバイトを指定すること。組合せのゲノム戦略（たとえば、複数の核酸配列のアセンブリまたは核酸配列の酵素ベース編集）を使用した塩基対塩基合成なしに一意の核酸配列を生成するための方法も開示される。

20

【 0 0 0 6 】

一態様では、本開示は、情報を核酸配列に書き込むための方法であって、（a）情報をシンボルの文字列（string）に翻訳することと、（b）シンボルの文字列を複数の識別子にマッピングすることであって、複数の識別子の個々の識別子は1つまたは複数の構成要素を含み、1つまたは複数の構成要素の個々の構成要素は核酸配列を含み、複数の識別子の個々の識別子はシンボルの文字列の個々のシンボルに対応する、マッピングすることと、（c）複数の識別子の少なくともサブセットを含む識別子ライブラリを構築することを含む方法を提供する。

30

【 0 0 0 7 】

一部の実施形態では、シンボルの文字列内の各シンボルは2つの可能なシンボル値のうちの1つである。一部の実施形態では、シンボルの文字列の各位置における1つのシンボル値は、識別子ライブラリ内の別個の識別子の不在によって表され得る。一部の実施形態では、2つの可能なシンボル値は0および1のビット値であり、シンボルの文字列内の0のビット値をもつ個々のシンボルは、識別子ライブラリ内の別個の識別子の不在によって表され得、シンボルの文字列内の1のビット値をもつ個々のシンボルは、識別子ライブラリ内の別個の識別子の存在によって表され得、逆も同様である。一部の実施形態では、シンボルの文字列の各シンボルは、1つまたは複数の可能なシンボル値のうちの1つである。一部の実施形態では、識別子ライブラリ内の個々の識別子の存在は2進文字列内の第1のシンボル値に対応し、個々の識別子の不在は2進文字列内の第2のシンボル値に対応する。一部の実施形態では、第1のシンボル値は1のビット値であり、第2のシンボル値は0のビット値である。一部の実施形態では、第1のシンボル値は0のビット値であり、第2のシンボル値は1のビット値である。

40

【 0 0 0 8 】

一部の実施形態では、識別子ライブラリ内の個々の識別子を構築することは、1つまたは複数の構成要素を1つまたは複数の層から組み立てる（assemble）ことを含み、1つまたは複数の層の各層は、構成要素の別個のセットを含む。一部の実施形態では、識別子ライブラリからの個々の識別子は、1つまたは複数の層の各層からの1つの構成要

50

素を含む。一部の実施形態では、1つまたは複数の構成要素は、固定順序で組み立てられる。一部の実施形態では、1つまたは複数の構成要素は、ランダムな順序で組み立てられる。一部の実施形態では、1つまたは複数の構成要素は、1つまたは複数の層のうちの異なる層からの2つの構成要素間に設置された1つまたは複数のパーティション構成要素と組み立てられる。一部の実施形態では、個々の識別子は、1つまたは複数の層のサブセットの各層からの1つの構成要素を含む。一部の実施形態では、個々の識別子は、1つまたは複数の層の各々からの少なくとも1つの構成要素を含む。一部の実施形態では、1つまたは複数の構成要素は、オーバーラップ伸長 (o v e r l a p - e x t e n s i o n) ポリメラーゼ連鎖反応 (P C R)、ポリメラーゼサイクリングアセンブリ、粘着末端ライゲーション、パイオブリックアセンブリ、ゴールドゲートアセンブリ、ギブソンアセンブリ、リコンビナーゼアセンブリ、リガーゼサイクリング反応、または鋳型指向ライゲーション (t e m p l a t e d i r e c t e d l i g a t i o n) を使用して組み立てられる。

10

【0009】

一部の実施形態では、識別子ライブラリ内の個々の識別子を構築することは、核酸編集酵素を親識別子に適用することによって親識別子内の少なくとも1つの構成要素を削除、交換、または挿入することを含む。一部の実施形態では、親識別子は、ヌクレアーゼ特異的標的部位、リコンビナーゼ認識部位、または別個のスペーサ配列によって隣接された複数の構成要素を含む。一部の実施形態では、核酸編集酵素は、C R I S P R - C a s、T A L E N、ジンクフィンガーヌクレアーゼ、リコンビナーゼ、およびその機能的変異体からなる群から選択される。

20

【0010】

一部の実施形態では、識別子ライブラリは複数の核酸配列を含む。一部の実施形態では、複数の核酸配列は、情報のメタデータを記憶する、および/または情報を隠蔽する。一部の実施形態では、メタデータは、情報の源、情報の意図されたレシピエント、情報の元のフォーマット、情報を符号化するために使用される機器類および方法、情報を識別子ライブラリに書き込む日付および時間、情報に対してなされる修正、ならびに/または他の情報への参照に対応する二次情報を含む。

【0011】

一部の実施形態では、1つまたは複数の識別子ライブラリが組み合わされ、この1つまたは複数の識別子ライブラリの各識別子ライブラリは、別個のバーコードでタグ付けされる。一部の実施形態では、識別子ライブラリ内の各個々の識別子は、別個のバーコードを含む。一部の実施形態では、複数の識別子は、読取り動作、書込み動作、アクセス動作、コピー動作、および削除動作を簡単にするために選択される。一部の実施形態では、複数の識別子は、書込みエラー、突然変異、分解、および読取りエラーを最小にするために選択される。

30

【0012】

別の態様では、本開示は、核酸配列内に符号化された情報をコピーするための方法であって、(a) シンボルの文字列を符号化する識別子ライブラリを提供することであって、この識別子ライブラリは複数の識別子を含み、この複数の識別子の個々の識別子は1つまたは複数の構成要素を含み、1つまたは複数の構成要素の個々の構成要素は核酸配列を含み、複数の識別子の個々の識別子はシンボルの文字列の個々のシンボルに対応する、提供することと、(b) 識別子ライブラリの1つまたは複数のコピーを構築することを含む方法を提供する。

40

【0013】

一部の実施形態では、複数の識別子は、1つまたは複数のプライマー結合部位を含む。一部の実施形態では、識別子ライブラリは、ポリメラーゼ連鎖反応 (P C R) を使用してコピーされる。一部の実施形態では、P C R は、従来の P C R またはリニア P C R であり、識別子ライブラリのコピーの数はそれぞれ、各 P C R サイクルとともに2倍になるまたは線形的に増加する。一部の実施形態では、識別子ライブラリ内の個々の識別子は、P C

50

Rの前に環状ベクターにライゲーションされ、この環状ベクターは、個々の識別子の各末端においてバーコードを含む。

【0014】

一部の実施形態では、識別子ライブラリは、複数の核酸配列を含む。一部の実施形態では、複数の核酸配列がコピーされる。一部の実施形態では、1つまたは複数の識別子ライブラリは、コピーの前に組み合わせられ、1つまたは複数の識別子ライブラリの各ライブラリは、別個のバーコードを含む。

【0015】

別の態様では、本開示は、核酸配列内に符号化された情報にアクセスするための方法であって、(a)シンボルの文字列を符号化する識別子ライブラリを提供することであって、この識別子ライブラリは複数の識別子を含み、複数の識別子の個々の識別子は1つまたは複数の構成要素を含み、1つまたは複数の構成要素の個々の構成要素は核酸配列を含み、複数の識別子の個々の識別子はシンボルの文字列の個々のシンボルに対応する、提供することと、(b)複数の識別子の標的サブセットを識別子ライブラリから抽出することを含む方法を提供する。

10

【0016】

一部の実施形態では、複数のプローブは、識別子ライブラリと組み合わせられる。一部の実施形態では、複数のプローブは、識別子ライブラリからの複数の識別子の標的サブセットと相補性を共有する。一部の実施形態では、複数のプローブは、識別子ライブラリ内の複数の識別子の標的サブセットをハイブリダイゼーションさせる。一部の実施形態では、複数のプローブは1つまたは複数の親和性タグを含み、1つまたは複数の親和性タグは、親和性ビーズまたは親和性カラムによってキャプチャされる。

20

【0017】

一部の実施形態では、識別子ライブラリは、複数のプローブの1つまたは複数のサブセットと順次組み合わせられ、識別子ライブラリの一部分は、複数のプローブの1つまたは複数のサブセットに結合する。一部の実施形態では、複数のプローブの1つまたは複数のサブセットに結合する識別子ライブラリの一部分は、識別子ライブラリへの複数のプローブの別のサブセットの追加の前に除去される。

【0018】

一部の実施形態では、複数の識別子の個々の識別子は、1つもしくは複数の共通プライマー結合領域、1つもしくは複数の可変プライマー結合領域、またはそれらの任意の組合せを含む。一部の実施形態では、識別子ライブラリは、1つもしくは複数の共通プライマー結合領域にまたは1つもしくは複数の可変プライマー結合領域に結合するプライマーと組み合わせられる。一部の実施形態では、1つまたは複数の可変プライマー結合領域に結合するプライマーは、識別子ライブラリの標的サブセットを選択的に増幅するために使用される。

30

【0019】

一部の実施形態では、識別子の一部分は、選択的ヌクレアーゼ切断によって識別子ライブラリから除去される。一部の実施形態では、識別子ライブラリはCas9およびガイドプローブと組み合わせられ、このガイドプローブは、指定された識別子を識別子ライブラリから除去するためにCas9をガイドする。一部の実施形態では、個々の識別子は一本鎖であり、識別子ライブラリは、一本鎖特異的エンドヌクレアーゼと組み合わせられる。一部の実施形態では、識別子ライブラリは、一本鎖特異的エンドヌクレアーゼの追加の前に標的の個々の識別子を分解から保護する個々の識別子の相補セットと混合される。一部の実施形態では、選択的ヌクレアーゼ切断によって切断されない個々の識別子は、サイズ選択的クロマトグラフィによって分離される。一部の実施形態では、選択的ヌクレアーゼ切断によって切断されない個々の識別子は増幅され、選択的ヌクレアーゼ切断によって切断される個々の識別子は増幅されない。一部の実施形態では、識別子ライブラリは複数の核酸配列を含み、この複数の核酸配列は、識別子ライブラリ内の複数の識別子の標的サブセットで抽出される。

40

50

【 0 0 2 0 】

別の態様では、本開示は、核酸配列内に符号化された情報を読み取るための方法であって、(a) 複数の識別子を含む識別子ライブラリを提供することであって、複数の識別子の個々の識別子は 1 つまたは複数の構成要素を含み、1 つまたは複数の構成要素の個々の構成要素は核酸配列を含む、提供することと、(b) 識別子ライブラリ内の複数の識別子を識別することと、(c) (b) において識別された複数の識別子から複数のシンボルを生成することであって、複数のシンボルの個々のシンボルは複数の識別子の個々の識別子に対応する、生成することと、(d) 情報を複数のシンボルからコンパイルすることを含む方法を提供する。

【 0 0 2 1 】

一部の実施形態では、シンボルの文字列内の各シンボルは、2 つの可能なシンボル値のうちの 1 つである。一部の実施形態では、シンボルの文字列の各位置における 1 つのシンボル値は、識別子ライブラリ内の別個の識別子の不在によって表され得る。一部の実施形態では、2 つの可能なシンボル値は 0 および 1 のビット値であり、シンボルの文字列内の 0 のビット値をもつ個々のシンボルは、識別子ライブラリ内の別個の識別子の不在によって表され得、シンボルの文字列内の 1 のビット値をもつ個々のシンボルは、識別子ライブラリ内の別個の識別子の存在によって表され得、逆も同様である。一部の実施形態では、識別子ライブラリ内の個々の識別子の存在は、2 進文字列内の第 1 のシンボル値に対応し、識別子ライブラリ内の個々の識別子の不在は、2 進文字列内の第 2 のシンボル値に対応する。一部の実施形態では、第 1 のシンボル値は 1 のビット値であり、第 2 のシンボル値は 0 のビット値である。一部の実施形態では、第 1 のシンボル値は 0 のビット値であり、第 2 のシンボル値は 1 のビット値である。

【 0 0 2 2 】

一部の実施形態では、複数の識別子を識別することは、識別子ライブラリ内の複数の識別子を配列決定することを含む。一部の実施形態では、配列決定することは、デジタルポリメラーゼ連鎖反応 (P C R)、定量的 P C R、マイクロアレイ、合成による配列決定、または超並列配列決定を含む。一部の実施形態では、識別子ライブラリは、複数の核酸配列を含む。一部の実施形態では、複数の核酸配列は、情報のメタデータを記憶する、および/または情報を隠蔽する。一部の実施形態では、1 つまたは複数の識別子ライブラリが組み合わせられ、1 つまたは複数の識別子ライブラリ内の各識別子ライブラリが、別個のバーコードを含む。一部の実施形態では、バーコードは、情報のメタデータを記憶する。

【 0 0 2 3 】

別の態様では、本開示は、核酸ベースのコンピュータデータ記憶のための方法であって、(a) コンピュータデータを受信することと、(b) コンピュータデータを符号化する核酸配列を含む核酸分子を合成することであって、コンピュータデータは、合成され核酸分子の各々の配列決定内にない核酸分子の少なくともサブセット内で符号化される、合成することと、(c) 核酸配列を有する核酸分子を記憶することを含む方法を提供する。

【 0 0 2 4 】

一部の実施形態では、核酸分子の少なくともサブセットは、一緒に群化される。一部の実施形態では、方法は、核酸配列を決定するために核酸分子を配列決定し、それによって、コンピュータデータを取り出すことをさらに含む。一部の実施形態では、(b) は、約 1 日より短い時間期間内に実行される。一部の実施形態では、(b) は、少なくとも約 9 0 % の精度で実行される。

【 0 0 2 5 】

別の態様では、本開示は、核酸ベースのコンピュータデータ記憶のための方法であって、(a) コンピュータデータを受信することと、(b) コンピュータデータを符号化する少なくとも 1 つの核酸配列を含む核酸分子を合成することであって、核酸分子を合成することは塩基単位核酸合成の不在下である、合成することと、(c) 少なくとも 1 つの核酸配列を含む核酸分子を記憶することを含む方法を提供する。

【 0 0 2 6 】

一部の実施形態では、方法は、核酸配列を決定するために核酸分子を配列決定し、それによって、コンピュータデータを取り出すことをさらに含む。一部の実施形態では、(b)は、約1日より短い時間期間内に実行される。一部の実施形態では、(b)は、少なくとも約90%の精度で実行される。

【0027】

別の態様では、本開示は、核酸を使用して2進配列データを符号化するためのシステムであって、識別子ライブラリを構築するように構成されたデバイスであって、この識別子ライブラリは複数の識別子を含み、複数の識別子の個々の識別子は1つまたは複数の構成要素を含み、1つまたは複数の構成要素の個々の構成要素は核酸配列である、デバイスと、このデバイスに動作可能に結合された1つまたは複数のコンピュータプロセッサであって、(i)情報をシンボルの文字列に翻訳すること、(ii)シンボルの文字列を複数の識別子にマッピングすることであって、複数の識別子の個々の識別子はシンボルの文字列の個々のシンボルに対応する、マッピングすること、および(iii)複数の識別子を含む識別子ライブラリを構築することを行うように個々にまたは集合的にプログラムされる、1つまたは複数のコンピュータプロセッサとを含むシステムを提供する。

10

【0028】

一部の実施形態では、デバイスは複数のパーティションを含み、識別子ライブラリは、複数のパーティションのうちの1つまたは複数の中で生成される。一部の実施形態では、複数のパーティションは、ウェルを含む。一部の実施形態では、識別子ライブラリ内の個々の識別子を構築することは、1つまたは複数の構成要素を1つまたは複数の層から組み立てることを含み、1つまたは複数の層の各層は、構成要素の別個のセットを含む。一部の実施形態では、1つまたは複数の層の各層はデバイスの別々の部分内に記憶され、デバイスは、1つまたは複数の層からの1つまたは複数の構成要素を組み合わせるように構成される。一部の実施形態では、識別子ライブラリは、複数の核酸配列を含む。一部の実施形態では、1つまたは複数の識別子ライブラリはデバイスの単一のエリア内で組み合わせられ、この1つまたは複数の識別子ライブラリの各識別子ライブラリは、別個のバーコードを含む。

20

【0029】

別の態様では、本開示は、核酸配列内で符号化される情報を読み取るためのシステムであって、複数の識別子を含む識別子ライブラリを記憶するデータベースであって、この複数の識別子の個々の識別子は1つまたは複数の構成要素を含み、1つまたは複数の構成要素の個々の構成要素は核酸配列を含む、データベースと、このデータベースに動作可能に結合された1つまたは複数のコンピュータプロセッサであって、(i)識別子ライブラリ内の複数の識別子を識別すること、(ii)(i)において識別された複数の識別子から複数のシンボルを生成することであって、複数のシンボルの個々のシンボルは複数の識別子の個々の識別子に対応する、生成すること、および(iii)複数のシンボルからの情報をコンパイルすることを行うように個々にまたは集合的にプログラムされる、1つまたは複数のコンピュータプロセッサとを含むシステムを提供する。

30

【0030】

一部の実施形態では、システムは、複数のパーティションをさらに含む。一部の実施形態では、パーティションは、ウェルである。一部の実施形態では、複数のパーティションの所与のパーティションは、1つまたは複数の識別子ライブラリを含み、この1つまたは複数の識別子ライブラリの各識別子ライブラリは、別個のバーコードを含む。一部の実施形態では、システムは、識別子ライブラリ内の複数の識別子を識別するように構成された検出ユニットをさらに含む。

40

【0031】

本開示の追加の態様および利点は、以下の詳細な説明から、当業者に容易に明らかになる。本開示の例示的な実施形態のみが、図示および説明される。諒解されるように、本開示は、他のおよび異なる実施形態が可能であり、そのいくつかの詳細は、すべて本開示から逸脱することなく、種々の明白な点における修正が可能である。したがって、図面お

50

よび説明は、本質的に、限定的ではなく、例示的とみなされるべきである。
本願明細書は、例えば、以下の項目も提供する。

(項目 1)

情報を核酸配列に書き込むための方法であって、

(a) 前記情報をシンボルの文字列に翻訳することと、

(b) 前記シンボルの文字列を複数の識別子にマッピングすることであって、前記複数の識別子の個々の識別子が 1 つまたは複数の構成要素を含み、前記 1 つまたは複数の構成要素の個々の構成要素が核酸配列を含み、前記複数の識別子の前記個々の識別子が前記シンボルの文字列の個々のシンボルに対応する、マッピングすることと、

(c) 前記複数の識別子の少なくともサブセットを含む識別子ライブラリを構築することと

10

を含む方法。

(項目 2)

前記シンボルの文字列の各シンボルが 1 つまたは複数の可能なシンボル値のうちの 1 つである、項目 1 に記載の方法。

(項目 3)

前記シンボルの文字列内の各シンボルが 2 つの可能なシンボル値のうちの 1 つである、項目 1 に記載の方法。

(項目 4)

前記シンボルの文字列の各位置における 1 つのシンボル値が、前記識別子ライブラリ内の別個の識別子の不在によって表され得る、項目 1 に記載の方法。

20

(項目 5)

前記 2 つの可能なシンボル値が 0 および 1 のビット値であり、前記シンボルの文字列内の 0 の前記ビット値をもつ前記個々のシンボルが、前記識別子ライブラリ内の別個の識別子の不在によって表され得、前記シンボルの文字列内の 1 の前記ビット値をもつ前記個々のシンボルが、前記識別子ライブラリ内の前記別個の識別子の存在によって表され得、逆も同様である、項目 3 に記載の方法。

(項目 6)

前記識別子ライブラリ内の前記個々の識別子を構築することが、前記 1 つまたは複数の構成要素を 1 つまたは複数の層から組み立てることを含み、前記 1 つまたは複数の層の各層が、構成要素の別個のセットを含む、項目 1 に記載の方法。

30

(項目 7)

前記識別子ライブラリからの前記個々の識別子が、前記 1 つまたは複数の層の各層からの 1 つの構成要素を含む、項目 6 に記載の方法。

(項目 8)

前記 1 つまたは複数の構成要素が、固定順序で組み立てられる、項目 7 に記載の方法。

(項目 9)

前記 1 つまたは複数の構成要素が、任意の順序で組み立てられる、項目 7 に記載の方法。

(項目 10)

前記 1 つまたは複数の構成要素が、前記 1 つまたは複数の層のうちの異なる層からの 2 つの構成要素間に設置された 1 つまたは複数のパーティション構成要素と組み立てられる、項目 7 に記載の方法。

40

(項目 11)

前記個々の識別子が、前記 1 つまたは複数の層のサブセットの各層からの 1 つの構成要素を含む、項目 6 に記載の方法。

(項目 12)

前記個々の識別子が、前記 1 つまたは複数の層の各々からの少なくとも 1 つの構成要素を含む、項目 6 に記載の方法。

(項目 13)

前記 1 つまたは複数の構成要素が、オーバーラップ伸長ポリメラーゼ連鎖反応 (PCR

50

)、ポリメラーゼサイクリングアセンブリ、粘着末端ライゲーション、バイオリックアセンブリ、ゴールデンゲートアセンブリ、ギブソンアセンブリ、リコンビナーゼアセンブリ、リガーゼサイクリング反応、または鋳型指向ライゲーションを使用して組み立てられる、項目 6 に記載の方法。

(項目 14)

前記識別子ライブラリ内の前記個々の識別子を構築することが、核酸編集酵素を親識別子に適用することによって前記親識別子内の少なくとも 1 つの構成要素を削除、交換、または挿入することを含む、項目 1 に記載の方法。

(項目 15)

前記親識別子が、ヌクレアーゼ特異的標的部位、リコンビナーゼ認識部位、または別個のスペーサ配列によって隣接された複数の構成要素を含む、項目 14 に記載の方法。

10

(項目 16)

前記核酸編集酵素が、CRISPR-Cas、TALEN、ジンクフィンガーヌクレアーゼ、リコンビナーゼ、およびその機能的変異体からなる群から選択される、項目 14 に記載の方法。

(項目 17)

前記識別子ライブラリが複数の核酸配列を含む、項目 1 に記載の方法。

(項目 18)

前記複数の核酸配列が、前記情報のメタデータを記憶する、および/または前記情報を隠蔽する、項目 17 に記載の方法。

20

(項目 19)

前記メタデータが、前記情報の源、前記情報の意図されたレシピエント、前記情報の元のフォーマット、前記情報を符号化するために使用される機器類および方法、前記情報を前記識別子ライブラリに書き込む日付および時間、前記情報に対してなされる修正、ならびに/または他の情報への参照に対応する二次情報を含む、項目 18 に記載の方法。

(項目 20)

1 つまたは複数の識別子ライブラリが組み合わせられ、前記 1 つまたは複数の識別子ライブラリの各識別子ライブラリが、別個のバーコードでタグ付けされる、項目 1 に記載の方法。

(項目 21)

30

前記識別子ライブラリ内の各個々の識別子が、前記別個のバーコードを含む、項目 20 に記載の方法。

(項目 22)

前記複数の識別子が、読取り動作、書込み動作、アクセス動作、コピー動作、および削除動作を容易にするために選択される、項目 1 に記載の方法。

(項目 23)

前記複数の識別子が、書込みエラー、突然変異、分解、および読取りエラーを最小にするように選択される、項目 1 に記載の方法。

(項目 24)

核酸配列内に符号化された情報をコピーするための方法であって、

40

(a) シンボルの文字列を符号化する識別子ライブラリを提供することであって、前記識別子ライブラリが複数の識別子を含み、前記複数の識別子の個々の識別子が 1 つまたは複数の構成要素を含み、前記 1 つまたは複数の構成要素の個々の構成要素が核酸配列を含み、前記複数の識別子の前記個々の識別子が前記シンボルの文字列の個々のシンボルに対応する、提供することと、

(b) 前記識別子ライブラリの 1 つまたは複数のコピーを構築することとを含む方法。

(項目 25)

前記複数の識別子が 1 つまたは複数のプライマー結合部位を含む、項目 24 に記載の方法。

50

(項目 26)

前記識別子ライブラリが、ポリメラーゼ連鎖反応 (PCR) を使用してコピーされる、項目 25 に記載の方法。

(項目 27)

前記 PCR が、従来の PCR またはリニア PCR であり、前記識別子ライブラリのコピーの数がそれぞれ、各 PCR サイクルとともに 2 倍になるまたは線形的に増加する、項目 26 に記載の方法。

(項目 28)

前記識別子ライブラリ内の前記個々の識別子が、PCR の前に環状ベクターにライゲーションされ、前記環状ベクターが、前記個々の識別子の各末端においてバーコードを含む、項目 25 に記載の方法。

10

(項目 29)

前記識別子ライブラリが複数の核酸配列を含む、項目 24 に記載の方法。

(項目 30)

前記複数の核酸配列がコピーされる、項目 29 に記載の方法。

(項目 31)

1 つまたは複数の識別子ライブラリが、コピーの前に組み合わせられ、前記 1 つまたは複数の識別子ライブラリの各ライブラリが、別個のバーコードを含む、項目 24 に記載の方法。

(項目 32)

核酸配列内に符号化された情報にアクセスするための方法であって、

(a) シンボルの文字列を符号化する識別子ライブラリを提供することであって、前記識別子ライブラリが複数の識別子を含み、前記複数の識別子の個々の識別子が 1 つまたは複数の構成要素を含み、前記 1 つまたは複数の構成要素の個々の構成要素が核酸配列を含み、前記複数の識別子の前記個々の識別子が前記シンボルの文字列の個々のシンボルに対応する、提供することと、

(b) 前記複数の識別子の標的サブセットを前記識別子ライブラリから抽出することを含む方法。

20

(項目 33)

複数のプローブが前記識別子ライブラリと組み合わせられる、項目 32 に記載の方法。

30

(項目 34)

前記複数のプローブが、前記識別子ライブラリの前記標的サブセットの前記個々の識別子の前記 1 つまたは複数の構成要素との相補性を共有する、項目 33 に記載の方法。

(項目 35)

前記複数のプローブが、前記識別子ライブラリ内の前記複数の識別子の前記標的サブセットの一部分とハイブリダイゼーションする、項目 34 に記載の方法。

(項目 36)

前記複数のプローブが 1 つまたは複数の親和性タグを含み、前記 1 つまたは複数の親和性タグが、親和性ビーズまたは親和性カラムによってキャプチャされる、項目 34 に記載の方法。

40

(項目 37)

前記識別子ライブラリが、前記複数のプローブの 1 つまたは複数のサブセットと順次組み合わせられ、前記識別子ライブラリの前記標的サブセットの一部分が、前記複数のプローブの前記 1 つまたは複数のサブセットに結合する、項目 33 に記載の方法。

(項目 38)

前記複数のプローブの前記 1 つまたは複数のサブセットに結合する前記識別子ライブラリの前記一部分が、前記識別子ライブラリへの前記複数のプローブの別のサブセットの追加の前に除去される、項目 37 に記載の方法。

(項目 39)

前記複数の識別子の前記個々の識別子が、1 つもしくは複数の共通プライマー結合領域

50

、1つもしくは複数の可変プライマー結合領域、またはそれらの任意の組合せを含む、項目32に記載の方法。

(項目40)

前記識別子ライブラリが、前記1つもしくは複数の共通プライマー結合領域にまたは前記1つもしくは複数の可変プライマー結合領域に結合するプライマーと組み合わせられる、項目39に記載の方法。

(項目41)

前記1つまたは複数の可変プライマー結合領域に結合する前記プライマーが、前記識別子ライブラリの前記標的サブセットを選択的に増幅するために使用される、項目40に記載の方法。

10

(項目42)

識別子の一部分が、選択的ヌクレアーゼ切断によって前記識別子ライブラリから除去される、項目32に記載の方法。

(項目43)

前記識別子ライブラリがCas9およびガイドプローブと組み合わせられ、前記ガイドプローブが、指定された識別子を前記識別子ライブラリから除去するために前記Cas9をガイドする、項目42に記載の方法。

(項目44)

前記個々の識別子が一本鎖であり、前記識別子ライブラリが、一本鎖特異的エンドヌクレアーゼと組み合わせられる、項目42に記載の方法。

20

(項目45)

前記識別子ライブラリが、前記一本鎖特異的エンドヌクレアーゼの追加の前に標的の個々の識別子を分解から保護する個々の識別子の相補セットと混合される、項目44に記載の方法。

(項目46)

前記選択的ヌクレアーゼ切断によって切断されない前記個々の識別子が、サイズ選択的クロマトグラフィによって分離される、項目42に記載の方法。

(項目47)

前記選択的ヌクレアーゼ切断によって切断されない前記個々の識別子が増幅され、前記選択的ヌクレアーゼ切断によって切断される前記個々の識別子が増幅されない、項目42に記載の方法。

30

(項目48)

前記識別子ライブラリが複数の核酸配列を含み、前記複数の核酸配列が、前記識別子ライブラリ内の前記複数の識別子の前記標的サブセットで抽出される、項目32に記載の方法。

(項目49)

核酸配列内に符号化された情報を読み取るための方法であって、

(a) 複数の識別子を含む識別子ライブラリを提供することであって、前記複数の識別子の個々の識別子が1つまたは複数の構成要素を含み、前記1つまたは複数の構成要素の個々の構成要素が核酸配列を含む、提供することと、

40

(b) 前記識別子ライブラリ内の前記複数の識別子を識別することと、

(c) (b)において識別された前記複数の識別子から複数のシンボルを生成することであって、前記複数のシンボルの個々のシンボルが前記複数の識別子の前記個々の識別子に対応する、生成することと、

(d) 前記情報を前記複数のシンボルからコンパイルすることと

を含む方法。

(項目50)

前記識別子ライブラリ内の個々の識別子の存在が2進文字列内の第1のシンボル値に対応し、前記識別子ライブラリからの前記個々の識別子の不在が2進文字列内の第2のシンボル値に対応する、項目49に記載の方法。

50

(項目 5 1)

前記第 1 のシンボル値が 1 のビット値であり、前記第 2 のシンボル値が 0 のビット値である、項目 5 0 に記載の方法。

(項目 5 2)

前記第 1 のシンボル値が 0 のビット値であり、前記第 2 のシンボル値が 1 のビット値である、項目 5 0 に記載の方法。

(項目 5 3)

前記複数の識別子を識別することが、前記複数の識別子を配列決定することを含む、項目 4 9 に記載の方法。

(項目 5 4)

配列決定することが、デジタルポリメラーゼ連鎖反応 (PCR)、定量的 PCR、マイクロアレイ、合成による配列決定、または超並列配列決定を含む、項目 5 3 に記載の方法。

(項目 5 5)

前記識別子ライブラリが複数の核酸配列を含む、項目 4 9 に記載の方法。

(項目 5 6)

前記複数の核酸配列が、前記情報のメタデータを記憶する、および / または前記情報を隠蔽する、項目 5 5 に記載の方法。

(項目 5 7)

1 つまたは複数の識別子ライブラリが組み合わせられ、前記 1 つまたは複数の識別子ライブラリ内の各識別子ライブラリが、別個のバーコードを含む、項目 4 9 に記載の方法。

(項目 5 8)

前記バーコードが前記情報のメタデータを記憶する、項目 5 7 に記載の方法。

(項目 5 9)

核酸を使用して 2 進配列データを符号化するためのシステムであって、
識別子ライブラリを構築するように構成されたデバイスであって、前記識別子ライブラリが複数の識別子を含み、前記複数の識別子の個々の識別子が 1 つまたは複数の構成要素を含み、前記 1 つまたは複数の構成要素の個々の構成要素が核酸配列である、デバイスと、
前記デバイスに動作可能にカップリングされた 1 つまたは複数のコンピュータプロセッサであって、(i) 前記情報をシンボルの文字列に翻訳すること、(ii) 前記シンボルの文字列を前記複数の識別子にマッピングすることであって、前記複数の識別子の前記個々の識別子が前記シンボルの文字列の個々のシンボルに対応する、マッピングすること、および (iii) 前記複数の識別子を含む識別子ライブラリを構築することを行うように個々にまたは集合的にプログラムされる、1 つまたは複数のコンピュータプロセッサとを含むシステム。

(項目 6 0)

前記デバイスが複数のパーティションを含み、前記識別子ライブラリが、前記複数のパーティションのうちの 1 つまたは複数の中で生成される、項目 5 9 に記載の方法。

(項目 6 1)

前記複数のパーティションがウェルを含む、項目 5 9 に記載の方法。

(項目 6 2)

前記識別子ライブラリ内の前記個々の識別子を構築することが、前記 1 つまたは複数の構成要素を 1 つまたは複数の層から組み立てることを含み、前記 1 つまたは複数の層の各層が、構成要素の別個のセットを含む、項目 5 9 に記載の方法。

(項目 6 3)

前記 1 つまたは複数の層の各層が前記デバイスの別々の一部分内に記憶され、前記デバイスが、前記 1 つまたは複数の層からの前記 1 つまたは複数の構成要素を組み合わせるように構成される、項目 6 2 に記載の方法。

(項目 6 4)

前記識別子ライブラリが複数の核酸配列を含む、項目 5 9 に記載の方法。

(項目 6 5)

10

20

30

40

50

1 つまたは複数の識別子ライブラリが前記デバイスの単一のエリア内で組み合わせられ、前記 1 つまたは複数の識別子ライブラリの各識別子ライブラリが、別個のバーコードを含む、項目 5 9 に記載の方法。

(項目 6 6)

核酸配列内に符号化された情報を読み取るためのシステムであって、

複数の識別子を含む識別子ライブラリを記憶するデータベースであって、前記複数の識別子の個々の識別子が 1 つまたは複数の構成要素を含み、前記 1 つまたは複数の構成要素の個々の構成要素が核酸配列を含む、データベースと、

前記データベースに動作可能にカップリングされた 1 つまたは複数のコンピュータプロセッサであって、(i) 前記識別子ライブラリ内の前記複数の識別子を識別すること、(i i) (i) において識別された前記複数の識別子から複数のシンボルを生成することであって、前記複数のシンボルの個々のシンボルが前記複数の識別子の前記個々の識別子に対応する、生成すること、および(i i i) 前記複数のシンボルからの前記情報をコンパイルすることを行うように個々にまたは集合的にプログラムされる、1 つまたは複数のコンピュータプロセッサと

を含むシステム。

(項目 6 7)

複数のパーティションを含むデバイスをさらに含む、項目 6 6 に記載の方法。

(項目 6 8)

前記パーティションがウェルである、項目 6 7 に記載の方法。

(項目 6 9)

前記複数のパーティションの所与のパーティションが 1 つまたは複数の識別子ライブラリを含み、前記 1 つまたは複数の識別子ライブラリの各識別子ライブラリが別個のバーコードを含む、項目 6 6 に記載の方法。

(項目 7 0)

前記システムが、前記識別子ライブラリ内の前記複数の識別子を識別するように構成された検出ユニットをさらに含む、項目 6 6 に記載の方法。

【 0 0 3 2 】

参照による組み込み

本明細書において論及されるすべての公報、特許、および特許出願は、各個々の公報、特許、または特許出願が具体的および個々に参照により組み込まれるように示されるかのようと同じ範囲まで、参照により本明細書において組み込まれる。参照により組み込まれる公報および特許または特許出願が、本明細書に含まれる開示に矛盾する程度まで、本明細書は、任意のそのような矛盾する資料に取って代わるおよび / またはこれに優先することが意図されている。

【 0 0 3 3 】

本発明の新しい特徴が、添付の請求の範囲において特定性 (p a r t i c u l a r i t y) とともに記載されている。本発明の特徴および利点のより良い理解は、本発明の原理が利用される例示的な実施形態について記載する以下の詳細な説明と、添付の図面 (本明細書における「図 (F i g u r e) 」および「図 (F I G .) 」も) とを参照して取得されよう。

【図面の簡単な説明】

【 0 0 3 4 】

【図 1】図 1 は、核酸配列内に記憶されたデジタル情報を符号化する、書き込む、アクセスする、読み取る、および復号する (d e c o d i n g) ためのプロセスの概要を概略的に例示する図である。

【 0 0 3 5 】

【図 2】図 2 A および図 2 B は、オブジェクトまたは識別子 (たとえば、核酸分子) を使用して、「アドレスにおけるデータ」と呼ばれるデジタルデータを符号化する例となる方法を概略的に例示する図である。図 2 A は、ランクオブジェクト (またはアドレスオブジ

10

20

30

40

50

エクト)をバイト値オブジェクト(またはデータオブジェクト)と組み合わせて、識別子を作成することを例示する図である。図2Bは、ランクオブジェクトおよびバイト値オブジェクトがそれ自体他のオブジェクトの組合せ連結である、アドレスにおけるデータ方法の一実施形態を例示する図である。

【0036】

【図3】図3Aおよび図3Bは、オブジェクトまたは識別子(たとえば、核酸配列)を使用してデジタル情報を符号化する例となる方法を概略的に例示する図である。図3Aは、ランクオブジェクトを識別子として使用してデジタル情報を符号化することを例示する図である。図3Bは、アドレスオブジェクトがそれ自体他のオブジェクトの組合せ連結である、符号化する方法の一実施形態を例示する図である。

10

【0037】

【図4】図4は、可能な識別子の組合せ空間(C、x軸)と所与のサイズの情報(輪郭線)を記憶するように構築され得る識別子の平均数(k、y軸)との間の関係の対数空間内の輪郭プロットを図示する図である。

【0038】

【図5】図5は、情報を核酸配列(たとえば、デオキシリボ核酸)に書き込むための方法の概要を概略的に例示する図である。

【0039】

【図6】図6Aおよび図6Bは、別個の構成要素(たとえば、核酸配列)を組合せにより組み立てることによって識別子(たとえば、核酸分子)を構築するための、「生産物方式(product scheme)」と呼ばれる例となる方法を例示する図である。図6Aは、生産物方式を使用して構築された識別子のアーキテクチャを例示する図である。図6Bは、生産物方式を使用して構築され得る識別子の組合せ空間の一例を例示する図である。

20

【0040】

【図7】図7は、識別子(たとえば、核酸分子)を構成要素(たとえば、核酸配列)から構築するためのオーバーラップ伸長ポリメラーゼ連鎖反応の使用を概略的に例示する図である。

【0041】

【図8】図8は、識別子(たとえば、核酸分子)を構成要素(たとえば、核酸配列)から構築するための粘着末端ライゲーションの使用を概略的に例示する図である。

30

【0042】

【図9】図9は、識別子(たとえば、核酸分子)を構成要素(たとえば、核酸配列)から構築するためのリコンビナーゼアセンブリの使用を概略的に例示する図である。

【0043】

【図10A】図10Aおよび図10Bは、鋳型指向ライゲーションを示す図である。図10Aは、識別子(たとえば、核酸分子)を構成要素(たとえば、核酸配列)から構築するための鋳型指向ライゲーションの使用を概略的に例示する図である。

【図10B】図10Aおよび図10Bは、鋳型指向ライゲーションを示す図である。図10Bは、各々が1つのプールされた鋳型指向ライゲーション反応において6つの核酸配列(たとえば、構成要素)から組合せにより組み立てられた256の別個の核酸配列のコピー数(存在度(abundance))のヒストグラムを示す図である。

40

【0044】

【図11A】図11Aおよび図11Gは、並べ替えられた構成要素(たとえば、核酸配列)をもつ識別子(たとえば、核酸分子)を構築するための、「並べ替え方式」と呼ばれる例となる方法を概略的に例示する図である。図11Aは、並べ替え方式を使用して構築された識別子のアーキテクチャを例示する図である。

【図11B】図11Aおよび図11Gは、並べ替えられた構成要素(たとえば、核酸配列)をもつ識別子(たとえば、核酸分子)を構築するための、「並べ替え方式」と呼ばれる例となる方法を概略的に例示する図である。図11Bは、並べ替え方式を使用して構築さ

50

れ得る識別子の組合せ空間の一例を例示する図である。

【図 1 1 C】図 1 1 A および図 1 1 G は、並べ替えられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「並べ替え方式」と呼ばれる例となる方法を概略的に例示する図である。図 1 1 C は、鋳型指向ライゲーションを有する並べ替え方式の例となる一実装形態を図示する図である。

【図 1 1 D】図 1 1 A および図 1 1 G は、並べ替えられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「並べ替え方式」と呼ばれる例となる方法を概略的に例示する図である。図 1 1 D は、どのように図 1 1 C からの実装形態が、並べ替えられた繰り返された構成要素をもつ識別子を構築するように修正され得るかの一例を図示する図である。

10

【図 1 1 E】図 1 1 A および図 1 1 G は、並べ替えられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「並べ替え方式」と呼ばれる例となる方法を概略的に例示する図である。図 1 1 E は、どのように図 1 1 D からの例となる実装形態が、核酸サイズ選択とともに除去され得る望ましくない副生産物につながり得るかを図示する図である。

【図 1 1 F】図 1 1 A および図 1 1 G は、並べ替えられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「並べ替え方式」と呼ばれる例となる方法を概略的に例示する図である。図 1 1 F は、並べ替えられた繰り返された構成要素をもつ識別子を構築するために鋳型指向ライゲーションおよびサイズ選択をどのように使用するべきかの別の例を図示する図である。

20

【図 1 1 G】図 1 1 A および図 1 1 G は、並べ替えられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「並べ替え方式」と呼ばれる例となる方法を概略的に例示する図である。図 1 1 G は、いつサイズ選択が、特定の識別子を望ましくない副生産物から隔離するのに失敗し得るかの一例を図示する図である。

【 0 0 4 5 】

【図 1 2 A】図 1 2 A ~ 図 1 2 D は、より大きな数 M の可能な構成要素からの任意の数 K の組み立てられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「M c h o o s e K」方式と呼ばれる例となる方法を概略的に例示する図である。図 1 2 A は、M c h o o s e K 方式を使用して構築された識別子のアーキテクチャを例示する図である。

30

【図 1 2 B】図 1 2 A ~ 図 1 2 D は、より大きな数 M の可能な構成要素からの任意の数 K の組み立てられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「M c h o o s e K」方式と呼ばれる例となる方法を概略的に例示する図である。図 1 2 B は、M c h o o s e K 方式を使用して構築され得る識別子の組合せ空間の一例を例示する図である。

【図 1 2 C】図 1 2 A ~ 図 1 2 D は、より大きな数 M の可能な構成要素からの任意の数 K の組み立てられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「M c h o o s e K」方式と呼ばれる例となる方法を概略的に例示する図である。図 1 2 C は、鋳型指向ライゲーションを使用した M c h o o s e K 方式の例となる一実装形態を図示する図である。

40

【図 1 2 D】図 1 2 A ~ 図 1 2 D は、より大きな数 M の可能な構成要素からの任意の数 K の組み立てられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「M c h o o s e K」方式と呼ばれる例となる方法を概略的に例示する図である。図 1 2 D は、どのように図 1 2 C からの例となる実装形態が、核酸サイズ選択とともに除去され得る望ましくない副生産物につながり得るかを図示する図である。

【 0 0 4 6 】

【図 1 3 A】図 1 3 A および図 1 3 B は、パーティションされた構成要素をもつ識別子を構築するための、「パーティション方式」と呼ばれる例となる方法を概略的に例示する図である。図 1 3 A は、パーティション方式を使用して構築され得る識別子の組合せ空間の一例を図示する図である。

50

【図 1 3 B】図 1 3 A および図 1 3 B は、パーティションされた構成要素をもつ識別子を構築するための、「パーティション方式」と呼ばれる例となる方法を概略的に例示する図である。図 1 3 B は、鋳型指向ライゲーションを使用したパーティション方式の例となる一実装形態を図示する図である。

【0 0 4 7】

【図 1 4 A】図 1 4 A および図 1 4 B は、いくつかの可能な構成要素からの構成要素の任意の文字列から構成される識別子を構築するための、「無制約文字列」（すなわち U S S）方式と呼ばれる例となる方法を概略的に例示する図である。図 1 4 A は、U S S 方式を使用して構築され得る識別子の組合せ空間の一例を図示する図である。

【図 1 4 B】図 1 4 A および図 1 4 B は、いくつかの可能な構成要素からの構成要素の任意の文字列から構成される識別子を構築するための、「無制約文字列」（すなわち U S S）方式と呼ばれる例となる方法を概略的に例示する図である。図 1 4 B は、鋳型指向ライゲーションを使用した U S S 方式の例となる一実装形態を図示する図である。

【0 0 4 8】

【図 1 5 A】図 1 5 A および図 1 5 B は、構成要素を親識別子から除去することによって識別子を構築するための、「構成要素削除」と呼ばれる例となる方法を概略的に例示する図である。図 1 5 A は、構成要素削除方式を使用して構築され得る識別子の組合せ空間の一例を図示する図である。

【図 1 5 B】図 1 5 A および図 1 5 B は、構成要素を親識別子から除去することによって識別子を構築するための、「構成要素削除」と呼ばれる例となる方法を概略的に例示する図である。図 1 5 B は、二本鎖標的切断および修復を使用した構成要素削除方式の例となる実装形態を図示する図である。

【0 0 4 9】

【図 1 6】図 1 6 は、リコンビナーゼを親識別子に適用することによってさらなる識別子が構築され得るリコンビナーゼ認識部位をもつ親識別子を概略的に例示する図である。

【0 0 5 0】

【図 1 7 A】図 1 7 A ~ 図 1 7 C は、より大きな数の識別子からのいくつかの特定の識別子にアクセスすることによって核酸配列内に記憶された情報の一部分にアクセスするための例となる方法の概要を概略的に例示する図である。図 1 7 A は、ポリメラーゼ連鎖反応、親和性タグプローブ、および分解標的プローブを使用して、指定された構成要素を含有する識別子にアクセスするための、例となる方法を図示する図である。

【図 1 7 B】図 1 7 A ~ 図 1 7 C は、より大きな数の識別子からのいくつかの特定の識別子にアクセスすることによって核酸配列内に記憶された情報の一部分にアクセスするための例となる方法の概要を概略的に例示する図である。図 1 7 B は、ポリメラーゼ連鎖反応を使用し、「OR」演算または「AND」演算を実行して、複数の指定された構成要素を含有する識別子にアクセスするための例となる方法を図示する図である。

【図 1 7 C】図 1 7 A ~ 図 1 7 C は、より大きな数の識別子からのいくつかの特定の識別子にアクセスすることによって核酸配列内に記憶された情報の一部分にアクセスするための例となる方法の概要を概略的に例示する図である。図 1 7 C は、親和性タグを使用し、「OR」演算または「AND」演算を実行して、複数の指定された構成要素を含有する識別子にアクセスするための例となる方法を図示する図である。

【0 0 5 1】

【図 1 8】図 1 8 A および図 1 8 B は、核酸分子内に符号化されたデータを符号化すること、書き込むこと、および読み取ることの一例を図示する図である。図 1 8 A は、5, 8 5 6 ビットのデータを符号化すること、書き込むこと、および読み取ることの一例を図示する図である。図 1 8 b は、6 2, 8 2 4 ビットのデータを符号化すること、書き込むこと、および読み取ることの一例を図示する図である。

【0 0 5 2】

【図 1 9】図 1 9 は、本明細書において提供される方法を実装するようにプログラムまたはそうでなければ構成されたコンピュータシステムを図示する図である。

10

20

30

40

50

【発明を実施するための形態】

【0053】

詳細な説明

本発明の種々の実施形態が本明細書において図示および説明されているが、そのような実施形態が単に例として提供されることは、当業者に明白であろう。数多くの変形、変更、および置換えが、本発明から逸脱することなく当業者に想到されよう。本明細書において説明される本発明の実施形態に対する種々の代替形態が用いられてよいことは理解されるべきである。

【0054】

本明細書において使用される「シンボル」という用語は、一般に、デジタル情報の単位の表現を指す。デジタル情報は、シンボルの文字列に分割または翻訳されることがある。一例では、シンボルはビットであってよく、ビットは、「0」または「1」の値を有してよい。

10

【0055】

本明細書において使用される「別個の」または「一意の」という用語は、一般に、群内の他のオブジェクトと区別可能なオブジェクトを指す。たとえば、別個の、または一意の、核酸配列は、他の任意の核酸配列と同じ配列を有さない核酸配列であってよい。別個の、または一意の、核酸分子は、他の任意の核酸分子と同じ配列をも有さなくてもよい。別個の、または一意の、核酸配列または分子は、別の核酸配列または分子との類似性の領域を共有することがある。

20

【0056】

本明細書において使用される「構成要素」という用語は、一般に、核酸配列を指す。構成要素は、別個の核酸配列であってよい。構成要素は、他の核酸配列または分子を生成するために、1つまたは複数の他の構成要素と連結されるまたは組み立てられることがある。

【0057】

本明細書において使用される「層」という用語は、一般に、構成要素の群またはプールを指す。各層は、1つの層における構成要素が別の層における構成要素と異なるように、別個の構成要素のセットを含んでよい。1つまたは複数の層からの構成要素は、1つまたは複数の識別子を生成するために組み立てられることがある。

【0058】

本明細書において使用される「識別子」という用語は、一般に、より大きなビット文字列内のビット文字列の位置および値を表す核酸分子または核酸配列を指す。より一般に、識別子は、シンボルの文字列内のシンボルを表すまたはこれに対応する任意のオブジェクトを指してよい。一部の実施形態では、識別子は、1つまたは複数の連結された構成要素を含むことがある。

30

【0059】

本明細書において使用される「組合せ空間」という用語は、一般に、構成要素などのオブジェクトの開始セットおよび識別子を形成するためにそれらのオブジェクトをどのように修正するかに関するルールの許容可能なセットから生成され得るすべての可能な別個の識別子のセットを指す。構成要素を組み立てるまたは連結することによって作られる識別子の組合せ空間のサイズは、構成要素の層の数、各層における構成要素の数、および識別子を生成するために使用される特定のアセンブリ方法に依存することがある。

40

【0060】

本明細書において使用される「識別子ランク」という用語は、一般に、セット内の識別子の順序を規定する関係を指す。

【0061】

本明細書において使用される「識別子ライブラリ」という用語は、一般に、デジタル情報を表すシンボル文字列内のシンボルに対応する識別子の集合を指す。一部の実施形態では、識別子ライブラリ内での所与の識別子の不在は、特定の位置におけるシンボル値を示すことがある。1つまたは複数の識別子ライブラリは、プール、群、または識別子のセッ

50

ト内で組み合わせられることがある。各識別子ライブラリは、識別子ライブラリを識別する一意のバーコードを含むことがある。

【 0 0 6 2 】

本明細書において使用される「核酸」という用語は、一般的な、デオキシリボ核酸 (DNA)、リボ核酸 (RNA)、またはそれらの変異体を指す。核酸は、アデノシン (A)、シトシン (C)、グアニン (G)、チミン (T)、およびウラシル (U)、またはそれらの変異体から選択される 1 つまたは複数のサブユニットを含み得る。ヌクレオチドは、A、C、G、T、または U、またはそれらの変異体を含むことができる。ヌクレオチドは、成長しつつある核酸鎖に組み込み可能な任意のサブユニットを含むことができる。そのようなサブユニットは、より相補的な A、C、G、T、もしくは U のうちの 1 つに固有であり得る、またはプリン (すなわち、A もしくは G、またはそれらの変異体) もしくはピリミジン (すなわち、C、T、もしくは U、またはそれらの変異体) に相補的であり得る、A、C、G、T、もしくは U、または他の任意のサブユニットであってよい。いくつかの例では、核酸は、一本鎖であってもよいし、二本鎖であってもよく、いくつかのケースでは、核酸は環状である。

10

【 0 0 6 3 】

本明細書において使用される「核酸分子」または「核酸配列」という用語は、一般に、デオキシリボヌクレオチド (DNA) もしくはリボヌクレオチド (RNA) またはそれらの類似体である、種々の長さを有することがある、ヌクレオチドのポリマー形式すなわちポリヌクレオチドを指す。「核酸配列」という用語は、ポリヌクレオチドのアルファベット表現を指すことがある。代替的に、その用語は、物理的ポリヌクレオチドそれ自体に適用されることもある。このアルファベット表現は、中央処理装置を有し、核酸配列または核酸分子をシンボルまたはビットにマッピングし、デジタル情報を符号化するために使用される、コンピュータ内のデータベースに入力可能である。核酸配列またはオリゴヌクレオチドは、1 つまたは複数の非標準的なヌクレオチド、ヌクレオチド類似体、および / または修飾ヌクレオチドを含むことがある。

20

【 0 0 6 4 】

本明細書において使用される「オリゴヌクレオチド」は、一般に、一本鎖核酸配列を指し、典型的には、4 つのヌクレオチド塩基、すなわち、アデニン (A)、シトシン (C)、グアニン (G)、およびチミン (T)、またはポリヌクレオチドが RNA であるときはウラシル (U) の固有配列から構成される。

30

【 0 0 6 5 】

修飾ヌクレオチドの例は、限定するものではないが、ジアミノプリン、5 - フルオロウラシル、5 - ブロモウラシル、5 - クロロウラシル、5 - ヨードウラシル、ヒポキサンチン、キサチン (xanthine)、4 - アセチルシトシン、5 - (カルボキシヒドロキシルメチル) ウラシル、5 - カルボキシメチルアミノメチル - 2 - チオウリジン、5 - カルボキシメチルアミノメチルウラシル、ジヒドロウラシル、ベータ - D - ガラクトシルキューエオシン (galactosyl queosine)、イノシン、N6 - イソペンテニルアデニン、1 - メチルグアニン、1 - メチルイノシン、2, 2 - ジメチルグアニン、2 - メチルアデニン、2 - メチルグアニン、3 - メチルシトシン、5 - メチルシトシン、N6 - アデニン、7 - メチルグアニン、5 - メチルアミノメチルウラシル、5 - メトキシアミノメチル - 2 - チオウラシル、ベータ - D - マンノシルキューエオシン (mannosyl queosine)、5' - メトキシカルボキシメチルウラシル、5 - メトキシウラシル、2 - メチルチオ - D46 - イソペンテニルアデニン、ウラシル - 5 - オキシ酢酸 (v)、ワイブトソシン (wybutoxosine)、偽ウラシル (pseudouracil)、キューエオシン (queosine)、2 - チオシトシン、5 - メチル - 2 - チオウラシル、2 - チオウラシル、4 - チオウラシル、5 - メチルウラシル、ウラシル - 5 - オキシ酢酸メチルエステル、ウラシル - 5 - オキシ酢酸 (v)、5 - メチル - 2 - チオウラシル、3 - (3 - アミノ - 3 - N - 2 - カルボキシプロピル) ウラシル、(acp3)w、2, 6 - ジアミノプリンなどを含む。核酸分子はまた、塩基部分 (たとえば、典型的には

40

50

相補的ヌクレオチドとの水素結合を形成するために利用可能である1つもしくは複数の原子において、および/または典型的には相補的ヌクレオチドとの水素結合を形成することが可能でない1つまたは複数の原子において)において修飾されることもあり、糖部分において修飾されることもあり、リン酸骨格において修飾されることもある。核酸分子は、N-ヒドロキシコハク酸イミドエステル(NHS)などのアミン反応部分の共有付着(attachment)を可能にするために、アミノアリル-dUTP(aa-dUTP)およびアミノヘキシルアクリルアミド(aminohexylacrylamide)-dCTP(aha-dCTP)などのアミン修飾基も含有することがある。

【0066】

本明細書において使用される「プライマー」という用語は、一般に、ポリメラーゼ連鎖反応(PCR)などの核酸合成の開始点として働くことができる核酸の鎖を指す。一例では、DNA試料の複製中に、複製を触媒する酵素は、DNA試料に付着されたプライマーの3'-末端において複製を開始し、反対の鎖をコピーする。

【0067】

本明細書において使用される「ポリメラーゼ」または「ポリメラーゼ酵素」という用語は、一般に、ポリメラーゼ反応を触媒することが可能である任意の酵素を指す。ポリメラーゼの例は、限定なしで、核酸ポリメラーゼを含む。ポリメラーゼは、天然に存在し得るか、または合成され得る。例となるポリメラーゼは、29ポリメラーゼまたはその誘導体である。いくつかのケースでは、転写酵素またはリガーゼは、新しい核酸配列を構築するために、ポリメラーゼとともに、またはポリメラーゼの代替として、使用される(すなわち、結合の形成を触媒する酵素)。ポリメラーゼの例は、DNAポリメラーゼ、RNAポリメラーゼ、耐熱性ポリメラーゼ、野生型リメラーゼ、修飾ポリメラーゼ、大腸菌DNAポリメラーゼI、T7DNAポリメラーゼ、バクテリオファージT4DNAポリメラーゼ29(phi29)DNAポリメラーゼ、Taqポリメラーゼ、Tthポリメラーゼ、Tliポリメラーゼ、PfuポリメラーゼPwoポリメラーゼ、VENTポリメラーゼ、DEEPVENTポリメラーゼ、Ex-Taqポリメラーゼ、LA-Taqポリメラーゼ、SsoポリメラーゼPocポリメラーゼ、Pabポリメラーゼ、MthポリメラーゼES4ポリメラーゼ、Truポリメラーゼ、Tacポリメラーゼ、Tneポリメラーゼ、Tmaポリメラーゼ、Tcaポリメラーゼ、Tihポリメラーゼ、Tfiポリメラーゼ、プラチナTaqポリメラーゼ、Tbrポリメラーゼ、Tflポリメラーゼ、Pfu tub oポリメラーゼ、Pyrobestポリメラーゼ、KODポリメラーゼ、Bstポリメラーゼ、Sacポリメラーゼ、3'~5'エキソヌクレアーゼ活性をもつKlenow断片ポリメラーゼ、ならびにそれらの変異体、修飾生産物、および誘導体を含む。

【0068】

バイナリコードの形式をしたコンピュータデータなどのデジタル情報は、配列またはシンボルの文字列を含むことができる。バイナリコードは、たとえば、ビットと呼ばれる、典型的には0および1である、2つの2進シンボルを有する2進数法を使用して、テキストまたはコンピュータプロセッサ命令を符号化するまたは表すことがある。デジタル情報は、非バイナリシンボルの配列を含むことができる非バイナリコードの形式で表されることがある。各符号化されたシンボルは、一意のビット文字列(すなわち「バイト」)に再度割り当て可能であり、一意のビット文字列またはバイトは、バイトの文字列またはバイトストリームへと配置可能である。所与のビットに対するビット値は、2つのシンボルのうちの1つ(たとえば、0または1)とすることができる。Nビットの文字列を含むことができるバイトは、合計 2^N の一意のバイト値を有することができる。たとえば、8つのビットを含むバイトは、合計 2^8 すなわち256の可能な一意のバイト値を生じさせることができ、256バイトの各々は、バイトを用いて符号化可能である256の可能な別個のシンボル、文字、または命令のうちの1つに対応することができる。生データ(たとえば、テキストファイルおよびコンピュータ命令)は、バイトの文字列すなわちバイトストリームとして表されることが可能である。生データを含むZipファイル、または圧縮されたデータファイルは、バイトストリーム内に記憶可能であり、これらのファイルは、圧

10

20

30

40

50

縮された形式でのバイトストリームとして記憶され、次いで、コンピュータによって読み取られる前に生データへと解凍可能である。

【0069】

本開示の方法およびシステムは、コンピュータデータまたは複数の識別子内の情報を符号化するために使用されることがあり、その各々は、元の情報の1つまたは複数のビットを表し得る。いくつかの例では、本開示の方法およびシステムは、各々が元の情報の2ビットを表す識別子を使用して、データまたは情報を符号化する。

【0070】

デジタル情報を核酸に符号化するための従前の方法は、コストが高く時間がかかることがある核酸の塩基単位合成に依拠している。代替方法は、効率を改善し、デジタル情報を符号化するための塩基単位核酸合成に対する依拠を減少させることによってデジタル情報記憶の商業的実行可能性を改善し、あらゆる新しい情報記憶要求に対する別個の核酸配列の新規合成をなくすことがある。

10

【0071】

新しい方法は、塩基単位核酸合成または新規核酸合成（たとえば、ホスホラミダイト合成）に依拠する代わりに、構成要素の組合せ配置を含む、複数の識別子内のデジタル情報（たとえば、バイナリコード）、または核酸配列を符号化することができる。したがって、新しい戦略は、情報記憶の第1の要求のために別個の核酸配列（または構成要素）の第1のセットを生じさせることがあり、その後、その後の情報記憶要求に同じ核酸配列（または構成要素）を再使用することができる。これらの手法は、情報-DNA符号化および書込みプロセスにおける核酸配列の新規合成の役割を減少させることによって、DNAベースの情報記憶のコストを大幅に減少させることができる。さらに、各伸長しつつある核酸への各塩基の周期的送達を使用することがある、ホスホラミダイト化学なしまたは鋳型なしポリメラーゼベース核酸伸長などの塩基単位合成の実装形態とは異なり、構成要素からの識別子構造を使用した情報-DNA書込みの新しい方法は、周期的核酸伸長を必ずしも使用するとは限らない、高度に並列化可能なプロセスである。したがって、新しい方法は、古い方法と比較して、デジタル情報をDNAに書き込むスピードを増加させ得る。

20

【0072】

情報を符号化し、核酸配列に書き込むための方法

一態様では、本開示は、情報を核酸配列に符号化するための方法を提供する。情報を核酸配列に符号化するための方法は、(a)情報をシンボルの文字列に翻訳することと、(b)シンボルの文字列を複数の識別子にマッピングすることと、(c)複数の識別子の少なくともサブセットを含む識別子ライブラリを構築することとを含むことがある。複数の識別子の個々の識別子は、1つまたは複数の構成要素を含むことがある。1つまたは複数の構成要素の個々の構成要素は、核酸配列を含むことがある。シンボルの文字列内の各位置における各シンボルは、別個の識別子に対応することがある。個々の識別子は、シンボルの文字列内の個々の位置における個々のシンボルに対応することがある。さらに、シンボルの文字列内の各位置における1つのシンボルは、識別子の不在に対応することがある。たとえば、「0」および「1」の2進シンボル（たとえば、ビット）の文字列内で、「0」の各発生は、識別子の不在に対応することがある。

30

40

【0073】

別の態様では、本開示は、核酸ベースのコンピュータデータ記憶のための方法を提供する。核酸ベースのコンピュータデータ記憶のための方法は、(a)コンピュータデータを受信することと、(b)コンピュータデータを符号化する核酸配列を含む核酸分子を合成することと、(c)核酸配列を有する核酸分子を記憶することとを含むことがある。コンピュータデータは、合成され、核酸分子の各々の配列内にない核酸分子の少なくともサブセット内に符号化されることがある。

【0074】

別の態様では、本開示は、情報を核酸配列内に書き込み、記憶するための方法を提供する。この方法は、(a)情報を表す仮想識別子ライブラリを受信または符号化することと

50

、(b) 識別子ライブラリを物理的に構築することと、(c) 識別子ライブラリの1つまたは複数の物理的コピーを1つまたは複数の別々のロケーション内に記憶することとを含むことがある。識別子ライブラリの個々の識別子は、1つまたは複数の構成要素を含むことがある。1つまたは複数の構成要素の個々の構成要素は、核酸配列を含むことがある。

【0075】

別の態様では、本開示は、核酸ベースのコンピュータデータ記憶のための方法を提供する。核酸ベースのコンピュータデータ記憶のための方法は、(a) コンピュータデータを受信することと、(b) コンピュータデータを符号化する少なくとも1つの核酸配列を含む核酸分子を合成することと、(c) 少なくとも1つの核酸配列を含む核酸分子を記憶することとを含むことがある。核酸分子を合成することは、塩基単位核酸合成の不在下であることがある。

10

【0076】

別の態様では、本開示は、情報を核酸配列内に書き込み、記憶するための方法を提供する。情報を核酸配列内に書き込み、記憶するための方法は、(a) 情報を表す仮想識別子ライブラリを受信または符号化することと、(b) 識別子ライブラリを物理的に構築することと、(c) 識別子ライブラリの1つまたは複数の物理的コピーを1つまたは複数の別々のロケーション内に記憶することとを含むことがある。識別子ライブラリの個々の識別子は、1つまたは複数の構成要素を含むことがある。1つまたは複数の構成要素の個々の構成要素は、核酸配列を含むことがある。

【0077】

20

図1は、情報を核酸配列に符号化し、情報を核酸配列に書き込み、核酸配列に書き込まれた情報を読み取り、読み取られた情報を復号するための概要プロセスを例示する。デジタル情報すなわちデータは、シンボルの1つまたは複数の文字列へと翻訳されることがある。一例では、シンボルはビットであり、各ビットは、「0」または「1」のどちらかの値を有することがある。各シンボルは、そのシンボルを表すオブジェクト（たとえば、識別子）にマップされることがあり、またはこれに符号化されることがある。各シンボルは、別個の識別子によって表されることがある。別個の識別子は、構成要素から構成される核酸分子であることがある。構成要素は、核酸配列であることがある。デジタル情報は、情報に対応する識別子ライブラリを生成することによって、核酸配列へと書き込まれることがある。識別子ライブラリは、デジタル情報の各シンボルに対応する識別子を物理的に構築することによって、物理的に生成されることがある。デジタル情報のすべてまたは任意の部分は、一度にアクセスされてよい。一例では、識別子のサブセットが、識別子ライブラリからアクセスされる。識別子のサブセットは、配列決定し、識別子を識別することによって読み取られることがある。識別された識別子は、デジタルデータを復号するために、対応するシンボルと関連づけられることがある。

30

【0078】

図1の手法を使用して情報を符号化し、読み取るための方法は、たとえば、ビットストリームを受信することと、識別子ランクまたは核酸インデックスを使用して、ビットストリーム内の各1つのビット（「1」のビット値をもつビット）を別個の核酸識別子にマッピングすることとを含むことができる。核酸試料プール、または識別子ライブラリを構築することは、1のビット値に対応する識別子のコピー（と、0のビット値に対する識別子を除外することと）を含む。試料を読み取ることは、分子生物学方法（たとえば、配列決定、ハイブリダイゼーション、PCRなど）を使用することと、どの識別子が識別子ライブラリ内で表されるかを決定することと、それらの識別子に対応するビットに「1」のビット値を、他のところに「0」のビット値を割り当てること（各識別子に対応する元のビットストリーム内のビットを識別するために識別子ランクを再び参照して）と、したがって情報を元の符号化されたビットストリームに復号することとを含むことができる。

40

【0079】

Nの別個のビットの文字列を符号化することは、等しい数の一意の核酸配列を可能な識別子として使用することができる。情報符号化に対するこの手法は、記憶することになる

50

情報（ N ビットの文字列）の各新しい項目に対する識別子（たとえば、核酸分子）の新規合成を使用することがある。他の事例では、記憶することになる情報の各新しい項目に対して識別子（数が N に等しいまたはそれよりも小さい）を新しく合成するコストは、1回限りの新規合成およびすべての可能な識別子のその後の維持によって減少可能であり、したがって、情報の新しい項目を符号化することは、識別子ライブラリを形成するためにあらかじめ合成された（または、あらかじめ製作された）識別子を機械的に選択し一緒に混合することを伴うことがある。他の事例では、（1）記憶することになる情報の各新しい項目に対する最大 N 個の識別子の新規合成または（2）記憶することになる情報の各新しい項目に対して N 個の可能な識別子を維持し、これから選択すること、またはそれらの組合せ、の両方のコストは、ある数（ N よりも小さく、いくつかのケースでは、 N よりもはるかに小さい）の核酸配列を合成および維持し、次いで、記憶することになる情報の各新しい項目のために最大 N 個の識別子を生成するために酵素反応を通してこれらの配列を修飾することによって、減少されることがある。

【0080】

識別子は、読取り動作、書込み動作、アクセス動作、コピー動作、および削除動作を簡単にするために、合理的に設計および選択されてよい。識別子は、書込みエラー、突然変異、分解、および読取りエラーを最小にするように設計および選択されてよい。

【0081】

図2Aおよび図2Bは、オブジェクトまたは識別子（たとえば、核酸分子）内のデジタルデータを符号化する、「アドレスにおけるデータ」と呼ばれる例となる方法を概略的に例示する。図2Aは、ビットストリームを識別子ライブラリに符号化することを例示し、個々の識別子は、識別子ランクを指定する単一の構成要素をバイト値を指定する単一の構成要素と連結するまたは組み立てることによって構築される。一般に、アドレスにおけるデータ方法は、2つのオブジェクト、すなわち、バイト値を識別する、1つのオブジェクトである「バイト値オブジェクト」（または「データオブジェクト」）と、識別子ランク（または元のビットストリーム内のバイトの相対的位置）を識別する、1つのオブジェクトである「ランクオブジェクト」（または「アドレスオブジェクト」）、を含むことによって情報をモジュール的に符号化する識別子を使用する。図2Bは、各ランクオブジェクトが構成要素のセットから組合せにより構築されることがあり、各バイト値オブジェクトが構成要素のセットから組合せにより構築されることがあるアドレスにおけるデータ方法の一例を例示する。ランクオブジェクトおよびバイト値オブジェクトのそのような組合せ構造は、単一の構成要素のみから作られたオブジェクト（たとえば、図2A）よりも多くの情報が識別子に書き込まれることを可能にする。

【0082】

図3Aおよび図3Bは、オブジェクトまたは識別子（たとえば、核酸配列）内のデジタル情報を符号化する別の例となる方法を概略的に例示する。図3Aは、ビットストリームを識別子ライブラリに符号化することを例示し、識別子は、識別子ランクを指定する単一の構成要素から構築される。特定のランク（またはアドレス）における識別子の存在は、「1」のビット値を指定し、特定のランク（またはアドレス）における識別子の不在は、「0」のビット値を指定する。このタイプの符号化は、ランク（元のビットストリーム内のビットの相対的位置）のみを符号化する識別子を使用し、それらの識別子の識別子ライブラリ内での存在または不在を使用して、「1」または「0」のビット値をそれぞれ符号化することがある。情報を読み取ることおよび復号することは、識別子ライブラリ内に存在する識別子を識別することと、「1」のビット値をそれらの対応するランクに割り当てることと、「0」のビット値を他のところに割り当てることとを含むことがある。図3Bは、各可能な組合せ構造がランクを指定するように各識別子が構成要素のセットから組合せにより構築されることがある例となる符号化する方法を例示する。そのような組合せ構造は、単一の構成要素のみから作られた識別子（たとえば、図3A）よりも多くの情報が識別子に書き込まれることを可能にする。たとえば、構成要素セットは、5つの別個の構成要素を含むことがある。5つの別個の構成要素は、各々が5つの構成要素のうちの2つ

10

20

30

40

50

を含む 10 の別個の識別子を生成するために組み立てられることがある。10 の別個の識別子は各々、ビットストリーム内のビットの位置に対応するランク（またはアドレス）を有することがある。識別子ライブラリは、ビット値「1」の位置に対応するそれらの 10 の可能な識別子のサブセットを含み、長さ 10 のビットストリーム内のビット値「0」の位置に対応するそれらの 10 の可能な識別子のサブセットを除外することがある。

【0083】

図 4 は、図 3 A および図 3 B に示される符号化する方法を使用して、可能な識別子の組合せ空間（C、x 軸）とビット単位の所与の元のサイズの情報（D、輪郭線）を記憶するように物理的に構築され得る識別子の平均数（k、y 軸）との間の関係の対数空間内の輪郭プロットを図示する。このプロットは、いくつかのビット k が「1」のビット値を有する場合、サイズ D の元の情報が、C ビットの文字列（ここで、C は D よりも大きくてよい）に再コードされると仮定する。さらに、プロットは、情報 - 核酸符号化が、再コードされたビット文字列に対して実行され、ビット値が「1」である位置に対する識別子が構築され、ビット値が「0」である位置に対する識別子は構築されないと仮定する。仮定に従って、可能な識別子の組合せ空間は、再コードされたビット文字列内のあらゆる位置を識別するためにサイズ C を有し、サイズ D のビット文字列を符号化するために使用される識別子の数は、 $D = \log_2(C \text{ choose } k)$ であるようなものであり、ここで、 $C \text{ choose } k$ は、C 個の可能性から k 個の無秩序な結果を取り上げる手段の数に関する数式であってよい。したがって、可能な識別子の組合せ空間が、情報の所与の項目のサイズ（ビット単位）を越えて増加するにつれて、所与の情報を記憶するために使用される物理的に構築される識別子の数が、減少することがある。

【0084】

図 5 は、情報を核酸配列に書き込むための概要方法を図示する。情報を書き込む前に、情報は、シンボルの文字列に翻訳され、複数の識別子に符号化されてよい。情報を書き込むことは、可能な識別子を生じさせるように反応を設定することを含むことがある。反応は、入力をコンパートメントに置くことによって設定されることがある。入力、核酸、構成要素、鋳型、酵素、または化学試薬を含むことがある。コンパートメントは、ウェル、チューブ、面上の位置、マイクロ流体デバイス内のチャンバ、またはエマルション内の液滴であってよい。複数の反応が、複数のコンパートメント内で設定されてよい。反応は、プログラムされた温度インキュベーションまたはサイクリングを通じて識別子を生じさせるように進んでよい。反応は、選択的にまたは遍在的に除去（たとえば、削除）されてよい。反応は、識別子を 1 つのプール内に収集するように、選択的にまたは遍在的に中断、合併、および精製されてもよい。複数の識別子ライブラリからの識別子は、同じプール内に収集されることがある。個々の識別子は、どの識別子ライブラリにそれが属するかを識別するためにバーコードまたはタグを含むことがある。代替的に、またはこれに加えて、バーコードは、符号化された情報のためのメタデータを含むことがある。補足的核酸または識別子が、識別子ライブラリと一緒に識別子プール内に含まれることもある。補足的核酸または識別子は、符号化された情報のためのメタデータを含むまたは符号化された情報を不明瞭にするまたは隠蔽するために働くことがある。

【0085】

識別子ランク（たとえば、核酸インデックス）は、識別子の順序づけを決定するための方法またはキーを含むことができる。この方法は、すべての識別子とそれらの対応するランクとをもつルックアップテーブルを含むことができる。この方法は、識別子を構成するすべての構成要素のランクと、それらの構成要素の組合せを含む任意の識別子の順序づけを決定するための関数とをもつルックアップテーブルも含むことができる。そのような方法は、辞書式順序づけと呼ばれることがあり、辞書内の単語がアルファベット順に順序づけられる様式に類似することがある。アドレスにおけるデータ符号化方法では、識別子ランク（識別子のランクオブジェクトによって符号化される）は、ビットストリーム内のバイトの位置（識別子のバイト値オブジェクトによって符号化される）を決定するために使用されることがある。代替の方法では、現在の識別子のための識別子ランク（識別子自体

全体によって符号化される)は、ビットストリーム内の「1」のビット値の位置を決定するために使用されることがある。

【0086】

キーは、別個のバイトを、試料内の識別子(たとえば、核酸分子)の一意のサブセットに割り当てることがある。たとえば、単純な形式では、キーは、バイト内の各ビットを、ビットの位置を指定する一意の核酸配列に割り当ててよく、次いで、試料内でのその核酸配列の存在または不在はそれぞれ、1または0のビット値を指定することがある。符号化された情報を核酸試料から読み取ることは、配列決定、ハイブリダイゼーション、またはPCRを含む任意の数の分子生物学技法を含むことができる。一部の実施形態では、符号化されたデータセットを読み取ることは、データセットの一部分を再構築すること、または符号化されたデータセット全体を各核酸試料から再構築することを含むことがある。配列が読み取られ得るとき、核酸インデックスは、一意の核酸配列の存在または不在とともに使用可能であり、核酸試料は、ビットストリーム(たとえば、ビットの各文字列、バイト、複数のバイト、またはバイトの文字列)へと復号(decode)可能である。

【0087】

識別子は、構成要素核酸配列を組合せにより組み立てることによって構築されることがある。たとえば、情報は、核酸分子(たとえば、識別子)のセットを分子の規定された群(たとえば、組合せ空間)から得られることによって符号化されることがある。分子の規定された群の各可能な識別子は、層に分割され得る構成要素のあらかじめ製作されたセットからの核酸配列(たとえば、構成要素)のアセンブリであってよい。各個々の識別子は、あらゆる層からの1つの構成要素を固定順序で連結することによって、構築されることがある。たとえば、M個の層があり、各層がn個の構成要素を有することがある場合、最大 $C = n^M$ 個の一意の識別子が構築されることがあり、最大 2^C 個の情報の異なる項目またはCビットが、符号化および記憶されることがある。たとえば、情報のメガビットの記憶は、 1×10^6 の別個の識別子またはサイズ $C = 1 \times 10^6$ の組合せ空間を使用することがある。この例の識別子は、異なる手段で編成されたさまざまな構成要素から組み立てられることがある。アセンブリは、各々が $n = 1 \times 10^3$ の構成要素を含有する、 $M = 2$ のあらかじめ製作された層から作られることがある。代替として、アセンブリは、各々が $n = 1 \times 10^2$ の構成要素を含有する、 $M = 3$ の層から作られることがある。この例が例示するように、より大きな数の層を使用して同じ量の情報を符号化することは、構成要素の総数をより小さくすることを可能にし得る。より小さい数の全構成要素を使用することは、書込みコストに関して有利であることがある。

【0088】

一例では、各々がそれぞれx個およびy個の構成要素(たとえば、核酸配列)をもつ、一意の核酸配列または層XおよびYの2つのセットで開始することができる。Xからの各核酸配列は、Yからの各核酸配列に組み立て可能である。2つのセット内で維持される核酸配列の総数はxとyの合計であり得るが、生成可能な核酸分子の総数、したがって可能な識別子の総数は、xとyの積であってよい。Xからの配列が、任意の順序でYの配列に組み立て可能である場合、はるかに多くの核酸配列(たとえば、識別子)が生成可能である。たとえば、生成される核酸配列(たとえば、識別子)の数は、アセンブリ順序がプログラム可能である場合、xとyの積の2倍であってよい。生成可能であるすべての可能な核酸配列のこのセットは、XYと呼ばれることがある。XY内の一意の核酸配列の組み立てられたユニットの順序は、別個の5'末端および3'末端をもつ核酸、ならびに制限消化、ライゲーション、ポリメラーゼ連鎖反応(PCR)を使用して制御可能であり、配列決定は、配列の別個の5'末端および3'末端に関して発生することがある。そのような手法は、それらのアセンブリ生産物の組合せおよび順序における情報を符号化することによって、N個の別個のビットを符号化するために使用される核酸配列(たとえば、構成要素)の総数を減少させることができる。たとえば、100ビットの情報を符号化するために、10の別個の核酸分子(たとえば、構成要素)の2つの層が、 10×10 すなわち100の別個の核酸分子(たとえば、識別子)を生じさせるために、固定順序で組み立てられる

ことがある、または5つの別個の核酸分子（たとえば、構成要素）の1つの層および10の別個の核酸分子（たとえば、構成要素）の別の層は、100の別個の核酸分子（たとえば、識別子）を生じさせるために、任意の順序で組み立てられることがある。

【0089】

各層内の核酸配列（たとえば、構成要素）は、一意の（または別個の）配列、またはバーコードを中央に、共通ハイブリダイゼーション領域を一方の末端上に、別の共通ハイブリダイゼーション領域を別の他方の末端上に含むことがある。バーコードは、層内のあらゆる配列を一意に識別するのに十分な数のヌクレオチドを含有することがある。たとえば、典型的には、バーコード内の各塩基位置に対して4つの可能なヌクレオチドがある。したがって、3つの塩基バーコードは、 $4^3 = 64$ の核酸配列を一意に識別することがある。バーコードは、ランダムに生成されるように設計されることがある。代替的に、バーコードは、識別子の構造化学または配列決定の複雑化を作成する配列を回避するように設計されることがある。加えて、バーコードは、各々が、他のバーコードから最小ハミング距離を有し、それによって、塩基分解能（resolution）突然変異または読取りエラーがバーコードの適切な識別に干渉し得る可能性を減少させ得るように設計されることがある。

10

【0090】

核酸配列（たとえば、構成要素）の一方の末端上のハイブリダイゼーション領域は、各層において異なってもよいが、ハイブリダイゼーション領域は、層内の各部材に対して同じであってもよい。隣接する層は、それらが互いに対話することを可能にする構成要素上に相補的ハイブリダイゼーション領域を有する層である。たとえば、層Xからの任意の構成要素は、層Yからの任意の構成要素に付着することを可能にすることがあるが、これは、それらが相補的ハイブリダイゼーション領域を有することがあるからである。反対側の末端上のハイブリダイゼーション領域は、第1の末端上のハイブリダイゼーション領域と同じ目的を果たすことがある。たとえば、層Yからの任意の構成要素は、一方の末端上の層Xの任意の構成要素および反対側の末端上の層Zの任意の構成要素に付着することがある。

20

【0091】

図6Aおよび図6Bは、各層からの別個の構成要素（たとえば、核酸配列）を固定順序で組合せにより組み立てることによって識別子（たとえば、核酸分子）を構築するための、「生産物方式」と呼ばれる例となる方法を例示する。図6Aは、生産物方式を使用して構築された識別子のアーキテクチャを例示する。識別子は、各層からの単一の構成要素を固定順序で組み合わせることによって構築されることがある。各々がN個の構成要素をもつM個の層の場合、 N^M 個の可能な識別子がある。図6Bは、生産物方式を使用して構築され得る識別子の組合せ空間の一例を例示する。一例では、組合せ空間は、各々が3つの別個の構成要素を含む3つの層から生成されることがある。構成要素は、各層からの1つの構成要素が固定順序で組み合わせられ得るように組み合わせられることがある。このアセンブリ方法に対する組合せ空間全体は、27個の可能な識別子を含むことがある。

30

【0092】

図7～図10は、生産物方式を実装するための化学的方法を例示する（図6を参照されたい）。図7～図10に描かれている方法は、2つまたはこれより多くの別個の構成要素を固定順序で組み立てるための他の任意の方法とともに、たとえば、識別子ライブラリ内の任意の1つまたは複数の識別子を生じさせるために使用されることがある。識別子は、本明細書で開示される方法またはシステム中の任意の時間に、図7～図10において説明される実装方法のいずれかを使用して構築されることがある。いくつかの事例では、可能な識別子の組合せ空間のすべてまたは一部分は、デジタル情報が符号化されるまたは書き込まれる前に構築されることがあり、その場合、書込みプロセスは、（情報を符号化する）識別子を既に存在するセットから機械的に選択およびプールすることを伴うことがある。他の事例では、識別子は、データ符号化または書込みプロセスの1つまたは複数のステップが発生した後（すなわち、情報が書き込まれているとき）に構築されることがある。

40

【0093】

50

酵素反応は、異なる層またはセットから構成要素を組み立てるために使用されることがある。各層の構成要素（たとえば、核酸配列）が、隣接する層の構成要素に対する特異的ハイブリダイゼーションまたは付着領域を有するので、アセンブリは、ワンポット反応において発生することができる。たとえば、層 X からの核酸配列（たとえば、構成要素） X_1 、層 Y からの核酸配列 Y_1 、および層 Z からの核酸配列 Z_1 は、組み立てられた核酸分子（たとえば、識別子） $X_1 Y_1 Z_1$ を形成することがある。加えて、複数の核酸分子（たとえば、識別子）は、各層からの複数の核酸配列を含むことによって、1つの反応において組み立てられることがある。たとえば、前の例のワンポット反応において Y_1 と Y_2 の両方を含むことが、2つの組み立てられた生産物（たとえば、識別子）すなわち $X_1 Y_1 Z_1$ および $X_1 Y_2 Z_1$ を生ずることがある。この反応多重化は、物理的に構築される複数の識別子に対する書込み時間をスピードアップするために使用されることがある。核酸配列のアセンブリは、約 1 日、12 時間、10 時間、9 時間、8 時間、7 時間、6 時間、5 時間、4 時間、3 時間、2 時間、または 1 時間よりも短いまたはこれに等しい時間期間内に実行されることがある。符号化されたデータの精度は、少なくとも約 90%、95%、96%、97%、98%、99%、またはこれらに等しいまたはこれらよりも大きいことがある。

10

【0094】

識別子は、図 7 において例示されるように、オーバーラップ伸長ポリメラーゼ連鎖反応（OEP CR）を使用して生産物方式により構築されることがある。各層内の各構成要素は、隣接する層からの構成要素の配列末端上の共通ハイブリダイゼーション領域と相同するおよび/または相補的であり得る共通ハイブリダイゼーション領域を配列末端上にもつ二本鎖または一方鎖（図に描かれるように）核酸配列を含むことがある。個々の識別子は、構成要素 $X_1 - X_A$ を含む層 X（または層 1）からの 1つの構成要素（たとえば、一意の配列）と、 $Y_1 - Y_A$ を含む層 Y（または層 2）からの第 2の構成要素（たとえば、一意の配列）と、 $Z_1 - Z_B$ を含む層 Z（または層 3）からの第 3の構成要素（たとえば、一意の配列）とを連結することによって構築されることがある。層 X からの構成要素は、層 Y からの構成要素上の 3' 末端との相補性を共有する 3' 末端を有することがある。したがって、層 X および Y からの一本鎖構成要素は、3' 末端において一緒にアニーリングされることがあり、二本鎖核酸分子を生成するために PCR を使用して延長されることがある。生成された二本鎖核酸分子は、層 Z からの構成要素の 3' 末端との相補性を共有する 3' 末端を生成するように融解されることがある。層 Z からの構成要素は、生成された核酸分子とともにアニーリングされることがあり、固定順序で層 X、Y、および Z からの単一の構成要素を含む一意の識別子を生成するように延長されることがある。DNA サイズ選択（たとえば、ゲル抽出を用いた）または最も外側の層と隣接するプライマーとのポリメラーゼ連鎖反応（PCR）は、反応中に形成し得る他の副生産物から識別子生産物を隔離するために実装されることがある。

20

30

【0095】

識別子は、図 8 において例示されるように、粘着末端ライゲーションを使用する生産物方式により組み立てられることがある。一本鎖 3' オーバーハングをもつ二本鎖構成要素（たとえば、二本鎖 DNA（dsDNA））を各々が含む 3つの層は、別個の識別子を組み立てるために使用可能である。たとえば、構成要素 $X_1 - X_A$ を含む層 X（または層 1）からの 1つの構成要素と、 $Y_1 - Y_B$ を含む層 Y（または層 2）からの第 2の構成要素と、 $Z_1 - Z_C$ を含む層 Z（または層 3）からの第 3の構成要素とを含む識別子。層 X からの構成要素を層 Y からの構成要素と組み合わせるために、層 X 内の構成要素は、図 8 では a と標識される共通 3' オーバーハングを含むことができ、層 Y 内の構成要素は、共通の相補的 3' オーバーハングすなわち a^* を含むことができる。層 Y からの構成要素を層 Z からの構成要素と組み合わせるために、層 Y 内のエレメントは、図 8 では b と標識される共通 3' オーバーハングを含むことができ、層 Z 内のエレメントは、共通の相補的 3' オーバーハングすなわち b^* を含むことができる。層 X 構成要素内の 3' オーバーハングは、層 Y 構成要素内の 3' 末端と相補的であることができ、層 Y 構成要素内の他の 3' オーバーハング

40

50

は、層 Z 構成要素内の 3' 末端と相補的であることができ、構成要素がハイブリダイゼーションおよびライゲーションすることを可能にする。したがって、層 X からの構成要素は、層 X または層 Z からの他の構成要素とハイブリダイゼーションすることはできず、同様に、層 Y からの構成要素は、層 Y からの他のエレメントとハイブリダイゼーションすることはできない。そのうえ、層 Y からの単一の構成要素は、層 X の単一の構成要素および層 Z の単一の構成要素にライゲーションし、完全な識別子の形成を保証することができる。DNA サイズ選択（たとえば、ゲル抽出を用いた）または最も外側の層と隣接するプライマーとの PCR は、反応中に形成し得る他の副生産物から識別子生産物を隔離するために実装されることがある。

【0096】

粘着末端ライゲーションのための粘着末端は、制限エンドヌクレアーゼで各層の構成要素を処理することによって生成されることがある。一部の実施形態では、複数の層の構成要素は、構成要素の 1 つの「親」セットから生成されることがある。たとえば、二本鎖構成要素の単一の親セットが各末端上に相補的な制限部位（たとえば、BamHI および BglII のための制限部位）を有することがある一実施形態。任意の 2 つの構成要素が、アセンブリのために選択され、1 つまたは他の相補的な制限酵素（たとえば、BglII または BamHI）を用いて個々に消化され、不活性瘢痕（scar）となる一緒にライゲーション可能な相補的な粘着末端になることがある。生産物核酸配列は、各末端上に相補的な制限部位を含むことがあり（たとえば、5' 末端上の BamHI および 3' 末端上の BglII）、同じプロセスに従う親セットからの別の構成要素にさらにライゲーション可能である。このプロセスは、無期限に循環することがある。親が N 個の構成要素を含む場合、各サイクルは、N 個の構成要素の余剰層を生産物方式に追加することに等しくてよい。

【0097】

ライゲーションを使用して、セット X（たとえば、dsDNA のセット 1）からのエレメントとセット Y（たとえば、dsDNA のセット 2）からのエレメントとを含む核酸の配列を構築するための方法は、二本鎖配列の 2 つまたはこれよりも多いプール（たとえば、dsDNA のセット 1 および dsDNA のセット 2）を取得または構築するステップであって、第 1 のセット（たとえば、dsDNA のセット 1）は粘着末端（たとえば、a）を含み、第 2 のセット（たとえば、dsDNA のセット 2）は、第 1 のセットの粘着末端に相補的である粘着末端（たとえば、a*）を含む、取得または構築するステップを含むことができる。第 1 のセット（たとえば、dsDNA のセット 1）からの任意の DNA および第 2 のセット（たとえば、dsDNA のセット 2）からの DNA の任意のサブセットは、組み合わせられ、組み立てられて、次いで、一緒にライゲーションされ、第 1 のセットからのエレメントと第 2 のセットからのエレメントとをもつ単一の二本鎖 DNA を形成することが可能である。

【0098】

識別子は、図 9 において例示されるように、部位特異的な組換えを使用する生産物方式により組み立てられることがある。識別子は、3 つの異なる層から構成要素を組み立てることによって構築されることがある。層 X（または層 1）内の構成要素は、分子の一方の側に attB_x リコンビナーゼ部位をもつ二本鎖分子を含むことがあり、層 Y（または層 2）からの構成要素は、一方の側に attP_x リコンビナーゼ部位を、他方の側に attB_y リコンビナーゼ部位をもつ二本鎖分子を含むことがあり、層 Z（または層 3）内の構成要素は、分子の一方の側に attP_y リコンビナーゼ部位を含むことがある。下付き文字によって示される、ペア内の attB 部位および attP 部位は、対応するリコンビナーゼ酵素の存在下で組み換えることが可能である。各層からの 1 つの構成要素は、層 X からの 1 つの構成要素が層 Y からの 1 つの構成要素と関連し、層 Y からの 1 つの構成要素が層 Z からの 1 つの構成要素と関連するように、組み合わせられてよい。1 つまたは複数のリコンビナーゼ酵素の適用は、構成要素を組み換えて、順序付き構成要素を含む二本鎖識別子を生成することがある。DNA サイズ選択（たとえば、ゲル抽出を用いた）または最も外側の層と隣接するプライマーとの PCR は、反応中に形成し得る他の副生産物から識別子生

10

20

30

40

50

産物を隔離するために実装されることがある。一般に、複数の直交する (o r t h o g o n a l) a t t B および a t t P ペアが使用されることがあり、各ペアは、余剰層からの構成要素を組み立てるために使用されることがある。リコンビナーゼの大型セリンファミリー (l a r g e - s e r i n e f a m i l y) の場合、リコンビナーゼあたり最大6つの直交する a t t B および a t t P ペアが生成されることがあり、複数の直交するリコンビナーゼもまた実装されることがある。たとえば、13の層が、12の直交する a t t B および a t t P ペア、すなわち、B x b I および P h i C 3 1 などの、2つの大型セリンリコンビナーゼの各々からの6つの直交するペアを使用することによって組み立てられることがある。a t t B および a t t P ペアの直交性 (o r t h o g o n a l i t y) は、1つのペアからの a t t B 部位が、別のペアからの a t t P 部位と反応しないことを保証する。これによって、異なる層からの構成要素が固定順序で組み立てられることが可能になる。リコンビナーゼ媒介性組換え反応は、実装されるリコンビナーゼシステムに応じて、可逆的であってもよいし、非可逆的であってもよい。たとえば、大型セリンリコンビナーゼファミリーは、高エネルギーコファクターを必要とすることなく非可逆的な組換え反応を触媒するが、チロシンリコンビナーゼファミリーは、可逆的な反応を触媒する。

【 0 0 9 9 】

識別子は、図 1 0 A において例示されるように、鋳型指向ライゲーション (T D L) を使用する生産物方式により構築されることがある。鋳型指向ライゲーションは、「鋳型」または「ステーブル (s t a p l e) 」と呼ばれる一本鎖状の核酸配列を利用して、識別子を形成するために構成要素の順序付きライゲーションを容易にする。鋳型は、隣接する層からの構成要素に同時にハイブリダイゼーションし、リガーゼがそれらをライゲーションする間、それらを互いに隣接して (5 ' 末端に対する 3 ' 末端) 保持する。図 1 0 A からの例では、一本鎖構成要素の3つの層またはセットが組み合わせられる。配列 a * に相補的である、3 ' 末端上の共通配列 a を共有する構成要素の第1の層 (たとえば、層 X または層 1)、配列 b * および c * に相補的である、それぞれ 5 ' 末端および 3 ' 末端上の共通配列 b および c を共有する構成要素の第2の層 (たとえば、層 Y または層 2)、配列 d * に相補的であり得る、5 ' 末端上の共通配列 d を共有する構成要素の第3の層 (たとえば、層 Z または層 3)、ならびに配列 a * b * (5 ' から 3 ') を含む第1のステーブルと配列 c * d * (' 5 から 3 ') を含む第2のステーブルとをもつ2つの鋳型または「ステーブル」のセット。この例では、各層からの1つまたは複数の構成要素が選択され、ステーブルとの反応へと混合されることがあり、それが、相補的アニーリングによって、識別子を形成するために、規定された順序での各層からの1つの構成要素のライゲーションを容易にすることがある。DNA サイズ選択 (たとえば、ゲル抽出を用いた) または最も外側の層と隣接するプライマーとの P C R は、反応中に形成し得る他の副生産物から識別子生産物を隔離するために実装されることがある。

【 0 1 0 0 】

図 1 0 B は、各々が6層 T D L と組み立てられた256個の別個の核酸配列のコピー数 (存在度) のヒストグラムを示す。縁部層 (第1の層および最後の層) は各々、1つの構成要素を有し、内部層 (残りの4つの層) の各々は、4つの構成要素を有する。各縁部層構成要素は、10塩基ハイブリダイゼーション領域を含む28の塩基であった。各内部層構成要素は、5 ' 末端上の10塩基共通ハイブリダイゼーション領域と、10塩基可変 (バーコード) 領域と、3 ' 末端上の10塩基共通ハイブリダイゼーション領域とを含む30塩基であった。3つの鋳型鎖の各々は、長さが20塩基であった。全256個の別個の配列は、多重なやり方で組み立てられ、1つの反応は、構成要素および鋳型、T4ポリヌクレオチドキナーゼ (構成要素をリン酸化するための)、ならびにT4リガーゼ、A T P、および他の適切な反応試薬のすべてを含有した。反応は、37度で30分間、次いで室温で1時間インキュベートされた。配列決定アダプタが、P C R との反応生産物に追加され、この生産物は、I l l u m i n a M i S e q 機器を用いて配列決定された。192910の全組み立てられた配列読取りからの各別個の組み立てられた配列の相対的コピー数が図示されている。この方法の他の実施形態は、二本鎖構成要素を使用することがあり、そ

の構成要素は、最初は、ステーブルにアニーリングすることができる一本鎖バージョンを形成するように融解される。この方法（すなわち、TDL）の他の実施形態または派生物は、生産物方式において達成され得るものよりも複雑な識別子の組合せ空間を構築するために使用されることがある。

【0101】

識別子は、ゴールデンゲートアセンブリ、ギブソンアセンブリ、およびリガーゼサイクリング反応アセンブリを含む種々の他の化学的実装形態を使用する生産物方式により構築されることがある。

【0102】

図11Aおよび図11Bは、並べ替えられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「並べ替え方式（permutation scheme）」と呼ばれる例となる方法を概略的に例示する。図11Aは、並べ替え方式を使用して構築された識別子のアーキテクチャを例示する。識別子は、各層からの単一の構成要素をプログラム可能な順序で組み合わせることによって構築されることがある。図11Bは、並べ替え方式を使用して構築され得る識別子の組合せ空間の一例を例示する。一例では、サイズ6の組合せ空間が、各々が1つの別個の構成要素を含む3つの層から生成されることがある。構成要素は、任意の順序で連結されてよい。一般に、各々がN個の構成要素をもつM個の層では、並べ替え方式は、 $N^M M!$ 個の総識別子の組合せ空間を可能にする。

【0103】

図11Cは、鋳型指向ライゲーション（TDL）を有する並べ替え方式の例となる一実装形態を例示する。複数の層からの構成要素は、縁部スキャフォールドと呼ばれる、固定された左末端構成要素と右末端構成要素の間に組み立てられる。これらの縁部スキャフォールドは、組合せ空間内のすべての識別子に対して同じであり、したがって、実装のための反応マスタ混合の一部として追加されることがある。鋳型またはステーブルは、異なる層からの構成要素が反応において識別子に組み込まれる順序が、反応のために選択された鋳型に依存するように、任意の2つの層またはスキャフォールドの間の任意の可能な接合に対して存在する。M個の層に対する層の任意の可能な並べ替えを可能にするために、あらゆる可能な接合（スキャフォールドとの接合を含む）に対して $M^2 + 2M$ 個の別個の選択可能なステーブルがあることがある。それらの鋳型（灰色で陰影がつけられている）のうちのM個は、層とそれら自体との間の接合を形成し、本明細書において説明される並べ替えアセンブリの目的のために除外されることがある。しかしながら、それらを含めることによって、図11D～図11Gにおいて例示される繰り返し構成要素を含む識別子をもつより大きな組合せ空間を可能にすることができる。DNAサイズ選択（たとえば、ゲル抽出を用いた）または縁部スキャフォールドを標的とするプライマーとのPCRは、反応中に形成し得る他の副生産物から識別子生産物を隔離するために実装されることがある。

【0104】

図11D～図11Gは、どのようにして並べ替え方式が、繰り返される構成要素をもつ識別子のいくつかのインスタンスを含むように拡張され得るかという例となる方法を例示する。図11Dは、どのように図11Cからの実装形態が、並べ替えられた繰り返された構成要素をもつ識別子を構築するために使用され得るかの一例を示す。たとえば、識別子は、2つの別個の構成要素から組み立てられた3つの全構成要素を含むことがある。この例では、層からの構成要素は、識別子内に複数回存在することがある。同じ構成要素の隣接する連結は、図における $a * b *$ （5'から3'）ステーブルなどの同じ構成要素の3'末端と5'末端の両方に対する隣接する相補的ハイブリダイゼーション領域をもつステーブルを使用することによって達成されることがある。一般に、M個の層の場合、M個のそのようなステーブルがある。この実装形態との繰り返される構成要素の組み込みは、図11Eにおいて示される縁部スキャフォールドの間に組み立てられる、1よりも長い長さ（すなわち、1つ、2つ、3つ、4つ、またはこれよりも多くの構成要素を含む）の核酸配列を生成することがある。図11Eは、図11Dからの例となる実装形態が、識別子以外の、

縁部スキャフォールドの間に組み立てられる非標的核酸配列にどのようにつながり得るかを示す。適切な識別子が、PCRを用いて非標的核酸配列から隔離されることはできないのは、それらが、縁部上で同じプライマー結合部位を共有するからである。しかしながら、この例では、DNAサイズ選択（たとえば、ゲル抽出を用いた）は、各組み立てられた核酸配列は、一意の長さを有するように設計可能であるので（たとえば、すべての構成要素が同じ長さを有する場合）、非標的配列から標的識別子（たとえば、上から第2の配列）を隔離するために実装されることがある。図11Fは、繰り返される構成要素を用いて識別子を構築することが、同じ反応において等しい縁部配列をもつが別個の長さをもつ複数の核酸配列を生成し得る別の例を図示する。この方法では、交互パターンで1つの層内の構成要素を他の層内の構成要素と組み立てる鑄型が使用されることがある。図11Eに示される方法と同様に、サイズ選択は、設計された長さの識別子を選択するために使用されることがある。図11Gは、繰り返される構成要素を用いて識別子を構築することが、等しい縁部配列をもち、いくつかの核酸配列（たとえば、上から第3の配列および第4の配列ならびに上から第6の配列および第7の配列）の場合、等しい長さをもつ、複数の核酸配列を生成することがある例を示す。この例では、等しい長さを共有するそれらの核酸配列は、PCRおよびDNAサイズ選択が実装される場合でも、他方も構築することなしに一方を構築することが可能でないことがあるため、個々の識別子である両方から除外されることがある。

【0105】

図12A～図12Dは、より大きな数M個の可能な構成要素からの任意の数K個の組み立てられた構成要素（たとえば、核酸配列）をもつ識別子（たとえば、核酸分子）を構築するための、「M choose K方式」と呼ばれる例となる方法を概略的に例示する。図12Aは、M choose K方式を使用して構築された識別子のアーキテクチャを例示する図である。この方法識別子を使用することは、すべての層の任意のサブセット内の各層からの1つの構成要素を組み立てる（たとえば、M個の可能な層からのk個の層から構成要素を選ぶ）ことによって構築される。図12Bは、M choose K方式を使用して構築され得る識別子の組合せ空間の一例を例示する図である。このアセンブリ方式では、組合せ空間は、M個の層に対する N^k M choose K可能な識別子と、層あたりN個の構成要素と、K個の構成要素の識別子長さとを含むことがある。一例では、各々が1つの構成要素を含む5つの層がある場合、各々2つの構成要素を含む最大10個の別個の識別子が組み立てられることがある。

【0106】

M choose K方式は、図12Cに図示されるように、鑄型指向ライゲーションを使用して実装されることがある。並べ替え方式（図11C）のためのTDL実装と同様に、この例における構成要素は、反応マスタ混合内に含まれても含まれなくてもよい縁部スキャフォールドの間に組み立てられる。構成要素は、M個の層、たとえば2からMのあらかじめ規定されたランクをもつ $M = 4$ の層に分割されることがあり、左縁部スキャフォールドはランク1であることがあり、右縁部スキャフォールドはランク $M + 1$ であることがある。鑄型は、より低いランクからより高いランクをそれぞれもつ任意の2つの構成要素の3'から5'ライゲーションのための核酸配列を含む。 $((M + 1^2) + M + 1) / 2$ のそのような鑄型がある。別個の層からの任意のK個の構成要素の個々の識別子は、ライゲーション反応においてそれらの選択された構成要素を、K個の構成要素をランク順に縁部スキャフォールドと一緒にするために使用される対応する $K + 1$ 個のステーブルと組み合わせることによって構築されることがある。そのような反応設定は、縁部スキャフォールド間の標的識別子に対応する核酸配列を生ずることがある。代替的に、すべての鑄型を含む反応混合は、標的識別子を組み立てるために選択構成要素と組み合わせられることがある。この代替方法は、図12Dにおいて例示されるように、同じ縁部配列をもつが別個の長さをもつ（すべての構成要素長さが等しい場合）種々の核酸配列を生成することがある。標的識別子（下部）は、サイズによって副生産物核酸配列から隔離されることがある。

【0107】

図 1 3 A および図 1 3 B は、パーティションされた構成要素をもつ識別子を構築するための、「パーティション方式」と呼ばれる例となる方法を概略的に例示する。図 1 3 A は、パーティション方式を使用して構築され得る識別子の組合せ空間の一例を図示する図である。個々の識別子は、各層からの 1 つの構成要素を固定順序で組み立てることによって構築されることがあり、任意のパーティション（特別に分類された構成要素）の任意選択の留置が、異なる層の任意の 2 つの構成要素の間にある。たとえば、構成要素のセットは、1 つのパーティション構成要素および各々 1 つの構成要素を含有する 4 つの層へと編成されることがある。各層からの構成要素は、固定順序で組み合わせられることがあり、単一のパーティション構成要素は、層間の種々のロケーション内に組み立てられることがある。この組合せ空間内の識別子は、8 つの可能な識別子の組合せ空間を作るために、パーティション構成要素、第 1 の層および第 2 の層からの構成要素間のパーティション構成要素、第 2 の層および第 3 の層からの構成要素間のパーティションなどを含まないことがある。一般に、各々が N 個の構成要素をもつ M 個の層と、 p 個のパーティション構成要素では、構築され得る $N^K (p + 1)^{M - 1}$ 個の可能な識別子がある。この方法は、種々の長さの識別子を生成し得る。

10

【0108】

図 1 3 B は、鋳型指向ライゲーションを使用したパーティション方式の例となる一実装形態を図示する。鋳型は、固定順序で M 個の層の各々からの 1 つの構成要素と一緒にライゲーションするための核酸配列を含む。各パーティション構成要素に対して、パーティション構成要素が任意の 2 つの隣接する層からの構成要素の間にライゲーションすることを可能にする鋳型の追加のペアが存在する。たとえば、ペア内の 1 つの鋳型（たとえば配列 $g * b * (5' \text{ から } 3')$ ）をもつ）が、層 1（配列 b をもつ）の $3'$ 末端がパーティション構成要素（配列 g をもつ）の $5'$ 末端にライゲーションすることを可能にし、ペア内の第 2 の鋳型（たとえば配列 $c * h * (5' \text{ から } 3')$ ）をもつ）が、パーティション構成要素（配列 h をもつ）の $3'$ 末端が層 2（配列 c をもつ）の $5'$ 末端にライゲーションすることを可能にするような 1 対の鋳型。隣接する層の任意の 2 つの構成要素間にパーティションを挿入するために、それらの層と一緒にライゲーションするための標準的な鋳型は、反応において除外されることがあり、その位置にあるパーティションをライゲーションするための鋳型のペアが、反応において選択されることがある。現在の例では、層 1 と層 2 との間のパーティション構成要素を標的とすることは、鋳型 $c * h * (5' \text{ から } 3')$ と $g * b * (5' \text{ から } 3')$ のペアを使用して、鋳型 $c * b * (5' \text{ から } 3')$ ではなく反応を選択することがある。構成要素は、（それぞれ第 1 の層および第 M の層にライゲーションするためのそれらの対応する鋳型とともに）反応混合内に含まれることがある縁部スキャフォールド間に組み立てられることがある。一般に、合計で約 $M - 1 + 2 * p * (M - 1)$ 個の選択可能な鋳型が、 M 個の層および p 個のパーティション構成要素のためのこの方法に使用されることがある。パーティション方式のこの実装形態は、同じ縁部配列だが異なる長さをもつ縁部配列との反応時に異なる長さの種々の核酸配列を生成することがある。標的識別子は、DNA サイズ選択によって副生産物核酸配列から隔離されることがある。具体的には、正確に M 個の層構成要素をもつ正確に 1 つの核酸配列生産物があることがある。パーティション構成要素と比較して十分大きな層構成要素が設計される場合、汎用サイズ選択領域を規定することが可能であることがあり、それによって、識別子は、識別子内の構成要素の特定のパーティションに関係なく選択されてよく（非標的副生産物のいずれも選択されない）、それによって、複数の反応からの複数のパーティションされた識別子が、同じサイズ選択ステップにおいて隔離されることが可能にする。

20

30

40

【0109】

図 1 4 A および図 1 4 B は、いくつかの可能な構成要素からの構成要素の任意の文字列から構成される識別子を構築するための、「無制約文字列」方式すなわち「USS」と呼ばれる例となる方法を概略的に例示する図である。図 1 4 A は、無制約文字列方式を使用して構築され得る 3 構成要素（または 4 スキャフォールド）長さ識別子の組合せ空間の一例を図示する。無制約文字列方式は、各々が 1 つまたは複数の層から取られた 1 つまたは

50

複数の別個の構成要素をもつ長さK構成要素の個々の識別子を構築し、各別個の構成要素は、識別子内のK個の構成要素位置のいずれかにおいて出現することができる（繰り返しを可能にする）。たとえば、各々が1つの構成要素を含む2つの層の場合、8つの可能な3構成要素長さ識別子がある。一般に、各々が1つの構成要素をもつM個の層では、長さK構成要素の M^K 個の可能な識別子がある。図14Bは、鋳型指向ライゲーションを使用した無制約文字列方式の例となる一実装形態を図示する。この方法では、 $K+1$ 個の一本鎖および順序付きスキャフォールドDNA構成要素（2つの縁部スキャフォールドと $K-1$ 個の内部スキャフォールドとを含む）が反応混合に存在する。個々の識別子は、隣接するスキャフォールドのあらゆるペアの間にライゲーションされた単一の構成要素を含む。たとえば、すべてのK個の隣接するスキャフォールドが接合するまで、スキャフォールドAとBとの間にライゲーションされた構成要素、スキャフォールドCとDとの間にライゲーションされた構成要素などは、構成要素によって占められる。反応時、異なる層から選択された構成要素は、適切なスキャフォールド上に組み立てるように指示する選択されたステープルのペアとともにスキャフォールドに導入される。たとえば、ステープル $a * L * (5' \text{ から } 3')$ と $A * b * (5' \text{ から } 3')$ のペアは、LスキャフォールドとAスキャフォールドの間にライゲーションするように、5'末端領域「a」と3'末端領域「b」とをもつ層1構成要素に指示する。一般に、M個の層および $K+1$ 個のスキャフォールドでは、 $2 * M * K$ 個の選択可能なステープルが、長さKの任意のUSS識別子を構築するために使用されることがある。構成要素を5'末端上のスキャフォールドに接続するステープルは、同じ構成要素を3'末端上のスキャフォールドに接続するステープルから解体されるので、核酸副生産物は、標的識別子と等しいが、K個よりも少ない構成要素（ $K+1$ 個よりも少ないスキャフォールド）またはK個よりも多い構成要素（ $K+1$ 個よりも多いスキャフォールド）をもつ縁部スキャフォールドとの反応において形成することがある。標的識別子は、正確にK個の構成要素（ $K+1$ 個のスキャフォールド）とともに形成することがあり、したがって、すべての構成要素が、長さが等しいように設計され、すべてのスキャフォールドが、長さが等しいように設計された場合、DNAサイズ選択のような技法を通して選択可能であることがある。層あたり1つの構成要素であることがある無制約文字列方式のある特定の実施形態では、その構成要素は、（1）識別バーコード、（2）スキャフォールドへの5'末端のステープル媒介ライゲーションのためのハイブリダイゼーション領域、および（3）スキャフォールドへの3'末端のステープル媒介ライゲーションのためのハイブリダイゼーション領域という3つの役割すべてを果たす単一の別個の核酸配列のみを含むことがある。

【0110】

図14Bにおいて例示される内部スキャフォールドは、構成要素へのスキャフォールドのステープル媒介5'ライゲーションと別の（必ずしも別個のとは限らない）構成要素へのスキャフォールドのステープル媒介3'ライゲーションとの両方に、同じハイブリダイゼーション配列を使用するように設計されることがある。したがって、図14Bに描かれた1つのスキャフォールド、2つのステープルが積層されたハイブリダイゼーションイベントは、スキャフォールドとステープルの各々との間で発生し、したがって、5'構成要素ライゲーションと3'構成要素ライゲーションの両方を可能にする、統計的前後ハイブリダイゼーションイベント（statistical back-and-forth hybridization event）を表す。無制約文字列方式の他の実施形態では、スキャフォールドは、2つの連結されたハイブリダイゼーション領域 - ステープル媒介3'ライゲーションのための別個の3'ハイブリダイゼーション領域およびステープル媒介5'ライゲーションのための別個の5'ハイブリダイゼーション領域 - とともに設計されることがある。

【0111】

図15Aおよび図15Bは、核酸配列（または構成要素）を親識別子から除去することによって識別子を構築するための、「構成要素削除方式」と呼ばれる例となる方法を概略的に例示する。図15Aは、構成要素削除方式を使用して構築され得る可能な識別子の組

合せ空間の一例を図示する図である。この例では、親識別子は、複数の構成要素を含むことがある。親識別子は、約 2 つ、3 つ、4 つ、5 つ、6 つ、7 つ、8 つ、9 つ、10、20、30、40、50、またはより多くの構成要素よりも多くまたはこれに等しく含むことがある。個々の識別子は、N 個の可能な構成要素から任意の数の構成要素を選択的に削除し、サイズ 2^N の「フル」組合せ空間に至ることによって、または固定数の K 個の構成要素を N 個の可能な構成要素から削除し、したがってサイズ $N \text{ choose } K$ の「 $N \text{ choose } K$ 」組合せ空間に至ることによって、構築されることがある。3 つの構成要素をもつ親識別子をもつ一例では、フル組合せ空間は 8 であってよく、 $3 \text{ choose } 2$ 組合せ空間は 3 であってよい。

【0112】

図 15 B は、二本鎖標的切断および修復 (DSTCR) を使用した構成要素削除方式の例となる実装形態を図示する。親配列は、ヌクレアーゼ特異的標的部位 (長さが 4 またはこれよりも少ない塩基とすることができる) によって隣接される構成要素を含み、親が、標的部位に対応する 1 つまたは複数の二本鎖固有ヌクレアーゼとともにインキュベートされることがある一本鎖 DNA 基質であってよい。個々の構成要素は、親上の構成要素 DNA (および隣接するヌクレアーゼ部位) に結合し、したがって、両方の末端上でヌクレアーゼによって切断され得る親上で安定した二本鎖配列を形成する、相補的一本鎖 DNA (または切断鋳型) を用いた削除の標的とされることがある。別の一本鎖 DNA (または修復鋳型) は、結果として生じる親の解体 (disjoint) 末端 (その間に、構成要素配列があった) にハイブリダイゼーションし、直接的にか、または置換配列によって架橋されるかのどちらかでライゲーションのためにそれらを一緒にし、したがって、親上のライゲーション配列は、アクティブなヌクレアーゼ標的部位をもはや含有しない。この方法を「二本鎖標的切断」(DSTC) と呼ぶ。サイズ選択は、ある数の削除された構成要素をもつ識別子を選択するために使用されることがある。

【0113】

代替的に、またはこれに加えて、親識別子は、2 つの構成要素が同じ配列によって隣接されないように、スペーサ配列によって分離される構成要素を含む二本鎖核酸基質または一本鎖核酸基質であってよい。親識別子は、Cas9 ヌクレアーゼとともにインキュベートされることがある。個々の構成要素は、構成要素の縁部に結合し、その隣接する部位における Cas9 媒介切断を可能にする、ガイドリボ核酸 (切断鋳型) を用いた削除の標的とされることがある。一本鎖核酸 (修復鋳型) は、結果として生じる親識別子の解体末端 (たとえば、構成要素配列があった末端の間に) にハイブリダイゼーションし、したがって、ライゲーションのためにそれらを一緒にすることがある。ライゲーションは、親上のライゲーション配列が、Cas9 によって標的とされ得るスペーサ配列をもはや含有しないように、直接的に、または置換配列をもつ末端を架橋することによって、なされてよい。この方法を、「配列固有標的切断および修復」すなわち「SSTCR」と呼ぶ。

【0114】

識別子は、DSTCR の派生物を使用して構成要素を親識別子に挿入することによって構築されることがある。親識別子は、各々が別個の核酸配列内に埋め込まれたヌクレアーゼ特異的標的部位 (長さが 4 またはそれより小さい塩基とすることができる) を含む一本鎖核酸基質であることがある。親識別子は、標的部位に対応する 1 つまたは複数の二本鎖固有ヌクレアーゼとともにインキュベートされることがある。親識別子上の個々の標的部位は、標的部位と親識別子上の別個の周囲核酸配列とを結合し、したがって二本鎖部位を形成する、相補的一本鎖核酸 (切断鋳型) をもつ構成要素挿入の標的とされることがある。この二本鎖部位は、ヌクレアーゼによって切断されることがある。別の一本鎖核酸 (修復鋳型) は、結果として生じる親識別子の解体末端にハイブリダイゼーションし、構成要素配列によって架橋されるライゲーションのためにそれらを一緒にすることがあり、したがって、親上のライゲーション配列は、アクティブなヌクレアーゼ標的部位をもはや含有しない。代替的に、SSTCR の派生物が、構成要素を親識別子に挿入するために使用されることがある。親識別子は、二本鎖核酸または一本鎖核酸であってよく、親は、Cas

10

20

30

40

50

9ヌクレアーゼとともにインキュベートされることがある。親識別子上の別個の部位は、ガイドRNA（切断鋳型）を用いた切断の標的とされることがある。一本鎖核酸（修復鋳型）は、親識別子の解体末端にハイブリダイゼーションし、それらをライゲーションのために一緒にし、構成要素配列によって架橋されることがあり、したがって、親識別子上のライゲーション配列は、アクティブなヌクレアーゼ標的的部位をもはや含有しない。サイズ選択は、ある数の構成要素挿入をもつ識別子を選択するために使用されることがある。

【0115】

図16は、リコンビナーゼ認識部位をもつ親識別子を概略的に例示する。異なるパターンの認識部位は、異なるリコンビナーゼによって認識可能である。リコンビナーゼの所与のセットに対するすべての認識部位は、リコンビナーゼが適用された場合にそれらの中間の核酸が切除され得るように配置される。図16に図示される核酸鎖は、それに適用されるリコンビナーゼのサブセットに応じて、 $2^5 = 32$ 個の異なる配列を採用することができる。一部の実施形態では、図16において描かれるように、一意の分子が、リコンビナーゼを使用してDNAのセグメントを切除、シフト、反転、および転位して異なる核酸分子を作成するために、生成可能である。一般に、N個のリコンビナーゼでは、親から作り上げられた 2^N 個の可能な識別子がある場合がある。一部の実施形態では、異なるリコンビナーゼからの認識部位の複数の直交するペアは、1つのリコンビナーゼの適用が、下流リコンビナーゼが適用されるときに発生する組換えイベントのタイプに影響するように、重複するやり方で、親識別子上で配置可能である（Roquetら、*Synthetic recombina*

【0116】

いくつかの事例では、単一の核酸配列は、別個の順序で数多くのリコンビナーゼを適用することによって、多数の別個の核酸配列になるようにプログラム可能である。リコンビナーゼの数すなわちMが、大型セリンリコンビナーゼファミリーに対して7よりも小さいまたはこれに等しいことがあるとき、約 $e^{-1}M!$ の別個の核酸配列が、異なるそれらのサブセット内で異なる順序でM個のリコンビナーゼを適用することによって生成されることがある。リコンビナーゼの数すなわちMが7よりも大きいことがあるとき、生じされることが可能な配列の数は、 $3 \cdot 9^M$ に近い。たとえば、Roquetら、*Synthetic recombina*

【0117】

符号化されることになる情報のビットストリームが、任意の単一の核酸分子によって符号化可能なものよりも大きい場合、情報は、分割され、核酸配列バーコードとともにインデックス付与可能である。さらに、N個の核酸分子のセットからのサイズkの核酸分子の任意のサブセットは、 $\log_2(N \text{ choose } k)$ ビットの情報を生じさせるように選

ることができる。バーコードは、はるかに長いビットストリームを符号化するために、サイズ k のサブセット内の核酸分子上に組み立てられることがある。たとえば、 M 個のバーコードは、 $M * \log_2(N \text{ choose } k)$ ビットの情報を生じさせるために使用されることがある。セット内のある数 N の利用可能な核酸分子およびある数 M の利用可能なバーコードを想定すると、サイズ $k = k_0$ のサブセットは、1つの情報を符号化するためにプール内の分子の総数を最小にするように選ばれることがある。デジタル情報を符号化するための方法は、ビットストリームを細分化し、個々のエレメントを符号化するためのステップを含むことができる。たとえば、6 ビットを含むビットストリームは、各構成要素が 2 ビットを含む 3 つの構成要素に分割可能である。各 2 ビット構成要素は、情報カセットを形成するためにバーコード化され、情報カセットのハイパープールを形成するために一緒に群化またはプール化可能である。

10

【0118】

バーコードは、符号化されることになるデジタル情報の量が、1つのプールのみに適応可能である量を超えると、情報インデックス付与を容易にすることができる。ビットのより長い文字列および/または複数のバイトを含む情報は、図3に開示されている手法を層化することによって、たとえば、核酸インデックスを使用して符号化された一意の核酸配列をもつタグを含むことによって、符号化可能である。情報カセットまたは識別子ライブラリは、所与の配列が対応するビットストリームの1つまたは複数の構成要素を示すバーコードまたはタグに加えてロケーションおよびビット値情報を提供する一意の核酸配列を含む窒素塩基または核酸配列を含むことができる。情報カセットは、1つまたは複数の一意の核酸配列ならびにバーコードまたはタグを含むことができる。情報カセット上のバーコードまたはタグは、情報カセットおよび情報カセット内に含まれる任意の配列の参照を提供することができる。たとえば、情報カセット上のタグまたはバーコードは、ビットストリームまたはビットスチームのビット構成要素のどの部分に対して一意の配列が情報(たとえば、ビット値およびビット位置情報)を符号化するかを示すことができる。

20

【0119】

バーコードを使用して、可能な識別子の組合せ空間のサイズよりも多くのビット内情報が、プール内で符号化可能である。10 ビットの配列は、たとえば、バイトの2つのセットへと分離されることができ、各バイトは5ビットを含む。各バイトは、5つの可能な別個の識別子のセットにマッピング可能である。最初、各バイトに対して生成される識別子は同じとすることができるが、これらの識別子は、別々のプール内に保たれてよく、そうでなければ、誰かが情報を読み取ることは、特定の核酸配列がどのバイトに属するかを知らせることを可能にしないことがある。しかしながら、各識別子は、符号化された情報が適用されるバイトに対応する標識とともにバーコード化またはタグ化可能であり(たとえば、バーコード1は、第1の5つのビットを提供するために核酸プール内の配列に付着されることがあり、バーコード2は、第2の5つのビットを提供するために核酸プール内の配列に付着されることがある)、次いで、2バイトに対応する識別子は、1つのプール(たとえば、「ハイパープール」または1つもしくは複数の識別子ライブラリ)へと組み合わせ可能である。1つまたは複数の組み合わせられた識別子ライブラリの各識別子ライブラリは、所与の識別子を所与の識別子ライブラリに属すると識別する別個のバーコードを含むことがある。バーコードを識別子ライブラリ内の各識別子に追加するための方法は、PCR、ギブソン、ライゲーション、または所与のバーコード(たとえば、バーコード1)が所与の核酸試料プールに(たとえば、バーコード1を核酸試料プール1に、バーコード2を核酸試料プール2に)付着することを可能にする他の任意の手法を使用することを含むことができる。ハイパープールからの試料は、配列決定方法を用いて読み取ることができ、配列決定情報は、バーコードまたはタグを使用して解析可能である。 M 個のバーコードおよび N 個の可能な識別子(組合せ空間)のセットとともに識別子ライブラリおよびバーコードを使用する方法は、 M と N の積に等しい長さをもつビットのストリームを符号化することができる。

30

40

【0120】

50

一部の実施形態では、識別子ライブラリは、ウェルのアレイ内に記憶されることがある。ウェルのアレイは、 n 個の列と q 個の行とを有すると規定されることがあり、各ウェルは、2 つまたはこれよりも多くの識別子ライブラリをハイパープール内に含むことがある。各ウェル内で符号化された情報は、ウェルの各々に含有される情報よりも大きい、サイズ $n \times q$ の情報の 1 つの大きな連続した項目を構成することがある。アリコートは、ウェルのアレイ内のウェルのうちの 1 つまたは複数から取られることがあり、符号化は、配列決定、ハイブリダイゼーション、または PCR を使用して読み取られることがある。

【0121】

核酸試料プール、ハイパープール、識別子ライブラリ、識別子ライブラリの群、または核酸試料プールもしくはハイパープールを含有するウェルは、情報のビットに対応する一意の核酸分子（たとえば、識別子）と、複数の補足的核酸配列とを含むことがある。補足的核酸配列は、符号化されたデータに対応しないことがある（たとえば、ビット値に対応しない）。補足的核酸試料は、試料プール内に記憶された情報をマスキングまたは暗号化することがある。補足的核酸配列は、生物学的源から派生されてもよいし、合成的に生じさせられてもよい。生物学的源から派生された補足的核酸配列は、ランダムに断片化された核酸配列または合理的に断片化された配列を含むことがある。生物学的に派生された補足的核酸は、特に、合成的に符号化された情報（たとえば、識別子の組合せ空間）が、自然遺伝情報（たとえば、断片化されたゲノム）に似るように作成された場合、合成的に符号化された情報とともに自然遺伝情報を提供することによって、試料プール内のデータ含有核酸を隠すまたは曖昧にすることがある。一例では、識別子は、生物学的源から派生され、補足的核酸は、生物学的源から派生される。試料プールは、識別子および補足的核酸配列の複数のセットを含有することがある。識別子および補足的核酸配列の各セットは、異なる有機体から派生されてよい。一例では、識別子は、1 つまたは複数の有機体から派生され、補足的核酸配列は、単一の、異なる有機体から派生される。補足的核酸配列は、1 つまたは複数の有機体から派生されることもあり、識別子は、補足的核酸が派生される有機体とは異なる単一の有機体から派生されることがある。識別子と補足的核酸配列の両方が、複数の異なる有機体から派生されることがある。キーは、識別子を補足的核酸配列と区別するために使用されることがある。

【0122】

補足的核酸配列は、書き込まれた情報についてのメタデータを記憶することがある。メタデータは、元の情報の源およびまたは元の情報の意図されたレシピエントを決定および/または許可するための余剰情報を含むことがある。メタデータは、元の情報のフォーマット、元の情報を符号化し書き込むために使用される機器および方法、ならびに元の情報を識別子に書き込む日付および時間についての余剰情報を含むことがある。メタデータは、元の情報のフォーマット、元の情報を符号化し書き込むために使用される機器および方法、ならびに元の情報を核酸配列に書き込む日付および時間についての追加情報を含むことがある。メタデータは、情報を核酸配列に書き込んだ後に元の情報に対してなされた修正についての追加情報を含むことがある。メタデータは、元の情報に対する注釈または外部情報への 1 つもしくは複数の参照を含むことがある。代替的に、またはこれに加えて、メタデータは、識別子に付着された 1 つまたは複数のバーコードまたはタグ内に記憶されることがある。

【0123】

識別子プール内の識別子は、同じ長さを有してもよいし、類似の長さを有してもよいし、互いとは異なる長さを有してもよい。補足的核酸配列は、識別子の長さよりも短い、これに実質的に等しい、またはこれよりも長い長さを有してよい。補足的核酸配列は、識別子の平均的な長さの 1 つの塩基以内、2 つの塩基以内、3 つの塩基以内、4 つの塩基以内、5 つの塩基以内、6 つの塩基以内、7 つの塩基以内、8 つの塩基以内、9 つの塩基以内、10 の塩基以内、またはこれよりも多くの塩基以内である、平均的な長さを有することがある。一例では、補足的核酸配列は、識別子と同じまたは実質的に同じ長さである。補足的核酸配列の濃度は、識別子ライブラリ内の識別子の濃度よりも小さくてもよいし、こ

10

20

30

40

50

れに実質的に等しくてもよいし、これよりも大きくてもよい。補足的核酸の濃度は、約 1 %、10 %、20 %、40 %、60 %、80 %、100 %、125 %、150 %、175 %、200 %、1000 %、 1×10^4 %、 1×10^5 %、 1×10^6 %、 1×10^7 %、 1×10^8 % よりも小さいまたはこれに等しくてもよいし、識別子の濃度よりも小さくてもよい。補足的核酸の濃度は、約 1 %、10 %、20 %、40 %、60 %、80 %、100 %、125 %、150 %、175 %、200 %、1000 %、 1×10^4 %、 1×10^5 %、 1×10^6 %、 1×10^7 %、 1×10^8 % よりも大きいまたはこれに等しくてもよいし、識別子の濃度よりも大きくてもよい。より大きな濃度が、不明瞭化またはデータを隠蔽するために有益であることがある。一例では、補足的核酸配列の濃度は、識別子プール内の識別子の濃度よりも実質的に大きい（たとえば、 1×10^8 % 大きい）。

10

【0124】

核酸配列内に記憶されたデータをコピーおよびアクセスするための方法

別の態様では、本開示は、核酸配列内で符号化される情報をコピーするための方法を提供する。核酸配列内に符号化された情報をコピーするための方法は、(a) 識別子ライブラリを提供することと、(b) 識別子ライブラリの 1 つまたは複数のコピーを構築することとを含むことがある。識別子ライブラリは、より大きな組合せ空間からの複数の識別子のサブセットを含むことがある。複数の識別子の各個々の識別子は、シンボルの文字列内の個々のシンボルに対応することがある。識別子は、1 つまたは複数の構成要素を含むことがある。構成要素は、核酸配列を含むことがある。

【0125】

20

別の態様では、本開示は、核酸配列内で符号化された情報にアクセスするための方法を提供する。核酸配列内に符号化された情報にアクセスするための方法は、(a) 識別子ライブラリを提供することと、(b) 識別子ライブラリ内に存在する識別子の一部分またはサブセットを識別子ライブラリから抽出することとを含むことがある。識別子ライブラリは、より大きな組合せ空間からの複数の識別子のサブセットを含むことがある。複数の識別子の各個々の識別子は、シンボルの文字列内の個々のシンボルに対応することがある。識別子は、1 つまたは複数の構成要素を含むことがある。構成要素は、核酸配列を含むことがある。

【0126】

情報は、本明細書において他のところで説明されるように、1 つまたは複数の識別子ライブラリへと書き込まれることがある。識別子は、本明細書において他のところで説明される任意の方法を使用して構築されてよい。記憶されたデータは、識別子ライブラリ内または 1 つもしくは複数の識別子ライブラリ内の個々の識別子のコピーを生成することによって、コピーされることがある。識別子の一部分がコピーされてもよいし、ライブラリ全体がコピーされてもよい。コピーは、識別子ライブラリ内の識別子を増幅することによって実行されることがある。1 つまたは複数の識別子ライブラリが組み合わせられるとき、単一の識別子ライブラリまたは複数の識別子ライブラリがコピーされることがある。識別子ライブラリが補足的核酸配列を含む場合、この補足的核酸配列は、コピーされてもよいし、コピーされなくてもよい。

30

【0127】

40

識別子ライブラリ内の識別子は、1 つまたは複数の共通プライマー結合部位を含むように構築されることがある。1 つまたは複数の結合部位は、各識別子の縁部に配されてもよいし、各識別子全体を通じて混交されてもよい。プライマー結合部位は、識別子ライブラリ固有プライマーペアまたは汎用プライマーペアが識別子に結合および増幅することを可能にすることがある。識別子ライブラリ内の識別子のすべてまたは 1 つもしくは複数の識別子ライブラリ内の識別子のすべてが、複数の PCR サイクルによって複数回複製されることがある。従来の PCR は、識別子をコピーするために使用されることがあり、識別子は、各 PCR サイクルとともに指数関数的に複製されることがある。識別子のコピーの数は、各 PCR サイクルとともに指数関数的に増加することがある。リニア PCR は、識別子をコピーするために使用されることがあり、識別子は、各 PCR サイクルとともに線形

50

的に複製されることがある。識別子コピーの数は、各PCRサイクルとともに線形的に増加することがある。識別子は、PCR増幅の前に環状ベクターへとライゲーションされることがある。円ベクトルは、識別子挿入部位の各末端にバーコードを含むことがある。識別子を増幅するためのPCRプライマーは、バーコード化された縁部が増幅生産物内に識別子とともに含まれるように、ベクトルにプライム (prime) するように設計されることがある。増幅中、識別子間の組換えは、各縁部上の無相関 (non-correlated) バーコードを含むコピーされた識別子という結果になることがある。無相関バーコードは、識別子を読み取るときに検出可能であることがある。無相関バーコードを含有する識別子は、偽陽性とみなされることがあり、情報復号プロセス中に無視されることがある。

10

【0128】

情報は、情報の各ビットを一意的な核酸分子に割り当てることによって符号化されることがある。たとえば、各々2つの核酸配列を含有する3つの試料セット (X、Y、およびZ) が、8つの一意的な核酸分子へと組み立てられ、データの8ビットを符号化することがある。

N 1 = X 1 Y 1 Z 1
N 2 = X 1 Y 1 Z 2
N 3 = X 1 Y 2 Z 1
N 4 = X 1 Y 2 Z 2
N 5 = X 2 Y 1 Z 1
N 6 = X 2 Y 1 Z 2
N 7 = X 2 Y 2 Z 1
N 8 = X 2 Y 2 Z 2

20

次いで、文字列内の各ビットが、対応する核酸分子に割り当てられることがある (たとえば、N 1 が、第1のビットを指定することがあり、N 2 が、第2のビットを指定することがあり、N 3 が、第3のビットを指定することがある、など)。ビット文字列全体が、「1」のビット値に対応する核酸分子が組合せまたはプール内に含まれる核酸分子の組合せに割り当てられることがある。たとえば、UTF-8コーディングでは、「K」という文字は、4つの核酸分子の存在によって符号化され得る8ビット文字列コード01001011によって表され得る (たとえば、上記の例におけるX 1 Y 1 Z 2、X 2 Y 1 Z 1、X 2 Y 2 Z 1、およびX 2 Y 2 Z 2)。

30

【0129】

情報は、配列決定またはハイブリダイゼーションアッセイを通じてアクセスされることがある。たとえば、プライマーまたはプローブは、共通領域または核酸配列のバーコード化領域に結合するように設計されることがある。これは、核酸分子の任意の領域の増幅を可能にすることがある。次いで、増幅生産物は、増幅生産物を配列決定することによって読み取られてもよいし、ハイブリダイゼーションアッセイによって読み取られてもよい。「1C」という文字を符号化する上記の例では、データの第1の半分が、X 1 核酸配列のバーコード領域に特異的な目的のプライマーである場合、Zセットの共通領域に結合するプライマーは、核酸分子を増幅するために使用されることがある。これは、配列Y 1 Z 2を返すことがあり、これは、0100に関して符号化されることがある。そのデータの部分文字列も、Y 1 核酸配列のバーコード領域に結合するプライマーおよびZセットの共通配列に結合するプライマーをもつ核酸分子をさらに増幅することによってアクセスされることがある。これは、部分文字列01を符号化するZ 2 核酸配列を返すことがある。代替的に、データは、配列決定なしに特定の核酸配列の存在または不在をチェックすることによってアクセスされることがある。たとえば、Y 2 バーコードに特異的なプライマーを用いた増幅は、Y 2 バーコードに対する増幅生産物を生成することがあるが、Y 1 バーコードに対する増幅生産物を生成しないことがある。Y 2 増幅生産物の存在は、「1」のビット値をシグナリングすることがある。代替的に、Y 2 増幅生産物の不在は、「0」のビット値をシグナリングすることがある。

40

50

【0130】

PCRベースの方法は、識別子または核酸試料プールからのデータにアクセスしてコピーするために使用されることができる。プールまたはハイパープール内の識別子と隣接する共通プライマー結合部位を使用して、情報を含む核酸は、容易にコピーされることができる。代替的に、等温増幅などの他の核酸増幅手法も、試料プールまたはハイパープール（たとえば、識別子ライブラリ）からデータを容易にコピーするために使用されることがある。試料がハイパープールを含む事例では、情報の特定のサブセット（たとえば、特定のバーコードに関係するすべての核酸）は、識別子の反対側の縁部上の共通配列を逆方向に結合する別のプライマーとともに、識別子の一方の縁部において固有バーコードを順方向に結合するプライマーを使用することによって、アクセスされ、取り出されることができる。種々の読出し方法が、符号化された核酸から情報を引き出すために使用可能である。たとえば、マイクロアレイ（または任意の種類の蛍光ハイブリダイゼーション）、デジタルPCR、定量的PCR（qPCR）、および種々の配列決定プラットフォームが、符号化された配列と、拡張によってデジタル的に符号化されたデータを読み出すためにさらに使用可能である。

10

【0131】

核酸分子（たとえば、識別子）内に記憶された情報にアクセスすることは、非標的識別子の一部分を識別子ライブラリまたは識別子のプールから選択的に除去する、または、たとえば、識別子ライブラリのすべての識別子を複数の識別子ライブラリのプールから選択的に除去することによって実行されることがある。データにアクセスすることは、識別子ライブラリまたは識別子のプールから標的識別子を選択的にキャプチャすることによっても実行されることがある。標的識別子は、より大きな情報の項目内の目的のデータに対応することがある。識別子のプールは、補足的核酸分子を含むことがある。補足的核酸分子は、符号化された情報についてのメタデータを含むしてもよいし、情報に対応する識別子を暗号化またはマスキングするために使用されてもよい。補足的核酸分子は、標的識別子にアクセスする間、抽出されてもよいし、抽出されなくてもよい。図17A～図17Cは、より大きな数の識別子からのいくつかの特定の識別子にアクセスすることによって核酸配列内に記憶された情報の一部分にアクセスするための例となる方法の概要を概略的に例示する。図17Aは、ポリメラーゼ連鎖反応、親和性タグプローブ、および分解標的プローブを使用して、指定された構成要素を含む識別子にアクセスするための、例となる方法を図示する。PCRベースのアクセスの場合、識別子のプール（たとえば、識別子ライブラリ）は、各末端における共通配列、各末端における可変配列、または各末端における共通配列もしくは可変配列のうちの1つをもつ識別子を含むことがある。共通配列または可変配列は、プライマー結合部位であってよい。1つまたは複数のプライマーは、識別子縁部上の共通領域または可変領域に結合することがある。結合されたプライマーをもつ識別子は、PCRによって増幅されることがある。増幅される識別子は、増幅されていない識別子より著しく多いことがある。読取中、増幅された識別子が識別されることがある。識別子ライブラリからの識別子は、そのライブラリに対して別個であるその末端の一方または両方の上に配列を含み、したがって、単一のライブラリが、プールまたは1つよりも多い識別子ライブラリの群から選択的にアクセスされることを可能にすることがある。

20

30

40

【0132】

親和性タグベースのアクセスの場合、プール内で識別子を構成する構成要素は、1つまたは複数のプローブとの相補性を共有することがある。1つまたは複数のプローブは、アクセスされることになる識別子に結合またはハイブリダイゼーションすることがある。プローブは、親和性タグを含むことがある。親和性タグは、ビーズに結合し、ビーズと、少なくとも1つのプローブと、少なくとも1つの識別子とを含む複合体を生成することがある。ビーズは磁氣的であってよく、磁石とともに、ビーズは、アクセスされることになる識別子を収集および隔離することがある。識別子は、読取の前に、変性条件下でビーズから除去されることがある。代替的に、またはこれに加えて、ビーズは、非標的識別子を収集し、それらを、別々の容器へと洗い流して読み取ることが可能なプールの残りから隔

50

絶させる (sequester) ことがある。親和性タグは、カラムに結合することがある。アクセスされることになる識別子は、キャプチャのためにカラムに結合することがある。カラムに結合した識別子は、その後、読取りの前にカラムから溶出または変性されることがある。代替的に、非標的識別子は、カラムに対して選択的に標的とされることがあるが、標的識別子は、カラムを流れることがある。標的識別子にアクセスすることは、1つもしくは複数のプローブを識別子のプールに同時に適用すること、または1つもしくは複数のプローブを識別子のプールに順次適用することを含むことがある。

【0133】

分解ベースのアクセスの場合、プール内の識別子を構成する構成要素は、1つまたは複数の分解標的プローブとの相補性を共有することがある。プローブは、識別子上の別個の構成要素に結合するまたはこれとハイブリダイゼーションすることがある。プローブは、エンドヌクレアーゼなどの分解酵素の標的となることがある。一例では、1つまたは複数の識別子ライブラリが組み合わせられることがある。プローブのセットは、識別子ライブラリの1つとハイブリダイゼーションすることがある。プローブのセットはRNAを含むことがあり、RNAは、Cas9酵素をガイドすることがある。Cas9酵素は、1つまたは複数の識別子ライブラリに導入されることがある。プローブとハイブリダイゼーションされた識別子は、Cas9酵素によって分解されることがある。アクセスされることになる識別子は、分解酵素によって分解されないことがある。別の例では、識別子は、一本鎖であることがあり、識別子ライブラリは、アクセスされないことになる識別子を選択的に分解させる、S1ヌクレアーゼなどの一本鎖特異的エンドヌクレアーゼと組み合わせられることがある。アクセスされることになる識別子は、それらを一本鎖特異的エンドヌクレアーゼによる分解から保護するために、識別子の相補的セットとハイブリダイゼーションされることがある。アクセスされることになる識別子は、サイズ選択クロマトグラフィ（たとえば、アガロースゲル電気泳動）などのサイズ選択によって、分解生産物から分離されることがある。代替的に、または加えて、分解されない識別子は、分解生産物が増幅されないように、（たとえば、PCRを使用して）選択的に増幅されることがある。非分解識別子は、非分解識別子の各末端にハイブリダイゼーションし、したがって、分解または切断された識別子の各末端にハイブリダイゼーションしないプライマーを使用して、増幅されることがある。

【0134】

図17Bは、ポリメラーゼ連鎖反応を使用し、「OR」演算または「AND」演算を実行して、複数の構成要素を含有する識別子にアクセスするための例となる方法を図示する。一例では、2つの順プライマーが、左末端上の識別子の別個のセットを結合した場合、識別子のこれらのセットの和集合の「OR」増幅は、右末端上の識別子のすべてに結合する逆プライマーとの多重PCR反応において2つの順プライマーを一緒に使用することによって達成されることがある。別の例では、1つの順プライマーが左末端上の識別子のセットに結合し、1つの逆プライマーが右末端上の識別子のセットに結合する場合、識別子のそれらの2つのセットの交差の「AND」増幅は、PCR反応におけるプライマーペアとして順プライマーと逆プライマーを一緒に使用することによって達成されることがある。

【0135】

図17Cは、親和性タグを使用し、「OR」演算または「AND」演算を実行して、複数の構成要素を含有する識別子にアクセスするための例となる方法を図示する。一例では、親和性プローブ「P1」が、構成要素「C1」をもつすべての識別子をキャプチャし、別の親和性プローブ「P2」が、構成要素「C2」をもつすべての識別子をキャプチャした場合、C1またはC2をもつすべての識別子のセットは、P1とP2を同時に使用する（「OR」演算に対応する）ことによってキャプチャ可能である。同じ構成要素およびプローブをもつ別の例では、C1およびC2をもつすべての識別子のセットは、P1とP2を順次使用する（「AND」演算に対応する）ことによってキャプチャ可能である。

【0136】

核酸配列内に記憶された情報を読み取るための方法

10

20

30

40

50

別の態様では、本開示は、核酸配列内に符号化された情報を読み取るための方法を提供する。核酸配列内に符号化された情報を読み取るための方法は、(a) 識別子ライブラリを提供することと、(b) 識別子ライブラリ内に存在する識別子を識別することと、(c) 識別子ライブラリ内に存在する識別子からシンボルの文字列を生成することと、(d) シンボルの文字列からの情報をコンパイルすることとを含むことがある。識別子ライブラリは、組合せ空間からの複数の識別子のサブセットを含むことがある。識別子のサブセットの各個々の識別子は、シンボルの文字列内の個々のシンボルに対応することがある。識別子は、1つまたは複数の構成要素を含むことがある。構成要素は、核酸配列を含むことがある。

【0137】

情報は、本明細書において他のところで説明されるように、1つまたは複数の識別子ライブラリへと書き込まれることがある。識別子は、本明細書において他のところで説明される任意の方法を使用して構築されてよい。記憶されたデータは、本明細書において他のところで説明される任意の方法を使用してコピーおよびアクセスされることがある。

【0138】

識別子は、符号化されたシンボルのロケーション、符号化されたシンボルの値、または符号化されたシンボルのロケーションと値の両方に関連する情報を含むことがある。識別子は、符号化されたシンボルのロケーションに関連する情報を含むことがあり、識別子ライブラリ内の識別子の存在または不在は、シンボルの値を示すことがある。識別子ライブラリ内の識別子の存在は、2進文字列内の第1のシンボル値（たとえば、第1のビット値）を示すことがあり、識別子ライブラリ内の識別子の不在は、2進文字列内の第2のシンボル値（たとえば、第2のビット値）を示すことがある。2進法では、識別子ライブラリ内の識別子の存在または不在に関するビット値をベースとすることは、組み立てられた識別子の数を減少させ、したがって、書き込み時間を減少させることがある。一例では、識別子の存在は、マッピングされたロケーションにおける「1」のビット値を示すことがあり、識別子の不在は、マッピングされたロケーションにおける「0」のビット値を示すことがある。

【0139】

1つの情報に対してシンボル（たとえば、ビット値）を生成することは、シンボル（たとえば、ビット）がマッピングまたは符号化されることがある識別子の存在または不在を識別することを含むことがある。識別子の存在または不在を決定することは、この識別子を配列決定すること、またはハイブリダイゼーションアレイを使用して識別子の存在を検出することを含むことがある。一例では、符号化された配列を復号し読み取ることは、配列決定プラットフォームを使用して実行されることがある。配列決定プラットフォームの例は、2014年8月21日に出願された米国特許出願第14/465,685号、2013年5月2日に提出された米国特許出願第13/886,234号、および2009年3月9日に提出された米国特許出願第12/400,593号に記載されており、これら出願の各々は、参照により全体が本明細書に組み込まれる。

【0140】

一例では、核酸符号化データを復号することは、Illumina（登録商標）Sequencingなどの核酸鎖の塩基単位配列決定によって、またはキャピラリー電気泳動によるフラグメンテーション解析などの固有核酸配列の存在もしくは不在を示す配列決定技法を利用することによって、達成されることがある。配列決定は、可逆的ターミネータの使用を用いることがある。配列決定は、天然または非天然（たとえば、人工）ヌクレオチドまたはヌクレオチド類似体の使用を用いることがある。代替的に、またはこれに加えて、核酸配列を復号することは、限定するものではないが、光信号、電気化学信号、または化学信号を生成する任意の方法を含む、さまざまな解析技法を使用して、実行されることがある。限定するものではないが、ポリメラーゼ連鎖反応（PCR）、デジタルPCR、サンガー配列決定、高スループット配列決定、sequencing-by-synthesis、単一分子配列決定、sequencing-by-ligation、RN

10

20

30

40

50

A - Seq (Illumina)、次世代配列決定、Digital Gene Expression (Helicos)、Clonal Single MicroArray (Solexa)、ショットガン配列決定、マキシム - ギルバート (Maxim - Gilbert) 配列決定、または超並列配列決定を含むさまざまな配列決定手法が使用されてよい。

【0141】

種々の読出し方法が、符号化された核酸から情報を引き出すために使用可能である。一例では、マイクロアレイ（または任意の種類の蛍光ハイブリダイゼーション）、デジタルPCR、定量的PCR (qPCR)、および種々の配列決定プラットフォームが、符号化された配列と、拡張によってデジタル的に符号化されたデータを読み出すためにさらに使用可能である。

10

【0142】

識別子ライブラリは、情報についてのメタデータを提供する、情報を暗号化もしくはマスキングする、またはメタデータの提供と情報のマスキングの両方を行う、補足的核酸配列をさらに含むことがある。補足的核酸は、識別子の識別と同時に識別されることがある。代替的に、補足的核酸は、識別子を識別する前または後に識別されることがある。一例では、補足的核酸は、符号化された情報の読取り中に識別されない。補足的核酸配列は、識別子と区別不可能であることがある。識別子インデックスまたはキーは、補足的核酸分子を識別子と見分けるために使用されることがある。

【0143】

データを符号化および復号することの効率は、より少ない核酸分子の使用を可能にするために、入力ビット文字列を記録することによって増加されることがある。たとえば、入力文字列が、符号化方法とともに3つの核酸分子（たとえば、識別子）にマッピングすることがある「111」部分文字列の高い発生とともに受信される場合、この入力文字列は、核酸分子のヌルセットにマッピングすることがある「000」部分文字列に記録されることがある。「000」の代替入力部分文字列も、「111」に記録されることがある。この記録する方法は、データセット内の「1」の数の減少があり得るので、データを符号化するために使用される核酸分子の総量を減少させることがある。この例では、データセットの全体的サイズは、新しいマッピング命令を指定するコードブックに対応するように増加されることがある。符号化および復号することの効率を増加させるための代替方法は、可変長を減少させるために入力文字列を記録することである。たとえば、「111」は、データセットのサイズを縮小し、データセット内の「1」の数を減少させることがある「00」に記録されることがある。

20

30

【0144】

データが符号化された核酸を復号するスピードおよび効率は、検出を簡単にするために識別子を具体的に設計することによって、制御される（たとえば、増加される）ことがある。たとえば、検出を簡単にするために設計された核酸配列（たとえば、識別子）は、光学的性質、電気化学的性質、化学的性質、または物理的性質に基づいて呼び出し、検出するのがより簡単なヌクレオチドの大多数を含む核酸配列を含むことがある。人工核酸配列は、一本鎖または二本鎖のどちらかであってよい。人工核酸配列は、核酸配列の検出可能な性質を改善する合成ヌクレオチドまたは非天然ヌクレオチドを含むことがある。人工核酸配列は、すべての天然ヌクレオチド、すべての合成または非天然ヌクレオチド、または天然ヌクレオチド、合成ヌクレオチド、および非天然ヌクレオチドの組合せを含むことがある。合成ヌクレオチドは、ペプチド核酸、ロックド (locked) 核酸、グリコール核酸、およびトレオース核酸などのヌクレオチド類似体を含むことがある。非天然ヌクレオチドは、dNaM、3 - メトキシ - 2 - ナフチリ (naphthyl) 基を含有する人工ヌクレオシド、およびd5SICS、6 - メチルイソキノリン - 1 - チオン - 2 - イル基を含有する人工ヌクレオシドを含むことがある。人工核酸配列は、強化された光学的性質などの単一の強化された性質のために設計されることがあり、または設計された核酸配列は、強化された光学的性質および電気化学的性質もしくは強化された光学的性質および

40

50

化学的性質などの、複数の強化された性質とともに設計されることがある。

【0145】

人工核酸配列は、核酸配列の光学的性質、電気化学的性質、化学的性質、または物理的性質を改善しない、反応性天然ヌクレオチド、合成ヌクレオチド、および非天然ヌクレオチドを含むことがある。核酸配列の反応性構成要素は、核酸配列への改善された性質を与える化学的部分の追加を可能にすることがある。各核酸配列は、単一の化学的部分を含んでもよいし、複数の化学的部分を含んでもよい。例となる化学的部分は、限定するものではないが、蛍光部分、化学発光部分、酸性または基本部分、疎水性または親水性部分、および核酸配列の酸化状態または反応性を変える部分を含むことがある。

【0146】

配列決定プラットフォームは、核酸配列へと符号化された情報を復号し読み取るために具体的に設計されることがある。配列決定プラットフォームは、一本鎖または二本鎖核酸分子を配列決定するのに専用であることがある。配列決定プラットフォームは、個々の塩基（たとえば、塩基単位配列決定）を読み取ることによって、または核酸分子（たとえば、識別子）内に組み込まれた核酸配列（たとえば、構成要素）全体の存在もしくは不在を検出することによって、データが符号化された核酸を復号することがある。配列決定プラットフォームは、乱雑な試薬の使用、読取り長さの増加、および検出可能な化学的部分の追加による固有核酸配列の検出を含むことがある。配列決定中のより乱雑な試薬の使用は、より高速な塩基呼出し（*calling*）を可能にすることによって読取り効率を増加させることがあり、それによって、配列決定時間を減少させることがある。読取り長さの増加の使用は、読取りごとに復号されることになる符号化された核酸のより長い配列を可能にすることがある。検出可能な化学的部分タグの追加は、化学的部分の存在または不在による核酸配列の存在または不在の検出を可能にすることがある。たとえば、情報のビットを符号化する各核酸配列は、一意の光学的信号、電気化学的信号、または化学的信号を生成する化学的部分とともにタグ付与されることがある。その一意の光学的信号、電気化学的信号、または化学的信号の存在または不在は、「0」または「1」のビット値を示すことがある。核酸配列は、単一の化学的部分を含んでもよいし、複数の化学的部分を含んでもよい。化学的部分は、データを符号化するための核酸配列の使用の前に核酸配列に追加されることがある。代替的に、またはこれに加えて、化学的部分は、データを符号化した後であるが、データを復号する前に、核酸配列に追加されることがある。化学的部分タグは、核酸配列に直接的に追加されることがあり、または、核酸配列は、合成または非天然のヌクレオチドアンカーを含むことがあり、化学的部分タグが、そのアンカーに追加されることがある。

【0147】

一意のコードは、符号化エラーおよび復号エラーを最小にするまたは検出するために適用されることがある。符号化エラーおよび復号エラーは、偽陰性（たとえば、ランダム試料採取に含まれない核酸分子または識別子）から発生することがある。エラー検出コードの一例は、識別子ライブラリに含まれる可能な識別子の連続したセット内の識別子の数をカウントするチェックサム配列であることがある。識別子ライブラリを読み取る間、チェックサムは、識別子のその連続したセットからのどれくらい多くの識別子を取り出すと予想されるかを示すことがあり、識別子は、予想された数が満たされるまで、読取りのために引き続き試料採取されることができる。一部の実施形態では、チェックサム配列は、R個の識別子のあらゆる連続したセットのために含まれることがあり、Rは、サイズが、1、2、5、10、50、100、200、500、または1000と等しいまたはこれよりも多くてもよいし、1000、500、200、100、50、10、5、または2よりも小さくてよい。Rの値が小さいほど、エラー検出が向上する。一部の実施形態では、チェックサムは、補足的核酸配列であってよい。たとえば、7つの核酸配列（たとえば、構成要素）を含むセットは、2つの群、すなわち、生産物方式を用いて識別子を構築するための核酸配列（層X内の構成要素X1～X3および層Y内のY1～Y3）、および補足的チェックサムのための核酸配列（X4～X7およびY4～Y7）に分割されることがあ

10

20

30

40

50

る。チェックサム配列 $X_4 \sim X_7$ は、層 X のゼロ、1つ、2つ、または3つの配列が層 Y の各部材と組み立てられるかどうかを示すことがある。代替的に、チェックサム配列 $Y_4 \sim Y_7$ は、層 Y のゼロ、1つ、2つ、または3つの配列が層 X の各部材と組み立てられるかどうかを示すことがある。この例では、識別子 $\{X_1 Y_1, X_1 Y_3, X_2 Y_1, X_2 Y_2, X_2 Y_3\}$ をもつ元の識別子ライブラリは、以下のプールすなわち $\{X_1 Y_1, X_1 Y_3, X_2 Y_1, X_2 Y_2, X_2 Y_3, X_1 Y_6, X_2 Y_7, X_3 Y_4, X_6 Y_1, X_5 Y_2, X_6 Y_3\}$ になるために、チェックサムを含むように補足されることがある。チェックサム配列は、誤り訂正にも使用されることがある。たとえば、上記のデータセットからの $X_1 Y_1$ の不在ならびに $X_1 Y_6$ および $X_6 Y_1$ の存在は、 $X_1 Y_1$ 核酸分子がデータセットから欠落しているという干渉を可能にすることがある。チェックサム配列は、識別子が識別子ライブラリの試料採取または識別子ライブラリのアクセスされる部分から欠落しているかどうかを示すことがある。欠落したチェックサム配列のケースでは、PCR または親和性タグプローブハイブリダイゼーションなどのアクセス方法は、それを増幅および/または隔離することがある。一部の実施形態では、チェックサムは、補足的核酸配列でなくてもよい。それらのチェックサムは、識別子によって表されるように、情報へと直接的にコードされることがある。

10

【0148】

データ符号化および復号化におけるノイズは、識別子を回文の形に構築することによって、たとえば、生産物方式内の単一の構成要素ではなく構成要素の回文的ペアを使用することによって、減少されることがある。次いで、異なる層からの構成要素のペアは、回文的様式で（たとえば、構成要素 X および Y に対して XY の代わりに YXY ）互いに組み立てられることがある。この回文的方法は、より大きな数の層（たとえば、 XYZ の代わりに $ZYXYZ$ ）に拡張されてよく、識別子間の誤った交差反応の検出を可能にすることがある。

20

【0149】

補足的核酸配列を過剰に（たとえば、非常に大きな過剰）識別子に追加することは、配列決定が符号化された識別子を回復するのを防止することがある。情報を復号する前に、識別子は、補足的核酸配列から濃縮される（*enriched*）ことがある。たとえば、識別子は、識別子末端に固有であるプライマーを使用した核酸増幅反応によって濃縮されることがある。代替的に、またはこれに加えて、情報は、固有プライマーを使用した配列決定（たとえば、合成による配列決定）によって試料プールを濃縮することなしに復号されることがある。両方の復号方法では、復号キーを有するまたは識別子の組成についての何かを知ることなしに情報を濃縮または復号することが困難なことがある。親和性タグベースのプローブを使用することなどの代替アクセス方法も用いられてよい。

30

【0150】

2進配列データを符号化するためのシステム

デジタル情報を核酸（たとえば、DNA）へと符号化するためのシステムは、システム、方法、およびファイルおよびデータ（たとえば、生データ、圧縮されたzipファイル、整数データ、および他の形式のデータ）をバイトへと変換し、このバイトを核酸、典型的にはDNAのセグメントもしくは配列、またはそれらの組合せへと符号化するためのデバイスを含むことができる。

40

【0151】

一態様では、本開示は、核酸を使用して2進配列データを符号化するためのシステムを提供する。核酸を使用して2進配列データを符号化するためのシステムは、デバイスと、1つまたは複数のコンピュータプロセッサとを含むことがある。デバイスは、識別子ライブラリを構築するように構成されることがある。1つまたは複数のコンピュータプロセッサは、(i) 情報をシンボルの文字列に翻訳し、(ii) シンボルの文字列を複数の識別子にマッピングし、(iii) 複数の識別子の少なくともサブセットを含む識別子ライブラリを構築するように個々にまたは集合的にプログラムされることがある。複数の識別子の個々の識別子は、シンボルの文字列の個々のシンボルに対応することがある。複数の識別子の個々の識別子は、1つまたは複数の構成要素を含むことがある。1つまたは複数の

50

構成要素の個々の構成要素は、核酸配列を含むことがある。

【0152】

別の態様では、本開示は、核酸を使用して2進配列データを読み取るためのシステムを提供する。核酸を使用して2進配列データを読み取るためのシステムは、データベースと、1つまたは複数のコンピュータプロセッサとを含むことがある。データベースは、情報を符号化する識別子ライブラリを記憶することがある。1つまたは複数のコンピュータプロセッサは、(i) 識別子ライブラリ内の識別子を識別し、(ii) (i)において識別された識別子から複数のシンボルを生成し、(iii) 複数のシンボルから情報をコンパイルするように個々にまたは集合的にプログラムされることがある。識別子ライブラリは、複数の識別子のサブセットを含むことがある。複数の識別子の各個々の識別子は、シンボルの文字列内の個々のシンボルに対応することがある。識別子は、1つまたは複数の構成要素を含むことがある。構成要素は、核酸配列を含むことがある。

10

【0153】

システムを使用してデジタルデータを符号化するための方法の非限定的な実施形態は、バイトストリームの形でデジタル情報を受信するためのステップを含むことができる。バイトストリームを個々のバイトへと解析すること、核酸インデックス(または識別子ランク)を使用してバイト内のビットのロケーションをマッピングすること、および1のビット値または0のビット値のどちらかに対応する配列を識別子へと符号化すること。デジタルデータを取り出すためのステップは、1つまたは複数のビットにマッピングする核酸(たとえば、識別子)の配列を含む核酸試料または核酸プールを配列決定すること、識別子ランクを参照して、識別子が核酸プールに存在するかどうかを確認すること、ならびに各配列に対するロケーションおよびビット値情報を、デジタル情報の配列を含むバイトへと復号することを含むことができる。

20

【0154】

符号化され核酸分子へと書き込まれた情報を符号化する、書き込む、コピーする、アクセスする、読み取る、および復号するためのシステムは、単一の統合されたユニットであってもよいし、前述の動作のうちの1つまたは複数の遂行するように構成された複数のユニットであってもよい。情報を符号化し核酸分子(たとえば、識別子)へと書き込むためのシステムは、デバイスと、1つまたは複数のコンピュータプロセッサとを含むことがある。1つまたは複数のコンピュータプロセッサは、情報をシンボルの文字列(たとえば、ビットの文字列)へと解析するようにプログラムされることがある。コンピュータプロセッサは、識別子ランクを生成することがある。コンピュータプロセッサは、シンボルを2つまたはそれより多くのカテゴリーへとカテゴリー化することがある。1つのカテゴリーは、識別子ライブラリ内での対応する識別子の存在によって表されることになるシンボルを含むことがあり、他のカテゴリーは、識別子ライブラリ内での対応する識別子の不在によって表されることになるシンボルを含むことがある。コンピュータプロセッサは、識別子ライブラリ内での識別子の存在に表されることになるシンボルに対応する識別子を組み立てるようにデバイスに指示することがある。

30

【0155】

デバイスは、複数の領域、セクション、またはパーティションを含むことがある。識別子を組み立てる試薬および構成要素は、デバイスの1つまたは複数の領域、セクション、またはパーティション内に記憶されることがある。層は、デバイスのセクションの別々の領域内に記憶されることがある。層は、1つまたは複数の一意の構成要素を含むことがある。1つの層内の構成要素は、別の層内の構成要素から一意であることがある。領域またはセクションは、容器を含むことがあり、パーティションは、ウェルを含むことがある。各層は、別々の容器またパーティション内に記憶されることがある。各試薬または核酸配列は、別々の容器またパーティション内に記憶されることがある。代替的に、またはこれに加えて、試薬は、識別子構造のためのマスタ混合を形成するように組み合わせられることがある。デバイスは、試薬、構成要素、および鋳型を、別のセクション内で組み合わせられることになるデバイスの1つのセクションから搬送することがある。デバイスは、アセン

40

50

ブリ反応を完了するための条件を提供することがある。たとえば、デバイスは、加熱、攪拌、および反応進捗の検出を提供することがある。構築された識別子は、1つまたは複数のその後の反応を受けて、バーコード、共通配列、可変配列、またはタグを識別子の1つまたは複数の末端に追加するように指示されることがある。次いで、識別子は、識別子ライブラリを生成するように領域またはパーティションに指示されることがある。1つまたは複数の識別子ライブラリは、デバイスの各領域、セクション、または個々のパーティション内に記憶されることがある。デバイスは、圧力、真空、または吸引を使用して、流体（たとえば、試薬、構成要素、鋳型）を搬送することがある。

【0156】

識別子ライブラリは、デバイス内に記憶されてもよいし、別々のデータベースに移されてもよい。データベースは、1つまたは複数の識別子ライブラリを含むことがある。データベースは、識別子ライブラリの長期記憶のための条件（たとえば、識別子の分解を減少させる条件）を提供することがある。識別子ライブラリは、粉末形式、液体形式、または固体形式で記憶されてよい。データベースは、紫外線光保護、温度の低下（たとえば、冷蔵または冷凍）、ならびに化学物質および酵素を分解させることからの保護を提供することがある。データベースに搬送される前に、識別子ライブラリは、凍結乾燥または凍結されてよい。識別子ライブラリは、核酸分子の安定性を維持するように、ヌクレアーゼおよび/または緩衝液を不活性化するために、エチレンジアミン四酢酸（EDTA）を含むことがある。

【0157】

データベースは、情報を識別子に書き込む、情報をコピーする、情報にアクセスする、または情報を読み取るデバイスにカップリングされてもよいし、これを含んでもよいし、これから分離されてもよい。識別子ライブラリの一部分が、コピー、アクセス、または読取の前にデータベースから除去されることがある。データベースから情報をコピーするデバイスは、情報を書き込むデバイスと同じデバイスであってもよいし、これと異なるデバイスであってもよい。情報をコピーするデバイスは、識別子ライブラリのアリコートデバイスをデバイスから抽出し、そのアリコートを試薬および構成物と組み合わせて、識別子ライブラリの一部分または全体を増幅することがある。デバイスは、増幅反応の温度、圧力、および攪拌を制御することがある。デバイスは、パーティションを含むことがあり、1つまたは複数の増幅反応は、識別子ライブラリを含むパーティション内で発生することがある。デバイスは、識別子の1つよりも多いプールを一度にコピーすることがある。

【0158】

コピーされた識別子は、コピーデバイスからアクセスデバイスに搬送されることがある。アクセスデバイスは、コピーデバイスと同じデバイスであってもよい。アクセスデバイスは、別々の領域、セクション、またはパーティションを含むことがある。アクセスデバイスは、親和性タグに結合された識別子を分離するための1つまたは複数のカラム、ピーズリザーバ、または磁気領域を有することがある。代替的に、またはこれに加えて、アクセスデバイスは、1つまたは複数のサイズ選択ユニットを有することがある。サイズ選択ユニットは、アガロースゲル電気泳動または核酸分子を選択するサイズのための他の任意の方法を含むことがある。コピーおよび抽出は、デバイスの同じ領域内で実行されてもよいし、デバイスの異なる領域内で実行されてもよい。

【0159】

アクセスされたデータは、同じデバイス内で読み取られてもよいし、アクセスされたデータは、別のデバイスに搬送されてもよい。読取りデバイスは、識別子を検出および識別するために、検出ユニットを含むことがある。検出ユニットは、シーケンサ、ハイブリダイゼーションアレイ、または識別子の存在もしくは不在を識別するための他のユニットの一部であってもよい。配列決定プラットフォームは、核酸配列へと符号化された情報を復号して読み取るために具体的に設計されることがある。配列決定プラットフォームは、一本鎖または二本鎖核酸分子を配列決定するのに専用であることがある。配列決定プラットフォームは、個々の塩基（たとえば、塩基単位配列決定）を読み取ることによって、または

10

20

30

40

50

核酸分子（たとえば、識別子）内に組み込まれた核酸配列（たとえば、構成要素）全体の存在もしくは不在を検出することによって、データが符号化された核酸を復号することがある。代替的に、配列決定プラットフォームは、Illumina（登録商標）Sequencingまたはキャピラリー電気泳動によるフラグメンテーション解析などのシステムであってよい。代替的に、またはこれに加えて、核酸配列を復号することは、限定するものではないが、光信号、電気化学信号、または化学信号を生成する任意の方法を含む、デバイスによって実装される、さまざまな解析技法を使用して、実行されることがある。

【0160】

核酸分子内の情報記憶は、限定するものではないが、長期情報記憶、機密情報記憶、および医療情報の記憶を含む種々の適用例を有することがある。一例では、人間の医療情報（たとえば、病歴および記録）が、核酸分子内に記憶され、その人間に携行されることがある。情報は、身体の外部に（たとえば、ウェアラブルデバイス内に）に記憶されてもよいし、身体の内部に（たとえば、皮下カプセル内に）記憶されてもよい。患者が診療所または病院に連れて来られたとき、試料が、デバイスまたはカプセルから取られることがあり、情報は、核酸シーケンサを使用して復号されることがある。核酸分子内での医療記憶の個人的記憶は、コンピュータおよびクラウドベース記憶システムに代替案を提供することがある。核酸分子内での医療記録の個人的記憶は、医療記録がハッキングされる事例または流行を減少させることがある。医療記録のカプセルベースの記憶に使用される核酸分子は、ヒトゲノム配列から派生されることがある。ヒトゲノム配列の使用は、カプセル損傷および漏洩のイベントでの核酸配列の免疫原性を減少させることがある。

【0161】

コンピュータシステム

本開示は、本開示の方法を実装するようにプログラムされるコンピュータシステムを提供する。図19は、デジタル情報を核酸配列へと符号化するおよび/または核酸配列から派生された情報を読み取る（たとえば、復号する）ようにプログラムまたはそうでなければ構成されたコンピュータシステム1901を示す。コンピュータシステム1901は、たとえば、符号化されたビットストリームまたはバイトストリームからの所与のビットまたはバイトのためのビット値およびビットロケーション情報などの、本開示の符号化手順および復号手順の種々の態様を調節することができる。

【0162】

コンピュータシステム1901は、中央処理ユニット（CPU、また本明細書において「プロセッサ」、および「コンピュータプロセッサ」）1905を含み、これは、シングルコアプロセッサもしくはマルチコアプロセッサ、または並列処理のための複数のプロセッサとすることができる。コンピュータシステム1901は、メモリまたはメモリロケーション1910（たとえば、ランダムアクセスメモリ、読出し専用メモリ、フラッシュメモリ）と、電子記憶ユニット1915（たとえば、ハードディスク）と、1つまたは複数の他のシステムと通信するための通信インターフェース1920（たとえば、ネットワークアダプタ）と、キャッシュ、他のメモリ、データ記憶および/または電子ディスプレイアダプタなどの周辺デバイス1925も含む。メモリ1910、記憶ユニット1915、インターフェース1920、および周辺デバイス1925は、マザーボードなどの、通信バス（実線）を通してCPU1905と通信する。記憶ユニット1915は、データを記憶するためのデータ記憶ユニット（またはデータリポジトリ）とすることができる。コンピュータシステム1901は、通信インターフェース1920を用いてコンピュータネットワーク（「ネットワーク」）1930に動作可能にカップリング可能である。ネットワーク1930は、インターネット、インターネットおよび/もしくはエクストラネット、またはインターネットと通信するイントラネットおよび/もしくはエクストラネットとすることができる。ネットワーク1930は、いくつかのケースでは、電気通信および/またはデータネットワークである。ネットワーク1930は、1つまたは複数のコンピュータサーバを含むことができ、これは、クラウドコンピューティングなどの分散コンピューティングを可能にすることができる。ネットワーク1930は、コンピュータシステム1901

を用いるいくつかのケースでは、ピアツーピアネットワークを実装することができ、このことによって、コンピュータシステム 1901 にカップリングされたデバイスがクライアントまたはサーバとして振る舞うことが可能になることがある。

【0163】

CPU 1905 は、プログラムまたはソフトウェア内で具現化され得る機械可読命令の配列を遂行することができる。命令は、メモリ 1910 などのメモリロケーション内に記憶されることがある。命令は、CPU 1905 に向けられることができ、命令は、その後、本開示の方法を実装するように CPU 1905 をプログラムまたはそうでなければ構成することができる。CPU 1905 によって実行される動作の例は、フェッチ、復号、遂行、およびライトバックを含むことができる。

10

【0164】

CPU 1905 は、集積回路などの回路の一部とすることができる。システム 1901 の 1 つまたは複数の他の構成要素は、回路内に含まれることができる。いくつかのケースでは、回路は、特定用途向け集積回路 (ASIC) である。

【0165】

記憶ユニット 1915 は、ドライバ、ライブラリ、および保存されたプログラムなどの、ファイルを記憶することができる。記憶ユニット 1915 は、ユーザデータ、たとえば、ユーザの好みおよびユーザプログラムを記憶することができる。いくつかのケースにおけるコンピュータシステム 1901 は、イントラネットまたはインターネットを通してコンピュータシステム 1901 と通信するリモートサーバ上に配されるなどの、コンピュータシステム 1901 の外部にある 1 つまたは複数の追加データ記憶ユニットを含むことができる。

20

【0166】

コンピュータシステム 1901 は、ネットワーク 1930 を通して 1 つまたは複数のリモートコンピュータシステムと通信することができる。たとえば、コンピュータシステム 1901 は、ユーザのリモートコンピュータシステムまたは核酸の配列内の符号化されたもしくは復号されたデータを解析中にユーザによって使用され得る他のデバイスおよびもしくは機械類 (たとえば、シーケンサまたは核酸配列内の窒素塩基の順序を化学的に決定するための他のシステム) と通信することができる。リモートコンピュータシステムの例としては、パーソナルコンピュータ (たとえば、ポータブル PC)、スレートもしくはタブレット PC (たとえば、Apple (登録商標) iPad (登録商標)、Samsung (登録商標) Galaxy Tab)、電話、スマートフォン (たとえば、Apple (登録商標) iPhone (登録商標)、Android 対応デバイス、BlackBerry (登録商標))、または携帯情報端末がある。ユーザは、ネットワーク 1930 を介してコンピュータシステム 1901 にアクセスすることができる。

30

【0167】

本明細書において説明される方法は、たとえば、メモリ 1910 または電子記憶ユニット 1915 上などの、コンピュータシステム 1901 の電子記憶ロケーション上に記憶された機械 (たとえば、コンピュータプロセッサ) 実行可能コードによって実装可能である。機械実行可能コードまたは機械可読コードは、ソフトウェアの形式で提供可能である。使用中、コードは、プロセッサ 1905 によって実行可能である。いくつかのケースでは、コードは、記憶ユニット 1915 から取り出され、プロセッサ 1905 による準備の整ったアクセスのためにメモリ 1910 上に記憶可能である。いくつかの状況では、電子記憶ユニット 1915 は排除可能であり、機械実行可能命令は、メモリ 1910 上に記憶される。

40

【0168】

コードは、あらかじめコンパイルされ、コードを遂行するように適合されたプロセッサを有する機械とともに使用するために構成可能である、またはランタイム中にコンパイル可能である。コードは、あらかじめコンパイルされた様式またはコンパイルされたような (as-compiled) 様式でコードが遂行されることを可能にするために選択可能で

50

あるプログラミング言語において供給可能である。

【 0 1 6 9 】

コンピュータシステム 1 9 0 1 などの、本明細書において提供されるシステムおよび方法の態様は、プログラミングにおいて具現化可能である。技術の種々の態様は、典型的には一種の機械可読媒体において携行または具現化される機械（またはプロセッサ）実行可能コードおよび／または関連づけられたデータの形式をした「生産物」または「製造品」とみなされることがある。機械実行可能コードは、メモリ（たとえば、読出し専用メモリ、ランダムアクセスメモリ、フラッシュメモリ）またはハードディスクなどの電子記憶ユニット上に記憶可能である。「記憶」タイプ媒体は、コンピュータの有形メモリ、プロセッサなど、またはソフトウェアプログラミングのための任意の時点で非一時的な記憶を提供し得る種々の半導体メモリ、テープドライブ、ディスクドライブなどの、その関連モジュールのいずれかまたはすべてを含むことができる。ソフトウェアのすべてまたは一部分は、時々、インターネットまたは種々の他の電気通信ネットワークを通して通信されることがある。そのような通信は、たとえば、1つのコンピュータまたはプロセッサから別のコンピュータまたはプロセッサへの、たとえば、管理サーバまたはホストコンピュータからアプリケーションサーバのコンピュータプラットフォームへの、ソフトウェアのロードを可能にすることがある。したがって、ソフトウェアエレメントを担持することがある別のタイプの媒体は、ローカルデバイス間の物理的インタフェースにわたって、ワイヤードおよび光学的地上ネットワークを通して、および種々のエアリンク上で、使用されるなどの、光学波、電気波、および電磁波を含む。ワイヤードリンクまたはワイヤレスリンク、光学的リンクなどの、そのような波を携行する物理的エレメントも、ソフトウェアを担持する媒体とみなされることがある。本明細書において使用されるように、非一時的な有形の「記憶」媒体に制限されない限り、コンピュータまたは機械「可読媒体」などの用語は、遂行のためにプロセッサに命令を提供することに関わる任意の媒体を指す。

【 0 1 7 0 】

したがって、コンピュータ実行可能コードなどの機械可読媒体は、限定するものではないが、有形記憶媒体、搬送波媒体、または物理的伝送媒体を含む多くの形式をとることがある。不揮発性記憶媒体としては、たとえば、図面に示されるデータベースなどを実装するために使用され得るなどの、任意のコンピュータ内などの記憶デバイスのいずれかなどの光ディスクまたは磁気ディスクがある。揮発性記憶媒体としては、そのようなコンピュータプラットフォームのメインメモリなどの動的メモリがある。有形伝送媒体としては、同軸ケーブル、コンピュータシステム内のバスを含むワイヤを含む銅線および光ファイバがある。搬送波伝送媒体は、電気信号もしくは電磁信号、または無線周波数（RF）および赤外線（IR）データ通信中に生成される波などの音響波もしくは光波の形式をとることがある。したがって、コンピュータ可読媒体の一般的な形式としては、たとえば、フロッピー（登録商標）ディスク、フレキシブルディスク、ハードディスク、磁気テープ、他の任意の磁気媒体、CD-ROM、DVDもしくはDVD-ROM、他の任意の光学的媒体、パンチカード紙テープ、穴のパターンをもつ他の任意の物理的記憶媒体、RAM、ROM、PROMおよびEPROM、FLASH-EPROM、他の任意のメモリチップもしくはカートリッジ、データまたは命令を輸送する搬送波、そのような搬送波を輸送するケーブルもしくはリンク、またはコンピュータがプログラミングコードおよび／もしくはデータを読み取り可能な他の任意の媒体がある。これらの形式のコンピュータ可読媒体の多くは、1つまたは複数の命令の1つまたは複数の配列を遂行のためにプロセッサに搬送することに伴われ得る。

【 0 1 7 1 】

コンピュータシステム 1 9 0 1 は、たとえば、核酸、生データ、ファイル、およびDNAが記憶されたデータへと符号化または復号されることになる圧縮または解凍されたzipファイルを符号化または復号している機械またはコンピュータシステムによって符号化されたまたはこれによって読み取られたクロマトグラフ、配列ならびにビット、バイト、またはビットストリームを含む配列出力データを提供するためのユーザインタフェース（

10

20

30

40

50

UI) 1940を含む電子ディスプレイ1935を含む、またはこれと通信することができる。UIの例としては、限定するものではないが、グラフィカルユーザインタフェース(GUI)およびウェブベースのユーザインタフェースがある。

本開示の方法およびシステムは、1つまたは複数のアルゴリズムによって実装可能である。アルゴリズムは、中央処理ユニット1905による遂行時にソフトウェアによって実装可能である。アルゴリズムは、たとえば、デジタル情報を符号化する前に、生データまたはzipファイル圧縮されたデータからデジタル情報をコードするためのカスタマイズされた方法を決定するために、DNAインデックスおよび生データまたはzipファイル圧縮または解凍されたデータとともに使用可能である。

【実施例】

【0172】

(実施例1)

DNA分子内の単一のポエム(poem)の符号化、書込み、および読取り

符号化されることになるデータは、ポエムを含有するテキストファイルである。データは、オーバーラップ伸長PCRとともに実装される生産物方式を使用して識別子を構築するように、96個の構成要素の2つの層からのDNA構成要素と一緒に混合するためにピペットを用いて手動で符号化される。第1の層すなわちXは、96個の全体的DNA構成要素を含む。第2の層すなわちYも、96個の全体的構成要素を含む。DNAを書き込む前に、データは、2進にマッピングされ、次いで、元のデータの61ビットのあらゆる連続した(隣接する解体)文字列が1の正確に17ビット値をもつ96ビット文字列に翻訳される均一な重みフォーマットに記録される。この均一な重みフォーマットは、天然誤りチェック品質を有することがある。次いで、データは、参照マップを形成するために96×96テーブルにハッシュされる。

【0173】

図18Aの中央パネルは、ポエムを複数の識別子へと符号化する、96×96テーブルの2次元参照マップを図示する。暗い点は「1」ビット値に対応し、白い点は「0」ビット値に対応する。データは、96個の構成要素の2つの層を使用して識別子へと符号化される。テーブルの各X値およびY値は、構成要素に割り当てられ、X構成要素およびY構成要素は、「1」値をもつ各(X,Y)座標にオーバーラップ伸長PCRを使用して、識別子へと組み立てられる。データは、各可能な(X,Y)アセンブリの存在または不在を決定するために、識別子ライブラリを配列決定することによって読み返された(たとえば、復号された)。

【0174】

図18Aの右パネルは、配列決定によって決定される識別子ライブラリに存在する配列の存在度の2次元ヒートマップ(heat map)を図示する。各ピクセルは、対応するX構成要素とY構成要素を含む分子を表し、そのピクセルにおけるグレイスケール強度は、他の分子と比較した、その分子の相対的存在度を表す。識別子は、各行の上位17の最も豊富な(X,Y)アセンブリとしてみなされる(均一な重み符号化は、96ビットの各連続した文字列が、正確に17個の「1」値、したがって17個の対応する識別子を有することがあることを保証するので)。

【0175】

(実施例2)

62824ビットテキストファイルのコード化

符号化されることになるデータは、合計で62824ビットになる、3つのポエムのテキストファイルである。データは、オーバーラップ伸長PCRを用いて実装される生産物方式を使用して識別子を構築するように、384個の構成要素の2つの層からのDNA構成要素と一緒に混合するために、Labcyte Echo(登録商標)液体処理装置を使用して符号化される。第1の層すなわちXは、384個の全体的DNA構成要素を含む。第2の層すなわちYも、384個の全体的構成要素を含む。DNAを書き込む前に、データは、2進にマッピングされ、次いで、重み(「1」のビット値の数)を減少させ、チ

10

20

30

40

50

チェックサムを含むために記録される。チェックサムは、データの192ビットのあらゆる連続した文字列に対してチェックサムに対応する識別子があるように確立される。再コードされたデータは、約10,100の重みを有し、これは、構築されることになる識別子の数に対応する。次いで、データは、参照マップを形成するために384×384テーブルにハッシュされることがある。

【0176】

図18Bの中央パネルは、テキストファイルを複数の識別子へと符号化する、384×384テーブルの2次元参照マップを図示する。各座標(X,Y)は、位置 $X + (Y - 1) * 192$ におけるデータのビットに対応する。黒い点は「1」のビット値に対応し、白い点は「0」のビット値に対応する。図の右側の黒い点はチェックサムであり、図の上部の黒い点のパターンはコードブック（たとえば、データを復号するための辞書）である。テーブルの各X値およびY値は、構成要素に割り当てられてよく、X構成要素およびY構成要素は、「1」値をもつ各(X,Y)座標にオーバーラップ伸長PCRを使用して、識別子へと組み立てられる。データは、各可能な(X,Y)アセンブリの存在または不在を決定するために、識別子ライブラリを配列決定することによって読み返された（たとえば、復号された）。

【0177】

図18Bの右パネルは、配列決定によって決定される識別子ライブラリに存在する配列の存在度の2次元ヒートマップを図示する。各ピクセルは、対応するX構成要素とY構成要素を含む分子を表し、そのピクセルにおけるグレイスケール強度は、他の分子と比較した、その分子の相対的存在度を表す。識別子は、各行における上位S個の最も豊富な(X,Y)アセンブリとみなされ、各行に対するSは、チェックサム値であってよい。

【0178】

本発明の好ましい実施形態が本明細書において図示および説明されているが、そのような実施形態が単に例として提供されることは、当業者に明白であろう。本発明が、本明細書において提供される特定の例によって制限されることは意図されていない。本発明は、前述の明細書を参照しながら説明されてきたが、本明細書における実施形態の説明および例示は、限定的な意味で解釈されることを意味したものではない。ここで、数多くの変形、変更、および置換えが、本発明から逸脱することなく当業者に想到される。そのうえ、本発明のすべての態様は、さまざまな条件および変数に依存する、本明細書において記載される特定の描写、構成、または相対的な比率に限定されないことが理解されるものとする。本発明を実践する際に本明細書において説明される本発明の実施形態に対する種々の代替形態が用いられてよいことは理解されるべきである。したがって、本発明は、任意のそのような代替形態、修正形態、変形形態、または均等物も包含するものとすることが意図される。以下の特許請求の範囲が本発明の範囲を規定し、これらの特許請求の範囲内の方法および構造ならびにその均等物がそれによって包含されることが意図されている。

10

20

30

40

50

【図面】

【図 1】

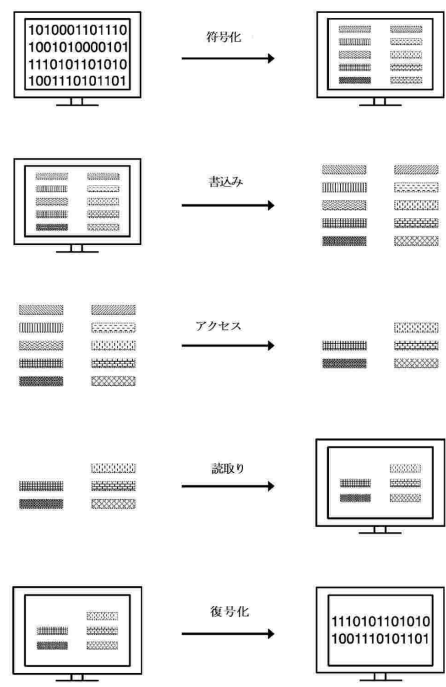


FIG. 1

【図 2】

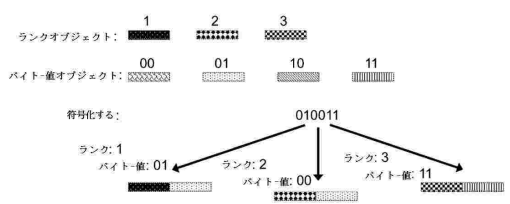


FIG. 2A

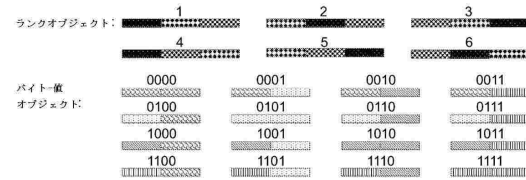


FIG. 2B

【図 3】

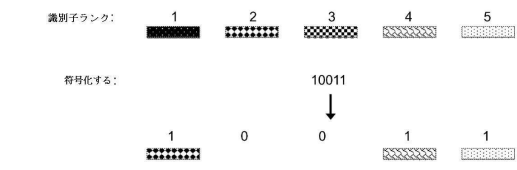


FIG. 3A

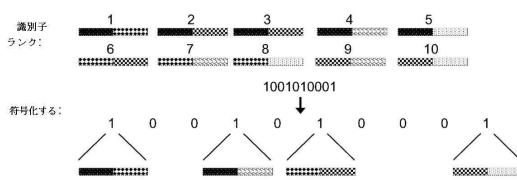


FIG. 3B

【図 4】

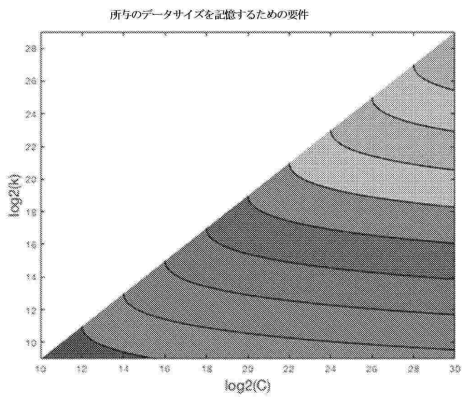


FIG. 4

10

20

30

40

50

【図 5】

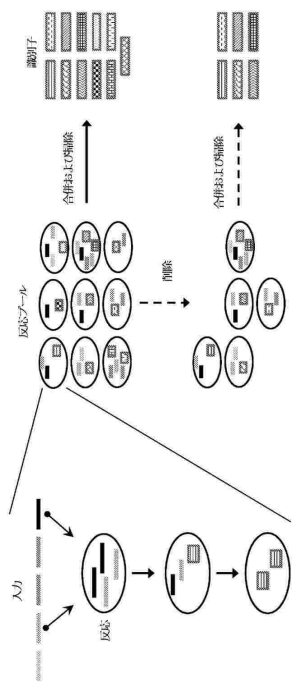


FIG. 5

【図 6】

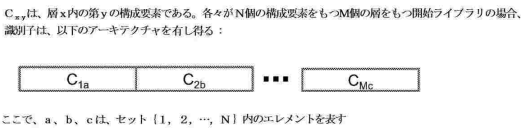


FIG. 6A

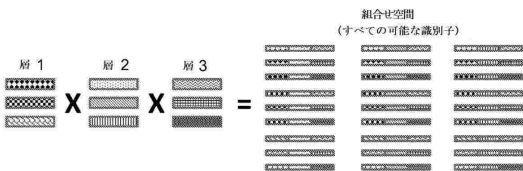


FIG. 6B

【図 7】

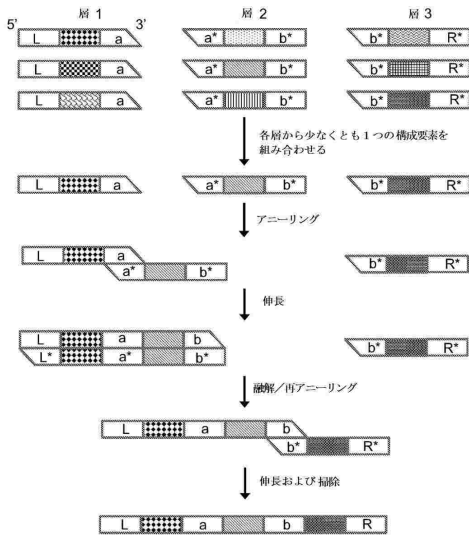


FIG. 7

【図 8】

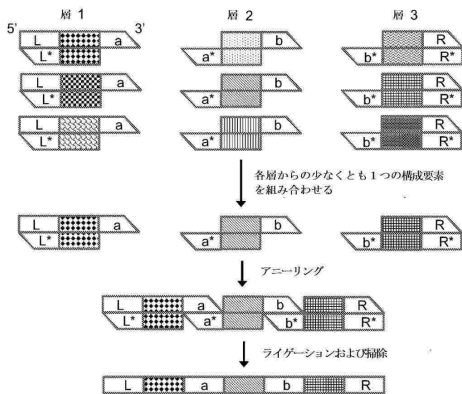


FIG. 8

10

20

30

40

50

【 図 9 】

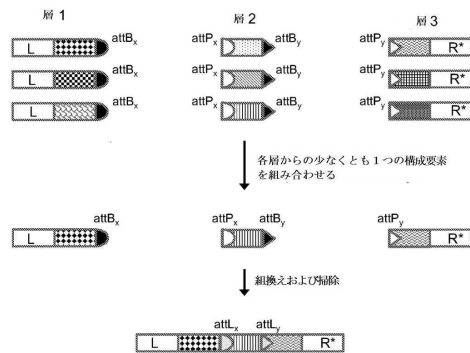


FIG. 9

【 図 1 0 A 】

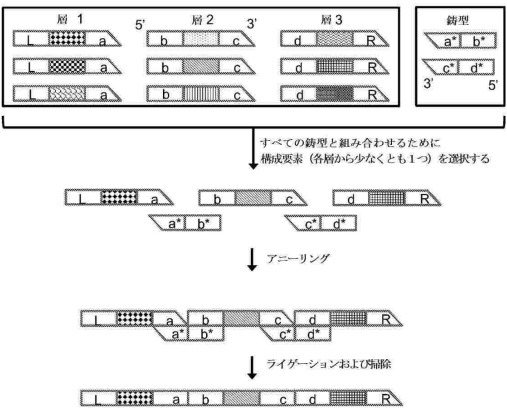


FIG. 10A

【 図 1 0 B 】

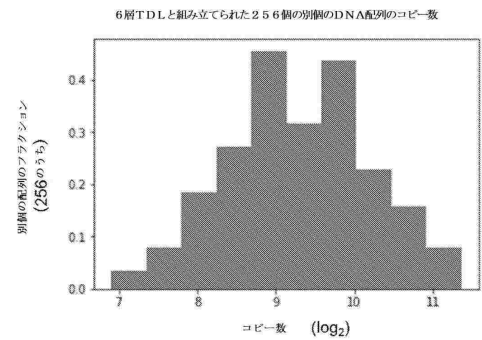


FIG. 10B

【 図 1 1 A 】

C_{xy} は、層x内の第yの構成要素である。各々がN個の構成要素をもつM個の層をもつ開始ライブラリの場合、識別子は、以下のアーキテクチャを有し得る：



ここで、b、e、hは、セット{1, 2, ..., N}内のエレメントであり、a、d、fは、セット{1, 2, ..., M}内の一意の要素である

可能な識別子の組合せ空間は $N^M M!$ 個である

FIG. 11A

10

20

30

40

50

【図 1 1 B】

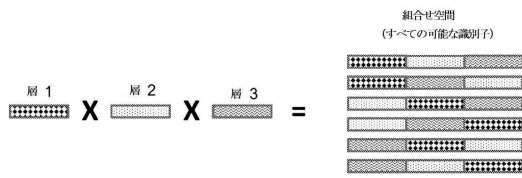


FIG. 11B

【図 1 1 C】

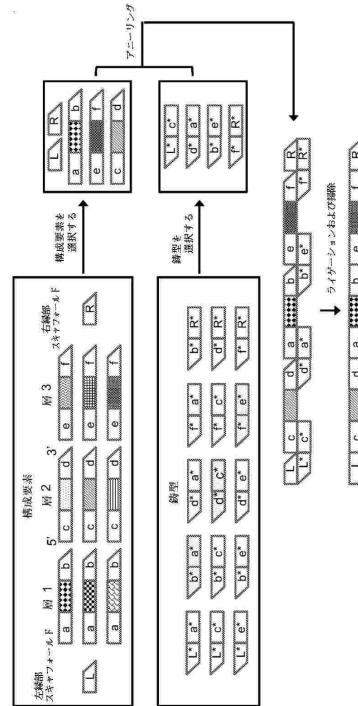


FIG. 11C

【図 1 1 D】

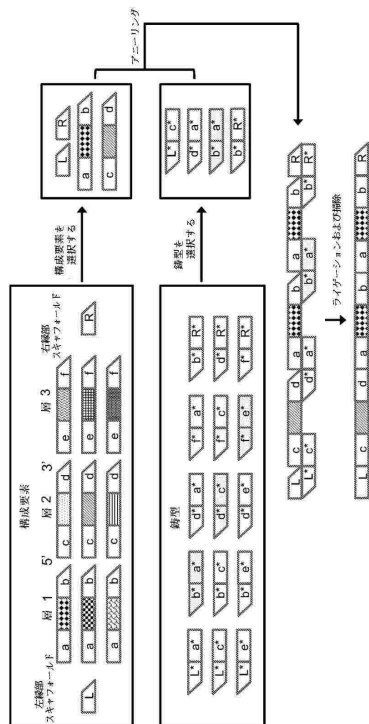


FIG. 11D

【図 1 1 E】

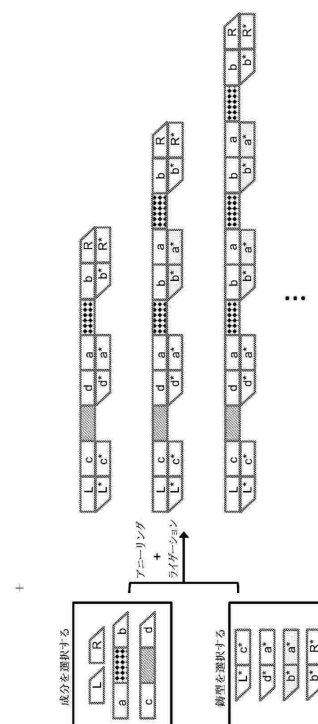


FIG. 11E

10

20

30

40

50

【図 1 1 F】

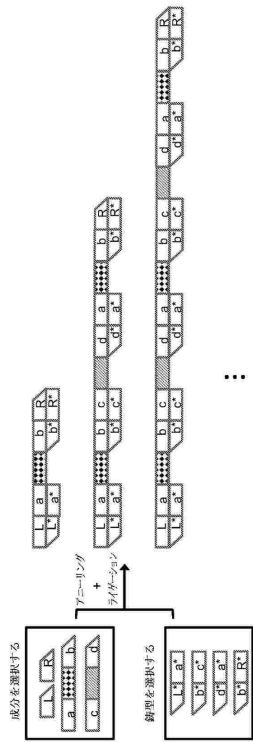


FIG. 11F

【図 1 1 G】

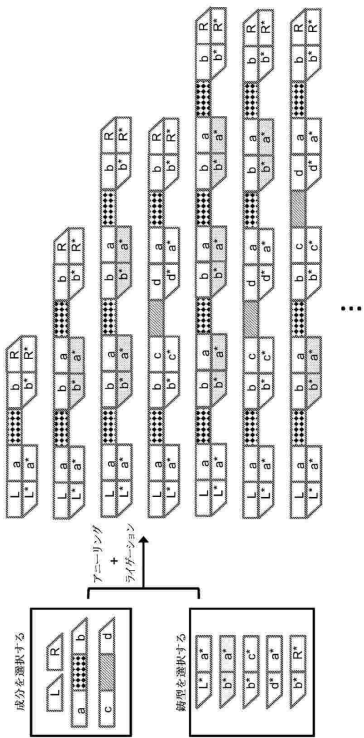


FIG. 11G

10

20

【図 1 2 A】

C_{xy} は、層x内の第yの構成要素である。各々がN個の構成要素をもつM個の層をもつ開始ライブラリの場合、識別子は、固定数の構成要素kを組み立て、以下のアーキテクチャを有する：

C_{ab} C_{de} ... C_{fh}

ここで、b、e、hは、セット{1, 2, ..., N}内のエレメントであり、a、d、fは、セット{1, 2, ..., M}内のk個の一意のランク順序 付き要素のうちである

可能な識別子の組合せ空間は、 $N^k M \text{choose} k$ である

FIG. 12A

【図 1 2 B】

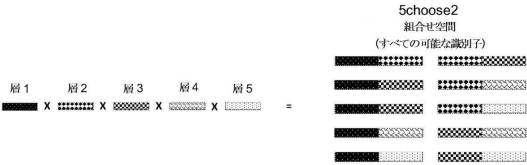


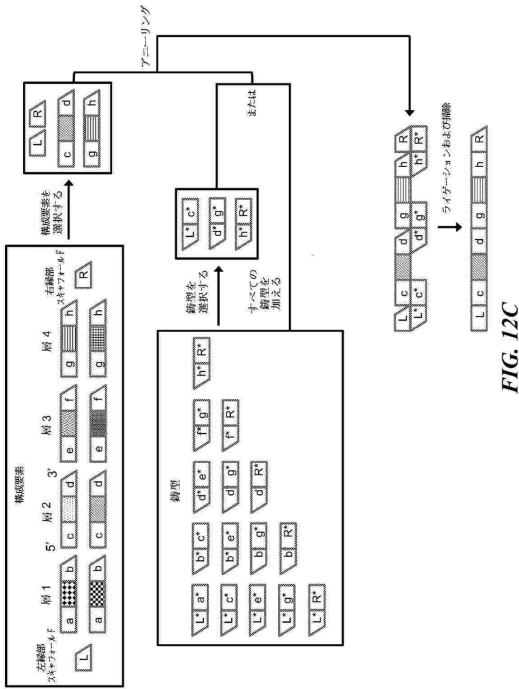
FIG. 12B

30

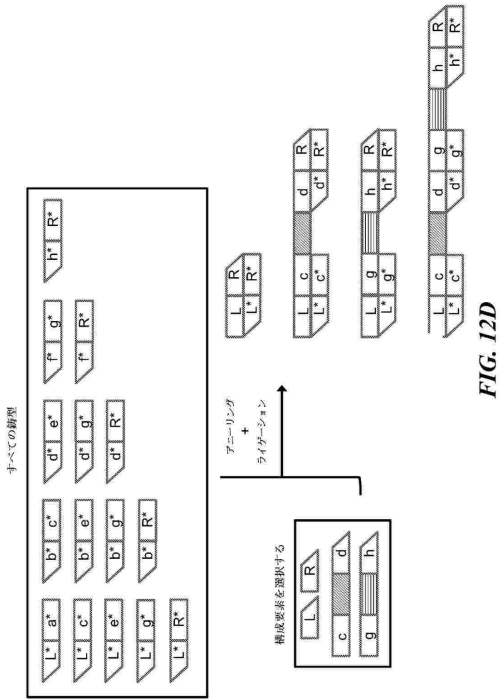
40

50

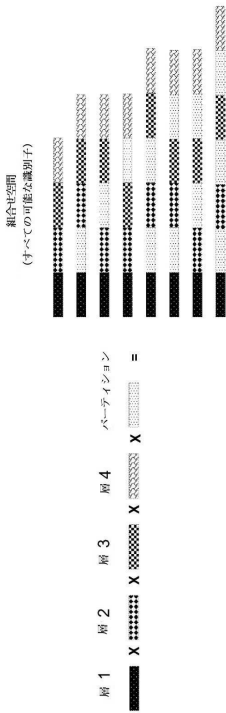
【図 1 2 C】



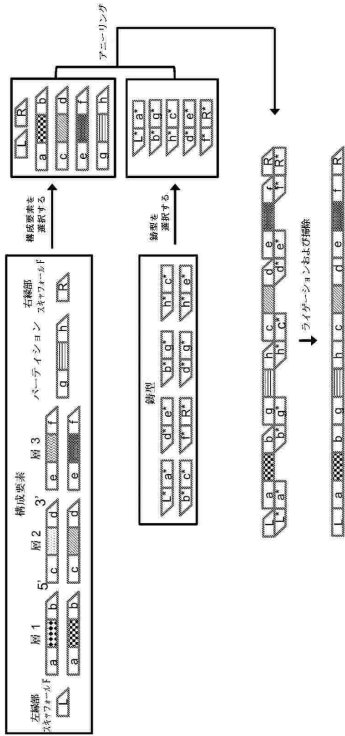
【図 1 2 D】



【図 1 3 A】



【図 1 3 B】



【図 1 4 A】

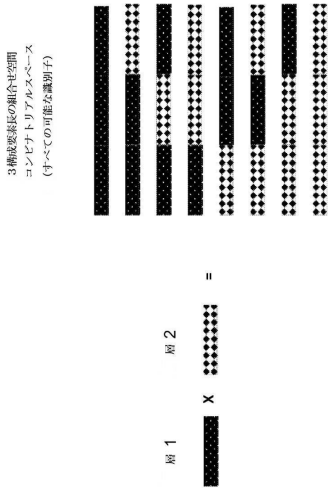


FIG. 14A

【図 1 4 B】

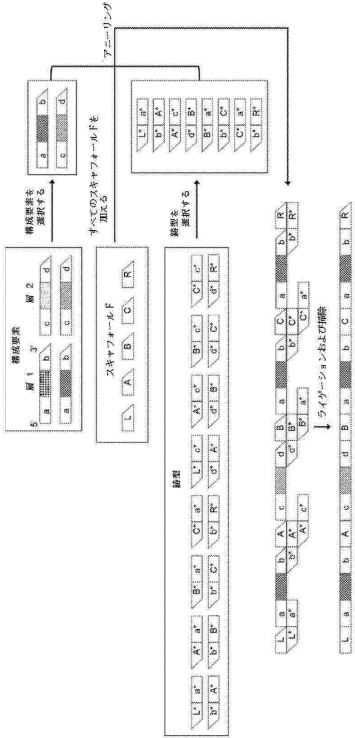


FIG. 14B

【図 1 5 A】

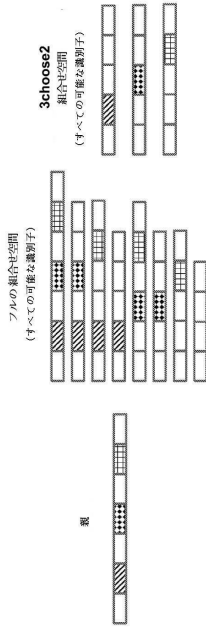


FIG. 15A

【図 1 5 B】

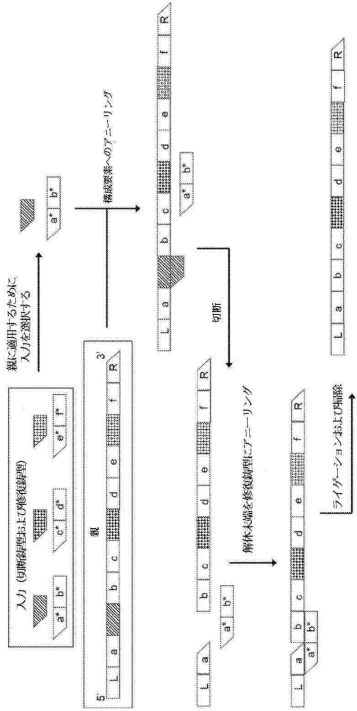


FIG. 15B

【図 16】



FIG. 16

【図 17 A】

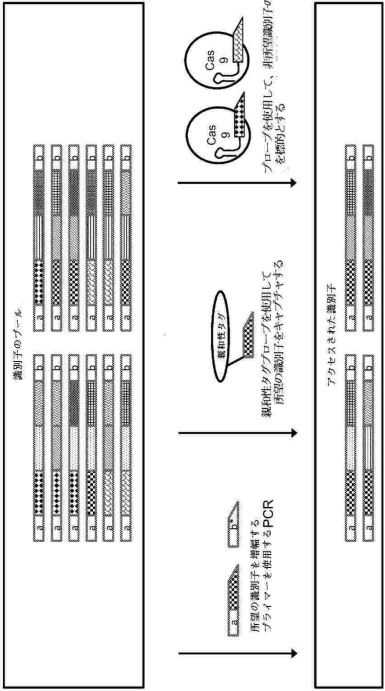


FIG. 17A

【図 17 B】

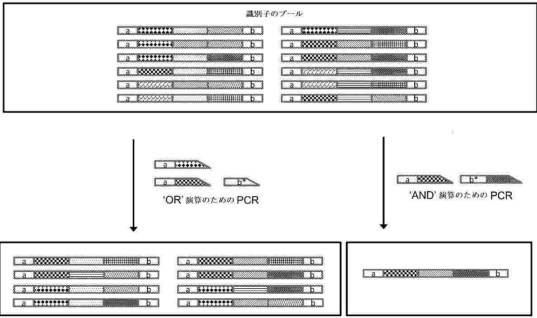


FIG. 17B

【図 17 C】

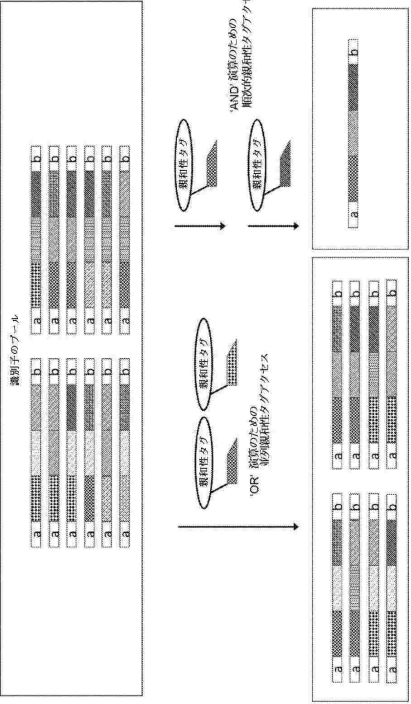


FIG. 17C

【 図 1 8 】

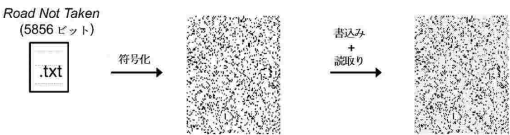


FIG. 18A

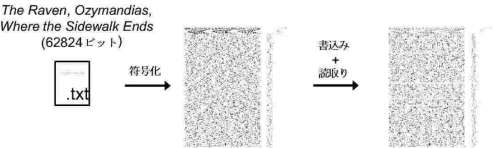


FIG. 18B

【 図 1 9 】

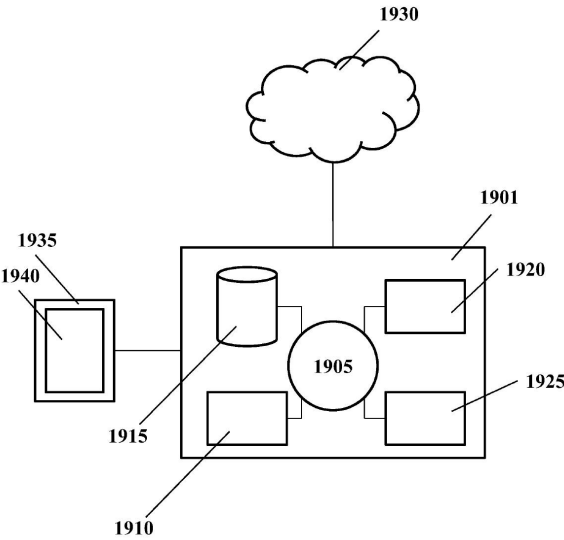


FIG. 19

10

20

30

40

50

フロントページの続き

米国(US)

(31)優先権主張番号 62/466,304

(32)優先日 平成29年3月2日(2017.3.2)

(33)優先権主張国・地域又は機関

米国(US)

アメリカ合衆国 マサチューセッツ 02139, ケンブリッジ, コロンビア ストリート 147, ユニット 3アール

(72)発明者 ヒョンジュン パク

アメリカ合衆国 マサチューセッツ 02134, ケンブリッジ, シドニー ストリート 91, スイート 609

(72)発明者 スワップニル ピー . パティア

アメリカ合衆国 マサチューセッツ 02134, ボストン, ウェスタン アベニュー 127

審査官 坂井田 京

(56)参考文献 特表2015-533077(JP, A)

国際公開第2004/009844(WO, A1)

特表2015-525560(JP, A)

特表2015-529864(JP, A)

A DNA-Based Archival Storage System, Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems, 2016年03月, pp.637-664

IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 2015, 1(3), pp.230-248

(58)調査した分野 (Int.Cl., DB名)

C40B 40/00 - 40/18

G16B 50/00 - 50/50

C12N 15/00 - 15/90

JSTPlus/JMEDPlus/JST7580(JDreamIII)

Caplus/MEDLINE/EMBASE/BIOSIS(STN)