

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2019-523940

(P2019-523940A)

(43) 公表日 令和1年8月29日(2019.8.29)

(51) Int.Cl.
G 1 6 B 20/20 (2019.01)F 1
G 1 6 B 20/20

テーマコード (参考)

審査請求 未請求 予備審査請求 未請求 (全 35 頁)

(21) 出願番号 特願2018-563706 (P2018-563706)
 (86) (22) 出願日 平成29年6月9日 (2017.6.9)
 (85) 翻訳文提出日 平成31年2月4日 (2019.2.4)
 (86) 国際出願番号 PCT/US2017/036868
 (87) 国際公開番号 W02017/214574
 (87) 国際公開日 平成29年12月14日 (2017.12.14)
 (31) 優先権主張番号 62/348,786
 (32) 優先日 平成28年6月10日 (2016.6.10)
 (33) 優先権主張国・地域又は機関
 米国 (US)
 (31) 優先権主張番号 62/375,858
 (32) 優先日 平成28年8月16日 (2016.8.16)
 (33) 優先権主張国・地域又は機関
 米国 (US)

(71) 出願人 516040017
 ツイスト バイオサイエンス コーポレー
 ション
 アメリカ合衆国 94158 カリフォル
 ニア州 サンフランシスコ ミッション・
 ベイ・ブールバード サウス 455 ス
 イート 545
 (74) 代理人 100082072
 弁理士 清原 義博
 (72) 発明者 ディガンズ, ジェームズ
 アメリカ合衆国 94070 カリフォル
 ニア州 サン・カルロス コーディレラス
 ・アベニュー 1324

最終頁に続く

(54) 【発明の名称】 生物学的配列の自動アノテーションとスクリーニングのためのシステムおよび方法

(57) 【要約】

本開示は、コミュニティ知識および参加に基づく有効なバイオセキュリティのためのソフトウェアツールを記載する。本明細書に記載されるアノテーションツールは、個々のタンパク質および負の結果の関係性に関する新しい科学を追跡するために、合成生物学コミュニティを支援する。実務者および生物学的配列または構築物の提供者が、合成や発現まで待つのではなく、オーダーリクエストの安全性を評価する権限を与えられるように、本明細書に記載されるスクリーニングツールによって、コミュニティはバイオセキュリティの対象および有効な実施の両方を広げることができる。加えて、本明細書に記載されるスクリーニングツールは、基準データベースからの有害な生物学的配列に関連する配列への同じか複数のオーダーでのポリヌクレオチドのスクリーニングを提供する。

【選択図】 図3A

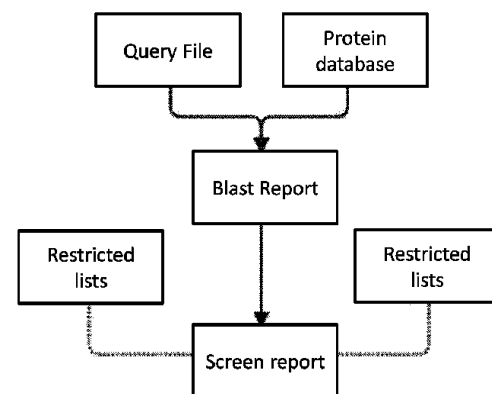


FIG. 3A

【特許請求の範囲】**【請求項 1】**

増強されたポリヌクレオチド合成を提供するためのコンピュータ化されたシステムであって：前記コンピュータ化されたシステムは、

a) 有害な生物学的配列のリストを表わすことに適しているデータベースをホストするためのサーバー；

b) ネットワーク接続；および

c) 汎用コンピュータに対する命令を含むコンピュータ可読媒体を含み；

前記コンピュータ化されたシステムは、：

i) 1つ以上の設計命令を受信する方法であり、ここで、設計命令は複数の生物学的配列を含み、生物学的配列の各々はせいぜい500の塩基の長さであり、そして複数の生物学的配列は核酸またはアミノ酸配列を含む、方法；

i i) 複数の生物学的配列の少なくとも2つの生物学的配列がまとめて、データベース中の有害な生物学的配列の少なくとも20%に相当するか否かを自動的に決定する方法；および、

i i i) 有害な生物学的配列の少なくとも20%が検出された場合に、自動的に警報を発生させる方法で動作するよう構成されるコンピュータ化されたシステム。

【請求項 2】

警報が発生されない場合、1つ以上の配列が合成されることを特徴とする、請求項1に記載のコンピュータ化されたシステム。

【請求項 3】

有害な生物学的配列を取り除くために、有害な生物学的配列の少なくとも20%に相当する複数の生物学的配列の少なくとも2つの生物学的配列を変更するための命令を受信することを特徴とする、請求項1に記載のコンピュータ化されたシステム。

【請求項 4】

複数の受信された設計命令が1つ以上の時点で受信されることを特徴とする、請求項1または3に記載のコンピュータ化されたシステム。

【請求項 5】

複数の受信された設計命令は、様々なソースからのものであることを特徴とする、請求項1～4のいずれか1つに記載のコンピュータ化されたシステム。

【請求項 6】

複数の受信された設計命令は、3つ以上の様々なソースからのものであることを特徴とする、請求項5に記載のコンピュータ化されたシステム。

【請求項 7】

複数の受信された設計命令は、5つ以上の様々なソースからのものであることを特徴とする、請求項5に記載のコンピュータ化されたシステム。

【請求項 8】

複数の受信された設計命令は、10以上の様々なソースからのものであることを特徴とする、請求項5に記載のコンピュータ化されたシステム。

【請求項 9】

1つ以上の生物学的配列は、各々せいぜい200の塩基の長さであることを特徴とする、請求項1～8のいずれか1つに記載のコンピュータ化されたシステム。

【請求項 10】

1つ以上の生物学的配列は、各々せいぜい100の塩基の長さであることを特徴とする、請求項9に記載のコンピュータ化されたシステム。

【請求項 11】

1つ以上の生物学的配列は、各々せいぜい50の塩基の長さであることを特徴とする、請求項9に記載のコンピュータ化されたシステム。

【請求項 12】

1つ以上の生物学的配列は、各々せいぜい20の塩基の長さであることを特徴とする、

10

20

30

40

50

請求項 9 に記載のコンピュータ化されたシステム。

【請求項 13】

増強されたポリヌクレオチド合成を提供するための方法であって：前記方法は

- a) 1 つ以上の設計命令を受信する工程であって、ここで、設計命令は複数の生物学的配列を含み、生物学的配列の各々はせいぜい 500 の塩基の長さであり、および複数の生物学的配列は核酸またはアミノ酸配列を含む、工程；
- b) 複数の生物学的配列の少なくとも 2 つの生物学的配列がまとめて、データベースにおける有害な生物学的配列の少なくとも 20 % に相当するか否かを自動的に決定する工程；および、
- c) 有害な生物学的配列の少なくとも 20 % が検出された場合に、自動的に警報を発生させる工程、を含む、方法。

10

【請求項 14】

警報が発生されない場合、1 つ以上の配列が合成されることを特徴とする、請求項 13 に記載の方法。

【請求項 15】

有害な生物学的配列を取り除くために、有害な生物学的配列の少なくとも 20 % に相当する複数の生物学的配列の少なくとも 2 つの生物学的配列を変更するための命令を受信することの特徴とする、含請求項 13 に記載の方法。

【請求項 16】

増強されたポリヌクレオチド合成を提供するためのコンピュータ化されたシステムであって：前記コンピュータ化されたシステムは、

20

- a) 配列のリストを表わすことに適しているデータベースをホストするためのサーバー；
- b) ネットワーク接続；および
- c) 汎用コンピュータに対する命令を含むコンピュータ可読媒体を含み、前記コンピュータ化されたシステムは、：

i) 1 つ以上の設計命令を受信する方法であり、ここで、設計命令は複数の生物学的配列を含み、複数の生物学的配列はベクター配列を含み、および複数の追加の挿入配列を含む、方法；

i i) ベクターおよび複数の挿入配列の少なくとも 1 つがまとめて、データベースにおける有害な生物学的配列の少なくとも 20 % に相当するか否かを自動的に決定する方法；および、

30

i i i) 有害な生物学的配列の少なくとも 20 % が検出された場合に、自動的に警報を発生させる方法で動作するよう構成されるコンピュータ化されたシステム。

【請求項 17】

警報が発生されない場合、1 つ以上の生物学的配列が合成されることを特徴とする、請求項 16 に記載のコンピュータ化されたシステム。

【請求項 18】

有害な生物学的配列を取り除くために、有害な生物学的配列の少なくとも 20 % に相当するベクターおよび複数の挿入配列少なくとも 1 つを変更するための命令を受信することの特徴とする、請求項 16 に記載のコンピュータ化されたシステム。

40

【請求項 19】

複数の受信された設計命令は、1 つ以上の時点において受信されることを特徴とする、請求項 16 ~ 18 に記載のいずれか 1 つのシステム。

【請求項 20】

複数の受信された設計命令は、様々なソースから受信されることを特徴とする、請求項 16 ~ 19 のいずれか 1 つに記載のシステム。

【請求項 21】

複数の受信された設計命令は、3 つ以上の様々なソースからのものであることを特徴とする、請求項 20 に記載のコンピュータ化されたシステム。

【請求項 22】

50

複数の受信された設計命令は、5つ以上の様々なソースからのものであることを特徴とする、請求項20に記載のコンピュータ化されたシステム。

【請求項23】

複数の受信された設計命令は、10以上の様々なソースからのものであることを特徴とする、請求項20に記載のコンピュータ化されたシステム。

【請求項24】

1つ以上の生物学的配列は、せいぜい200の塩基の長さであることを特徴とする、請求項16～23のいずれか1つに記載のシステム。

【請求項25】

1つ以上の生物学的配列は、各々せいぜい100の塩基の長さであることを特徴とする、請求項24に記載のコンピュータ化されたシステム。

【請求項26】

1つ以上の生物学的配列は、各々せいぜい50の塩基の長さであることを特徴とする、請求項24に記載のコンピュータ化されたシステム。

【請求項27】

1つ以上の生物学的配列は、各々せいぜい20の塩基の長さであることを特徴とする、請求項24に記載のコンピュータ化されたシステム。

【請求項28】

増強されたポリヌクレオチド合成を提供するための方法であって：前記方法は

- a) 1つ以上の設計命令を受信する工程であって、ここで、設計命令はベクター配列である複数の生物学的配列および複数の追加の挿入配列を含む、工程；
- b) ベクターおよび少なくとも1つの複数の挿入配列がまとめて、データベースにおける有害な生物学的配列の少なくとも20%に相当するか否かを自動的に決定する工程；および、
- c) 有害な生物学的配列の少なくとも20%が検出された場合に、自動的に警報を発生させる工程、を含む、方法。

【請求項29】

生物学的配列は、物理的な核酸またはタンパク質サンプルの配列決定から得られることを特徴とする、請求項28に記載の方法。

【請求項30】

有害な生物学的配列を取り除くために、有害な生物学的配列の少なくとも20%に相当するベクターおよび複数の挿入配列の少なくとも1つを変更するための命令を受信することを特徴とする、請求項28に記載の方法。

【請求項31】

警報が発生されない場合、1つ以上の生物学的配列が合成されることを特徴とする、請求項28～30のいずれか1つに記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

相互参照

本出願は、2016年6月10日に出願された米国仮特許出願番号62/348,786、および2016年8月16日に出願された米国仮特許出願番号62/375,858の利益を主張し、各々の利益は全体において、参照により明細書に組み込まれる。

【背景技術】

【0002】

公共の安全および/または環境に対して潜在的な脅威を生み出すことができる、個々のタンパク質および生物システムに関する私たちの集合知の成長速度はすさまじい。しかし、この知識は、種々の研究組織、機関およびジャーナルなどによって広く知れ渡っている。所定のタンパク質が害を引き起こす可能性、及びその害がどのような状況で発生するかについてアノテーションすることに重点が置かれた一元化された情報源がない。したが

10

20

30

40

50

って、新しいシステムおよび方法がその課題に取り組むために必要である。

【発明の概要】

【0003】

本明細書には、データベースをホストするためのサーバーを含む、増強されたポリヌクレオチド合成を提供するためのコンピュータ化システムが提供され、ここで、そのデータベースは、有害な生物学的配列のリストと；ネットワーク接続と；汎用コンピュータに対する命令を含むコンピュータ可読媒体と；を表わすことに適しており、ここで、上記コンピュータ化システムは：1) 1つ以上の設計命令を受信する方法であって、ここで、設計命令は複数の生物学的配列を含み、生物学的配列の各々はせいぜい500の塩基の長さであり、および複数の生物学的配列は核酸またはアミノ酸配列を含む、方法；2) 複数の生物学的配列の少なくとも2つの生物学的配列がまとめて、データベース中の有害な生物学的配列の少なくとも20%に相当するか否かを自動的に決定する方法；および3) 有害な生物学的配列の少なくとも20%が検出された場合に、自動的に警報を発生させる方法で動作するよう構成されるコンピュータ化されたシステム。さらに本明細書には、警告が発生されない場合、1つ以上の配列が合成されることを含むコンピュータ化システムが提供される。さらに本明細書には、有害な生物学的配列を取り除くために、有害な生物学的配列の少なくとも20%に相当する複数の生物学的配列の少なくとも2つの生物学的配列を変更するための命令を受信することを含む、コンピュータ化システムが提供される。さらに本明細書には、複数の受信された設計命令が1つ以上の時点で受信されるコンピュータ化システムが提供される。さらに本明細書には、複数の受信された設計命令が3つ以上の様々なソースからのものであるコンピュータ化システムが提供される。さらに本明細書には、複数の受信された設計命令が5つ以上の様々なソースからのものであるコンピュータ化システムが提供される。さらに本明細書には、複数の受信された設計命令が10以上の様々なソースからのものであるコンピュータ化システムを提供する。さらに本明細書には、1つ以上の生物学的配列が、各々せいぜい200の塩基の長さであるコンピュータ化システムが提供される。さらに本明細書には、1つ以上の生物学的配列が、各々せいぜい100の塩基の長さであるコンピュータ化システムが提供される。さらに本明細書には、1つ以上の生物学的配列が、各々せいぜい50の塩基の長さであるコンピュータ化システムが提供される。さらに本明細書には、1つ以上の生物学的配列が、各々せいぜい20の塩基の長さであるコンピュータ化システムが提供される。

10

20

30

【0004】

本明細書には、増強されたポリヌクレオチド合成を提供するための方法が記載され、その方法は：1) 1つ以上の設計命令を受信する工程であって、ここで、設計命令は複数の生物学的配列を含み、生物学的配列の各々はせいぜい500の塩基の長さであり、および複数の生物学的配列は核酸またはアミノ酸配列を含む工程；2) 複数の生物学的配列の少なくとも2つの生物学的配列がまとめて、データベース中の有害な生物学的配列の少なくとも20%に相当するか否かを自動的に決定する工程；および3) 有害な生物学的配列の少なくとも20%が検出された場合に、自動的に警報を発生させる工程を含む、方法。さらに本明細書には、警報が発生されない場合、1つ以上の配列が合成されることを含む方法が提供される。さらに本明細書には、有害な生物学的配列を取り除くために、有害な生物学的配列の少なくとも20%に相当する複数の生物学的配列の少なくとも2つの生物学的配列を変更するための命令を受信する工程を含む方法が提供される。

40

【0005】

本明細書には、データベースをホストするためのサーバーを含む、増強されたポリヌクレオチド合成を提供するためのコンピュータ化システムが提供され、ここで、そのデータベースは、配列のリストと；ネットワーク接続と；汎用コンピュータに対する命令を含むコンピュータ可読媒体と；を表すことに適しており、ここで、上記コンピュータ化システムは：1) 1つ以上の設計命令を受信する方法であって、ここで、設計命令はベクター配列である複数の生物学的配列および複数の追加の挿入配列を含む、方法；2) ベクター配列及び複数の生物学的配列の少なくとも1つがまとめて、データベース中の有害な生物学

50

的配列の少なくとも20%に相当するか否かを自動的に決定する方法；および3) 有害な生物学的配列の少なくとも20%が検出された場合に、自動的に警報を発生させる方法で動作するよう構成されるコンピュータ化されたシステム。さらに本明細書には、物理的核酸サンプルの配列決定から、生物学的配列が取得されるコンピュータ化システムが提供される。さらに本明細書には、警報が発生されない場合、1つ以上の生物学的配列が合成されることを含むコンピュータ化システムが提供される。さらに本明細書には、有害な生物学的配列を取り除くために、有害な生物学的配列の少なくとも20%に相当するベクター配列および複数の挿入配列の少なくとも1つを変更するための命令を受信することを含むコンピュータ化システムを提供する。さらに本明細書には、増強されたポリヌクレオチド合成を提供するためのコンピュータ化システムが提供され、ここで、複数の受信された設計命令は一つ以上の時点で受信される。さらに本明細書には、複数の受信された設計命令が、様々なソースから受信されるコンピュータ化システムが提供される。さらに本明細書には、複数の受信された設計命令が、3つ以上の様々なソースからのものであるコンピュータ化システムが提供される。さらに本明細書には、複数の受信された設計命令が、5つ以上の様々なソースからのものであるコンピュータ化システムが提供される。さらに本明細書には、複数の受信された設計命令が、10以上の様々なソースからのものであるコンピュータ化システムが提供される。さらに本明細書には、1つ以上の生物学的配列が、各々せいぜい200の塩基の長さであるコンピュータ化システムが提供される。さらに本明細書には、1つ以上の生物学的配列が、各々せいぜい100の塩基の長さであるコンピュータ化システムが提供される。さらに本明細書には、1つ以上の生物学的配列が、各々せいぜい50の塩基の長さであるコンピュータ化システムが提供される。さらに本明細書には、1つ以上の生物学的配列が、各々せいぜい20の塩基の長さであるコンピュータ化システムが提供される。

10

20

30

40

50

【0006】

本明細書には、増強されたポリヌクレオチド合成を提供するための方法が記載され、その方法は：1) 1つ以上の設計命令を受信する工程であって、ここで、設計命令はベクター配列である複数の生物学的配列および複数の追加の挿入配列を含む、工程；2) ベクター配列及び複数の生物学的配列の少なくとも1つがまとめて、データベース中の有害な生物学的配列の少なくとも20%に相当するか否かを自動的に決定する工程；および3) 有害な生物学的配列の少なくとも20%が検出された場合に、自動的に警報を発生させる工程、を含む。本明細書には、物理的核酸またはタンパク質のサンプルの配列決定から生物学的配列が取得される方法が提供される。さらに本明細書には、警告が発生されない場合、1つ以上の生物学的配列が合成されることを含む方法が提供される。さらに本明細書には、有害な生物学的配列を取り除くために、有害な生物学的配列の少なくとも20%に相当するベクター配列、および複数の挿入配列の少なくとも1つを変更するための命令を受信する工程を含む方法を提供する。

【0007】

引用による組み込み

個々の刊行物、特許、または特許出願が全体として、参照により組み込まれることが具体的にかつ個々に示されたのと同じ程度に、本明細書で言及される全ての刊行物、特許、および特許出願は、参照により本明細書に組み込まれる。

【図面の簡単な説明】

【0008】

本開示の技術的特徴は、添付の請求項で詳細に記載される。本開示の特徴および利点のより良い理解は、開示の原理が利用される例示的な実施形態を記載する下記の詳細な記述、および付随する下記の図面への参照によって得られる。

【0009】

【図1】タンパク質配列および関連する種、宿主、病原体、害をもたらす経路、結果およびタンパク質の種類の情報を含むユーザーインターフェースを例示する。さらに、配列登録番号、同一のタンパク質のリスト、配列記録を含むデータベースへのリンク、および類

似のタンパク質へのリンクが含まれる。

【図 2】タンパク質変異体および典型的なタンパク質“ヘマグルチニンノイラミニダーゼ・ニューカッスル病ウイルス”の部分的なリストを含むユーザーインターフェースを例示する。

【図 3 A】クエリファイル、タンパク質データベース、B L A S T レポート、制限されたリスト（有害な配列のリスト）およびスクリーニングレポートからの情報を含むフローチャートを図示する。

【図 3 B】インプット（核酸材料、核酸あるいはタンパク質配列）、意思決定（制限されたリスト、制限されていないリスト、エキスパートレビュー）およびアウトプット（警告の発生）の様々な形態を含むフローチャートを図示する。

【図 4】スクリーニングで検索するためのデータベースのリストを含むユーザーインターフェースを例示する。役割、型、名称、記載、加えられた日付およびアクティブ状態の欄が含まれる。

【図 5】配列提出スクリーンを含むユーザーインターフェースを例示する。名称、データベース、記載および F A S T F A ファイルの記入フォーム、および「Submit」ボタンが含まれる。データベースフォームは、「Seq shield」、「nr」および「Personal Database」を含むサブカテゴリがクリック時に現われるドロップダウン列を含む。

【図 6】スクリーニングステータスの概要を含むユーザーインターフェースを例示する。

【図 7】スクリーニングされた配列が、「未チェック」、「懸念がある」または「懸念がない」ものかを選択するためのプルダウンメニューを含むユーザーインターフェースを例示する。

【図 8】コンピューティングシステムを例示する。

【図 9】コンピュータシステムを例示する。

【図 10】コンピュータシステムのアーキテクチャーを例示するブロック図である。

【図 11】複数のコンピュータシステム、複数の携帯電話と携帯情報端末、およびネットワークアタッチトストレージ（NAS）を組込むように構成されたネットワークを例証する図である。

【図 12】共有される仮想アドレスメモリ空間を用いた、マルチプロセッサコンピュータシステムのブロック図である。

【発明を実施するための形態】

【0010】

合成生物学における設計機能の急成長に伴い、元々由来した引用配列に直接的には似ていない突然変異の大きい配列をしばしば使用して、多くの構築物を生成することが現在可能である。同時に、（様々な宿主および生物学的背景における）病原性の背後にあるプロセスについての理解における科学的な進歩は、文脈依存方法で、ヒト、特定の植物または動物に害を与えうるタンパク質配列、または環境に害をより広くもたらしうるタンパク質配列についての新しい知識を急速に生み出している。

【0011】

倫理的な責任のある合成生物学者は、害をもたらすことができる構築物を無意識に生み出すかもしれないが、生物系の合成設計を実証する前に、その機能を予測すること、または理解することができないかもしれない。一次配列のみから機能を予測することが実現可能でないため、以下のものを利用することで、これらの科学者の役に立つだろう：1) どのような配列が、制限ステータスと共に害をもたらすのかについてのメタデータのレポジトリ、および2) そのメタデータに対してDNAまたはタンパク質配列をチェックし、いかなる潜在的な懸念もユーザーに警告するための効果的なスクリーニングシステム。さらに、これらのニーズに対処することが可能なスクリーニングシステムは、それ自体がハイスループット設計/構造/テスト・ワークフローにシームレスに適合するように自動化に適していなければならない。本開示は、病原性に関する公的に利用可能な遺伝子レベルのメタデータの不足、および有効なスクリーニングのためのオープンソース・ツールの不足

10

20

30

40

50

の両方に対処するためのソフトウェアツールを提供する。

【0012】

定義

【0013】

様々な実施形態が本明細書中に示され記述されている一方、そのような実施形態が一例として提供されているにすぎないことは当業者にとって明白だろう。本明細書に開示される装置、システム及び方法から逸脱することなく、当業者は多くの変更、変化および置換を思い浮かべることができる。本明細書に記載される実施形態への様々な代替案が利用され得ることを理解されたい。

【0014】

特に定義されない限り、本明細書に使用される全ての技術用語は、本発明が属する分野における当業者によって一般的に理解されるのと同じ意味を有する。明細書および添付の請求項内で用いられる通り、単数形「a」、「an」および「the」は、文脈が他に明確に命令していない限り、複数の引用文を含む。本明細書の「または」へのあらゆる言及は、特に他に明記されていない限り、「および/または」を包含するように意図される。

【0015】

具体的に明示されていない、又は文脈から明白ではない限り、本明細書で使用されるように、数又はその範囲に関して用語「約」とは、明示された数及びその $+/-10\%$ の数を意味し、あるいは、範囲について列挙された値に対する下限より 10% 低いかつ上限より 10% 高い数を意味する。

【0016】

配列アノテーション

ある種の害をもたらす任意の単一の配列の能力に関する知識は、極端に知れ渡ってしまうこともある。研究者の個々のコミュニティは、有機体が宿主細胞に侵入し、宿主細胞機構を乗っ取り、宿主免疫システムから隠れる能力、そしてさらに宿主免疫応答を増強する能力を含む病原性の広く様々な態様に注目する。典型的な有害な生物学的配列は、病原性の配列をコードするもの、例えば、有害でありかつウイルス性起原のもの、細菌性起原のもの、あるいは寄生性起原のものを含む。有害な生物学的配列は、病原性の効果があると知られている野生型配列の突然変異型を含んでもよい。有害な生物学的配列は、転写または翻訳後に有害な配列産物を産生するか、有害な配列産物への前駆体として作用する配列を含む。有害な生物学的配列は、有害なタンパク質をコードする配列を含む。

【0017】

他の様相のうち、本発明は、病原性における役割のタグベースのアノテーションと共にユーザーが配列を提出することを可能にする、MediaWikiベースのユーザーインターフェースを提供する。ユーザーは、次のようにモデル化された所定の配列に関連する害の一般的なパターンを記述するために、各々の配列についていくつかのタグを提出するように促され得る：

宿主 + 背景 = 結果 + 懸念の程度

【0018】

単一の統制語彙を先験的に課さないために、本システムはタグベースのアプローチをとることができる。コミュニティアノテーションから生じるタグの収集は、より長期間にわたってそのような統制語彙の基礎を形成することができるだろう。

【0019】

各々の配列がアップロードされる時、ユーザーは4つのカテゴリーのそれぞれにタグを加えるように求められることもある。「宿主」及び「懸念の程度」をタグ付けすることは必須である；さらなる複雑さおよびドメイン知識が要求されることを考慮すると、「背景」および「結果」に対するタグを加えることは任意である。

【0020】

一例として、毒素リシンをコードする配列は、ユーザーによって以下のようにタグ付けされることもある：

10

20

30

40

50

【 0 0 2 1 】

【 表 1 】

タグ	評価
宿主	ヒト
背景	摂取、吸入
結果	熱、咳、呼吸不全、死亡
懸念の程度	最も高い

10

目標は、一般的な完全性よりも、長期にわたるメタデータの蓄積である。システムは一元的にホストされ、スクリーニングで使用されるFASTAとしてダウンロードされるキュレートされた配列の全体集合（またはタグによるクエリに基づく部分集合）を提供する。

【 0 0 2 2 】

本明細書には、アノテーション配列のための方法が提供され、ここで、データベースは、生物学的配列または生物学的な構築物（例えば、ヌクレオチド配列またはタンパク質配列）に関連する特性のリストを受信する。典型的な特性としては、限定されないが以下が挙げられる：核酸配列、タンパク質配列、タンパク質の名称、菌株の起源、配列データベースへのリンク（例えばNCBI）、配列データベース登録番号、同一の配列（タンパク質または核酸）、類似の配列（タンパク質または核酸）、疾病の種類（例えばウイルス、細菌あるいは真菌）、宿主情報（例えばヒト、哺乳動物、鳥、昆虫）、有害な相互作用の背景または経路（例えば摂取、吸入）及び懸念の程度。さらに本明細書には、各特性またはそのような特性の追加情報へのリンクを表すユーザーインターフェースが提供される。図1を参照。場合によっては、特定の菌株に対するウイルス配列が選択される。例えば、図2は、アノテーション用の赤血球凝集素ノイラミニダーゼ・ニューカッスル病ウイルスの679つの利用可能な菌株の一部を示す。

20

【 0 0 2 3 】

典型的な種は、動物種を含む。本明細書で使用されるような「動物」としては、限定されないが、哺乳動物、有袋動物、鳥、昆虫、節足動物、両生動物および爬虫類動物が挙げられる。典型的な哺乳動物としては、限定されないが、ヒツジ、ウシ、ヤギ、ブタ、ウサギ、野ウサギ、シカ、ヤギ、マウス、ラット、コウモリおよびオポッサムなどが挙げられる。典型的な疾病の種類は、以下のクラスからの病原体を含む：ウイルス、細菌、真菌および他の有害な病原体。有害な発現産物を有する典型的なウイルスとしては、限定されないが、マールブルグウイルス、エボラウイルス、ハンタウイルス、鳥インフルエンザ（例えばH5N1株）、ラッサ熱ウイルス、フニンウイルス、クリミア・コンゴ出血熱、マクポウイルス、キャサヌル森林病ウイルス、デング熱およびチクングニヤウイルスが挙げられる。有害な発現産物を有する典型的な細菌としては、限定されないが、多剤耐性黄色ブドウ球菌（MRSA）、大腸菌、リステリア、サルモネラ菌、淋菌、連鎖球菌およびブドウ球菌が挙げられる。有害な発現産物を有する典型的な真菌としては、限定されないが、アマニタ・アーキア（*Amanita archocheae*）、アマニタ・ビスポリゲラ（*Amanita bisporigera*）、アマニタ・エクシティアリス（*Amanita exitialis*）、アマニタ・マグニヴェラリス（*Amanita magnivelaris*）、アマニタ・オクレアータ（*Amanita ocreata*）、シロタマゴテングタケ、クリトサイブ・デールバータ（*Clitocybe dealbata*）、コートイナリウス・ジェンティルズ（*Cortinarius gentilis*）、レピオタ・ブルネオインカーナタ（*Lepiota brunneoincarnata*）、レピオタ・ブルネオインカーナタ（*Lepiota brunneoincarnata*）、レピオタ・ブルネオインカーナタ（*Lepiota brunneoincarnata*）

30

40

50

carnata)、レピオタ・ブルネオインカーナタ(Lepiota brunneo incarnata)が挙げられる。典型的な害をもたらす経路としては、限定されないが、摂取、吸入、皮膚接触および性感染が挙げられる。典型的な結果としては、限定されないが、熱、頭痛、吐き気、めまいおよび下痢が挙げられる。典型的なタンパク質データベースは、米国国立医療図書館および国立衛生研究所のタンパク質および遺伝子データベースを含む。典型的な疾病の懸念の程度は、低い、中程度、高いおよび最も高いを含む。

【0024】

本明細書には、有機体の名称およびまたは分類群によってクエリに関連する配列を同定するといった、基本的なキュレーションのための方法が提供される。一旦同定されれば、配列アノテーションは任意に更新され、そして任意に、特定の記述的な特徴について再分類され得る。同定された配列は、任意にFASTAフォーマットを用いて、単数またはバッチ形式でのダウンロードにさらに利用可能である。

10

【0025】

データ品質および市民参加は両方とも、公的に利用可能なデータベースに関連する懸念になり得る。有用性を即座に最大化するために、開示されたシステムは、制限されている可能性が最も高い配列または有害であると知られる他の配列を含もうとして、多くの病原性のタンパク質をデータベースに加える初期のキュレーションプロセスを実行することができる。該システムは、無害であると考えられ得る遺伝子に対応するNCBI GID同定子の「制限されていない」リストをキュレーションすることができる。さらに、その制限されていないリストも、キュレーションに利用できる。

20

【0026】

CAPTCHAの仕組みは、ボットで駆動されるキュレーションを防ぐために、そしてページの作成前または編集前に、ユーザー登録を要求するために使用され得る。GID同定子は、(存在有無について)定期的に確認され、そして失敗した場合、人間によるレビューのためにレコードがタグ付けされ得る。さらにユーザーは、コミュニティーあるいは管理者によるレビューを要求するために、レコードにフラグを立てることができる。

【0027】

本開示は、少なくとも1つの生物学的配列をアノテートするおよび/またはスクリーニングするシステム及び方法を提供する。場合によっては、生物学的配列は核酸配列である。核酸配列は、1; 10; 100; 200; 300; 400; 500; 600; 700; 800; 900; 1,000; 2,000; 5,000; 7,000; 10,000、またはそれ以上の核酸残基を含んでもよい。場合によっては、核酸配列は100~500の核酸残基を含む。場合によっては、核酸配列は50~1000の核酸残基を含む。場合によっては、核酸配列は20~200の核酸残基を含む。場合によっては、核酸配列は200の残基を含む。場合によっては、生物学的配列はDNAまたはRNAであってもよい。場合によっては、生物学的配列はタンパク質配列である。生物学的配列はアデニン(A)、シトシン(C)、グアニン(G)、チミン(T)あるいはウラシル(U)を含んでもよい。場合によっては、生物学的配列はタンパク質配列である。タンパク質は、1; 10; 100; 200; 300; 400; 500; 600; 700; 800; 900; 1,000; 2,000、またはそれ以上のアミノ酸を含んでもよい。場合によっては、タンパク質配列は100~300のアミノ酸を含む。場合によっては、核酸配列は50~500のアミノ酸を含む。場合によっては、核酸配列は10~200のアミノ酸を含む。場合によっては、核酸配列は60のアミノ酸を含む。場合によっては、せいぜい2、5、10、20、50あるいは200残基の核酸断片が、インシリコで核酸配列にアセンブルされる。場合によっては、核酸断片は、1つ以上のソース、あるいは同じソースからの1つ以上のオーダーから得られる。

30

40

【0028】

スクリーニングツール

所定の配列がバイオセキユリティーリスクをもたらすか否かを決定することができるスクリーニングシステムの構築は、時間へのある程度の投資およびすべての合成生物学者、

50

あるいは合成生物学に関連するすべての企業でも利用でない専門知識を含み得る。危険な配列のデータベースを利用できると仮定したとしても、アライナーの基礎的なパラメータ化および結果処理（より短い領域への相同性を隠さないように、類似領域へのアライメント数（`alignment counts`）を選び取ることを含む）は、ドメイン専門知識を含み得る。

【0029】

例示的なワークフローが、図3Aで提供される。図3Aを参照すると、プロセッサは、生物学的配列情報を含むクエリファイルを受信し、さらに同定された配列情報を有するタンパク質データベースと通信する。照会された生物学的配列と関連する、同定された同一および類似の配列の一部あるいは全部を列挙して、BLASTレポートが作成される。それから、BLASTレポートは、「制限された」リストとも呼ばれる有害な生物学的配列（タンパク質または核酸）に関連する配列を同定する配列アノテーションを含むデータベースに照会される。スクリーニングレポートは、これらのプロセスの結果を要約するユーザーインターフェースの形で作成される。

10

【0030】

例示的なロジックワークフローが、図3Bで提供される。図3Bを参照すると、（配列決定され得る）物理的核酸あるいはタンパク質材料などのデータ入力ソース、（タンパク質配列に翻訳され得る）核酸配列、あるいはタンパク質配列などのデータ入力ソースが、制限されたリストにあるか否かを決定するために、1つ以上のデータベースを検索するアルゴリズムを用いて評価され得る。典型的なアルゴリズムとしては、制限されないが、BLAST、DIAMOND、スミス・ウォーターマン（`Smith-Waterman`）、あるいは配列情報を比較するための他のアルゴリズムが挙げられる。制限されたリストにあることが分かった配列は、既知の偽陽性を含む制限されていないリストに対してさらに評価される。偽陽性と同定されない場合、配列はエキスパートレビューの対象となる。配列が無害であるとわかった場合、偽陽性としてさらに同定されることを防ぐために、該配列は制限されないリストに配置される。配列が有害であるとわかった場合、アウトプット警報が発生される。場合によっては、無害な配列が合成される。場合によっては、配列は有害な配列を取り除くために修飾される。場合によっては、修飾された配列は再度スクリーニングされる。場合によっては、修飾された無害な配列が発見されるまで、このプロセスは反復して繰り返される。場合によっては、修飾された無害な配列が合成される。

20

30

【0031】

図4を参照すると、ユーザーインターフェースは、スクリーニングプロセスの選択に利用可能な制限されたリストを表示する。図5を参照すると、例示的なユーザーインターフェースが、提出フォーム「`Submit a screen`」を示す。該フォームは、オープン・データベース、例えば公的に利用可能な情報の集合、に対するスクリーニング、または公的に利用可能でない選択基準に基づき得る個人データベースに対するスクリーニングの選択を可能にする。さらに、提出フォームは、アップロードするための生物学的配列ファイルの選択を可能にする。

【0032】

図6を参照すると、例示的なユーザーインターフェースは、ステータス情報、スクリーニングされた配列、レビューステータス、懸念の有無についてのステータス、配列追加の日付、およびBLAST結果を見るためのリンクを伴う、実施されたバイオセキュリティスクリーニングの概要を示す。図7を参照すると、例示的なユーザーインターフェースは、スクリーン中にアクセスされたリストの概要、スクリーニングされた配列、および配列に対する有害な（制限された）配列の割り当てを表示する。

40

【0033】

本明細書に開示される技術は、スクリーニングシステムのPythonベースのレファレンス実装を含み得る。クエリヌクレオチド配列を考慮すると、該システムは、前段落で説明されたインターフェースによって生成される、アノテーションされたコレクションに由来するタンパク質配列のセットと（例えばBLAST経由の）配列を、比較することが

50

できる。

【0034】

結果は、相同性、E - スコアおよびアライメントの長さの程度によってフィルタリングされ得る。パッシングヒット (Passing hits) は、それらの配列に関連するタグの分布、そして問題が発見されたクエリの領域によって集約され得る。ユーザーがより詳細なフォローアップができるように、元のデータベースエントリへのリンクが提供され得る。あらかじめ定義されたガイダンスに従って、いくつかの例は、アルゴリズムが100%の感度であることを示し、レポートがアーカイブの使用のためにダウンロードされる。短い (例えば、約200塩基未満の) 配列のスクリーニングは、多くの偽陽性の発見をもたらす得る。より短いポリヌクレオチド配列の効果的なスクリーニングは、アルゴリズム的アプローチを含み得る。

10

【0035】

スクリーニングシステムはデータベース上に位置し、グラフィカルユーザインターフェースと同様に、スクリーンリクエスト提出 (screen request submission) および結果検索のためのRESTfulアプリケーションプログラミングインターフェース (API) を含み得る。該アプリケーションは、ラップトップコンピュータにインストールされ、動作することができ、さらにAPIコールを介してハイスループットの使用にまで合理的に拡張することができる。

【0036】

累積的生物学的配列または構築物スクリーニング

20

特に、複数のソースを通じ、かつ複数の時点で生物学的配列あるいは構築物が得られる場合、個々にスクリーニングされた時に有害な配列の同定をもたらさない生物学的配列の断片および構築物を得ることが可能である。場合によっては、ソースは顧客であってもよい。例えば、指定病原体 (select agent) が規制する細菌あるいはウイルスのいずれかのゲノムの実質的な部分の蓄積は、より小さな断片で得られ、次いで、有害な生物学的配列または構築物がアセンブルされ得る。これに対処するために、場合によっては、各々のリクエストが受信された後のバックグラウンドプロセスは、生物学的配列または構築物の要求元からの以前のオーダーすべてについてデータベースへ照会し、任意の有害な生物学的配列または構築物への高い相同性を持った任意のセグメントの記録を収集する。これにより、たとえそれらのセグメントが個々のオーダーの間に正式な警報あるいは所有の拒否を引き起こすのに不十分だったとしても、評価および警報が保証される。場合によっては、これらの高い相同性のセグメントは、懸念のある指定病原体 (select agent) のゲノムに区間として表わされ、次いで、生物学的配列または構築物の要求元ごとおよびゲノムごとのすべての区間の結合が、生物学的配列または構築物の要求元ごとにこれらの有機体の最大の理論構成を決定するために生成される。場合によっては、一旦、生物学的配列または構築物の要求元が、所定の指定病原体 (select agent) ゲノムの20%以上を設計しようとする、上記生物学的配列または構築物の要求元を用いた人間によるレビューおよびフォローアップのために警報が意図的に発生される。場合によっては、一旦、生物学的配列または構築物の要求元が、少なくとも5%、10%、20%、30%、40%、50%、60%、70%、80%、90%、あるいはそれ以上の有害な生物学的配列または構築物を生成することができると、配列構成を認可する前に、人間によるレビューのための警報が発生される。場合によっては、一旦、生物学的配列または構築物の要求元が、5%~50%、10%~75%、20%~90%、30%~100%、10%~30%、5%~50%、あるいは15%~60%の有害な生物学的配列または構築物を生成することができると、配列構成を認可する前に、人間によるレビューのための警報が発生される。

30

40

【0037】

本明細書に記載された核の設計および/または構築する方法およびシステムについてスクリーニングされた生物学的配列は、1つ以上の核酸あるいはタンパク質配列を含み得る。せいぜい200塩基しか含まないような、より短い核酸配列については、既存のスクリ

50

ーニング方法は非常に高い偽陽性率を有する。場合によっては、より短い核酸配列は、せいぜい2000、1000、500、200、100、75、50、40、30の塩基を含むか、あるいはせいぜい20の塩基を含む。場合によっては、より短い核酸配列は、10~1000、20~500、30~300、40~200、50~200、20~200、10~100、あるいは100~300の塩基を含む。場合によっては、核酸配列は、せいぜい300、200、100、75、50、40、30、20、10、5、あるいは5以下のアミノ酸を含むより短いタンパク質をコードする。場合によっては、より短い核酸配列は10~300、20~200、30~100、10~200、20~100、5~50、10~100、あるいは25~75のアミノ酸を含む。一実施例では、生物学的配列または構築物の要求元が、制限されたあるいは有害な生物学的配列を潜在的にアセンブルするのに十分なポリヌクレオチドへの要求をいつ提出したか決定するために、ポリヌクレオチドのセットを調べる代替スクリーニングアプローチが用いられる。場合によっては、オーダーの間に、1つ以上のソース内のバックグラウンドプロセスは、アセンブリアルゴリズムを用いて、指定有害有機体のゲノムに対するオーダーでポリヌクレオチドをアセンブルする。場合によっては、アセンブリアルゴリズムは次世代シーケンシング・アセンブリアルゴリズムを含む。これらのアセンブリは、1つ以上のオーダーを1つ以上のソースに結び付ける仮説生成を可能にする。例えば、ソースAおよびBからのオーダーX、YおよびZは、有害な有機体から1つ以上の遺伝子をアセンブルするために組み合わせられる。場合によっては、ソースの数は少なくとも2、3、4、5、8、10、15、20、30、あるいはそれ以上である。場合によっては、ソースの数は、2~30、5~50、10~100、5~20、2~10、4~40、あるいは15~75である。場合によっては、仮説生成は、人間によるレビューのために警告し、また生物学的配列または構築物の要求元との後続議論を任意に引き起こし、あるいは法執行機関(law enforcement)へ直接報告する。遺伝子長の配列への高い相同性の可能性が低いことを考慮すると、偽陽性率は低いままでなければならない。場合によっては、さらなる偽陽性の削減は、適切な重複が1つ以上の有害な生物学的配列または構築物のアセンブリを許可するか否かを決定するために、仮定された配列の集合のアライメント構造を評価する形でもたらされる。

10

20

30

40

50

【0038】

場合によっては、ベクターあるいはインサートのような物理的核酸サンプルは、合成される1つ以上の核酸配列を伴うアセンブリのためのソースによって提供される。場合によっては、これらの物理的核酸材料は、NGSを用いるなどして初めに配列決定され、そして1つ以上のベクターおよびインサートの配列の仮説アセンブリが、スクリーニングの対象となる。場合によっては、少なくとも2つの配列の組み合わせがスクリーニングされる。場合によっては、少なくとも2、3、4、5、10、15、20、30、あるいは30より多い配列の組み合わせが、有害な生物学的配列または構築物についてスクリーニングされる。場合によっては、スクリーニングされた配列の数が、2~30、5~50、10~100、5~20、2~10、4~40、あるいは15~75の配列であり、有害な生物学的配列または構築物についてスクリーニングされる。

【0039】

デジタル処理装置

ある実施形態では、本明細書に記載されたプラットフォーム、システム、媒体、および方法は、デジタル処理装置、あるいはその使用を含み得る。ある実施形態では、デジタル処理装置は、装置の機能を実行する1つ以上のハードウェア中央処理装置(CPU)あるいは汎用グラフィック処理装置(GPGPU)を含み得る。ある実施形態では、デジタル処理装置はさらに、実行可能な命令を実施するように構成されたオペレーティングシステムを含み得る。デジタル処理装置は、任意にコンピュータネットワークに接続されてもよい。デジタル処理装置は、ワールドワイドウェブにアクセスするように、インターネットに任意に接続されてもよい。デジタル処理装置は、クラウドコンピューティング・インフラストラクチャに任意に接続されてもよい。デジタル処理装置は、イントラネットに任意

に接続されてもよい。デジタル処理装置は、データ記憶装置に任意に接続されてもよい。

【0040】

本明細書の記載に従って、適切なデジタル処理装置は、非限定的な例として、サーバーコンピュータ、デスクトップコンピュータ、ラップトップコンピュータ、ノート型コンピュータ、サブノート型コンピュータ、ネットブックコンピュータ、ネットパッドコンピュータ、セットトップコンピュータ、メディアストリーミングデバイス、ハンドヘルドコンピュータ、インターネットアプライアンス、モバイルスマートフォン、タブレットコンピュータ、携帯情報端末、ビデオゲーム機および媒体を含み得る。多くのスマートフォンが、本明細書に記載されたシステムの使用に適し得る。任意のコンピュータネットワーク接続を伴う、テレビ、ビデオプレーヤーおよびデジタル音楽プレーヤーは、本明細書に記載されたシステムの使用に適し得る。適切なタブレットコンピュータは、当業者に既知のブックレット、スレートおよび変換可能な構成を有するものを含んでもよい。

10

【0041】

デジタル処理装置は、実行可能命令を実施するように構成されたオペレーティングシステムを含んでもよい。オペレーティングシステムは、例えば、装置のハードウェアを制御し、アプリケーションの遂行のためのサービスを提供する、プログラムおよびデータを含むソフトウェアであってもよい。適切なサーバーオペレーティングシステムは、非限定的な例として、FreeBSD、OpenBSD、NetBSD（登録商標）、リナックス（登録商標）、Apple（登録商標）、Mac OS X Server（登録商標）、Oracle（登録商標）、Solaris（登録商標）、Windows Server（登録商標）、およびNovell（登録商標）、NetWare（登録商標）を含み得る。適切なパソコンオペレーティングシステムは、非限定的な例として、Microsoft（登録商標）、Windows（登録商標）、Apple（登録商標）、Mac OS X（登録商標）、UNIX（登録商標）、およびGNU/Linux（登録商標）などのUNIX（登録商標）のようなオペレーティングシステムを含み得る。いくつかの実施例において、オペレーティングシステムはクラウドコンピューティングによって提供され得る。装置は、記憶装置および/またはメモリ装置を含んでもよい。記憶装置および/またはメモリ装置は、一時的または永久的にデータまたはプログラムを保存するために使用される1つ以上の物理的な機器であってもよい。装置は、揮発性メモリであってもよいし、記憶された情報を維持するパワーを必要することもある。装置は、不揮発性メモリであってもよいし、パワーがデジタル処理装置に供給されない時、記憶された情報を保持する。不揮発性メモリは、フラッシュメモリ、ダイナミック・ランダム・アクセス・メモリ（DRAM）、強誘電体メモリ（FRAM（登録商標））、相変化メモリ（PRAM）を含んでもよい。

20

30

【0042】

デジタル処理装置は、ユーザーへ視覚情報を送るディスプレイを含み得る。ディスプレイは、陰極線管（CRT）、液晶ディスプレイ（LCD）、TFT液晶（TFT-LCD）、有機発光ダイオード（OLED）ディスプレイ、パッシブマトリクスOLED（PMOLED）あるいはアクティブマトリクスOLED（AMOLED）ディスプレイ、プラズマ・ディスプレイ、および/または、ビデオプロジェクターであってもよい。

40

【0043】

デジタル処理装置は、ユーザーからの情報を受信するための入力装置を含み得る。入力装置は、キーボードであってもよい。入力装置は、非限定的な例として、マウス、トラックボール、トラックパッド、ジョイスティック、ゲームコントローラーあるいはスタイラスを含むポインティングデバイスであってもよい。入力装置は、タッチスクリーンまたはマルチタッチスクリーンであってもよい。入力装置は、音声入力あるいは他の音入力をとらえるためのマイクロフォンであってもよい。入力装置は、動作入力あるいは視覚入力をとらえるためのビデオカメラあるいは他のセンサーであってもよい。入力装置は、Kinect（キネクト）、Leap Motion（リープモーション）などであってもよい。入力装置は、本明細書に開示されるもののような装置の組み合わせであってもよい。

50

【 0 0 4 4 】

図 8 を参照すると、特定の実施形態では、典型的なデジタル処理装置（ 8 0 1 ）はプログラムされるか、そうでなければアノテーションまたはスクリーニングを実行するように構成される。本実施例では、デジタル処理装置（ 8 0 1 ）は、シングルコアまたはマルチコア・プロセッサであり得る中央処理装置（ C P U 、さらに本明細書における「プロセッサ」および「コンピュータプロセッサ」）（ 8 0 5 ）、あるいは並行処理のための複数のプロセッサを含む。デジタル処理装置（ 8 0 1 ）は、メモリまたは記憶場所（ 8 1 0 ）（例えばランダムアクセスメモリ、読み出し専用メモリ、フラッシュメモリ）、電子記憶装置（ 8 1 5 ）（例えばハードディスク）、1 つ以上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（ 8 2 0 ）（例えばネットワークアダプタ）、およびキャッシュ、他のメモリ、データ記憶装置、および / または電子ディスプレイアダプターなどの周辺機器（ 8 2 5 ）も含有する。メモリ（ 8 1 0 ）、記憶装置（ 8 1 5 ）、インターフェース（ 8 2 0 ）、および周辺機器（ 8 2 5 ）は、マザーボードなどの通信バス（実線）を通じて、C P U （ 8 0 5 ）と通信する。記憶装置（ 8 1 5 ）は、データを記憶するためのデータ記憶装置（またはデータリポジトリ）であってもよい。デジタル処理装置（ 8 0 1 ）は、通信インターフェース（ 8 2 0 ）の助けを借りてコンピュータネットワーク（「ネットワーク」）（ 8 3 0 ）に動作可能に結合される。ネットワーク（ 8 3 0 ）は、インターネット、インターネットおよび / またはエクストラネット、またはインターネットと通信しているイントラネットおよび / またはエクストラネット、であり得る。ネットワーク（ 8 3 0 ）は、場合によっては、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（ 8 3 0 ）は、クラウドコンピューティングのような分散コンピューティングを可能にすることができる 1 つ以上のコンピュータサーバを含むことができる。ネットワーク（ 8 3 0 ）は、場合によっては、装置（ 8 0 1 ）の助けを借りて、装置（ 8 0 1 ）に連結された装置が、クライアントまたはサーバーとして挙動することを可能にするピアツーピア・ネットワークを実装することができる。

10

20

【 0 0 4 5 】

引き続き図 8 を参照すると、C P U （ 8 0 5 ）は、プログラムがソフトウェアで具体化することができる機械可読な命令の配列を実行することができる。命令は、メモリ（ 8 1 0 ）などの記憶場所に保存されてもよい。命令は C P U （ 8 0 5 ）に向けられ、これは本開示の方法を実行するための C P U （ 8 0 5 ）を引き続きプログラムするか、そうでなければ構成することができる。C P U （ 8 0 5 ）によって実行された動作の例は、フェッチ、デコード、実行、ライトバックを含む。C P U （ 8 0 5 ）は、集積回路などの回路の一部であってもよい。装置（ 8 0 1 ）の 1 つ以上の他のコンポーネントが、回路に含まれる。場合によっては、回路は、特定用途向け集積回路（ A S I C ）あるいはフィールドプログラマブルゲートアレイ（ F P G A ）である。

30

【 0 0 4 6 】

引き続き図 8 を参照すると、記憶装置（ 8 1 5 ）は、ドライバ、ライブラリーおよび保存されたプログラムのようなファイルを保存することができる。記憶装置（ 8 1 5 ）は、ユーザーデータ、例えばユーザーの好み、およびユーザープログラムを保存できる。場合によっては、デジタル処理装置（ 8 0 1 ）は、イントラネットまたはインターネットを通じて通信するリモートサーバ上に位置するような、外部の 1 つ以上の追加のデータ記憶装置を含むことができる。

40

【 0 0 4 7 】

引き続き図 8 を参照すると、デジタル処理装置（ 8 0 1 ）はネットワーク（ 8 3 0 ）を通じて、1 つ以上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。例えば、装置（ 8 0 1 ）は、ユーザー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。リモートコンピュータシステムの例は、パーソナルコンピュータ（例えば、ポータブル P C ）、スレート P C またはタブレット P C （例えば、A p p l e（登録商標）i P a d（登録商標）、S a m s u n g（登録商標）G a l a x y T a b）、電話、スマートフォン（例えば、A p p l e（登録商標）i P h o n e（登録商標）、アンドロイド対応の装置

50

、Blackberry（登録商標）、または携帯情報端末を含有する。

【0048】

本明細書に記載される方法は、例えばメモリ（810）あるいは電子記憶装置ユニット（815）などの、デジタル処理装置（801）の電子記憶場所に保存された機械（例えばコンピュータプロセッサ）実行可能なコードによって実行され得る。機械実行可能なまたは機械可読なコードは、ソフトウェアの形で提供されうる。使用中に、コードはプロセッサ（805）によって実行され得る。場合によっては、コードは、記憶装置（815）から検索され、プロセッサ（805）によって、容易なアクセスのためにメモリ（810）に保存されうる。ある状況においては、電子記憶装置（815）は排除され、記憶実行可能命令がメモリ（810）に保存され得る。

10

【0049】

追加のコンピュータシステム

【0050】

本明細書に記載されるシステムのいずれも、コンピュータに動作可能に接続されてもよいし、コンピュータを介してローカルまたはリモートで自動化されても良い。様々な場合において、本開示の方法およびシステムは、さらにコンピュータシステム上のソフトウェアプログラムおよびその使用を含むこともある。従って、堆積材料装置（the material deposition device）の動作、分配動作および真空操作を調整し同期するというような、分配／真空／補充機能の同期のためのコンピュータ制御は、本開示の範囲内である。コンピュータシステムは、基板の指定領域に適切な試剤を提供するために、ユーザー指定の塩基配列と材料堆積装置（a material deposition device）の位置をインターフェースで接続するようプログラムされ得る。

20

【0051】

図9に例示されるコンピュータシステム（900）は、固定媒体（912）を有するサーバー（909）に任意に接続され得るネットワークポート（905）および／または媒体（911）からの命令を読み取ることができる論理装置として理解され得る。図9に示されるようなシステムは、CPU（901）、ディスクドライブ（903）、キーボード（915）および／またはマウス（916）などの任意の入力装置、ならびに任意のモニター（907）を含んでも良い。示された通信媒体を介して、ローカル位置またはリモート位置でのサーバーへのデータ通信が可能になる。通信媒体は、データの送信および／または受信のための任意の手段を含むことができる。例えば、通信媒体は、ネットワーク接続、無線接続またはインターネット接続であってもよい。そのような接続は、ワールドワイドウェブ上での通信を提供することができる。本開示に関連するデータは、図9に例示される当事者（922）による受信および／またはチェックのために、そのようなネットワークあるいは接続を通じて送信され得ると想定される。

30

【0052】

図10は、本開示の実施例に関連して使用され得るコンピュータシステム（1000）の第1の典型的な例となるアーキテクチャーを例示するブロック図である。図10に図示されるように、例示的なコンピュータシステムは、命令を処理するためのプロセッサ（1002）を含むことができる。プロセッサの限定しない例としては、以下が挙げられる：Intel Xeon（商標）プロセッサ、AMD Opteron（商標）プロセッサ、Samsung 32-bit RISC ARM 1176JZ（F）-Sv1.0（商標）プロセッサ、ARM Cortex-A8 Samsung S5PC100（商標）プロセッサ、ARM Cortex-A8 Apple A4（商標）プロセッサ、Marvell PXA 930（商標）プロセッサ、または機能的に同等のプロセッサ。実行の複数のスレッドは、並列処理に使用され得る。場合によっては、単一のコンピュータシステム内、クラスタ内、あるいは複数のコンピュータ、携帯電話、および／または携帯情報端末装置を含むネットワーク上の分散型システム内であろうと、複数のプロセッサあるいは複数のコアを有するプロセッサが使用され得る。

40

50

【 0 0 5 3 】

図 1 0 に例示されるように、プロセッサ (1 0 0 2) によって最近使用された、または頻繁に使用される命令あるいはデータのための高速メモリを提供するために、高速キャッシュ (1 0 0 4) がプロセッサ (1 0 0 2) に接続されるか、組み込まれ得る。プロセッサ (1 0 0 2) は、プロセッサバス (1 0 0 8) によってノースブリッジ (1 0 0 6) に接続される。ノースブリッジ (1 0 0 6) はメモリーバス (1 0 1 2) によってランダムアクセスメモリ (R A M) (1 0 1 0) に接続され、プロセッサ (1 0 0 2) によって R A M (1 0 1 0) へのアクセスを管理する。また、ノースブリッジ (1 0 0 6) は、チップセットバス (1 0 1 6) によってサウスブリッジ (1 0 1 4) に接続される。次に、サウスブリッジ (1 0 1 4) は、周辺バス (1 0 1 8) に接続される。周辺バスは、例えば、 P C I 、 P C I - X 、 P C I E x p r e s s 、または他の周辺バスであってもよい。ノースブリッジおよびサウスブリッジは、しばしばプロセッサチップセットと呼ばれ、プロセッサ、 R A M 、および周辺バス (1 0 1 8) 上の周辺コンポーネントの間のデータ転送を管理する。幾つかの代替的なアーキテクチャーでは、ノースブリッジの機能性は、別のノースブリッジチップを使用する代わりに、プロセッサに組み込まれ得る。場合によっては、システム (1 0 0 0) は、周辺バス (1 0 1 8) に取り付けられたアクセラレータカード (1 0 2 2) を含むことができる。アクセラレータは、フィールドプログラマブルゲートアレイ (F P G A) または特定の処理を加速するための他のハードウェアを含むことができる。例えば、アクセラレータは、適応データの再構成のために、または拡張された設定処理に使用される代数式を評価するために使用されてもよい。

10

20

【 0 0 5 4 】

ソフトウェアおよびデータは、外部記憶装置 (1 0 2 4) に保存され、プロセッサによる使用のために、 R A M (1 0 1 0) および / またはキャッシュ (1 0 0 4) にロードされ得る。システム (1 0 0 0) は、システムリソースを管理するためのオペレーティングシステムを含み、オペレーティングシステムの非限定な例としては、以下が挙げられる： L i n u x (登録商標) 、 W i n d o w s (商標) 、 M A C O S (商標) 、 B l a c k B e r r y O S (商標) 、 i O S (商標) 、および他の機能的に同等のオペレーティングシステム、ならびに本開示の典型的な例に従ってデータの記憶および最適化を管理するためのオペレーティングシステム上で実行するアプリケーションソフトウェア。本実施例では、システム (1 0 0 0) はさらに、ネットワークアタッチトストレージ (N A S) および分散並行処理に使用される他のコンピュータシステムなどの、外部記憶装置へネットワークインターフェースを提供するための周辺バスに接続されたネットワークインターフェースカード (N I C) (1 0 2 0) および (1 0 2 1) を含む。

30

【 0 0 5 5 】

図 1 1 は、複数のコンピュータシステム (1 1 0 2 a) および (1 1 0 2 b) 、複数の携帯電話および携帯情報端末 (1 1 0 2 c) 、ならびにネットワークアタッチトストレージ (N A S) (1 1 0 4 a) および (1 1 0 4 b) を備えるネットワーク (1 1 0 0) を示す図である。典型的な例では、システム (1 1 0 2 a) 、 (1 1 0 2 b) および (1 1 0 2 c) は、データ記憶を管理し、ネットワークアタッチトストレージ (N A S) (1 1 0 4 a) および (1 1 0 4 b) に保存されたデータのためにデータアクセスを最適化することができる。数学的モデルがデータに使用され、コンピュータシステム (1 1 0 2 a) および (1 1 0 2 b) 、および携帯電話と携帯情報端末のシステム (1 1 0 2 c) にわたる分散並行処理を用いて評価され得る。コンピュータシステム (1 1 0 2 a) と (1 1 0 2 b) および携帯電話と携帯情報端末システム (1 1 0 2 c) はさらに、ネットワークアタッチトストレージ (N A S) (1 1 0 4 a) および (1 1 0 4 b) に保存されたデータの適応データ再構成 (a d a p t i v e d a t a r e s t r u c t u r i n g) のために平行処理を提供することができる。図 1 1 は、一例を例示したものにすぎず、多種多様な他のコンピューターアーキテクチャーおよびシステムが、本開示の様々な例と共に使用され得る。例えば、並列処理を提供するためにブレードサーバーが使用されても良い。並列処理を提供するために、プロセッサブレードがバックプレーンで接続されても良い

40

50

。また記憶装置は、バックプレーンに接続されても良いし、あるいは別のネットワークインターフェースを介してネットワークアタッチトストレージ（NAS）として接続されても良い。いくつかの典型的例では、プロセッサは、別のメモリ空間を維持することができ、ネットワークインターフェース、バックプレーン、また他のプロセッサによる並列処理のための他のコネクタを介してデータを送信することができる。他の例では、プロセッサの一部または全てが、共有の仮想アドレスメモリ空間に使用することができる。

【0056】

図12は、典型的な例に従って共有の仮想アドレスメモリ空間を使用するマルチプロセッサコンピュータシステム（1200）のブロック図である。システムは、共有メモリサブシステム（1204）にアクセスすることができる複数のプロセッサ（1202a） - （1202f）を含む。システムは、複数のプログラマブルハードウェアメモリアルゴリズムプロセッサ（MAP）（1206a） - （1206f）を共有メモリサブシステム（1204）に組み込む。MAP（1206a） - （1206f）はそれぞれ、メモリ（1208a） - （1208f）および1つ以上のフィールドプログラマブルゲートアレイ（FPGA）（1210a） - （1210f）を含むことができる。MAPは、設定可能な機能ユニットを提供し、特定のアルゴリズムまたはアルゴリズムの一部が、それぞれのプロセッサと緊密に連携して処理を行うためにFPGA（1210a） - （1210f）に提供され得る。例えば、MAPは、データモデルに関する代数式を評価するために、および例における適応データ再編成を実行するために使用され得る。この例では、各々のMAPは、これらの目的のための全てのプロセッサによって地球規模で利用可能である。1つの構成では、各々のMAPは、関連するメモリ（1208a） - （1208f）にアクセスするために直接メモリアクセス（DMA）を用いることができ、それぞれのマイクロプロセッサ（1202a） - （1202f）とは無関係および非同期にタスクを実行することを可能にする。この構成では、MAPは、アルゴリズムのパイプライン処理および並列実行のために、他のMAPに結果を直接供給することができる。

【0057】

上記のコンピュータアーキテクチャーおよびシステムは一例にすぎず、汎用プロセッサ、コプロセッサ、FPGAおよび他のプログラマブルロジックデバイス、システムオンチップ（SOC）、特定用途向け集積回路（ASIC）、ならびに他の処理素子および論理素子の任意の組み合わせを用いたシステムを含む、多種多様な他のコンピュータ、携帯電話、携帯情報端末アーキテクチャーおよびシステムが、例に関連して使用され得る。場合によっては、コンピュータシステムのすべてあるいは一部が、ソフトウェアまたはハードウェアにおいて実装されることもある。ランダムアクセスメモリ、ハードドライブ、フラッシュメモリ、テープドライブ、ディスクアレイ、ネットワークアタッチトストレージ（NAS）および他のローカルまたは分散型データストレージデバイスおよびシステムを含む、あらゆる種類のデータ記憶媒体が例に関連して使用されても良い。

【0058】

典型的な例では、コンピュータシステムは、上記あるいは他のコンピュータアーキテクチャーおよびシステムのいずれかで実行するソフトウェアモジュールを用いて実装されても良い。他の場合では、システムの機能は、ファームウェア、図12で参照されるフィールドプログラマブルゲートアレイ（FPGA）などのプログラマブルロジックデバイス、システムオンチップ（SOC）、特定用途向け集積回路（ASIC）、あるいは他の処理素子および論理素子において部分的にまたは完全に実装され得る。例えば、図10に例示されるアクセラレータカード（1022）などのハードウェアアクセラレータカードを用いることで、セットプロセッサおよびオブティマイザがハードウェアアクセラレーションで実装されることもある。

【0059】

非一時的コンピュータ可読記憶媒体

本明細書に開示されるプラットフォーム、システム、媒体および方法は、任意にネットワーク化されたデジタル処理デバイスのオペレーティングシステムによって実行可能な命

10

20

30

40

50

令を含むプログラムでコードされた1つ以上の非一時的なコンピュータ可読記憶媒体を含み得る。コンピュータ可読記憶媒体は、デジタル処理デバイスの有形要素であっても良い。コンピュータ可読記憶媒体は、デジタル処理デバイスから任意に取り外し可能である。コンピュータ可読記憶媒体は、非限定的な例として、CD-ROM、DVD、フラッシュメモリ装置、ソリッドステートメモリ、磁気ディスクドライブ、磁気テープドライブ、光ディスクドライブ、クラウドコンピューティングシステムおよびサービスなどを含む。場合によっては、プログラムおよび命令は、媒体上で永久に、実質的に永久に、半永久に、または非一時的にコードされる。

【0060】

コンピュータプログラム

いくつかの実施形態では、本明細書に開示されるプラットフォーム、システム、媒体および方法は、少なくとも1つのコンピュータプログラム、あるいはその使用を含むこともある。コンピュータプログラムは、特定のタスクを実行するために書き込まれた、デジタル処理デバイスのCPUにおいて実行可能な一連の命令を含む。コンピュータ可読命令は、特定のタスクを実行するか特定の抽象データ型を実施する、機能、オブジェクト、アプリケーションプログラムインターフェース(API)、データ構造などのプログラムモジュールとして実行されてもよい。本明細書で提供される開示に照らして、コンピュータプログラムは、様々な言語の様々なバージョンで書き込まれても良い。

【0061】

ウェブアプリケーション

コンピュータプログラムは、ウェブアプリケーションを含んでも良い。本明細書で提供される開示に照らして、ウェブアプリケーションは1つ以上のソフトウェアフレームワークおよび1つ以上のデータベースシステムを利用しても良い。ウェブアプリケーションは、Microsoft(登録商標)、.NET、またはRuby on Rails(ROR)などのソフトウェアフレームワーク上で作成され得る。ウェブアプリケーションは、非限定的な例として、リレーショナルデータベース、非リレーショナルデータベース、オブジェクト指向データベース、連想型データベース、およびXMLデータベースのシステムを含む、1つ以上のデータベースシステムを利用することもある。さらなる実施形態では、適切なリレーショナルデータベースシステムは、非限定的な例として、Microsoft(登録商標)SQL Server、mySQL(商標)、およびOracle(登録商標)を含む。当業者は、ウェブアプリケーションが様々な実施形態において1つ以上の言語の1つ以上のバージョンで書かれていることを認識するであろう。ウェブアプリケーションは、1つ以上のマークアップ言語、プレゼンテーション定義言語、クライアント側スクリプト言語、サーバー側コーディング言語、データベースクエリ言語、またはこれらの組み合わせで書かれてもよい。いくつかの実施形態では、ウェブアプリケーションは、ハイパーテキストマークアップ言語(HTML)、拡張可能ハイパーテキストマークアップ言語(XHTML)、または拡張可能マークアップ言語(XML)などのマークアップ言語である程度までは書かれている。ウェブアプリケーションは、カスケーディング・スタイル・シート(CSS)などのプレゼンテーション定義言語である程度まで書かれても良い。ウェブアプリケーションは、エイジャックス(Asynchronous JavaScript and XML(AJAX))、Flash(登録商標)ActionScript、JavaScript、またはSilverlight(登録商標)などの、クライアント側スクリプト言語である程度まで書かれても良い。ウェブアプリケーションは、アクティブサーバーページ(ASP)、ColdFusion(登録商標)、Perl、Java(商標)、JavaServer Pages(JSP)、ハイパーテキストプリプロセッサ(PHP)、Python(商標)、Ruby、Tcl、スモールトーク、WebDNA(登録商標)、またはGroovyなどのサーバー側コーディング言語である程度までは書かれても良い。ウェブアプリケーションは、構造化クエリ言語(SQL)などのデータベースクエリ言語である程度までは書かれても良い。

【0062】

モバイルアプリケーション

コンピュータプログラムは、モバイルデジタル処理デバイスに提供されるモバイルアプリケーションを含むこともある。モバイルアプリケーションは、製造時にモバイルデジタル処理デバイスに提供されても良い。モバイルアプリケーションは、本明細書に記載されるコンピュータネットワークを介してモバイルデジタル処理デバイスに提供されても良い。

【0063】

モバイルアプリケーションは、例えばハードウェア、言語および開発環境を用いて作成されても良い。モバイルアプリケーションは、様々なプログラミング言語で書かれても良い。適切なプログラミング言語は、非限定的な例として、C、C++、C#、Objective-C、Java（商標）、JavaScript、Pascal、Object Pascal、Python（商標）、Ruby、VB.NET、WML、およびCSSを含むまたは含まないXHTML/HTML、あるいはこれらの組み合わせを含む。

10

【0064】

適切なモバイルアプリケーションの開発環境は、いくつかのソースから利用可能である。市販の開発環境は、非限定的な例として、Airplay SDK、alchemo、Appcelerator（登録商標）、Celsius、Bedrock、Flash Lite、.NET Compact Framework、Rhobile、およびWorklight Mobile Platformを含む。他の開発環境は、コストをかけずに利用可能であり、非限定的な例として、Lazarus、Mobiflex、MoSync、およびPhonegapを含む。さらに、モバイルデバイスのメーカーは、非限定的な例として、iPhone（登録商標）およびiPad（登録商標）（iOS）SDK、アンドロイド（商標）SDK、BlackBerry（登録商標）SDK、BREW SDK、Palm（登録商標）OS SDK、Symbian SDK、webOS SDK、およびWindows（登録商標）モバイルSDKを含むソフトウェア開発者キットを販売する。

20

【0065】

スタンドアロンアプリケーション

コンピュータプログラムは、既存プロセスへのアドオン、例えばプラグイン、ではない独立したコンピュータプロセスとして実行されるプログラムである、スタンドアロンアプリケーションを含むことができる。スタンドアロンアプリケーションは、コンパイルされ得る。コンパイラは、プログラミング言語で書かれたソースコードをアセンブリ言語または機械コードなどのバイナリーオブジェクトコードに変換するコンピュータプログラムである。適切なコンパイルされたプログラミング言語は、非限定的な例として、C、C++、Objective-C、COBOL、Delphi、Eiffel、Java（商標）、Lisp、Python（商標）、Visual Basic、およびVB.NET、またはこれらの組み合わせを含む。コンパイラは、実行可能プログラムを作成するために少なくとも部分的に実行されることが多い。

30

【0066】

ウェブブラウザプラグイン

コンピュータプログラムは、ウェブブラウザプラグインを含むことができる。コンピューティングにおいて、プラグインは、より大きなソフトウェアアプリケーションに特定の機能を加える1つ以上のソフトウェアコンポーネントであっても良い。ソフトウェアアプリケーションのメーカーは、第三者の開発者が、アプリケーションを拡張する能力を生み出し、容易に新しい特徴の追加を支援し、そしてアプリケーションのサイズを縮小することができるようにプラグインをサポートする。プラグインがサポートされている場合、プラグインはソフトウェアアプリケーションの機能のカスタマイズを可能にし得る。例えば、プラグインは、ビデオの再生、インタラクティビティの生成、ウイルスのスキャン、および特定のファイルタイプの表示のために、ウェブブラウザで一般的に使用される。ウェブブラウザプラグインとしては、限定されないが、Adobe（登録商標）Flash

40

50

h (登録商標) プレーヤー、Microsoft (登録商標) Silverlight (登録商標)、および Apple (登録商標) QuickTime (登録商標) が挙げられる。ツールバーは、1つ以上のウェブブラウザ拡張機能、アドイン、またはアドオンを含むことができる。いくつかの実施形態では、ツールバーは、1つ以上のエクスプローラバー、ツールバンド、またはデスクバンドを含む。

【0067】

非限定的な例として、C++、Delphi、Java (商標)、PHP、Python (商標)、また VB.NET、あるいはそれらの組み合わせを含む、様々なプログラミング言語におけるプラグインの開発を可能にし得るいくつかのプラグイン・フレームワークが利用できる。

10

【0068】

ウェブブラウザ (インターネットブラウザとも呼ばれる) は、ワールドワイドウェブ上の情報資源を検索し、表示し、トラバースするためのソフトウェアアプリケーションであり、それは、ネットワーク接続されたデジタル処理デバイスと共に使用するために構成されても良い。適切なウェブブラウザは、非限定的な例として、Microsoft (登録商標) Internet Explorer (登録商標)、Mozilla (登録商標) Firefox (登録商標)、Google (登録商標) Chrome、Apple (登録商標) Safari (登録商標)、Opera Software (登録商標) Opera (登録商標)、および KDE Konqueror を含む。いくつかの実施形態では、ウェブブラウザは、モバイルのウェブブラウザである。モバイルのウェブブラウザ (マイクロブラウザ、ミニブラウザ、およびワイヤレスブラウザとも呼ばれる) は、限定しない例として、ハンドヘルドコンピュータ、タブレットコンピュータ、ネットブックコンピュータ、サブノートブックコンピュータ、スマートフォン、ミュージックプレーヤー、携帯情報端末 (PDA)、およびハンドヘルドビデオゲームシステムを含む、モバイルのデジタル処理デバイス上で使用されるために設計され得る。適切なモバイルのウェブブラウザは、非限定的な例として、Google (登録商標) Android (登録商標) ブラウザ、RIM BlackBerry (登録商標) Browser、Apple (登録商標) Safari (登録商標)、Palm (登録商標) Blazer、Palm (登録商標) WebOS (登録商標) Browser、携帯用の Mozilla (登録商標) Firefox (登録商標)、Microsoft (登録商標) Internet Explorer (登録商標) Mobile、Amazon (登録商標) Kindle (登録商標) Basic Web、Nokia (登録商標) Browser、Opera Software (登録商標) Opera (登録商標) Mobil、および Sony (登録商標) PSP (商標) ブラウザを含む。

20

30

【0069】

ソフトウェアモジュール

本明細書に記載されるシステム、媒体、ネットワークおよび方法は、ソフトウェア、サーバー、および/またはデータベースモジュールあるいはそれらの使用を含むこともある。ソフトウェアモジュールは、様々な機械、ソフトウェアおよびプログラミング言語を用いて作成されても良い。本明細書に開示されるソフトウェアモジュールは、多くの方法で実施される。ソフトウェアモジュールは、ファイル、コードのセクション、プログラミングオブジェクト、プログラミング構造、またはそれらの組み合わせを含んでも良い。ソフトウェアモジュールは、複数のファイル、コードの複数のセクション、複数のプログラミングオブジェクト、複数のプログラミング構造、またはそれらの組み合わせを含んでも良い。1つ以上のソフトウェアモジュールは、非限定的な例として、ウェブアプリケーション、モバイルアプリケーション、およびスタンドアロンアプリケーションを含んでも良い。いくつかの実施形態では、ソフトウェアモジュールは、1つのコンピュータプログラムまたはアプリケーション中にある。ソフトウェアモジュールは、1つを超えるコンピュータプログラムあるいはアプリケーション中にある。ソフトウェアモジュールは、1つの機械上でホストされても良い。ソフトウェアモジュールは、1つを超える機械上で

40

50

ホストされても良い。ソフトウェアモジュールは、クラウドコンピューティングプラットフォーム上でホストされても良い。ソフトウェアモジュールは、1つの場所において1つ以上の機械上でホストされても良い。ソフトウェアモジュールは、1つを超える場所において1つ以上の機械上でホストされても良い。

【0070】

データベース

本明細書に開示されるプラットフォーム、システム、媒体および方法は、1つ以上のデータベース、あるいはそれらの使用を含むこともある。本明細書に提供される開示を考慮すると、多くのデータベースが生理学的データの記憶および検索に適している。様々な実施形態では、適切なデータベースは、非限定的な例として、リレーショナルデータベース、非リレーショナルデータベース、オブジェクト指向型データベース、オブジェクトデータベース、実体関連モデルデータベース、連想型データベース、およびXMLデータベースを含む。さらに非限定的な例としては、SQL、PostgreSQL、MySQL、Oracle、DB2、およびSybaseが挙げられる。いくつかの実施形態では、データベースはインターネットを利用したものである。データベースは、ウェブを利用したものであっても良い。データベースは、クラウドコンピューティングを利用したものであっても良い。データベースは、1つ以上のローカルコンピュータ記憶デバイスに基づくこともある。

【0071】

以下の実施例は、本明細書に開示される実施形態の原理および実践をより明確に当業者に例証するために示されており、任意の主題の実施形態の範囲を限定するものとして解釈されるべきではない。特段の定めのない限り、すべての部およびパーセントは重量ベースである。

【0072】

アルゴリズム

本明細書に開示されるプラットフォーム、システム、媒体、および方法は、1つ以上のアルゴリズムまたはその使用を含む。本明細書で提供される本開示の観点から、多くのアルゴリズムが配列データの検索および比較に適している。様々な実施形態では、適切なアルゴリズムとしては、非限定的な例として、BLAST、DIAMOND、BLAT、BWT、PLAST、スミス-ウォーターマン、あるいは他の配列の検索とアライメントのためのアルゴリズムが挙げられる。アルゴリズムは、既存のアルゴリズムの高速化版または拡張版、あるいはこうしたアルゴリズムを使用するソフトウェアツールを含むこともある。いくつかの例では、適切な高速化あるいは拡張化アルゴリズムおよびソフトウェアツールとしては、非限定的な例として、CS-BLAST、Tera-BLAST、GPU-Blast、G-BLASTN、MPIBLAST、Parace1 BLAST、CaBLAST、あるいはBLASTアルゴリズムを高速化する任意の他のさらなるアルゴリズムまたはソフトウェアツールが挙げられる。

【0073】

バイオセイフティーとバイオセキュリティを增強させた生物学的配列あるいは構築物を設計および合成するためのシステムおよび方法が本明細書で提供される。いくつかの例では、バイオセイフティーとは、例えば、製造中または製造に由来する有害な生物学的製剤との接触を防ぐことを目的とする予防手段による個々の增強された安全性を指す。いくつかの例では、バイオセキュリティとは、例えば、有害な生物学的製剤の使用または拡散を防ぐことを目的とする予防手段によって集団の安全を守ることが指す。いくつかの例では、1以上の生物学的配列を含む1以上の生物学的構築物が受信され、データベースを用いてバイオセキュリティのリスクについてスクリーニングされ、および、生物学的配列あるいは構築物の1以上が有害な発現構築物または有害な生成物であると判定されると警報が発生される。いくつかの例では、生物学的配列または構築物とは合成配列を指す。いくつかの例では、生物学的配列または構築物とは自然発生配列を指す。いくつかの例では、生物学的配列または構築物とは核酸またはアミノ酸を含む。いくつかの例では、生物

10

20

30

40

50

学的配列とは合成配列を指す。いくつかの例では、生物学的配列とは自然発生配列を指す。いくつかの例では、生物学的配列は核酸またはアミノ酸を含む。いくつかの例では、ユーザーアノテーションは、データベース中の生物学的配列または構築物の特性に関する追加情報を提供するために使用される。いくつかの例では、该方法および該システムは、ハイスループット設計/構築/テストのワークフローにシームレスに適合するように自動化に適している。いくつかの例では、生物学的構築物のスクリーニングは、複数の時点で単一のソースあるいは複数のソースから得られた小さな生物学的配列の組み合わせの比較を含む。いくつかの例では、有害であると判定された生物学的配列あるいは構築物は、将来の偽陽性を減らすために人間の専門家によってさらに評価される。いくつかの例では、こうしたシステムと方法は、コンピュータ、ソフトウェアアプリケーション、およびユーザー並びにデータベースとインターフェース接続するためのネットワークを含む。

10

【0074】

以下を含むシステムが、本明細書で提供される：プロセッサおよびメモリ；生物学的な構築物のバイオセキュリティを評価するための機械の命令であって、生物学的な構築物に関連する複数のタグのデータベース、を含む機械の命令；アノテーションツール；および、随意に、スクリーニングツール。さらに、生物学的配列あるいは構築物が1以上の生物学的配列を含むシステムが、本明細書で提供される。さらに、生物学的配列が核酸配列であるシステムが、本明細書で提供される。さらに、生物学的配列がタンパク質配列であるシステムが、本明細書で提供される。さらに、アノテーションツールが、生物学的な構築物の配列の1以上のアノテーションされたタグをユーザーが提供することができるように構成されたシステムが、本明細書で提供される。さらに、1以上のアノテーションされたタグが少なくとも宿主と懸念の程度を含むシステムが、本明細書で提供される。さらに、1以上のアノテーションされたタグが結果を含むシステムが、本明細書で提供される。さらに、結果が疾患を含むシステムが、本明細書で提供される。さらに、1以上のアノテーションされたタグが背景を含むシステムが、本明細書で提供される。さらに、1以上のアノテーションされたタグが病原性を含むシステムが、本明細書で提供される。さらに、1以上のアノテーションされたタグが害を含むシステムが、本明細書で提供される。さらに、1以上のアノテーションされたタグが1以上のターム (t e r m s) に基づくシステムが、本明細書で提供される。さらに、1以上のアノテーションされたタグが1以上の文の記載に基づくシステムが、本明細書で提供される。さらに、アノテーションツールが1以上のアノテーションされたタグの統制語彙を作成するようにさらに構成されるシステムが、本明細書で提供される。さらに、アノテーションツールがキュレーションプロセスを含むシステムが、本明細書で提供される。さらに、キュレーションプロセスが外部データベースからデータベースまでの生物学的配列または構築物の情報を統合する工程を含むシステムが、本明細書で提供される。さらに、キュレーションプロセスが生物学的な構築物の無害な特徴を判定する工程を含むシステムが、本明細書で提供される。さらに、アノテーションツールが配列をデータベース中の生物学的配列または構築物の配列とアライメントする工程を含むシステムが、本明細書で提供される。さらに、スクリーニングツールが生物学的な構築物の所定の配列のバイオセキュリティリスクをユーザーが調べることができるように構成されるシステムが、本明細書で提供される。さらに、所定の配列がヌクレオチド配列を含むシステムが、本明細書で提供される。さらに、所定の配列がタンパク質配列を含むシステムが、本明細書で提供される。さらに、スクリーニングツールが所定の配列をデータベース中の生物学的配列あるいは構築物の配列とアライメントするための配列アライナー (s e q u e n c e a l i g n e r) を含むシステムが、本明細書で提供される。さらに、バイオセキュリティリスクの調査が相同性の程度によるフィルタリングことを含む検索システムが、本明細書で提供される。さらに、バイオセキュリティリスクの調査が配列アラインメント長さを評価することを含む検索システムが、本明細書で提供される。さらに、バイオセキュリティリスクが評価スコアを生成することを含む検索システムが、本明細書で提供される。さらに、スクリーニングツールがアプリケーションプログラマブルインタフェースをさらに含むシステムが、本明細書で提供され

20

30

40

50

る。さらに、機械の命令がアノテーションとスクリーニングのためのグラフィカルユーザインターフェースを含むシステムが、本明細書で提供される。

【0075】

バイオセキュリティーリスクを評価するためのコンピュータで実施される方法が本明細書で提供され、該方法は：生物学的な構築物に関連する複数のタグを保存するデータベースをプロセッサによって使用する工程と、生物学的な構築物の特徴をアノテーションするためのアノテーションツールをプロセッサによって使用する工程と、随意に、生物学的な構築物の特徴を調べるためのスクリーニングツールをプロセッサによって使用する工程とを含む。さらに、生物学的な構築物が生物学的配列を含む方法が、本明細書で提供される。さらに、生物学的配列が核酸配列である方法が、本明細書で提供される。さらに、生物学的配列がタンパク質配列である方法が、本明細書で提供される。さらに、アノテーションツールが、生物学的な構築物の配列の1以上のアノテーションされたタグをユーザーが提供することができるように構成された方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが少なくとも宿主と懸念の程度を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが結果を含む方法が、本明細書で提供される。さらに、結果が疾患を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが背景を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが病原性を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが害を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが1以上の項に基づく方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが1以上の文の記載に基づく方法が、本明細書で提供される。さらに、アノテーションツールが1以上のアノテーションされたタグの統制語彙を作成するようにさらに構成される方法が、本明細書で提供される。さらに、アノテーションツールがキュレーションプロセスを含む方法が、本明細書で提供される。さらに、キュレーションプロセスが外部データベースからデータベースまでの生物学的配列または構築物の情報を統合する工程を含む方法が、本明細書で提供される。さらに、キュレーションプロセスが生物学的な構築物の無害な特徴を判定する工程を含む方法が、本明細書で提供される。さらに、アノテーションツールが配列をデータベース中の生物学的な構築物の配列とアライメントする工程を含む方法が、本明細書で提供される。さらに、スクリーニングツールが生物学的な構築物の所定の配列のバイオセキュリティーリスクをユーザーが調べることを可能にするように構成される方法が、本明細書で提供される。さらに、所定の配列がヌクレオチド配列を含む方法が、本明細書で提供される。さらに、所定の配列がタンパク質配列を含む方法が、本明細書で提供される。さらに、スクリーニングツールが所定の配列をデータベース中の生物学的な構築物の配列とアライメントするための配列アライナー(sequence aligner)を含む方法が、本明細書で提供される。さらに、バイオセキュリティーリスクの調査が相同性の程度によるフィルタリングことを含む検索システムが、本明細書で提供される。さらに、バイオセキュリティーリスクの調査が配列アラインメント長さを評価することを含む検索システムが、本明細書で提供される。さらに、バイオセキュリティーリスクが評価スコアを生成することを含む検索システムが、本明細書で提供される。さらに、スクリーニングツールがアプリケーションプログラマブルインタフェースを含む方法が、本明細書で提供される。さらに、機械の命令がアノテーションとスクリーニングのためのグラフィカルユーザインターフェースを含む方法が、本明細書で提供される。

【0076】

バイオセキュリティーリスクを評価するためのコンピュータで実施される方法が本明細書で提供され、該方法は、生物学的な構築物に関連する複数のタグを保存するデータベースに、プロセッサによってアクセスする工程と、生物学的な構築物の特徴を調べるためのスクリーニングツールに、プロセッサによってアクセスする工程と、スクリーニングツールの調査結果を送るための報告ツールを、プロセッサによって送信する工程とを含む。さらに、生物学的な構築物が生物学的配列を含む方法が、本明細書で提供される。さらに、

10

20

30

40

50

生物学的配列が核酸配列である方法が、本明細書で提供される。さらに、生物学的配列がタンパク質配列である方法が、本明細書で提供される。さらに、生物学的な構築物の配列の1以上のアノテーションされたタグをユーザーが提供することができるように構成されたアノテーションツールを含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが少なくとも宿主と懸念の程度を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが結果を含む方法が、本明細書で提供される。さらに、結果が疾患を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが背景を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが病原性を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが害の程度を含む方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが1以上の項に基づく方法が、本明細書で提供される。さらに、1以上のアノテーションされたタグが1以上の文の記載に基づく方法が、本明細書で提供される。さらに、アノテーションツールが1以上のアノテーションされたタグの統制語彙を作成するようにさらに構成される方法が、本明細書で提供される。さらに、アノテーションツールがキュレーションプロセスを含む方法が、本明細書で提供される。さらに、キュレーションプロセスが外部データベースからデータベースまでの生物学的配列または構築物の情報を統合する工程を含む方法が、本明細書で提供される。さらに、キュレーションプロセスが生物学的な構築物の無害な特徴を判定する工程を含む方法が、本明細書で提供される。さらに、アノテーションツールが配列をデータベース中の生物学的な構築物の配列とアライメントする工程を含む方法が、本明細書で提供される。さらに、スクリーニングツールが生物学的な構築物の所定の配列のバイオセキュリティリスクをユーザーが調べることを可能にするように構成される方法が、本明細書で提供される。さらに、所定の配列がヌクレオチド配列を含む方法が、本明細書で提供される。さらに、所定の配列がタンパク質配列を含む方法が、本明細書で提供される。さらに、スクリーニングツールが所定の配列をデータベース中の生物学的な構築物の配列とアライメントするための配列アライナー(sequence aligner)を含む方法が、本明細書で提供される。さらに、バイオセキュリティリスクの調査が相同性の程度によるフィルタリングを含む検索システムが、本明細書で提供される。さらに、バイオセキュリティリスクの調査が配列アラインメント長さを評価することを含む検索システムが、本明細書で提供される。さらに、バイオセキュリティリスクが評価スコアを生成することを含む検索システムが、本明細書で提供される。さらに、スクリーニングツールがアプリケーションプログラマブルインタフェースを含む方法が、本明細書で提供される。さらに、アノテーションのためのグラフィカルユーザインターフェース用の機械の命令を送信する工程をさらに含む方法が、本明細書で提供される。さらに、スクリーニングのためのグラフィカルユーザインターフェース用の機械の命令を送信する工程をさらに含む方法が、本明細書で提供される。さらに、報告のためのグラフィカルユーザインターフェース用の機械の命令を送信する工程をさらに含む方法が、本明細書で提供される。さらに、生物学的な構築物が有害な発現産物(例えば翻訳に起因するタンパク質)あるいは有害な生成物(例えば転写に起因するRNA)に関連する生物学的配列を含む方法が、本明細書で提供される。さらに、生物学的配列がウイルス、細菌、または真菌である方法が、本明細書で提供される。さらに、生物学的な構築物に関連する複数のタグを保存するためにデータベースにアクセスするという受信された機械の命令を含む方法が、本明細書で提供される。さらに、機械の命令が生物学的な構築物に関連する情報を含む方法が、本明細書で提供される。さらに、生物学的配列あるいは構築物に関連する情報が核酸配列あるいはタンパク質配列を含む方法が、本明細書で提供される。さらに、生物学的配列または構築物に関連する情報がデータベース登録番号を含む方法が、本明細書で提供される。

【0077】

本開示の様々な態様は個々に、まとめて、あるいは互いに組み合わせて評価可能であることが理解されよう。本明細書に記載される本開示の様々な態様は以下で説明される特定の用途のいずれかに適用され得る。本開示の他の目的と特徴は、明細書、請求項、および

添付の図面を精査することによって明白になる。

【実施例】

【0078】

実施例1：配列アノテーション

【0079】

生物学的配列はプロセッサユニットによって受信された。本実施例では、生物学的配列はタンパク質配列である。プロセッサユニットはタンパク質データベースにアクセスし、受信されたタンパク質配列に一致するタンパク質配列を同定した。プロセッサユニットは、タンパク質配列の様々な特性に関連する情報を受信した。特性は以下のものを含んでいた：タンパク質配列に関連する核酸配列、タンパク質配列、タンパク質の名称、株ソース情報、配列データベース（例えば、NCBI）へのリンク、配列データベース登録番号、同一配列（タンパク質または核酸）、類似配列（タンパク質または核酸）、疾患源（例えば、ウイルス、細菌）、有機体の分類学的な記載（例えば、界、門、綱、目、科、属、種）、宿主情報（例えば、ヒト、哺乳動物、鳥、昆虫）、有害な相互作用の文脈または経路（例えば、摂取、吸入）、徴候、および懸念の程度。本実施例では、利用したタンパク質はニューカッスル病ウイルス-3であった。典型的なユーザーインターフェースにより提供されたアノテーションのための特性が図1で提供されている。機械の命令が生物学的配列に関連する特性の情報とともにプロセッサによって受け取られると、生物学的配列に関連するタグ情報が更新された。例えば、図1を参照すると、ニューカッスル病ウイルス-3は、タンパク質配列、同一タンパク質（AHL4519.1.1とAHL45193.1）、宿主型（鳥）、有害な相互作用の手段（吸入）、および徴候（呼吸不全）のタグ情報を有する。

10

20

【0080】

プロセッサユニットが「赤血球凝集素ノイラミニダーゼ-ニューカッスル病ウイルス」ファミリーに関する選択を受信すると、ウイルス株情報のリストがアクセスされ、随意に、菌株を表示するというユーザーインターフェースに対する機械の命令とともに送信された。例えば、アノテーションのための赤血球凝集素ノイラミニダーゼ-ニューカッスル病ウイルスの679の利用可能な菌株の一部のリストを提供している図2を参照する。

【0081】

本明細書と一致するさらなるタグ情報は、限定されないが、連邦政府の指定生物剤プログラム（FSAP）管理あるいは輸出管理（Export Control）を含むいくつかの例でも使用される。

30

【0082】

実施例2：配列スクリーニング

【0083】

図3Aを参照すると、プロセッサは、生物学的配列情報、この場合は核酸情報を含むクエリファイルの形態の機械の命令を受信した。プロセッサは核酸やタンパク質データベースとも通信していた。プロセッサは核酸とタンパク質データベースにアクセスした。BLAST処理されたレポートは、照会された生物学的配列と部分的または全体的に関連があると同定された同じおよび類似する配列を表記して作成される。BLAST処理されたレポートからの配列は、その後、「制限された」リストとも呼ばれる、有害な生物学的配列（タンパク質または核酸）に関連する配列を同定する配列アノテーションを含むデータベースに照会された。スクリーニングレポートは、これらの処理の結果を要約するユーザーインターフェースの形態で作成された。スクリーニングレポートはユーザーインターフェースに対する機械の命令の形態で送信された。プロセッサは、制限されたりリスト情報にアクセスするというデータベースに対する特別な命令を受けた。図4を参照。制限されたりリストは、インターネット上で開いていることもあれば、閉じられていることもあり、認証を用いてのみアクセス可能であることもある。スクリーニングレポートも生物学的配列のスクリーニングの要約を含むように作成された。5回のスクリーニングを行った。図6を参照。スクリーニングレポートは「制限された割り当て」つまり、有害な生物学的配列の

40

50

リストを含むように作成された。図7を参照。スクリーニングレポートは - B r u c e l l a s u i s - 2 タンパク質を同定した。

【0084】

実施例3：特定のゲノムに対するプレスクリーニング

【0085】

大痘瘡または小痘瘡のゲノムの500を超えるヌクレオチドへのアクセスは、世界保健機構(WHO)の政策によって制限されている。もっと長い配列を希望する者は申請を行い、合成の前にWHOにより許可を受けなければならない。痘瘡の固有の性質ゆえに、ワクシニアおよび他の密接に関連するオルソポックスウイルスとともに、大痘瘡と小痘瘡のゲノムのみに対するプレスクリーニングを行う。核酸配列は、実施例2の一般的なバイオセキュリティースクリーニング法とオルソポックスウイルスのゲノムを使用して評価された。このスクリーニングは(商品ハードウェア上でblastxによって)1秒未満で実行された。ワクシニアと他のオルソポックスの基準配列は、要求された配列の相同性が警報前に痘瘡に対して最も大きくなる(2010HHSガイダンス「最良」基準に類似する)ことを確かめるために含まれた。これは、有害な配列が検出されると製造を開始する前に人間によるレビューを求める警報が発生されるオーダー見積(order quote)生成プロセスの間に、随意に実施することが可能である。

【0086】

実施例4：ライブラリー鑄型スクリーニング

【0087】

約200のアミノ酸をコードする遺伝子をコードする約600のヌクレオチドの遺伝子長さの核酸配列を、変異体ライブラリーの産生のために選択した。配列を得て、実施例2の一般的なバイオセキュリティースクリーニング法に晒すことで、変異体ライブラリーが確実に有害な配列を含まないようにした。プログラムは、有害な配列が検出されると人間によるレビューを求める警報を発生させるように設計された。

【0088】

実施例5：カスタム核酸スクリーニング

【0089】

ベクターなどの物理的な核酸含有材料が次世代シーケンシング(NGS)により得られ、配列決定された。NGSから得られたコンセンサス配列データを、実施例2の一般的なバイオセキュリティースクリーニング法に晒した。これにより、核酸材料は、使用するための挿入部位から離れてベクター骨格中の毒素の発現をコードするなどしてバイオセキュリティーまたはバイオセイフティーの懸念を引き起こさず、大腸菌への形質転換が毒素などの有害な薬剤の発現を引き起こすことになる。プログラムは、有害な配列が検出されると人間によるレビューを求める警報を発生させるように設計された。

【0090】

実施例6：指定病原体のゲノムに対する同じクエリ内のクロスオーダーアセンブリ

【0091】

要求者(顧客などの生物学的配列あるいは構築物の要求元)が、経時的に、および個々のオーダーに沿って、指定病原体(select agent)により規制される細菌あるいはウイルスのいずれかのゲノムのかなりの部分を蓄積させる可能性があるというリスクを管理するために、各要求者の後のバックグラウンドプロセスは、その要求者からのすべての事前のオーダーのデータベースを照会し、かつ、実施例2の一般的な方法を用いて指定病原体の細菌あるいはウイルスのいずれかに対する相同性が高い任意のセグメントの記録を集める。これにより、たとえ上記のような領域が個々のオーダーの間に正式な警報あるいは所有の拒否を引き起こすのに不十分であったとしても、評価および警報が保証される。これらの相同性の高いセグメントは、懸念のある指定病原体のゲノム上の区間として表され、その後、要求者ごとおよびゲノムごとのすべての区間の結合が、要求者ごとにこうした有機体の最大の理論構成を決定するために生成される。いったん任意の要求者が所定の指定病原体のゲノムの20%以上を生成することができるようになると、人間によ

るレビューと要求者のフォローアップを求める警報が意図的に発生される。

【0092】

実施例7：仮説生成のための指定病原体のゲノムに対するポリヌクレオチドプールアセンブリ

【0093】

せいぜい200の塩基しか含んでいない配列などの短いポリヌクレオチド配列について、既存のスクリーニング方法は非常に高い偽陽性率を誇る。要求者（生物学的配列または構築物の要求元、つまり、顧客）が、制御されたあるいは有害な配列を潜在的にアセンブルするのに十分なポリヌクレオチドをいつオーダーしたのかを判定するために、ポリヌクレオチドのセットを調べる代替的なスクリーニングアプローチが用いられる。オーダーの間に、バックグラウンドプロセスは、1以上の要求元の内部で、NGSからのアセンブリアルゴリズムを使用して、指定病原体の細菌とウイルスのゲノムに対するオーダーにわたってポリヌクレオチドをアセンブルする。これらのアセンブリは、「要求者AおよびBのオーダーX、Y、およびZを組み合わせると、痘瘡の3つの遺伝子を完全にアセンブルすることができる」などといった仮説の生成を可能にする。こうした仮説は、人間によるレビューを求める警報を発生させ、随意に、要求者との継続的な議論を引き起こし、あるいは法執行機関（law enforcement）へ直接報告する。遺伝子長の配列に対する高い相同性の可能性が低いことを考慮すると、偽陽性率は低いままでなければならず、さらなる偽陽性の減少は、容易なアセンブリを可能にする適切な重複が存在する（つまり、それを念頭に設計されたように見える）かどうかを判定するために、仮定されたポリヌクレオチド集合のアライメント構造を評価する形態でもたらされる。

10

20

【0094】

実施例8：機械学習によりガイドされたりスクアノテーション

【0095】

スクリーニングプラットフォームと人間によるレビューにより、広範な制限されていないリストと、真陽性の警報のケースのセットとが作られ、ここで、生物学的配列または構築物の要求元が懸念のある制限された配列をオーダーしたことが確認された。機械学習アルゴリズムを、配列自体（例えば、隠れマルコフモデル（HMM）タイプの文脈を意識した状態モデル）および/またはGenBank記録アノテーション（例えば、あらかじめ制限されなかった配列を列挙した記録を用いて、共有された言語および意味に基づいて、将来の制限されていない配列割り当ての可能性を予測するための自然言語処理（NLP）タイプのモデル）の両方で学習させる。

30

【0096】

本開示の好ましい実施形態が本明細書で示され記載されているが、こうした実施形態はほんの一例として与えられているに過ぎないということは当業者には明白であろう。多くの変形、変更、および置換が、本開示から逸脱することなく、当業者によって想到される。本明細書に記載される本開示の実施形態の様々な代案が本開示の実施の際に利用され得ることを理解されたい。

[illegible]

FIG. 2

Pages in category "Hemagglutinin Neuraminidase-Newcastle Disease virus"

[illegible]

```
graph TD; A[クエリファイル] --> C[BLASTレポート]; B[タンパク質データベース] --> C; C --> D[スクリーニングレポート]; E[制限されたリスト] -.-> D; F[制限されたリスト] -.-> D;
```

```

graph LR
    A[物理的核酸材料] -- 配列決定 --> B[クエリ配列]
    B -- ORF検索 --> C[タンパク質データベース]
    B -- BLAST分析 --> D{制限されたリストにあるか?}
    C -- BLAST分析 --> D
    D -- はい --> E{制限されていないリストにあるか?}
    D -- いいえ --> F[警告が発生されない]
    E -- はい --> F
    E -- いいえ --> G[専門家によるチェック]
    G --> H{配列または配列産物  
が有害であるか?}
    H -- はい --> I[警告が発生される]
    H -- いいえ --> J[制限されていないリスト  
に追加する]
    J --> D
  
```

【 図 4 】

Home Screen+ Database+ Lists+ Logout Stepblast

Lists

Only active lists will be used for a new screen

Show 10 entries Search:

Role	Type	Name	Description	Date added	Active
Restricted	Accession	IGSC		2016-08-10 19:47:49	True
Restricted	Accession	Select Agents		2016-08-10 19:47:53	False
Restricted	Taxon	Example list		2016-08-10 22:49:29	True

Showing 1 to 3 of 3 entries

Previous 1 Next

FIG. 4

【 図 5 】

Home Screen+ Database+ Lists+ Logout Stepblast

Submit a screen

Name

Database

Seqfield

Personal Database

Description

fasta

Choose File: No file chosen

Submit

FIG. 5

【 図 6 】

Home Screen+ Database+ Lists+ Logout Stepblast

Biosecurity Screens

Show 10 entries Search:

Name	User	Status	Sequences	Unreviewed	No concern	Of concern	Date added	Blast
Screen 1	Ben	complete	16	16	0	0	2016-06-11 09:56:30	View
Screen 2	Ben	complete	16	11	0	0	2016-08-11 16:44:07	View
Screen 3	Ben	complete	16	16	0	0	2016-08-11 18:44:22	View
Screen 4	Ben	complete	6990	6990	0	0	2016-08-11 18:45:09	View
Screen 5	Ben	blast running	0	0	0	0	2016-08-11 16:50:29	View

Showing 1 to 5 of 5 entries

Previous 1 Next

FIG. 6

【 図 7 】

Home Screen+ Database+ Lists+ Logout Stepblast

Lists used for the screen

Show 10 entries Search:

Role	Type	Name	Description	Date added	Active
Restricted	None	IGSC		2016-08-11 16:40:37	True
Restricted	None	Example list	A toxin list	2016-08-11 16:11:28	True

Showing 1 to 2 of 2 entries

Previous 1 Next

Screened sequences

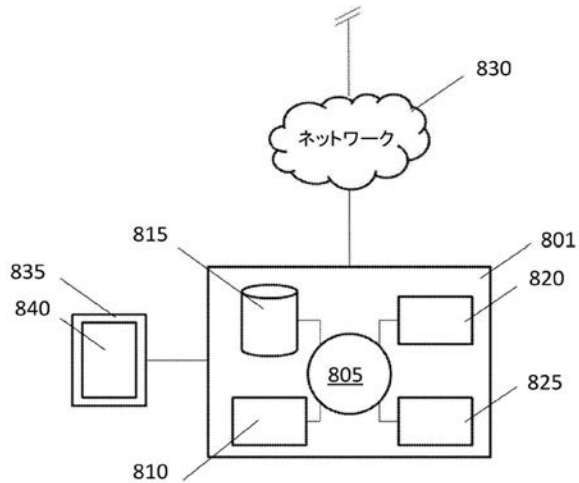
Sequence ID	Description	Concern	Reason	Hits	Unreviewed	Safe	Hazardous
end_Gora	end_Gora	Unknown		1	1	0	0

Restricted alignments

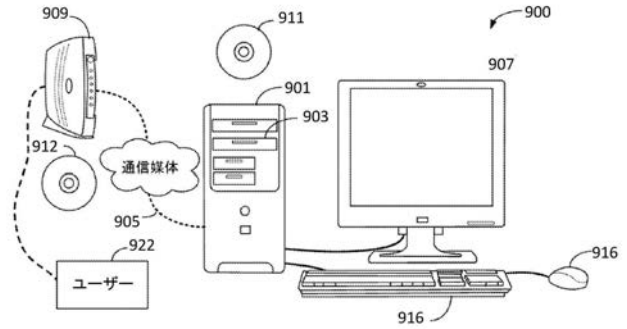
Accession	Taxon	Description	Concern	Reason	Start	End	Length	Strand	Type
672735645	None	Gen Cell Cycle Regulator Family: Brucella sub-2	Unreviewed		32	64	33	1	Accession

FIG. 7

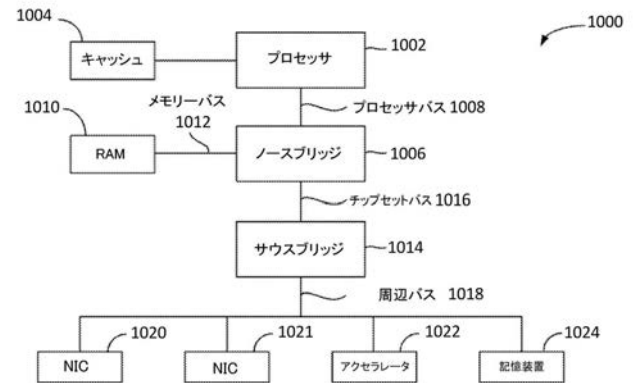
【図 8】



【図 9】



【図 10】



【図 11】

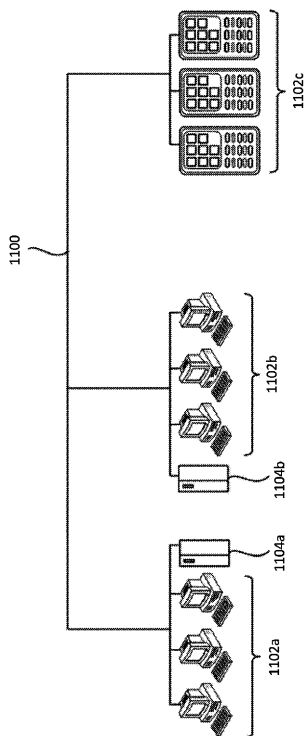
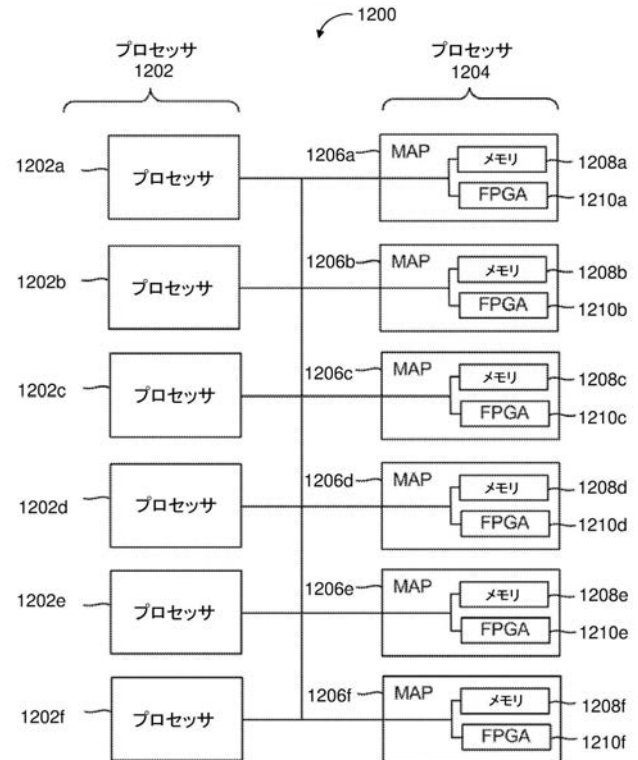


FIG. 11

【図 12】



【配列表】

20195239400000001.app

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2017/036868

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - G06F 17/50; G06F 19/22; G06F 19/28; C12N 15/10 (2017.01)

CPC - G06F 17/50; G06F 19/22; G06F 19/28; C12N 15/10 (2017.05)

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

USPC - 703/11; 506/10; 506/26; 506/17 (keyword delimited)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2015/0120265 A1 (AMIRAV-DRORY et al) 30 April 2015 (30.04.2015) entire document	16, 18, 19, 28, 30
Y		1-4, 13-15, 17, 29, 31
Y	US 2016/0096160 A1 (TWIST BIOSCIENCE CORPORATION) 07 April 2016 (07.04.2016) entire document	1-4, 13-15
Y	US 2010/0292102 A1 (NOURI) 18 November 2010 (18.11.2010) entire document	2, 14, 17, 31
Y	US 2009/0170802 A1 (STAHLER et al) 02 July 2009 (02.07.2009) entire document	29
A	US 2011/0172127 A1 (JACOBSON et al) 14 July 2011 (14.07.2011) entire document	1-4, 13-19, 28-31

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

30 July 2017

Date of mailing of the international search report

11 AUG 2017

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, VA 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Blaine R. Copenheaver

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/036868

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 5-12, 20-27
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

フロントページの続き

(81)指定国・地域 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, RU, TJ, TM), EP(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT

(特許庁注：以下のものは登録商標)

1 . J A V A S C R I P T

2 . V I S U A L B A S I C