

(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2013-531983

(P2013-531983A)

(43) 公表日 平成25年8月15日(2013.8.15)

(51) Int.Cl.

C 12 Q 1/68

C 12 N 15/09

(2006.01)

(2006.01)

F 1

C 12 Q 1/68

C 12 N 15/00

Z N A A

A

テーマコード(参考)

4 B 0 2 4

4 B 0 6 3

審査請求 未請求 予備審査請求 未請求 (全 94 頁)

(21) 出願番号 特願2013-514408 (P2013-514408)
 (86) (22) 出願日 平成23年6月10日 (2011.6.10)
 (85) 翻訳文提出日 平成25年2月1日 (2013.2.1)
 (86) 國際出願番号 PCT/US2011/040106
 (87) 國際公開番号 WO2011/156795
 (87) 國際公開日 平成23年12月15日 (2011.12.15)
 (31) 優先権主張番号 61/439,167
 (32) 優先日 平成23年2月3日 (2011.2.3)
 (33) 優先権主張国 米国(US)
 (31) 優先権主張番号 61/374,041
 (32) 優先日 平成22年8月16日 (2010.8.16)
 (33) 優先権主張国 米国(US)
 (31) 優先権主張番号 61/354,011
 (32) 優先日 平成22年6月11日 (2010.6.11)
 (33) 優先権主張国 米国(US)

(71) 出願人 512316910
 パソジエニカ、インコーポレイテッド
 PATHOGENICA, INC.
 アメリカ合衆国 マサチューセッツ州 0
 2142, ケンブリッジ, ファーストスト
 リート 245
 (74) 代理人 110001302
 特許業務法人北青山インターナショナル
 (72) 発明者 ダイアモンド, リサ
 アメリカ合衆国 カリフォルニア州 94
 501, アラメダ, ベイビュードライヴ
 3000

最終頁に続く

(54) 【発明の名称】多重生物検出のための核酸ならびにその使用および製造方法

(57) 【要約】

本発明は、複数の標的生物から1つまたは複数の対象の配列を大量に多重捕獲することができる環状化「捕獲」プローブを含む線状核酸プローブの混合物を提供する。本発明によって提供される方法は、一般的な病原体などの1つまたは複数の対象の生物の迅速、正確、かつ経済的な検出を可能にする。

【選択図】図25

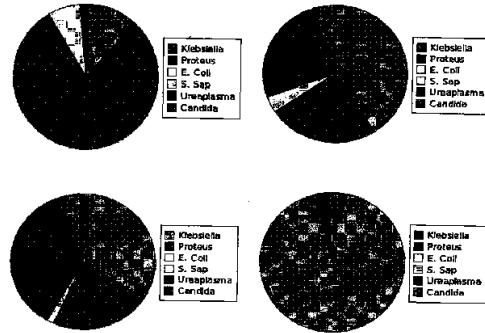


FIG. 25

【特許請求の範囲】**【請求項 1】**

被験者の中の少なくとも 1 つの標的生物を検出するための複数のプローブを含む混合物において、各プローブが、

a . 第 1 の末端において、前記少なくとも 1 つの標的生物のゲノム中に存在する第 1 の標的配列と特異的にハイブリッド形成する第 1 の相同プローブ配列と、

b . 第 2 の末端において、前記少なくとも 1 つの標的生物のゲノム中に存在する第 2 の標的配列と特異的にハイブリッド形成する第 2 の相同プローブ配列と、

c . 前記第 1 の末端と前記第 2 の末端との間の、検出可能な部分およびプライマーを含む骨格配列と

を含み、前記第 1 の標的配列および前記第 2 の標的配列が、少なくとも 2 つのヌクレオチドを含む対象の領域によって分離されおり、各プローブ中の前記第 1 および第 2 の相同プローブ配列のそれぞれが、

i . 前記標的生物と特異的にハイブリッド形成し、

i i . 50 ~ 72 の範囲の T_m を有し、

i i i . (a) 前記混合物中の他のいかなる相同プローブ配列、(b) いかなる骨格配列、(c) 前記被験者のゲノム中に存在するいかなるヌクレオチド配列、または(d) 前記標的生物以外の所定の配列決定された生物セットのゲノム中に存在するいかなるヌクレオチド配列とも特異的にハイブリッド形成せず、

i v . 前記少なくとも 1 つの標的ゲノムにおいて反復閾値未満で発生し、前記反復閾値が 20 であり、そして

v . 4 つを超える連続した同一ヌクレオチドを含有せず、実質的に二次構造を有しないことを特徴とする混合物。

【請求項 2】

請求項 1 に記載の混合物において、各プローブ中の前記第 1 および第 2 の相同プローブ配列のそれぞれが、前記対象の領域に隣接する前記対象の生物の配列決定された変異体のゲノムと特異的にハイブリッド形成することを特徴とする混合物。

【請求項 3】

請求項 1 に記載の混合物において、前記反復閾値が 2 であることを特徴とする混合物。

【請求項 4】

請求項 1 に記載の混合物において、前記対象の領域が、前記標的生物の配列決定された変異体の間で多形性であることを特徴とする混合物。

【請求項 5】

請求項 4 に記載の混合物において、前記対象の領域が、毒素産生または抗生物質耐性に関連することを特徴とする混合物。

【請求項 6】

請求項 1 に記載の混合物において、前記少なくとも 1 つの標的生物が病原体を含むことを特徴とする混合物。

【請求項 7】

請求項 1 に記載の混合物において、前記少なくとも 1 つの標的生物が細菌を含むことを特徴とする混合物。

【請求項 8】

請求項 1 に記載の混合物において、前記少なくとも 1 つの標的生物がウイルスを含むことを特徴とする混合物。

【請求項 9】

請求項 1 に記載の混合物において、前記少なくとも 1 つの標的生物が真菌を含むことを特徴とする混合物。

【請求項 10】

請求項 1 に記載の混合物において、前記少なくとも 1 つの標的生物が古細菌を含むことを特徴とする混合物。

【請求項 1 1】

請求項 1 に記載の混合物において、前記少なくとも 1 つの標的生物が真核生物を含むことを特徴とする混合物。

【請求項 1 2】

請求項 1 に記載の混合物において、前記骨格が切断部位をさらに含むことを特徴とする混合物。

【請求項 1 3】

請求項 1 2 に記載の混合物において、前記切断部位が、制限エンドヌクレアーゼ認識部位であることを特徴とする混合物。

【請求項 1 4】

請求項 1 に記載の混合物において、前記骨格が第 2 のプライマーをさらに含むことを特徴とする混合物。

【請求項 1 5】

請求項 1 に記載の混合物において、前記検出可能な部分がバーコード配列であることを特徴とする混合物。

【請求項 1 6】

請求項 1 に記載の混合物において、前記混合物が、少なくとも 4、10、15、20、30、40、60、80、100、150、200、250、300、400、500、1000、2000、4000、8000、10000、15000、または 20000 の異なる標的生物に対して少なくとも 1 つのプローブを含むことを特徴とする混合物。

【請求項 1 7】

請求項 1 に記載の混合物において、前記混合物が、少なくとも 10、20、30、40、60、80、100、200、250、500、1000、2000、4000、8000、10000、20000、30000、40000、50000、60000、70000、80000、90000、または 100000 のプローブを含むことを特徴とする混合物。

【請求項 1 8】

請求項 1 に記載の混合物において、前記混合物がさらに、少なくとも 1 つの被験者特異的プローブを含むことを特徴とする混合物。

【請求項 1 9】

請求項 1 に記載の混合物において、前記被験者が哺乳類であることを特徴とする混合物。

。

【請求項 2 0】

請求項 1 9 に記載の混合物において、前記被験者がヒトであることを特徴とする混合物。

。

【請求項 2 1】

請求項 1 乃至 2 0 の何れか 1 項に記載の混合物において、各プローブの前記相同プローブ配列が、18 ~ 50 のヌクレオチドであることを特徴とする混合物。

【請求項 2 2】

請求項 1 乃至 2 0 の何れか 1 項に記載の混合物において、各プローブの前記相同プローブ配列が、18 ~ 36 のヌクレオチドであることを特徴とする混合物。

【請求項 2 3】

請求項 1 乃至 2 0 の何れか 1 項に記載の混合物において、各プローブの前記相同プローブ配列が、20 ~ 32 のヌクレオチドであることを特徴とする混合物。

【請求項 2 4】

請求項 2 1 に記載の混合物において、前記第 1 および第 2 の相同プローブ配列が、22 ~ 28 のヌクレオチドであることを特徴とする混合物。

【請求項 2 5】

請求項 1 乃至 2 0 の何れか 1 項に記載の混合物において、各プローブの前記第 1 および第 2 の相同プローブ配列が、50 ~ 65 の T_m を有することを特徴とする混合物。

10

20

30

40

50

【請求項 2 6】

請求項 1 乃至 2 0 の何れか 1 項に記載の混合物において、各プローブの前記第 1 および第 2 の相同プローブ配列が、同一の T_m を有することを特徴とする混合物。

【請求項 2 7】

請求項 1 乃至 2 0 の何れか 1 項に記載の混合物において、前記第 1 の相同プローブ配列が、前記第 2 の相同プローブ配列よりも低い T_m を有することを特徴とする混合物。

【請求項 2 8】

請求項 1 乃至 2 0 の何れか 1 項に記載の混合物において、前記第 1 の相同プローブ配列が、前記第 2 の相同プローブ配列よりも高い T_m を有することを特徴とする混合物。

【請求項 2 9】

請求項 1 に記載の混合物において、試験サンプルから抽出された核酸をさらに含むことを特徴とする混合物。

【請求項 3 0】

請求項 2 9 に記載の混合物において、前記抽出された核酸が、生物サンプルに由来することを特徴とする混合物。

【請求項 3 1】

請求項 3 0 に記載の混合物において、前記生物サンプルが患者に由来することを特徴とする混合物。

【請求項 3 2】

請求項 3 1 に記載の混合物において、前記患者が哺乳類であることを特徴とする混合物。

【請求項 3 3】

請求項 3 2 に記載の混合物において、前記哺乳類がヒトであることを特徴とする混合物。

【請求項 3 4】

請求項 1 に記載の混合物において、少なくとも 1 つのサンプル内部較正標準核酸をさらに含むことを特徴とする混合物。

【請求項 3 5】

請求項 3 4 に記載の混合物において、前記サンプル内部較正標準核酸と特異的にハイブリッド形成する少なくとも 1 つのプローブをさらに含むことを特徴とする混合物。

【請求項 3 6】

請求項 3 4 に記載の混合物において、試験サンプルから抽出された核酸をさらに含むことを特徴とする混合物。

【請求項 3 7】

請求項 1 乃至 3 6 の何れか 1 項に記載の混合物において、前記混合物が、表 4、5、6、8、または 9 のいずれか 1 つからの少なくとも 1 つの相同プローブ配列を含むことを特徴とする混合物。

【請求項 3 8】

請求項 1 に記載の混合物において、前記対象の領域が、少なくとも 2、4、8、10、20、40、60、80、100、125、150、200、250、300、350、400、500、600、700、800、900、1000、1200、1400、1600、1800、または 2000 のヌクレオチドであることを特徴とする混合物。

【請求項 3 9】

請求項 1 乃至 3 8 の何れか 1 項に記載の混合物と、使用説明書とを含むことを特徴とするキット。

【請求項 4 0】

請求項 3 9 に記載のキットにおいて、DNA 抽出のための試薬をさらに含むことを特徴とするキット。

【請求項 4 1】

1 つまたは複数の標的生物の存在を検出する方法において、

10

20

30

40

50

a) 標的生物を含有する疑いがある試験サンプルを、請求項 1 乃至 3 8 の何れか 1 項に記載の混合物と接触させるステップと、

b) 第 1 および第 2 の標的配列とハイブリッド形成して環状化プローブを形成する少なくとも 1 つのプローブによって、対象の領域を捕獲するステップと、

c) 前記捕獲された対象の領域を検出し、それにより、1 つまたは複数の標的生物の存在を検出するステップと

を含むことを特徴とする方法。

【請求項 4 2】

請求項 4 1 に記載の方法において、前記対象の領域が、相同プローブ配列のポリメラーゼ依存性伸長によって捕獲されることを特徴とする方法。

10

【請求項 4 3】

請求項 4 1 に記載の方法において、前記対象の領域が、結合オリゴヌクレオチドの配列特異的連結によって捕獲されることを特徴とする方法。

【請求項 4 4】

請求項 4 1 に記載の方法において、前記環状化プローブを增幅して、前記捕獲された対象の領域を含有する複数の增幅産物を形成するステップをさらに含むことを特徴とする方法。

【請求項 4 5】

請求項 4 1 に記載の方法において、ステップ (b) とステップ (c) との間に、ヌクレアーゼ処理をして線状核酸を除去するステップをさらに含むことを特徴とする方法。

20

【請求項 4 6】

請求項 4 1 に記載の方法において、部位特異的エンドヌクレアーゼによる切断によって前記環状化プローブを線状化するステップをさらに含むことを特徴とする方法。

【請求項 4 7】

請求項 4 1 乃至 4 6 の何れか 1 項に記載の方法において、前記対象の領域を配列決定するステップをさらに含むことを特徴とする方法。

【請求項 4 8】

請求項 4 7 に記載の方法において、前記配列決定が、ジデオキシ配列決定であることを特徴とする方法。

30

【請求項 4 9】

請求項 4 7 に記載の方法において、前記捕獲された対象の領域の配列を、既知のゲノムの配列または既知の突然変異のデータベースと比較するステップをさらに含むことを特徴とする方法。

【請求項 5 0】

請求項 4 7 乃至 4 9 の何れか 1 項に記載の方法において、前記標的生物が、対象の配列を有する任意の標的生物であることを特徴とする方法。

【請求項 5 1】

請求項 5 0 に記載の方法において、前記対象の配列が、遺伝子、座位、1 つまたは複数の单ヌクレオチド多型 (S N P) を含有する配列、および 1 つまたは複数の挿入、欠失、または挿入欠失を含有する配列から選択されることを特徴とする方法。

40

【請求項 5 2】

請求項 4 7 に記載の方法において、既知のゲノムの配列および配列決定エラーのモデルに関して、前記捕獲された対象の領域の配列を分析して、前記サンプル中に存在する種々の生物の割合または存在量を推定するステップをさらに含むことを特徴とする方法。

【請求項 5 3】

請求項 4 1 に記載の方法において、前記環状化プローブが、ハイブリダイゼーションによって検出されることを特徴とする方法。

【請求項 5 4】

請求項 5 3 に記載の方法において、前記ハイブリダイゼーションが、前記環状化プローブと特異的にハイブリッド形成する少なくとも 1 つの特徴を含むマイクロアレイに対する

50

ものであることを特徴とする方法。

【請求項 5 5】

請求項 4 1 に記載の方法において、前記試験サンプルが、哺乳類被験者から得られることを特徴とする方法。

【請求項 5 6】

請求項 5 5 に記載の方法において、前記哺乳類がヒトであることを特徴とする方法。

【請求項 5 7】

請求項 5 5 に記載の方法において、前記試験サンプルがバイオプシーであることを特徴とする方法。

【請求項 5 8】

請求項 4 1 乃至 5 7 の何れか 1 項に記載の方法において、サンプル内部較正標準を前記試験サンプルに添加するステップをさらに含むことを特徴とする方法。

【請求項 5 9】

請求項 5 8 に記載の方法において、前記サンプル内部較正標準と特異的にハイブリッド形成するプローブを添加するステップと、前記サンプル内部較正標準を検出するステップとをさらに含むことを特徴とする方法。

【請求項 6 0】

請求項 4 1 乃至 5 9 の何れか 1 項に記載の方法において、結果をフォーマットして、医師の意思決定を通知するステップをさらに含むことを特徴とする方法。

【請求項 6 1】

請求項 6 0 に記載の方法において、前記フォーマッティングが、対象の生物の推定量を提供することを含むことを特徴とする方法。

【請求項 6 2】

請求項 6 1 に記載の方法において、前記フォーマッティングが、対象の生物の量をカラーコーディングすることを含むことを特徴とする方法。

【請求項 6 3】

請求項 6 0 乃至 6 2 の何れか 1 項に記載の方法において、前記フォーマットされた結果が、前記検出された少なくとも 1 つの標的生物に基づいた治療的勧告を含むことを特徴とする方法。

【請求項 6 4】

請求項 4 1 乃至 5 7 の何れか 1 項に記載の方法を含む、病原体に感染した被験者の治療方法において、少なくとも 1 つの病原体の存在を検出するステップと、前記検出された少なくとも 1 つの病原体に基づいて前記被験者に適切な予防を施すステップとをさらに含むことを特徴とする方法。

【請求項 6 5】

請求項 1 に記載の混合物の製造方法において、

a) 対象の生物のための少なくとも 1 つの基準ゲノム、少なくとも 1 つの非ハイブリッド形成ゲノム、および場合により、前記基準ゲノムと同一でない少なくとも 1 つのハイブリッド形成ゲノムを提供するステップと、

b) 前記基準ゲノムを $n - m e r$ ヘスライスするステップであって、 n が 18 ~ 50 の範囲であるステップと、

c) 前記スライスされた基準ゲノムからスクリーニングされた $n - m e r$ のセットを同定するステップであって、前記スクリーニングされた $n - m e r$ のセットが、

i) 非反復性であり、

i i) 実質的に二次構造を有しない $n - m e r$ からなり、

i i i) 4 つを超える連続した同ースクレオチドを含有する $n - m e r$ を有さず、

i v) 50 ~ 72 の範囲の T_m を有する $n - m e r$ からなるステップと、

d) 相同プローブ配列のセットを同定するステップであって、前記相同プローブ配列が、スクリーニングされた $n - m e r$ からなり、

i) 前記 $n - m e r$ が、いかなる非ハイブリッド形成ゲノムとも特異的にハイブリッ

10

20

30

40

50

ド形成せず、

i i) 前記 n - m e r が、前記基準ゲノムおよび場合により少なくとも 1 つのハイブリッド形成ゲノムにおいて 1 ~ 2 0 回発生するステップと、

e) 第 1 の相同プローブ配列および第 2 の相同プローブ配列を含む複数のプローブを構築するステップであって、

i) 前記第 1 および第 2 の相同プローブ配列がそれぞれ、前記対象の生物のゲノム中の第 1 および第 2 の標的配列と特異的にハイブリッド形成し、前記第 1 および第 2 の標的配列が少なくとも 2 つのスクレオチドを含む対象の領域によって分離されており、

i i) 前記複数のプローブが、互いに特異的にハイブリッド形成せず、そして

i i i) 前記複数のプローブが、実質的に二次構造を有しないステップと
を含むことを特徴とする方法。 10

【請求項 6 6】

請求項 6 5 に記載の方法において、2 つ以上の基準ゲノムが提供され、少なくとも 1 つのプローブが、前記基準ゲノムの少なくとも 1 つとハイブリッド形成することを特徴とする方法。

【請求項 6 7】

請求項 6 6 に記載の方法において、構築された各プローブが、対象の領域のゲノム配列のセット内の既知の配列の対の間で明らかにされた S N P の総数に基づいてスコア化されることを特徴とする方法。 20

【請求項 6 8】

請求項 6 7 に記載の方法において、前記混合物中の前記プローブが、対象の領域のゲノム配列のセット内の既知の配列間に存在する S N P の閾値数に基づいて選択されることを特徴とする方法。

【請求項 6 9】

請求項 6 6 に記載の方法において、構築された各プローブが、前記プローブによって捕獲される対象の標的座位の総数に基づいてスコア化されることを特徴とする方法。

【請求項 7 0】

請求項 6 9 に記載の方法において、前記混合物中のプローブの数が、特定の数の対象の標的座位を捕獲するプローブの閾値数に基づいて選択されることを特徴とする方法。 30

【請求項 7 1】

請求項 6 5 乃至 7 0 の何れか 1 項に記載の方法において、相同プローブ配列が、排除ゲノムのセットに対して指定の長さよりも長い完全な一致を含有しないように各プローブが変更され、前記変更されたプローブが、1 つまたは複数の標的ゲノムとハイブリッド形成した後にまだ環状化し得ることを特徴とする方法。

【請求項 7 2】

請求項 6 5 乃至 7 2 の何れか 1 項に記載の方法において、前記 n - m e r が、少なくとも閾値回数の基準ゲノムセットにおけるその発生に基づいて選択されることを特徴とする方法。

【請求項 7 3】

請求項 6 5 乃至 7 2 の何れか 1 項に記載の方法において、追加の対象の生物のそれぞれの数 m に対して、ステップ (a) ~ (e) を繰り返すことをさらに含むことを特徴とする方法。 40

【請求項 7 4】

請求項 7 3 に記載の方法において、m が 4 、 1 0 、 1 5 、 2 0 、 3 0 、 4 0 、 6 0 、 8 0 、 1 0 0 、 1 5 0 、 2 0 0 、 2 5 0 、 3 0 0 、 4 0 0 、 5 0 0 、 1 0 0 0 、 2 0 0 0 、 4 0 0 0 、 8 0 0 0 、 1 0 0 0 0 、 1 5 0 0 0 、 または 2 0 0 0 0 よりも大きいことを特徴とする方法。

【請求項 7 5】

請求項 6 5 に記載の方法において、前記少なくとも 1 つの非ハイブリッド形成ゲノムが、ヒトゲノムを含むことを特徴とする方法。 50

【請求項 7 6】

請求項 6 5 に記載の方法において、前記少なくとも 1 つの非ハイブリッド形成ゲノムが、前記標的生物以外の所定の配列決定された生物セットを含むことを特徴とする方法。

【請求項 7 7】

請求項 6 5 に記載の方法において、前記少なくとも 1 つのハイブリッド形成ゲノムが、前記基準ゲノムの同じ種、株、亜株、または血清型の配列決定された変異体を含むことを特徴とする方法。

【請求項 7 8】

請求項 6 5 乃至 7 5 の何れか 1 項に記載の方法において、前記ゲノムの n - m e r へのスライシングが、1 と n の間のオフセットを有することを特徴とする方法。

【請求項 7 9】

請求項 6 4 に記載の方法において、n が 1 8 ~ 3 5 のヌクレオチドであることを特徴とする方法。

【請求項 8 0】

請求項 7 9 に記載の方法において、n が 2 0 ~ 3 2 のヌクレオチドであることを特徴とする方法。

【請求項 8 1】

請求項 8 0 に記載の方法において、n が 2 2 ~ 2 8 のヌクレオチドであることを特徴とする方法。

【請求項 8 2】

請求項 6 5 乃至 8 1 の何れか 1 項に記載の方法において、前記方法が、少なくとも 1 0 、 9 、 8 、 7 、 6 、 5 、 4 、 3 、または 2 メガベースの標的ゲノムにおいて、シングルコア Pentium Xeon 2.5 ghz プロセッサを用いて 1 6 、 1 4 、 1 2 、 1 0 、 8 、 6 、もしくは 4 日間、または 7 2 、 4 8 、 3 6 、 2 4 、 1 2 、 1 0 、 8 、 6 、もしくは 4 時間からないことを特徴とする方法。

【請求項 8 3】

1 つまたは複数の対象の生物を検出するための、請求項 1 乃至 3 8 の何れか 1 項に記載の混合物。

【請求項 8 4】

1 つまたは複数の対象の生物を検出するための、請求項 1 乃至 3 8 の何れか 1 項に記載の混合物の使用。

【発明の詳細な説明】**【技術分野】****【0 0 0 1】**

本発明は、病原体を含む対象の生物を多重検出するための核酸プローブのセットと、本プローブの製造および使用方法とに関する。

【0 0 0 2】

配列決定技術の進歩は、塩基配列決定コストにおける急激な低下を促進し続けている。しかしながら、U. S. National Human Genome Research Institute (NHGR) によって提案された \$ 1 , 0 0 0 のパーソナルゲノム基準は、依然として捉えどころのないままである。さらに、患者の完全なゲノムでも、患者の現在の病状（進行中の感染など）への洞察をほとんどあるいは全く提供しない。そして、感染症は、ウイルス、細菌、古細菌、真菌、およびその他の真核生物（単細胞および多細胞の両方）を含む様々な種類の病原体によって発生され得るが、その多くは多大な苦労がなければ培養できないか、あるいは全く培養することができず、適切な臨床的介入の検出および選択が妨げられる。

【0 0 0 3】

患者のミクロビオーム（患者の中および上に存在する全ての微生物の集合（例えば、Frederich MJ, JAMA 300 (7) : 777 - 8 (2008) を参照））は、患者の現在の病状を明らかにすると共に、介護者がその将来の疾患、感染、または臨

10

20

30

40

50

床的合併症のリスクを予測するのに役立つことができる。しかしながら、人体の単一の微小環境においても観察され得る微生物多様性によって明示されるように、ミクロビオームは極めて複雑である。例えば、Hyman et al., PNAS 102 (22) : 7952-7 (2005) (ヒト膣上皮における微生物多様性を研究) を参照されたい。生物検出のために現存するモダリティは、通常、高価で時間のかかる単一病原体アッセイに限定されるので、患者サンプルなどの複雑なサンプル中の生物を検出するのにはあまり適していない。

【0004】

さらに、病原体検出のための核酸プローブを設計するための現存のプラットフォームは、DNAの単一の短い領域（数百または数千の塩基長）を入力として必要とする。従って、これらのプラットフォームは、生物を検出および区別するために16SリボソームDNA領域などの非常に限られたゲノム領域の選択を提供し、従って、最も広い可能な配列の範囲から最適なプライマー候補を同定することができない。加えて、現存の試験は、多くの場合、単一の標的病原体の単一の標的座位だけを調べることに基づいているので、これらの試験は、多くの場合、特定の生物の密接に関連した種または株の変異体（これらの病原性、抗生物質に対する感受性、または毒素の產生（介護者の決定に著しく影響を与える因子）はかなり異なり得る）を区別することができない。

10

【0005】

現存のアッセイが複雑なサンプル混合物中の対象の生物を検出することの難しさと、プライマー設計のための現存のプラットフォームが最も広い可能な配列の範囲から最適なプライマー候補を同定できないこととを考慮すると、培養を必要とせずに複雑な混合物中の複数の生物を検出する迅速な多重アッセイに対する必要性が存在する。

20

【0006】

本発明の実施形態は、培養を必要とすることなく当業者が複雑な混合物中の複数の生物を同時に検出できるようにする最適化核酸プローブ、ならびにその製造および使用方法を含む。本発明は、少なくとも部分的に、全ゲノムなどの大きいクエリー配列のセットから配列を迅速に同定することができるプロセスの発見に基づく。配列は、従来の診断法と比較してアッセイ時間およびコストを著しく削減する多重診断アッセイにおいて使用することができる。本発明の核酸および方法は、当業者が感染性病原体の種を同定し、さらに例えば抗生物質耐性に関連する領域の配列に基づいて密接に関連した株を区別することができるようとする。

30

【0007】

本発明の方法のさらなる利点は、例えば、宿主遺伝子型同定のために、感染性病原体の検出と並行して特異的な宿主座位を調べる能力である。有利に、本発明の方法は、さらに多重化され、集中研究室、病院、および/または診断施設による多数のサンプルのハイスクループ処理のためにマイクロプレートなどの自動システムにおいて使用され得る。さらに、本発明の混合物および方法は、給水、食料品、および農業サンプルのモニタリングなどの様々な種類の付加的な用途において使用することができる。

【0008】

従って、本発明の態様は、対象の領域を環状化捕獲（circularizing capture）することができる複数の核酸プローブを含む混合物を提供する。いくつかの実施形態では、混合物中のプローブはそれぞれ、少なくとも1つの標的生物の中のゲノム中の第1および第2の標的配列とそれぞれ特異的にハイブリッド形成する第1および第2の相同プローブ配列（骨格配列によって分離される）を含む。いくつかの実施形態では、第1および第2の相同プローブ配列は標的配列に対して相補的でないが、標的核酸、例えば、microRNAの5'および3'末端に連結し、リン酸化またはアデニル化5'末端および遊離3'ヒドロキシル基などの核酸連結酵素との適合性のための適切な化学基を有する。いくつかの実施形態では、第1および第2の標的配列は、少なくとも2つのヌクレオチドの対象の領域によって分離される。特定の実施形態では、これらは、少なくとも5、6、7、8、9、10、12、14、18、20、25、30、50、75、10

40

50

0、150、200、300、400、600、1200、1500、2500、またはそれ以上のヌクレオチドによって分離される。いくつかの実施形態では、第1および第2の標的配列は、5、6、7、8、9、10、12、14、18、20、25、30、50、75、100、150、200、300、400、600、1200、1500、または2500以下のヌクレオチドによって分離される。

【0009】

いくつかの実施形態では、混合物中の相同プローブ配列は、そのそれぞれの標的生物のゲノム中の標的配列と特異的にハイブリッド形成するが、所定の配列決定された生物セット（排除セット）のゲノム中のいかなる配列とも特異的にハイブリッド形成しない。捕獲標的と直接ハイブリッド形成しないプローブに関する実施形態では、「相同プローブ配列」は、特に、定義されたゲノムセット、すなわち排除セット内のいかなる配列とも実質的にハイブリッド形成しないように設計される。被験者からの生物サンプルの場合、排除セットは、宿主のゲノムを含む。特定の実施形態では、排除セットは、複数のウイルス、真核生物、原核生物、および古細菌のゲノムも含む。より特定の実施形態では、排除セット中の複数のウイルス、真核生物、原核生物、および古細菌のゲノムは、共生、非毒性（non-virulent）、または非病原性の生物からの配列決定されたゲノムを含んでいてもよい。さらにより特定の実施形態では、混合物中の全てのプローブに対する排除セットは、例えば、宿主ゲノムおよび共生、非毒性、または非病原性の生物を含む、配列決定されたゲノムの共通のサブセットを共有する。一般に、混合物中の各プローブが、混合物中のいかなる他のプローブの標的配列とも特異的にハイブリッド形成しないように、排除セットは混合物のプローブ間で異なる。

10

20

30

30

【0010】

1つの態様では、本発明は、相同プローブ配列をそれぞれが含む複数の核酸プローブを包含し、相同プローブ配列は、実質的に二次構造を有さず、單一ヌクレオチドの長いストリングを含有せず（例えば、7、6、5、4、3、または2未満の連続した同一塩基を有する）、少なくとも約8塩基（例えば、8、10、12、14、16、18、20、22、24、25、26、27、28、30、または32塩基の長さ）であり、そして50～72の範囲のT_m（例えば、約53、54、55、56、57、58、59、60、61、または62）を有する。いくつかの実施形態では、第1および第2の相同プローブ配列はほぼ同じ長さであり、同じT_mを有する。他の実施形態では、第1および第2の相同プローブ配列の長さおよびT_mは異なる。各プローブ中の相同プローブ配列は、標的生物のゲノムにおいて特定の閾値回数未満で発生する（例えば、20、10、5、4、3、または2回未満）ように選択されてもよい。

【0011】

特定のプローブに対する標的生物は、任意の生物でよい。特定の実施形態では、ウイルス、細菌、真菌、古細菌、または真核生物（単細胞および多細胞真核生物を含む）であり得る。特定の実施形態では、標的生物は病原体である。

【0012】

本発明の混合物は、例えば、10、20、30、40、50、100、200、400、500、1000、2000、3000、4000、5000、10000、20000、40000、80000またはそれ以上の多数のプローブを含むことができる。混合物は、多数の異なる標的生物、例えば、少なくとも10、20、40、60、80、100、150、200、250、またはそれ以上の異なる標的生物に向けられた1つまたは複数のプローブを含むことができる。いくつかの実施形態では、複数の標的生物に対する1つまたは複数のプローブを含む混合物は、標的生物に対してただ1つのプローブを含有する。他の実施形態では、混合物は、標的生物に対して2つ以上のプローブ、例えば、標的生物に対して約2、3、4、5、6、7、8、9、または10のプローブを含有する。患者の試験サンプルと共に使用するために設計された実施形態などの特定の実施形態では、混合物はさらに、宿主の遺伝子型同定などの用途のために、宿主ゲノムと特異的にハイブリッド形成する相同プローブ配列を有するプローブを含む。いくつかの実施形態では、

40

50

本発明の混合物はさらにサンプル内部較正標準を含む。

【0013】

本発明によって提供される混合物中のプローブの骨格配列は、検出可能な部分およびプライマー結合配列を含むことができる。いくつかの実施形態では、プローブの骨格配列は、第2のプライマーを含む。特定の実施形態では、検出可能な部分はバーコードである。特定の実施形態では、骨格はさらに、制限エンドヌクレアーゼ認識配列などの切断部位を含む。特定の実施形態では、骨格は、例えば、脱塩基フラン部分などを含む非ワトソン-クリックヌクレオチドを含有する。

【0014】

別の態様では、本発明は、本発明によって提供されるプローブの混合物と、使用説明書とを含むキットを提供する。特定の実施形態では、キットは、サンプルを得るための試薬（例えば、スワブ）、および／またはDNAを抽出するための試薬、および／または対象の領域を捕獲するための酵素（ポリメラーゼおよび／またはリガーゼなど）を含むことができる。

【0015】

別の態様では、本発明は、少なくとも1つの標的生物を含有する疑いがあるサンプルを本発明のプローブの混合物の何れかと接触させ、少なくとも1つの標的生物の対象の領域を捕獲（例えば、重合および／または連結によって）して環状化プローブを形成し、捕獲された対象の領域を検出し、それにより、1つまたは複数の標的生物の存在を検出することによって、1つまたは複数の標的生物の存在を検出するための方法を提供する。特定の実施形態では、捕獲された対象の領域は、複数の増幅産物を形成するために増幅され得る（例えば、PCRによって）。特定の実施形態では、サンプルは、対象の領域のプローブ環状化捕獲の後に線状核酸を除去するために、ヌクレアーゼにより処理される。いくつかの実施形態では、環状化プローブは、例えば、ヌクレアーゼ処理によって線状化される。他の実施形態では、環状化プローブ分子は、当該技術分野において既知の任意の手段によって、増幅されずに直接配列決定される。特定の実施形態では、環状化プローブは、分子のポリメラーゼ媒介性の伸長を刺激して、元の環状プローブの少なくとも1つ～100万またはそれ以上の鎖状体化（concatemerized）コピーを含む、環状化プローブの配列に相補的な配列を生成するオリゴヌクレオチドによって接触される。特定の実施形態では、環状化プローブ分子は、二次捕獲オリゴヌクレオチド捕獲プローブによって、反応溶液から濃縮される。二次捕獲オリゴヌクレオチド捕獲プローブは、ビオチン分子などの捕獲されるように設計された部分と、環状化プローブの少なくとも6つのヌクレオチドとハイブリッド形成するように設計された核酸配列とを含むことができる。環状化プローブの少なくとも6つのヌクレオチドとハイブリッド形成するように設計された核酸配列は、ポリメラーゼ伸長捕獲産物の1、2、4、8、16、32またはそれ以上のヌクレオチドを含み得る。特定の実施形態では、プローブおよび／または捕獲された対象の領域は、ポリメラーゼ依存性配列決定（ジデオキシ配列決定、ピロシーケンス、および合成による配列決定を含む）またはリガーゼベースの配列決定（例えば、ポロニー配列決定）などの、当該技術分野において既知の任意の手段によって配列決定される。特定の実施形態では、サンプルは生物サンプルである。より特定の実施形態では、生物サンプルは、ヒトなどの哺乳類に由来する。

【0016】

いくつかの実施形態では、1つまたは複数の標的生物の存在の検出方法は、例えば、1つまたは複数のグラフ表示を提供することによって医師の意思決定を容易にするために、結果をフォーマットするステップをさらに含む。

【0017】

従って、別の態様では、本発明は、病原体に感染した疑いのある被験者の治療方法を提供し、本方法は、本発明の方法によって少なくとも1つの標的生物（例えば、病原体）を検出することと、検出された少なくとも1つの生物に基づいて適切な治療処置を施すこととを含む。

10

20

30

40

50

【0018】

本発明のさらなる態様は、本発明によって提供されるプローブの混合物の製造方法を提供する。本方法は、基準ゲノムおよび排除ゲノムセットを提供することを含む。基準ゲノムの配列は、約 18 ~ 50 ヌクレオチドの n - m e r ストリングにスライスされる（インシリコで）。スライスされた n - m e r ストリングは、重複（redundant）配列、二次構造を有する配列、反復配列（例えば、4つを超える連続した同一ヌクレオチドを有するストリング）、および所定範囲以外（例えば、50 ~ 72 以外）の T_m を有する配列を排除するようにスクリーニングされる。スクリーニングされた n - m e r は、さらに、ゲノムの排除セット内のゲノム中の配列と特異的にハイブリッド形成する n - m e r（例えば、25 - m e r などの n - m e r においてペアサイズアライメントが 20 のうちの 19 の一致を含有する場合）、あるいは、標的生物のゲノムにおいて指定の回数よりも多く発生する n - m e r を排除することによって、相同プローブ配列を同定するようにスクリーニングされる。特定の実施形態では、相同プローブ配列は、標的生物のゲノムにおいて 1 回だけ発生する。一本鎖ゲノムを有する標的生物の場合、相同プローブ配列は、標的生物のゲノムの補体において 1 回だけ発生し得る。標的生物の配列決定された変異体が利用可能である（例えば、同じ種、属、または血清型）1つの実施形態では、相同プローブ配列は、追加の配列決定された変異体のゲノムと特異的にハイブリッド形成するようにフィルタリングされ、関連生物をグループ化するプローブをもたらす。代替の実施形態では、相同プローブ配列は、配列決定された変異体のゲノム（例えば、配列決定された変異体は排除セットの一部である）と特異的にハイブリッド形成しないようにフィルタリングされ、関連生物を区別するプローブをもたらし得る。これらのフィルタプロセスは、特定の混合物によって検出される各標的生物に対して繰り返される。いくつかの実施形態では、候補相同プローブ配列は、混合物中の他のプローブと特異的にハイブリッド形成し得るものを排除するようにスクリーニングされる。

10

20

30

【0019】

各標的生物に対して、相同プローブ配列は、例えば、特定のサイズの対象の領域を捕獲するように、あるいは特定の実施形態では、所定の対象の領域（薬物耐性、毒性、または毒素産生に関する領域など）を捕獲するように、あるいは被験者の遺伝子型同定の場合には被験者のゲノム中の座位を捕獲するように設計されたプローブに結合される。対象の領域は、指示された人による入力、統計的方法、配列データマイニング、文献データマイニング、またはこれらの組み合わせによって定義され得る。

40

【0020】

本発明のさらなる目的および利点は、一部は以下の記載において説明され、一部は記載から明らかであるか、あるいは本発明の実施によって分かるであろう。本発明の目的および利点は、特許請求の範囲において特に指摘される要素および組み合わせによって実現および達成されるであろう。

40

【0021】

上記の一般的な記載および以下の詳細な記載がいずれも単に例示的および説明的なだけであって、特許請求されるように、本発明を限定するものでないことは理解されるべきである。

【0022】

本明細書に組み込まれてその一部を構成する添付図面は本発明のいくつかの実施形態を示しており、記載と共に本発明の原理を説明するのに役立つ。

【図面の簡単な説明】

【0023】

【図1】図1は、本発明によって提供される1つの例示的なプローブの概略図である。

【図2】図2A、図2B、および図2Cは、対象の領域を捕獲するための本明細書に記載されるようなプローブの3つの代替的な使用方法の略図である。

【図3】図3は、本明細書に記載されるようなプローブを用いる、小さい核酸クローニングのための例示的な戦略を示す。

50

【図4】図4は、PCR增幅のために従来のプライマー対を用いる、本発明の特定の方法の図解である。

【図5】図5は、治療および診断方法を含む、本発明によって提供される方法の例示的なフローチャートを示す。

【図6】図6は、医師の意思決定を通知するようにフォーマットされた、可能性のあるアッセイ結果の説明的な表示である。

【図7】図7は、プローブ設計のための方法の例示的な実施形態のフローチャートである。

【図8】図8は、融解温度(T_m)の関数として、二本鎖型で存在する相同プローブ配列の集団の割合のプロットを示す。

【図9】図9は、特定の融解温度における読み取りカウントによって決定されるように、プローブの効率に対する融解温度の効果を示す。

【図10】図10は、特定の融解温度における読み取りカウントによって決定されるように、プローブの効率に対する融解温度の効果を示す。

【図11】図11は、特に、配列決定結果を処理、分析、および出力するための方法の例示的な実施形態のフローチャートである。

【図12】図12は、分析を実行し、配列決定データをフォーマットするためのシステムアーキテクチャの例示的な実施形態の略図である。

【図13】図13は、部分Aおよび部分Bを含む図13は、配列決定装置からの生FAS

TQデータの処理および基準ゲノムに対する定量化のための例示的なワークフローを示す。

【図14】図14は、次世代配列決定読み取りから得られた配列の例示的なアライメントを示す。

【図15】図15は、サンプル中の株を同定するための、基準株のデータベースに対する配列読み取りアライメントの使用の概略図である。

【図16】図16は、次世代配列決定による正確な多型モデリングおよび検出の方法を示す。

【図17】図17は、346のプローブおよび高リスクHPV株のセット(HPV16、18、31、33、35、39、45、51、52、56、58、59)を用いるHPV株検出のシミュレーションにおいて、どのHPVプローブ(x軸)がどのHPV株(y軸)を検出するかのマトリックスを示す。白色領域は、対応する株を検出するプローブを示す。

【図18】図18は、標的HPV株ゲノムに対する20のHPVプローブ群の標的マトリックスを示す。

【図19】図19は、27の特異的HPVプローブのそれぞれによって同定されたSNPの数およびタイプを示すように拡張された標的マトリックスを示す。

【図20】図20は、PCR増幅されたHPVプローブ環状化捕獲反応のアガロースゲルで分解されたサンプルを示す。

【図21】図21は、環状化捕獲反応産物および既知の細菌ゲノム配列のアライメントを示す。

【図22】図22は、PCR増幅された細菌または細菌遺伝子検出プローブ環状化捕獲反応のアガロースゲルで分解されたサンプルを示す。

【図23】図23は、PCR増幅された環状化プローブの観察されたSanger配列決定読み取りと、ゲノムのスタフィロコッカス・アウレウス(*Staphylococcus aureus*)配列とのアライメントを示す。

【図24】図24は、5つの個々の分子反転プローブを用いてRNAから逆転写されたcDNAの検出、ならびに通常のSanger(N)または次世代配列決定(T、テールド(tailed)プライマー)(198、256、292、293、および462で示されるプローブ)のための増幅を示す。

【図25】図25は、4つの尿路感染症患者サンプルにおいてプローブにより検出された

10

20

30

40

50

異なる感染性種の割合を示す。

【図26】図26は、種々の数の(i)PCRサイクル、(ii)ギャップ充填および連結のための様々な長さの時間、ならびに(iii)様々なハイブリダイゼーション温度を用いて実施される比較的な環状化捕獲プロトコルを示す。

【発明を実施するための形態】

【0024】

1. プローブ

本発明の1つの態様は、複雑なサンプル中の1つまたは複数の生物の高感度で迅速かつ非常に特異的な検出に適した環状化「捕獲」プローブの混合物を提供する。「プローブ」は、骨格配列によって分離された2つの相同プローブ配列を含む線状の非分枝ポリ核酸を指し、ここで、第1の相同プローブ配列は核酸の第1の末端にあり、第2の相同プローブ配列は核酸の第2の末端にあり、プローブは、少なくとも2ヌクレオチドの対象の領域を環状化捕獲することができる。「環状化捕獲」は、対象の領域に相補的な配列を取り込むことによって環状になるプローブを指す。単純な分子反転プローブ(MIP)および関連の捕獲プローブなどの環状化プローブの基本的な設計原理は当該技術分野において知られており、例えば、Nilsson et al., *Science*, 265:2085-88 (1994)、Hardenbol et al., *Genome Res.*, 15:269-75 (2005)、Akbaras et al., *PLOS One*, 9:e915 (2007)、Porecca et al., *Nature Methods*, 4:931-36 (2007)、Deng et al., *Nat. Biotechnol.*, 27(4):353-60 (2009)、米国特許第7,700,323号明細書および同第6,858,412号明細書、ならびに国際公開第1999/049079号パンフレットおよび同第1995/022623号パンフレットに記載されている。

10

20

30

30

【0025】

本発明の特定の態様は2つの相同プローブ配列を含むプローブを包含し、これらはそれぞれ、少なくとも2つのヌクレオチドを含む対象の領域に隣接する標的生物のゲノム中の異なる標的配列と特異的にハイブリッド形成し得る。プローブは、さらに、検出可能な部分およびプライマーを含有する骨格配列を、相同プローブ配列の間に含むことができる。通常、プローブの3'端部の相同プローブ配列は、H1(または伸長アーム)と呼ばれ、プローブの5'端部の相同プローブ配列は、H2(連結またはアンカーアーム)と呼ばれる。対象のゲノム中の標的部位とのハイブリダイゼーションの際、プローブ/標的二本鎖は、プローブにおける少なくとも2つのヌクレオチドのポリメラーゼ依存性取込み(伸長アームにおける)、および/またはプローブのリガーゼ依存性の環状化(ポリメラーゼ伸長プローブの環状化、または対象の領域に架かる連結ポリヌクレオチドの配列依存性連結のいずれかによる)のために適切な基質である。

【0026】

「捕獲反応」は、試験サンプルと接触される1つまたは複数のプローブが対象の領域の環状化捕獲を受けたプロセスを指し、プローブ中の第1および第2の相同プローブ配列は、プローブの第1の標的配列と第2の標的配列との間の対象の領域を捕獲するために、試験サンプル中のそのそれぞれの標的配列と特異的にハイブリッド形成されている。「捕獲反応産物」は、試験サンプルによる捕獲反応を完了することによって產生される核酸の混合物を指す。「増幅反応」は、捕獲反応産物を増幅するプロセスを指す。「増幅反応産物」は、捕獲反応産物による増幅反応を完了することによって產生される核酸の混合物を指す。

40

【0027】

いくつかの実施形態では、第1および第2の相同プローブ配列は、標的配列に対して相補的でないが、標的核酸、例えば、小さいRNAおよびmicroRNAの5'および3'末端に連結し、リン酸化またはアデニル化5'末端および遊離3'ヒドロキシル基などの核酸連結酵素との適合性のための適切な化学基を有する。小さい核酸クローニングのための例示的な戦略は図3に示される。いくつかの実施形態では、アデニル化5'端部およ

50

び遊離 3' - OH を有するプローブは、一段階で、適合性の連結端部を含有する小さい RNA 断片とほぼ同時に連結する（図 3 (i)）。さらなる実施形態では、プローブは、小さい標的核酸を二段階プロセスで捕獲することができ、アデニル化 5' 端部およびプロックされた 3' 端部（例えば、ジデオキシヌクレオチドプロック端部）を有するプローブが、標的の小さい RNA と連結され得る（図 3 (ii)、(iii) の 2 つのプローブ図のうちの最初の図）。これは、誘導されるリボヌクレアーゼ H 2 消化によって最初にプローブ内の RNA 塩基を除去し、続いて、ここでは 3' - OH - 末端であるプローブを小さい RNA にほぼ同時に連結させることによって起こり得る。代替の二段階プロセスでは、プローブは 5' - アデニル化プローブ部位に連結され、次にプローブのプロックされた 3' 端部がリボヌクレアーゼ H 2 によって消化されて、連結のための遊離 3' - OH が生成され得る（図 3 (ii)、(iii) の 2 つのプローブ図のうちの 2 番目の図）。

10

【0028】

1.1 相同プローブ配列

「相同プローブ配列」は、対象の生物のゲノム中に存在する標的配列と特異的にハイブリッド形成する本発明によって提供されるプローブの一部である。「相同プローブ配列」、「プローブアーム」、「ホーマー（homomer）」および「プローブ相同性領域」という用語はそれぞれ、標的ゲノム配列と特異的にハイブリッド形成し得る相同プローブ配列を指し、本明細書では交換可能に使用される。「標的配列」は、対象の生物のゲノム中の核酸の一本鎖における核酸配列を指す。いくつかの実施形態では、プローブ中の相同プローブ配列は、それぞれ、少なくとも 8、10、12、14、16、18、20、22、24、26、28、30、32、34、36、38、40、45、50、55、60、65、70、80、90、100、110、120、またはそれ以上のヌクレオチドの長さである。特定の実施形態では、相同プローブ配列は、18～50、18～36、20～32、または 22～28 ヌクレオチドの長さである。より特定の実施形態では、相同プローブ配列は 22～28 ヌクレオチドの長さである。特定の実施形態では、プローブ中の 2 つの相同プローブ配列は同じ長さであり、他の実施形態では、これらは異なる長さである。特定の実施形態では、プローブの相同プローブ配列は長さが異なるが、10、9、8、7、6、5、4、3、または 2 ヌクレオチド未満だけ異なる。

20

【0029】

いくつかの実施形態では、相同プローブ配列は、長く続く連結した同一ヌクレオチドを含有しない。いくつかの実施形態では、相同プローブ配列は、10、9、8、7、6、5、4、または 3 つ未満の連結した同一ヌクレオチドを含有する。より特定の実施形態では、これらは 6 つ未満の連結した同一ヌクレオチドを含有し、より特定の実施形態では、これらは 4 つ未満の連結した同一ヌクレオチドを含有する。

30

【0030】

相同プローブ配列は、ヘアピンなどの二次構造を実質的に有しないこともある。n が 7 であるときに、相同プローブ配列の逆補体の n - m e r が少なくとも 5 塩基離れた相同プローブ配列中の n - m e r に対して完全に相補的でない場合に、相同プローブ配列は「二次構造を実質的に有しない」。いくつかの実施形態では、n は、15、14、13、12、11、10、9、8、6、5、4、または 3 である。特定の実施形態では、n は 3～7 である。いくつかの実施形態では、配列の融解温度 (T_m) で 0.25 μM の濃度において、50 mM の Na⁺ を含有し Mg⁺⁺ を含有しない水溶液中の分子の 30% 未満が安定な分子内ヘアピンまたは分子間二量体である（ここで、溶液は他の配列を含まない）場合に、配列、例えば、相同プローブ配列、骨格配列、またはプローブは、二次構造を実質的に有しない。いくつかの実施形態では、配列の T_m よりも 15、10、8、6、4、または 2 低い温度で 0.25 μM の DNA 濃度において、50 mM の Na⁺ を含有し Mg⁺⁺ を含有しないときに分子の 30% 未満が安定な分子内ヘアピンまたは分子間二量体である（ここで、溶液は他の配列を含まない）場合に、配列は二次構造を実質的に有しない。いくつかの実施形態では、0.5 mM の Mg⁺⁺ の存在下、配列の T_m よりも 15、0、8、6、4、または 2 低い温度で 0.25 μM の DNA 濃度において、50 mM の Na⁺ を含有し Mg⁺⁺ を含有しないときに分子の 30% 未満が安定な分子内ヘアピンまたは分子間二量体である（ここで、溶液は他の配列を含まない）場合に、配列は二次構造を実質的に有しない。

40

【0031】

特定の実施形態では、配列の T_m よりも 15、0、8、6、4、または 2 低い温度で 0.25 μM の DNA 濃度において、50 mM の Na⁺ を含有し Mg⁺⁺ を含有しないときに分子の 30% 未満が安定な分子内ヘアピンまたは分子間二量体である（ここで、溶液は他の配列を含まない）場合に、配列は二次構造を実質的に有しない。

50

⁺ および 0.5 mM の Mg⁺⁺ を含有するときに分子の 30% 未満が安定な分子内ヘアピンまたは分子間二量体である場合に、配列は二次構造を実質的に有しない。二次構造を検出する他の方法は当該技術分野において知られており、本発明において使用することができ、例えば、Zuker, Nucleic Acids Res., 31: 3406-15 (2003)、Mathews et al., J. Mol. Biol., 288: 911-940 (1999)、Hilbers et al., Anal. Chem. 327: 70 (1987)、Serra et al., Nucleic Acids Res., 21: 3845-3849 (1993)、および Vallone et al., Biopolymers., 50: 425-442 (1999) に記載されている。

【0031】

10

いくつかの実施形態では、相同プローブ配列は、0.5 mM の Mg⁺⁺ の存在下、50 ~ 72、例えば、約 50、52、54、56、57、58、59、60、61、62、63、64、65、66、67、68、69、70、71、または 72 の融解温度 (T_m) を有するように設計される。特定の実施形態では、 T_m は、0.5 mM の Mg⁺⁺ の存在下、50 ~ 65 である。いくつかの実施形態では、 T_m は、Mg⁺⁺ の非存在下、38 ~ 72 である。特定の実施形態では、プローブ中の相同プローブ配列はほぼ同じ T_m を有し、他の実施形態では、これらは異なる T_m を有するが、互いに 10、9、8、7、6、5、4、3、2、または 1 の違いである。特定の実施形態では、第 1 の相同プローブ配列（すなわち、プローブ中の最も 5' 側）は第 2 の相同プローブ配列よりも低い T_m を有し、他の実施形態では、第 2 の相同プローブ配列よりも高い T_m を有する。

20

【0032】

「融解温度」（「 T_m 」）は、溶液中の DNA 分子の 50% がその相補的配列との二本鎖としてハイブリッド形成され、半分が解離されている温度を指す。他に記載されない限り、 T_m は、Mg⁺⁺ を含有せずに、0.25 μM の DNA 濃度および 50 mM のナトリウム濃度で決定される。 T_m は、経験的な測定または推定を含む当業者に知られている様々な方法によって決定され得る。特定の実施形態では、 T_m は、配列中の G および C ヌクレオチドの数または割合をカウントすることによって推定される。特定の実施形態では、相同プローブ配列中の G および C ヌクレオチドの数は、配列中のヌクレオチドの 30 ~ 60% の間、例えば約 30、35、40、45、50、または 55% などである。より特定の実施形態では、相同プローブ配列中の G および C ヌクレオチドの数は、相同プローブ配列中のヌクレオチドの 38 ~ 44% である。

30

【0033】

特定の実施形態では、隣接するヌクレオチド間の塩基スタッキングの説明となる T_m の最近傍推定値が使用される。最近傍計算は、例えば、Breslauer et al., PNAS, 83: 3746 - 3750 (1986) に記載されており、Santaluca, PNAS, 95 (4): 1460 - 65 (1998) において概説されており（いくつかの経験的な最近傍研究が概説され、特に、表 2 において DNA / DNA 二本鎖のための H および S マスター テーブルが提供される）、これは、参照によって本明細書中に援用される。

40

【0034】

相同プローブ配列は、標的生物のゲノム中の標的配列と特異的にハイブリッド形成するように設計されてもよい。「ハイブリッド形成する」という用語は、ワトソン - クリック 塩基対形成（A と T または U、および G と C）による核酸間の配列特異的相互作用を指す。「特異的にハイブリッド形成する」は、核酸が標的配列に対する完全補体の T_m よりも 8 以下だけ低い T_m で標的配列とハイブリッド形成することを意味する。特定の実施形態では、配列は、標的配列に対する完全補体の T_m よりも 7、6、5、4、3、2、または 1 以下だけ低い T_m で標的配列と特異的にハイブリッド形成する。いくつかの実施形態では、配列は、標的配列に対する完全補体である場合に、標的配列と特異的にハイブリッド形成する。他の実施形態では、配列は、標的配列の完全補体と約 99、98、97、96、95、94、93、92、91、90、85、80、75、70、または 65 % 同

50

ーである場合に、標的配列と特異的にハイブリッド形成する。いくつかの実施形態では、相同プローブ配列は標的配列と特異的にハイブリッド形成するが、ミスマッチ、例えば、約 18、20、22、24、25、26、28、30、35、40、または 45 の連続塩基のウインドウ内に約 1、2、3、4、5、またはそれ以上のミスマッチを含有する。

【0035】

特定の実施形態では、プローブは、標的ゲノムの DNA または RNA 成分に付加されているか、あるいは DNA または RNA 成分に相補的な配列に付加されている核酸配列とハイブリッド形成し得る。このような付加された核酸配列には、例えば、連結により付加されたオリゴヌクレオチドアダプター、あるいはポリメラーゼまたはヌクレオチド末端トランスフェラーゼ活性によって生成されるポリヌクレオチドラン（例えば、「AAAAAA」または「CCCCCC」）が含まれる。

10

【0036】

さらに特定の実施形態では、ブリッジ核酸が使用されてもよく、ここで、ブリッジ核酸の少なくとも第 1 の部分は捕獲プローブとハイブリッド形成することができ、ブリッジ核酸の少なくとも第 2 の部分（第 1 の部分と重複してもよい）は、同時または順次、標的核酸とハイブリッド形成することができ、これにより、捕獲プローブの標的への連結の効率が高められる。

【0037】

特定の実施形態では、a) プローブ中の両方の相同プローブ配列が、そのそれぞれの標的配列と、相同プローブ配列の全長にわたって少なくとも 60、65、70、75、80、85、90、95、または 100 % 正しい対形成でハイブリッド形成する場合、b) 第 1 の相同プローブ配列が、H1 の 3' 端部（第 2 の相同プローブ配列の最も 3' 側）の 8、7、6、5、4、3、または 2 つ塩基において 100 % 正しい対形成でハイブリッド形成する場合、そして c) 第 2 の相同プローブ配列が、H2 の 5' 端部（相同プローブ配列の最も 5' 側）の最初の 8、7、6、5、4、3、または 2 つの塩基とハイブリッド形成する場合に、プローブは特異的にハイブリッド形成する。さらにより特定の実施形態では、a) プローブ中の両方の相同プローブ配列が、そのそれぞれの標的配列と、相同プローブ配列の全長にわたって少なくとも 80 % 正しい対形成でハイブリッド形成する場合、b) 第 1 の相同プローブ配列が、H1 の 3' 端部の最初の 6 つの塩基と 100 % 正しい対形成でハイブリッド形成する場合、そして c) 第 2 の相同プローブ配列が、H2 の 5' 端部の最初の 6 つの塩基と 100 % 正しい対形成でハイブリッド形成する場合に、プローブは特異的にハイブリッド形成する。

20

【0038】

2 つの配列（例えば、相同プローブ配列および標的配列の補体）間の相同性は、ペアワイヤライメント、ドット - マトリックス、およびダイナミックプログラミングを含む当該技術分野において既知の任意の方法によって決定することができ、特定の実施形態では、FASTA (Lipman and Pearson, Science, 227:1435-41 (1985) および Lipman and Pearson, PNAS, 85:2444-48 (1998))、BLAST (McGinnis & Madden, Nucleic Acids Res., 32:W20-W25 (2004) (現在の BLAST 参考文献、特に Megablast について記載)、Zhang et al., J. Comput. Biol., 7(1-2):203-14 (2000) (Megablast において実行される「欲張リアルゴリズム (greedy algorithm)」について記載)、Altschul et al., J. Mol. Biol., 215:403-410 (1990) (最初の BLAST 刊行物))、Needleman - Wunsch (Needleman and Wunsch, J. Molec. Biol., 48(3):443-53 (1970))、Sellers (Sellers, Bull. Math. Biol., 46:501-14 (1984))、および Smith - Waterman (Smith and Waterman, J. Molec. Biol., 147:195-197 (1981))、ならびに他のアルゴリズム (Gerhard

30

40

50

et al., Genome Res., 14 (10B): 2121-27 (2004) に記載されるものを含む) によって決定することができ、これらは参考によって本明細書中に援用される。特定の実施形態では、本発明によって提供される方法は、1つまたは複数の注釈付きのゲノムに対する Megablast により配列の候補セットをスクリーニングすることを含む。

【0039】

いくつかの実施形態では、配列は、ストリンフェントなハイブリダイゼーション条件下で標的配列とハイブリッド形成する場合に「特異的にハイブリッド形成する」。「ストリンフェントなハイブリダイゼーション条件」は、65において $6 \times SSC$ および 1% SDS 中で核酸をハイブリッド形成し、まず約 42 において $0.1 \times SSC$ 中約 20% (v/v) のホルムアミドで 10 分間洗浄し、続いて 65 において $0.2 \times SSC$ および 0.1% SDS で洗浄することを指す。特定の実施形態では、代替のハイブリダイゼーション条件は、約 55、56、57、58、59、60、61、62、63、64、66、67、68、69、または 70 の異なるハイブリダイゼーションおよび/または洗浄温度、あるいは参考によって本明細書中に Sambrook and Russell, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, 3rd edition (2001) に開示されるような他のハイブリダイゼーション条件を含むことができる。特定の実施形態では、ハイブリダイゼーション温度は 60 よりも高く、例えば、60~65 である。

10

20

30

【0040】

相同プローブ配列は、特定の生物のゲノム中、あるいは特定の実施形態では、密接に関連した生物の群のゲノム中の標的配列と特異的にハイブリッド形成するように選択され得る。従って、いくつかの実施形態では、相同プローブ配列は、配列決定されたゲノムの排除セット中に含有される配列と特異的にハイブリッド形成しない。「排除セット」は、相同プローブ配列が特異的にハイブリッド形成しない所定の配列決定されたゲノムのセットを指す。捕獲標的に直接ハイブリッド形成しないプローブを包含する実施形態では、相同プローブ配列は、特に、排除セット内のいかなる配列とも実質的にハイブリッド形成しないように設計される。いくつかの実施形態では、相同プローブ配列は、排除セット内の配列に対して、約 15、20、21、22、23、24、25、26、27、28、29、30、35、または 40 の連続塩基のウインドウ内に少なくとも 1、2、3、4、5、6、7、8、9、または 10 のミスマッチを含有する。より特定の実施形態では、プローブ中の相同プローブ配列はそれぞれ、排除セット中のいかなる配列に対しても、20 塩基中に少なくとも 1 つのミスマッチを有する。

30

【0041】

「生物」は、ウイルス、細菌、古細菌、および真核生物（植物界、真菌、原生生物、および動物を含む）を含む、ゲノムを有する任意の生物である。

【0042】

「配列決定された生物」は、他の生物と区別することができるようそのゲノムの十分な部分が配列決定されている生物である。「配列決定されたゲノム」または「配列決定された生物のゲノム」は、配列決定された生物のゲノムのヌクレオチド配列である。いくつかの実施形態では、配列決定された生物は、完全または部分的に配列決定されている（例えば、ショットガンまたは cDNA 配列決定、ライプラリ配列決定、BAC または YAC 配列決定によって）。特定の実施形態では、生物のゲノムは、少なくとも 5、10、20、30、40、50、60、70、80、90、95、または 99% が配列決定されている。配列決定されたゲノムは、様々なレベルの包含範囲、例えば約 0.1、0.5、0.8、1、2、3、4、5、10、20×、またはそれ以上の包含範囲で配列決定されてもよい。いくつかの実施形態では、病原体などの対象の生物のゲノムサイズは、少なくとも 1、5、10、50、100、200、300、400、500、600、700、800、900、1000、2000、5000、10000、20000、50000、100000、200000、500000 である。

40

50

、100000万塩基またはそれ以上であり得る。特定の実施形態では、標的ゲノムは、少なくとも1~1000万塩基である。

【0043】

特定の実施形態では、排除セットは、試験サンプルが得られる被験生物のゲノムを含む。特定の実施形態では、排除セットはヒトゲノムを含む。より特定の実施形態では、排除セットはさらに、共通のヒトミクロフローラまたは共生生物のゲノムを含む。さらにより特定の実施形態では、排除セットはさらに、混合物中の他のプローブの標的生物のゲノム、例えば、パネルを含む（例えば、任意の所与の標的生物に対して混合物中のただ1つのプローブが特異的にハイブリッド形成するように）。いくつかの実施形態では、排除セットは、複数のウイルス、真核生物、原核生物、および古細菌のゲノムを含んでいてもよい。より特定の実施形態では、排除セット中の複数のウイルス、真核生物、原核生物、および古細菌のゲノムはさらに、共生生物、非毒性生物、または非病原性生物から配列決定されたゲノムを含んでいてもよい。さらにより特定の実施形態では、排除セットはさらに、配列決定された病原体を含む標的生物以外の生物の配列決定されたゲノムを含む。いくつかの実施形態では、混合物中の全てのプローブに対する排除セットは、例えば、宿主ゲノムおよび共生、非毒性、または非病原性生物を含む、配列決定されたゲノムに共通のサブセットを共有する。さらなる実施形態では、混合物中の各プローブが混合物中の任意の他のプローブの標的領域または相同プローブ配列のいずれかと特異的にハイブリッド形成しないように、排除セットは混合物中のプローブ間で異なる。

10

【0044】

本発明によって提供されるプローブは、対象の生物のゲノム中の第1および第2の標的配列と特異的にハイブリッド形成する第1および第2の相同プローブ配列を含むことができる。第1および第2の標的配列は、少なくとも2、3、4、5、6、7、8、9、10、15、20、30、40、50、60、80、100、125、150、200、250、300、350、400、500、600、700、800、900、1000、1200、1400、1600、1800、または2000スクレオチドを含む対象の領域によって分離される。「対象の領域」は、プローブ中の相同プローブ配列の2つの標的配列の最も近い末端間の配列を指す。特定の実施形態では、特定の標的領域は、人による入力またはコンピュータによるデータマイニング（統計的配列および/または文献データマイニングを含む）に基づいて選択され得る。ある特定の実施形態では、1つまたは複数の対象の領域は、密接に関連した生物の間（例えば、同一属の種の間、同一種の亜種の間、または同一種もしくは同一亜種の株の間）で多形性である。より特定の実施形態では、多型は、薬物耐性、毒素産生、または他の毒性因子に関連する。さらにより特定の実施形態では、対象の領域は、例えば、Arnold, Methods Mol Biol., 642:217-23(2010) (M.ツベルクリンス (M. tuberculosis) の多剤耐性 (MDR) 株においてリファンピシン感受性に関連するRNAポリメラーゼB遺伝子について考察)、Kurt et al., J. Clin Microbiol., 47:577-85(2009) (メチシリン耐性に関連するS.アウレウス (S. aureus) の遺伝子型同定領域)、Akhras et al., PLOS ONE, 2(9):e915(2007) (シプロフロキサシンへの耐性に関連するN.ゴノリニア (N. gonorrhoeae) からの領域について記載)、およびPourmand et al., PLoS One., 1(1):e95.(2006) (赤血球凝集素遺伝子における切断部位、グリコシリ化部位、ノイラミニダーゼにおけるオセルタミビル耐性部位を同定するH5N1ウイルスのための迅速アッセイについて記載)に開示されるもののうちの1つまたは複数を含む。

20

30

【0045】

本発明によって提供されるプローブ中の第1および第2の相同プローブ配列は、ポリメラーゼ連鎖反応 (PCR) において使用して、対象の生物から対象の領域を特異的に增幅するための一対の従来のプライマー対として使用するために容易に適合させることができる。「従来のプライマー対」は一対の線状核酸プライマーを指し、その各メンバーは、本

40

50

発明によって提供されるプローブ中の2つの相同プローブ配列のうちの1つに対応する配列を含み、少なくとも2つのヌクレオチドを含む対象の領域を指数関数的に増幅することができる。これらの従来のプライマー対は本発明によって包含され、本発明の一部である。従って、本発明によって提供される従来のプライマー対は、相同プローブ配列に対して上記で提供された同じ基準（例えば、長さ、T_m、ハイブリダイゼーション特異性、および介在する対象の領域の長さを含む）によって特徴付けられる。対象の領域に相補的な配列を環状化捕獲することができる本発明によって提供されるプローブとは対照的に、従来のプライマー対は、互いに向かい合うその3'端部によって方向付けられて、指数関数的な増幅を容易にする。図4は、従来のプライマー対を用いる本発明の特定の方法の図解である。特定の実施形態では、従来のプライマー対は、バーコード配列を含む。いくつかの実施形態では、従来のプライマー対は、例えば、アダプタマー（adaptor）プライマーとハイブリッド形成する配列を含む普遍的配列を含む。

10

【0046】

本発明によって提供されるプローブおよび従来のプライマー対は、天然に存在する従来のヌクレオチドA、C、G、T、およびU（デオキシリボース型および/またはリボース型）と、2' O-メチル修飾ヌクレオチド（Dunlap et al., Biochemistry. 10 (13) : 2581 - 7 (1971)）、IsodCもしくはIsodGなどの人工的な塩基対、または標準的なワトソン-クリック水素結合を形成しない脱塩基フラン（Spacerなど）（Chakravorty, et al. Methods Mol Biol. 634 : 175 - 85 (2010)）、ビオチン化ヌクレオチド、アデニル化ヌクレオチド、ブロッキング基（光切断可能なブロッキング基を含む）を含むヌクレオチド、およびロックド核酸（LNA、ポリ核酸において増強された塩基スタッキング相互作用を提供する修飾リボヌクレオチド、例えば、Levin et al. Nucleic Acid Res. 34 (20) : 142 (2006)を参照されたい）などの修飾ヌクレオチドと、ペプチド核酸骨格とを含むことができる。特定の実施形態では、本発明によって提供されるプローブの5'または3'相同プローブ配列は、そのそれぞれの末端にPC-ビオチンなどの光切断可能なブロッキング基を含む。より特定の実施形態では、本発明によって提供されるプローブは、その5'末端に、連結を光活性化までブロックするための光切断可能なブロッキング基を含む。他の特定の実施形態では、本発明によって提供されるプローブは、その3'末端に、ポリメラーゼ依存性伸長またはn-merオリゴヌクレオチド連結を光活性化までブロックするための光切断可能なブロッキング基を含む。

20

30

【0047】

他の実施形態では、本発明によって提供されるプローブの最も5'側のヌクレオチドは、連結および/またはハイブリダイゼーション効率を改善するために、アデニル化ヌクレオチドを含む。他の実施形態では、相同プローブ領域は、ハイブリダイゼーションおよび/または連結効率を改善するために、あるいはポリメラーゼ媒介性のストランド置換またはヌクレアーゼ開裂などの酵素活性への耐性を提供するために、1つまたは複数の2' Oメチル、IsodCもしくはIsodGなどの人工的な塩基対、または脱塩基フラン（Spacerなど）、あるいは2' Oメチル、脱塩基フラン、またはLNAヌクレオチド、例えば、1、2、3、4、5、6、7、8、9、10、またはそれ以上のLNA、あるいは10、20、30、40、50、60、70、80、90、または100%の2' Oメチル、脱塩基フラン、またはLNAヌクレオチドを含む。例えば、Hogrefe et al., J. Biol. Chem. 265 (10) : 5561 - 5566 (1990)を参照されたい。より特定の実施形態では、5'相同プローブ領域の5'端部（例えば、H2、連結アーム）は少なくとも1つのLNAを含み、さらにより特定の実施形態では、5'末端ヌクレオチドはLNAである。

40

【0048】

1.2 骨格配列

本発明によって提供されるプローブは、第1の相同プローブ配列と第2の相同プローブ

50

配列との間に、検出可能な部分および1つまたは複数のプライマー結合配列を含み得るプローブ骨格配列を含む。骨格配列は、少なくとも15、20、25、30、35、40、45、50、70、90、100、12、140、150、160、180、200、400塩基、またはそれ以上の塩基であり得る。より特定の実施形態では、骨格は第2のプライマーを含む。各骨格プライマーは、例えば、混合物中の全ての環状化プローブを増幅するために使用することができる1つまたは複数の普遍的配列を含むことができる。いくつかの実施形態では、プライマーは、特異的プローブまたはプローブのセットの同定および/または増幅のためのバーコードなどのプローブ特異的配列を含有してもよい。いくつかの実施形態では、骨格配列は、1つまたは複数の非ワトソン-クリックヌクレオチドを含む。さらなる実施形態では、骨格は、ハイブリダイゼーション反応においてより大きい反応性または不活性を付与するため、ポリメラーゼ媒介のストランド置換またはヌクレアーゼ開裂などの酵素活性への耐性を提供するため、誤った増幅事象の阻害剤としての機能を果たすため、あるいはPCRプライマーまたはビオチン化捕獲プローブなどのトランス作用性核酸オリゴヌクレオチドのための標的部位の役割を果たすために、1つまたは複数の2O'メチルヌクレオチド残基、IsodCもしくはIsodGなどの人工的な塩基対、または脱塩基フラン(dSpacerなど)、あるいは2O'メチル、脱塩基フラン、またはLNAヌクレオチド、例えば、1、2、3、4、5、6、7、8、9、10、またはそれ以上のLNA、あるいは10、20、30、40、50、60、70、80、90、または100%の2O'メチル、脱塩基フラン、またはLNAヌクレオチドを含む。

【0049】

「バーコード」という用語は、分子または関連分子の種類を独自に同定するヌクレオチド配列を指すために使用される。本発明のプローブにおいて使用するための適切なバーコード配列は、例えば、米国特許第5,445,934号明細書(Fodor et al.)および米国特許第5,635,400号明細書(Brenner)に記載されるn-merのアレイなどの、カスタマイズまたは既製された核酸アレイに対応する配列を含むことができる。特定の実施形態では、n-merバーコードは、少なくとも3、4、5、6、7、8、9、10、12、15、16、17、18、19、20、21、22、23、24、25、30、35、40、50、60、70、80、90、100、200、300、400または500ヌクレオチド、例えば、18~20、21、22、23、24、または25ヌクレオチドであり得る。特定の実施形態では、バーコードは、このバーコードが不注意に誤った別のものとして読み取られるのを可能にするために、1、2、3、4または5よりも多い配列決定エラーを要求するように設計された配列を含む。

【0050】

バーコード配列を生成するために、各バーコードサイズKに対して、パールスクリプトを用いて4つのDNAヌクレオチド、A、T、G、Cから 4^k ランダムバーコードが生成され得る。このバーコードのセットは、4つのヌクレオチドの変動を用いてKの長さの配列に対して可能な独特の配列の組み合わせの総数を表す。次に、1つのヌクレオチドが100%の長さを含む(例えば、TTTTTTT)場合のバーコードは、場合により、パターンマッチングパールスクリプトを用いて除去される。さらに、フィルタリングステップは、3よりも多いヌクレオチドのラン(run)、例えばTGGGGT、またはただ1つのヌクレオチドによって中断されるラン、例えばGGGTGGを含有するバーコードの除去を含むことができる。セルフハイブリダイゼーションにより二次構造を形成する傾向のあるパリンドロームまたは逆方向反復を含有するバーコードは、このような自己相補性を同定するように設計されたパールスクリプトを用いてフィルタリングされ得る。

【0051】

患者からのサンプルを試験するために使用されるプローブの混合物中で利用され得るバーコードの選択は、プール内のバーコード配列中の各位置における5%よりも多く、50%以下の特定のヌクレオチドの表示を提供し得るバーコードの組み合わせを選択することを含み得る。これは、パールスクリプトを用いて指定された条件が満たされるまで、プールされたセットにバーコードをランダムに付加および除去することによって達成される。

バーコードプール内に逆補体配列も存在する場合のバーコードも排除され得る。

【0 0 5 2】

適切なバーコード配列は、例示的な 3 - m e r、4 - m e r、5 - m e r、6 - m e r、7 - m e r、8 - m e r、9 - m e r、および 10 - m e r バーコードを説明する配列表 1 に記載されるようなバーコード配列を含む。表 1 において「1 ヌクレオチド距離」n - m e r として示される配列は、互いに少なくとも 1 の配列距離を有する、説明的な配列であり、ここで、「距離」は、同じカテゴリーの配列のそれぞれの間の配列決定の相違の最小数を指す。「2 ヌクレオチド距離」配列は、互いに少なくとも 2 ヌクレオチドの「距離」を有する。

表 1：例示的なバーコード配列

3-mer バーコード -1 ヌクレオチド距離 aaa 配列番号 (以下に付加) aac aag aat aca acc	
3-mer バーコード -2 ヌクレオチド距離 acg aga atc cag ccc cgt	10
4-mer バーコード -1 ヌクレオチド距離 aaaa aaac aaag aaat aaca aacc	20
4-mer バーコード -2 ヌクレオチド距離 aagg aatt acat accg acgc acta	
5-mer バーコード -1 ヌクレオチド距離 aaaaa aaaac aaaag aaaat aaaca aaacc	30
6-mer バーコード -1 ヌクレオチド距離 aaaaaa aaaaag aaaaat aaaaca aaaact aaaaga	40

7-mer バーコード -1 ヌクレオチド距離
 aaaaaaa
 aaaaaac
 aaaaaag
 aaaaaat
 aaaaacg
 aaaaagc

8-mer バーコード -1 ヌクレオチド距離
 aaaaaaaaa
 aaaaaaat
 aaaaaaga
 aaaaaatg
 aaaaagcg
 aaaaatct

9-mer バーコード -1 ヌクレオチド
 aaaaaaaaaa
 aaaaaaaaaac
 aaaaacggg
 aaaaagagg
 aaaaaggac
 aaaaattgc

10-mer バーコード -1 ヌクレオチド距離
 aaaaaactgg (配列番号 1)
 aaaaaagcat (配列番号 2)
 aaaaaatatc (配列番号 3)
 aaaaacactc (配列番号 4)
 aaaaacttg (配列番号 5)
 aaaaagggtt (配列番号 6)

10

20

30

40

50

【 0 0 5 3 】

特定の実施形態では、本発明によって提供されるプローブにおいて使用されるバーコードは、AFFYMETRIX (商標)によるTag3またはTag4バーコードアレイ上のものに相当する。バーコードシステムのさらなる考察は、Frank, BMC Bioinformatics, 10:362 (2009, 13頁)、Pierce et al., Nature Methods, 3:601-03 (2006) (ウェブ補足を含む)、およびPierce et al., Nature Protocols, 2:2958-74 (2007)において見出すことができる。

【 0 0 5 4 】

いくつかの実施形態では、骨格は、1つまたは複数のサンプル核酸特異的バーコード、例えば、1つまたは複数の患者特異的バーコードを含む。特定の実施形態では、1つの患者サンプルにつき2つ以上のバーコードが割り当てられ、各患者の複製サンプルが同じ配列決定反応において実施されることが可能になる。サンプル核酸特異的バーコードを用いることによって、本出願において記載されるような多重反応と、特異的バーコードの定義されたレパートリーを使用しない試験サンプル間の相互汚染との両方を検出することができる。特定の実施形態では、骨格は、時間バーコード、例えば、特定の期間を指定するバーコードを含むこともできる。時間バーコードを用いることにより、異なる日のラン間で、配列決定装置などのアッセイ装置におけるキャリーオーバーまたは汚染を検出することが可能である。より特定の実施形態では、サンプルおよび/または時間バーコードを使用して、サンプルおよび/または日の間の相互汚染を自動的に検出し、例えば、装置のオペレータに配列決定装置などのサンプル処理システムを浄化および/または除染するように指示することができる。

【0055】

特定の実施形態では、バーコード配列は、プライマー結合配列でもある。いくつかの実施形態では、骨格プライマーは、普遍的プライマーおよびプローブ特異的配列の両方を含む。いくつかの実施形態では、普遍的配列はプローブ特異的領域の内側（すなわち、3'）であり、他の実施形態では、普遍的配列は外側（すなわち、プローブ特異的領域に対し5'）である。いくつかの実施形態では、普遍的配列およびプローブ特異的配列は隣接している。他の実施形態では、これらは、少なくとも1、2、3、4、5、6、7、8、9、10、15、20、25、30、40、もしくは50、またはそれ以上のヌクレオチドによって分離されている。

【0056】

特定の実施形態では、骨格配列中の普遍的プライマー配列は、より長い「アダプタマー」プライマーのためのハイブリッド形成の鋳型の役割を果たす。「アダプタマープライマー」は、捕獲反応産物の増幅を容易にするために捕獲反応産物中の普遍的プライマー配列とハイブリッド形成するプライマーであり、さらに、サンプル特異的バーコード配列、例えば、アダプタマープライマーの普遍的プライマーハイブリッド形成領域の5'側の配列を含む。アダプタマープライマーを使用して、例えば、増幅反応産物にサンプル特異的バーコードを取り込んで、捕獲反応および増幅反応を完了した後に、サンプルのさらなる多重化を可能にすることができる。サンプル特異的バーコードの付加により、例えば、配列決定による検出の前に、多重捕獲および/または増幅反応産物がプールされることが可能になる。より特定の実施形態では、アダプタマープライマーはさらに、配列決定プライマーとハイブリッド形成する普遍的配列を含む。

10

20

30

40

50

【0057】

検出可能な部分は、骨格配列に関連していてもよい。蛍光（例えば、量子ドット、小分子、または蛍光タンパク質）標識、化学標識またはタンパク質ベースの標識などの直接標識の場合のように、ポリヌクレオチド配列に結合されていてもよい。あるいは、修飾ヌクレオチドなどの核酸標識、またはバーコードなどのプローブ特異的配列の場合のように、検出可能な部分は、ポリヌクレオチド配列内に取り込まれていてもよい。量子ドットは当該技術分野において知られており、例えば、国際公開第03/003015号パンフレットに記載されている。量子ドットを生体分子にカップリングさせる手段は、例えば、Medintz et al., Nature Materials 4: 235-46 (2005)、ならびに米国特許出願公開第2006/0068506号明細書および同第2008/0087843号明細書（それぞれ、2006年5月30日および2008年4月17日に公開）において概説されているように、当該技術分野において知られている。

【0058】

2 プローブ混合物

2.1 プローブおよび較正標準

本発明は、部分的に、標的生物（または、例えば、種、属、または血清型によって関連される生物の群）のゲノム中の標的配列に特異的にハイブリッド形成することができ、そして排除セット中のいかなる配列、例えば、少なくとも1つの非ハイブリッド形成ゲノム（宿主ゲノムおよび/または標的生物と異なる生物の所定のセット、例えば、病原性生物を含むが標的生物または標的生物群は含まない配列決定された細菌、ウイルス、真核生物、および古細菌生物の注釈付きのデータベースなど）とも特異的にハイブリッド形成しないプローブの集合を提供することに基づく。

【0059】

本発明の態様は、患者からの生物サンプル中の病原体の検出など、試験サンプルの多重分析のためのプローブの混合物を提供する。本発明によって提供される混合物は、少なくとも2、3、4、5、6、7、8、9、10、20、30、40、60、80、100、200、250、500、1000、2000、4000、8000、10000、20000、30000、40000、50000、60000、70000、80000、90000、または100000のプローブを含む。いくつかの実施形態では、混合物は

、特定の生物から複数の配列を捕獲するように設計される。特定の実施形態では、混合物は、少なくとも2、3、4、5、6、7、8、9、10、15、20、30、40、60、80、100、150、200、250、300、400、500、1000、2000、4000、8000、10000、15000、または20000の異なる標的生物のそれぞれに対して少なくとも1つの配列を捕獲することができる。特定の実施形態では、混合物は、表4、6、8、10、11のいずれか1つからの少なくとも1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、60、65、70、75、または80の相同プローブ配列、あるいは特定の配列、m t b - 37 r v - i n h a - p r - 01 - H 1、m t b - H 37 R v - r p o B - p r - 01 - H 1、m t b - H 37 R v - r p o B - p r - 01 - H 2、m t b - H 37 R v - r p o B - p r - 02 - H 1、m t b - H 37 R v - r p o B - p r - 02 - H 2、またはm t b - 37 r v - i n h a - p r - 01 - H 2、およびこれらの組み合わせを含む。特定の実施形態では、混合物は、表4、6、8、10、および11の何れかに記載される相同プローブ配列対を含む、少なくとも1、2、3、4、5、6、7、8、9、10、15、20、25、30、35、40、45、50、60、65、70、75、または80のプローブを含む。

【0060】

混合物中のプローブは、通常、同様のバルク特性（相同プローブ配列の長さ、相同プローブ配列のT_m、および捕獲される対象の領域の長さ、ならびに二次構造の欠如など）を有するか、あるいは同様の値の範囲内に含まれるであろう。いくつかの実施形態では、プローブの混合物中の相同プローブ配列のT_mは、互いに、10、9、8、7、6、5、4、3、2、または1の違いであり、特定の実施形態では、同じT_mを有する。いくつかの実施形態では、プローブの混合物中の相同プローブ配列は全て、互いに、10、9、8、7、6、5、4、3、2、または1ヌクレオチドの長さの違いであり、特定の実施形態では、これらは同じ長さである。プローブの標的配列間の対象の領域の長さは、混合物中の全てのプローブに共通でもよいし、2～20、20～100、20～200、40～300、100～300ヌクレオチドなどの範囲の値にわたって異なっていてもよい。特定の実施形態では、対象の領域は、互いに、100、90、80、70、60、50、40、30、20、または10ヌクレオチドの長さの違いである。より特定の実施形態では、対象の領域は同じ長さである。バーコードの長さも異なり得るが、一般には、互いに25、20、15、10、または5ヌクレオチドの違いである。特定の実施形態では、バーコードは同じ長さを有する。

【0061】

いくつかの実施形態では、本発明によって提供される混合物は、以下にさらに記載されるように、異なる試験サンプルからの捕獲反応産物および増幅反応産物を含む。簡単に言うと、異なる捕獲反応産物および/または増幅反応産物は、検出の前に、すなわち同時検出のために、結合および多重化され得る。これは、試験サンプルを同定するバーコード配列を用いて達成される。例えば、試験サンプルAからの捕獲反応産物は、サンプルA特異的バーコードを含み、サンプルBからの捕獲反応産物サンプルB特異的バーコードを含み得る。サンプルAおよびサンプルBからの捕獲反応産物が配列決定のために結合される場合、サンプルA捕獲反応産物中の全ての配列は、サンプルA特異的バーコード配列の存在によって同定される。

【0062】

特定の実施形態では、本発明の混合物は、サンプル内部較正核酸（S I C）を含有する。特定の実施形態では、既知の量の1つまたは複数のS I Cが、本発明によって提供される混合物中に含まれる。特定の実施形態では、少なくとも1、2、3、4、5、6、7、8、10、15、20、25、または30の異なるS I Cが混合物中に含まれる。特定の実施形態では、約4つの異なるS I Cが混合物中に存在する。いくつかの実施形態では、S I Cは、病原性DNA標的に特徴的なヌクレオチド組成を有し、品質管理のため、例えば、個々の試験サンプルの処理および配列決定ステップのための較正曲線の再構築を可能

10

20

30

40

50

にする特定のモル量で存在する。特定の実施形態では、S I Cは、混合物中約10%（モル量）の核酸、例えば、混合物中2、4、6、8、10、12、14、16、18、または20%（モル濃度）の核酸を構成する。特定の実施形態では、異なるS I Cは、例えば、1、2、3、4、5、6、7、8、9、10、20、50、100、200、500、1000、5000、10000、50000、または100000倍の、最も薄いS I Cから最も濃いS I Cまでの濃度範囲にわたる希釈系列で、2、3、4、5、6、7、8、9、10、20、30、40、または50段階において異なる濃度で存在する。特定の実施形態では、S I Cは、5、25、100、および250コピー／mLの濃度で、サンプル（例えば、プローブの混合物および試験サンプル、捕獲反応、捕獲反応産物、増幅反応、または増幅反応産物）中に存在する。所定濃度のS I Cを検出する（例えば、S I Cに向けられたプローブを用いることにより）ことによって、当業者は、試験サンプル中の対象の生物の濃度を推定することができる。特定の実施形態では、これは、捕獲された配列が検出される頻度を、核酸が得られたサンプルの容積に相關させることによって達成される。従って、単位容積当たりの生物カウント（例えば、血液または尿などの液体サンプルについてはコピー／mL）を、検出される各生物に対して推定することができる。

【0063】

特定の実施形態では、S I CおよびS I Cに向けられるプローブの濃度は、捕獲反応産物および／または増幅反応産物中で検出されるS I Cの配列が、混合物中の配列の約2、4、6、8、10、12、14、16、18、20、25、または30%を構成するように、経験的に調整される。特定の実施形態では、S I Cは、配列読取りの10～20%を構成する。特定の実施形態では、配列決定反応におけるS I C配列読取りの数は、予め定義されたパラメータ内でサンプル処理が発生することを保証するために、定量的に評価される。特定の実施形態では、予め定義されたパラメータには、以下のうちの1つまたは複数が含まれる：特定のランの間に配列決定される全てのサンプルに対する2つの標準偏差内での再現性、信頼性のある配列決定データのために経験的に決定された基準（例えば、塩基要求の信頼性、エラースコア、標的生物当たりの各プローブの全配列決定読取りの組成率）、配列決定ランにおいて約15%以下のG CまたはA U リッチS I Cの偏差。多重配列決定のためのブーリングを可能にするために患者サンプルがバーコード化される実施形態において、サンプル中のS I CのDNAは、独特なサンプル、例えば、特定の患者サンプルに相当する同じバーコードも含むであろう。

【0064】

より特定の実施形態では、S I Cは、上記で定義されたような対象の領域を含むことができ、ここで、対象の領域は、対象の領域に異種の配列をさらに含むように修飾されている。より特定の実施形態では、S I C中の対象の領域に異種の配列は、少なくとも1、2、3、4、5、6、7、8、9、10、15、20、25、30、40の近接する塩基、またはそれ以上である。修飾された対象の領域を含むS I Cを用いることによって、単一のプローブは、サンプル内の対象の生物と、定量化および検証のための内部対照を提供するS I Cとを両方とも検出するために使用することができる。従って、試験サンプル中で検出されるS I C配列および対象の生物からの対象の領域は、例えば、配列決定または配列特異的定量PCRによって、対象の領域に異種の配列を検出することによって区別することができる。

【0065】

2.2 サンプル

いくつかの実施形態では、本発明の混合物は、サンプル核酸を含有する。核酸は、生物サンプルなどの任意の試験サンプルから得ることができる。試験サンプルから得られる核酸は、少なくとも1、2、3、4、5、10、20、30、40、50、60、70、80、85、90、95、96、97、98、99重量%の有機物などの、様々な純度を有することができる。特定の実施形態では、サンプル核酸は、試験サンプルから抽出される。いくつかの実施形態では、サンプル核酸はさらに処理されて、例えば、メチル化状態の検出を可能にすることができます。ゲノム全域わたるメチル化部位の検出の概説については

10

20

30

40

50

、Deng(2009)(メチル化部位を位置付けるために、CpGアイランドのMIP捕獲および重硫酸塩配列決定について記載)を参照されたい。

【0066】

試験サンプルは任意の源に由来し、食料品サンプル(安全性試験、タギング、およびトラッキング)、農業サンプル(例えば、土壤サンプル、病原体検出および/またはGM作物の検出のため)、薬物ロット(例えば、小分子と、血液供給を含む生物学とのロットリリースアッセイのため)、水サンプル(給水の生物多様性の分析、農業、商業、政府、病院、工業、研究室、軍隊、家庭、または家畜の給水の安全性試験(例えば、生物テロ防御)、ならびに水泳または水浴びのための安全性試験を含む)、表面のスワブもしくは抽出物、空気品質のモニタリング、または患者サンプルなどの生物サンプルを含み得る。

10

【0067】

患者は、ヒト、または家畜、飼いならされた動物、および野生動物などの動物を含むことができる。いくつかの実施形態では、動物は、トリ、ウシ、イヌ、ウマ、ネコ、ヒツジ、魚類/魚、ブタ、靈長類、齧歯類、または有蹄動物である。患者は、成人、若者、胎児、または胎芽を含む任意の発達段階であり得る。特定の実施形態では、患者は哺乳類であり、より特定の実施形態ではヒトである。

【0068】

被験者または患者からの生物サンプルは、全細胞、組織、もしくは臓器、または3つの始原的な胚葉(外胚葉、中胚葉もしくは内胚葉)のいずれかに由来する組織を含むバイオプシーを含み得る。例示的な細胞または組織源には、皮膚、心臓、骨格筋、平滑筋、腎臓、肝臓、肺、骨、臍臓、中枢神経組織、末梢神経組織、循環組織、リンパ系組織、腸、脾臓、甲状腺、結合組織、または生殖腺が含まれる。試験サンプルは、入手してすぐにアッセイされてもよいし、あるいは、混合、化学処理、固定化/保存、凍結、または培養によって処理されてもよい。生物サンプルは、血液、胸水、乳、初乳、リンパ液、血清、血漿、尿、脳脊髄液、滑液、唾液、精液、涙、および糞便も含む。その他のサンプルは、スワブ、洗液(wash)、洗浄液(lavage)、排出物、または吸引液(鼻、経口、鼻咽頭の、中咽頭、食道、胃、直腸、または腔の、スワブ、洗液、洗浄液(ravage)、排出物、または吸引液など)、およびこれらの組み合わせ(前述のバイオプシー材料の何れかとの組み合わせを含む)を含む。

20

【0069】

2.3 パネル

特定の実施形態では、本発明の混合物は、特定の苦痛(例えば、呼吸器、血液、尿路感染症)またはサンプルタイプ(例えば、バイオプシー、水、食料品、または農業)のために、一般的な病原体などの生物のパネルを検出するように設計されたプローブを含む。「パネル」は本発明によって提供される混合物を指し、特定の苦痛またはサンプルタイプと間連される1つまたは複数の病原体に向けられた複数のプローブを含む。特定の実施形態では、本発明の混合物は、多重パネルを含有する。特定の病原体に向けられるプローブを含むパネルは、本出願の教示に従うことにより、ルーチン的な技能だけを用いて製造することができる。いくつかの実施形態では、本発明によって提供されるパネルは、米国特許出願公開第2010/0098680号明細書(特に、段落160、参照によって本明細書中に援用される)に記載されるものなどの複数の病原体に関する。特定の実施形態では、パネルは、米国特許出願公開第2010/0098680号明細書の段落160に記載される少なくとも1、2、3、4、5、0、15、20、25、30、35、40、または50の病原体のそれぞれに向けられる少なくとも1つのプローブを含有する。

30

【0070】

いくつかの実施形態では、パネルは脳脊髄液(CSF)パネルであり、ナイセリア・メンギティディス(*Neisseria meningitidis*)(例えば、ゲノム受入番号NC_008767、NC_010120、NC_003116、NC_003112、NC_013016、またはNC_004758、特定の実施形態では、ctrA遺伝子に向けられたプローブを含む)、HHV6(ヒトヘルペスウイルス6、例えば、

40

50

ゲノム受入番号NC_001664またはNC_000898、特定の実施形態では、メジャーカプシドタンパク質遺伝子に向けられたプローブを含む)、JCV (JCポリオーマウイルス、例えば、ゲノム受入番号NC_001699.1、特定の実施形態では、ラージT抗原遺伝子に向けられたプローブを含む)、BKV (BKポリオーマウイルス、例えば、ゲノム受入番号NC_001538、特定の実施形態では、調節領域に向けられたプローブを含む)、HSV1 (ヒトヘルペスウイルス1、例えば、ゲノム受入番号NC_001806またはX14112、特定の実施形態では、gD遺伝子 (X14112における位置138333~141048) に向けられたプローブを含む)、HSV2 (ヒトヘルペスウイルス2、例えば、ゲノム受入番号NC_001798またはZ86099、特定の実施形態では、gG遺伝子 (Z86099における位置137878~139977) に向けられたプローブを含む)、ストレプトコッカス・ニューモニエ (Streptococcus pneumoniae) (例えば、ゲノム受入番号NC_012469、NC_012468、NC_012467、NC_008533、NC_012466、NC_010380、またはNC_011072、特定の実施形態では、ply遺伝子に向けられたプローブを含む)、ヘモフィルス・インフルエンザエ (Haemophilus influenzae) (例えば、ゲノム受入番号NC_007146、NC_00907、NC_009566、NZ_AAZE00000000、NZ_AAZJ00000000、NC_009567、またはDQ115375、特定の実施形態では、bexA遺伝子に向けられたプローブを含む) に向けられたプローブを含む。特定の実施形態では、本発明によって提供されるパネルは、これらの生物のうちの1、2、3、4、5、6、7、または8全てのそれぞれに対して、より特定の実施形態では、生物の例示的な遺伝子に対して1つまたは複数のプローブを含む。

【0071】

いくつかの実施形態では、パネルは髄膜炎パネルであり、B群連鎖球菌、エシェリキア・コリ (Escherichia coli)、リステリア・モノサイトゲネス (Listeria monocytogenes)、ナイセリア・メンингイティディス (Neisseria meningitidis)、ストレプトコッカス・ニューモニエ (Streptococcus pneumoniae) (血清型6、9、14、18および23)、ヘモフィルス・インフルエンザエ (Haemophilus influenzae) B型、ブドウ球菌、シュードモナス、マイコバクテリウム・ツベルクローシス (Mycobacterium tuberculosis)、トレポネーマ・パリダム (Treponema pallidum)、ボレリア・ブルグドルフェリ (Borrelia burgdorferi)、クリプトコッカス・ネオフォルマンス (Cryptococcus neoformans)、ネグレリア・フォーレリ (Naegleria fowleri)、エンテロウイルス、単純疱疹ウイルス1型および2型、水痘带状疱疹ウイルス、流行性耳下腺炎ウイルス、HIV、LCMV、アンギオストロンギルス・カントネンシス (Angiostrongylus cantonensis)、ナソトーマ・スピニゲルム (Gnathostoma spinigerum)、結核、梅毒、クリプトコッカス症、およびコクシジオイデス症のうちの1つまたは複数に向けられた1つまたは複数のプローブを含む。特定の実施形態では、パネルは、これらの生物のうちの1、2、3、4、5、6、7、8、9、10、15、20、25、30、または31の生物に向けられたプローブを含む。

【0072】

いくつかの実施形態では、パネルは尿路感染症 (UTI) パネルであり、S.サプロフィカス (S. saprophyticus) (ATCC15305) (例えば、ゲノム受入番号AP008934またはAP008935、特定の実施形態では、gyrB遺伝子に向けられたプローブを含む)、エンテロコッカス・フェカリス (Enterococcus faecalis) (MMH594) (例えば、ゲノム受入番号AF034779、特定の実施形態では、esp遺伝子に向けられたプローブを含む、例えば、参照される)、大腸菌 (E. coli) (CFT073) (例えば、ゲノム受入番号NC_004

431.1、特定の実施形態では、fimH遺伝子に向けられたプローブを含む)、大腸菌(E. coli)(IAI39)(例えば、ゲノム受入番号NC_01750.1、特定の実施形態では、papG遺伝子に向けられたプローブを含む)、大腸菌(E. coli)(CFT073)(例えば、ゲノム受入番号NC_004431.1、特定の実施形態では、papX遺伝子に向けられたプローブを含む)、ウレアプラズマ・ウレアリチカム(Ureaplasma urealyticum)(血清型10str. ATCC33699)(例えば、ゲノム受入番号UUR10_0078、特定の実施形態では、hly遺伝子に向けられたプローブを含む)、ウレアプラズマ・バルバム(Ureaplasma parvum)(血清型3str. ATCC27815)(例えば、ゲノム受入番号CP000942、特定の実施形態では、hly遺伝子に向けられたプローブを含む)、エンテロコッカス・フェシウム(Enterococcus faecium)(CV133)(例えば、ゲノム受入番号AF544400、特定の実施形態では、hy1(efm)遺伝子に向けられたプローブを含む)、およびエンテロコッカス・フェシウム(Enterococcus faecium)(例えば、ゲノム受入番号AF034779、特定の実施形態では、esp遺伝子に向けられたプローブを含む)に向けられたプローブを含む。特定の実施形態では、本発明によって提供される核酸プローブの混合物は、これらの生物のうちの1、2、3、4、5、6、7、8、または9全てのそれぞれに対して、より特定の実施形態では、生物の例示的な遺伝子に対して1つまたは複数のプローブを含む。

10

20

30

40

【0073】

いくつかの実施形態では、パネルは、エシェリキア・コリ(Escherichia coli)、スタフィロコッカス・サプロフィチカス(Staphylococcus saprophyticus)、プロテウス種、クレブシエラ種、エンテロコッカス種、カンジダ・アルビカンス(Candida albicans)、ウレアプラズマ、およびマイコプラズマ種を含む1つまたは複数の生物に対する1つまたは複数のプライマーを含む、代替のUTIパネルである。特定の実施形態では、本発明によって提供される核酸プローブの混合物は、これらの生物のうちの1、2、3、4、5、6、7、または8全てのそれぞれに対する1つまたは複数のプローブを含む。

【0074】

さらに別の実施形態では、UTIパネルは、大腸菌(E. coli)に向けられた1つまたは複数のプローブを含む。より特定の実施形態では、パネルはさらに、クレブシエラ種、セラチア種、シトロバクター種、およびエンテロバクター種などの他の腸内細菌科、シュードモナス・エルジノーサ(Pseudomonas aeruginosa)などの非発酵菌、ならびにコアグラーーゼ陰性ブドウ球菌およびエンテロコッカス種を含むグラム陽性球菌に向けられた1つまたは複数のプローブを含む。さらにより特定の実施形態では、パネルはさらに、カンジダ・アルビカンス(Candida albicans)などのカンジダに向けられた1つまたは複数のプローブを含む。特定の実施形態では、本発明によって提供される核酸プローブの混合物は、これらの生物のうちの1、2、3、4、5、6、7、8、9、10、または11のそれぞれに対する1つまたは複数のプローブを含む。

40

【0075】

いくつかの実施形態では、パネルは、大腸菌(E. coli)、クラミジア、マイコプラズマ、スタフィロコッカス・サプロフィチカス(Staphylococcus saprophyticus)、およびスタフィロコッカス・エピデルミデス(Staphylococcus epidermidis)に向けられた1つまたは複数のプローブを含むUTIパネルである。特定の実施形態では、本発明によって提供される核酸プローブの混合物は、これらの生物のうちの1、2、3、4、または5のそれぞれに対する1つまたは複数のプローブを含む。

【0076】

特定の実施形態では、パネルは呼吸器パネルであり、スタフィロコッカス・アウレウス

50

(*Staphylococcus aureus*)、シュードモナス・エルジノーサ(*Pseudomonas aeruginosa*)、クレブシエラ・ニューモニエ(*Klebsiella pneumoniae*)、ヘモフィルス・インフルエンザエ(*Haemophilus influenzae*)、ブランハメラ(モラクセラ)カタラーリス(*Branhamella (Moraxella) catarrhalis*)、ストレプトコッカス・ピオゲネス(*Streptococcus pyogenes*) (A群)、コリネバクテリウム・ジフテリエ(*Corynebacterium diphtheriae*)、SARS-CoV、ボルダテラ・バータシス(*Bordatella pertussis*)、インフルエンザウイルス(A、B、C型)、ライノウイルス、コロナウイルス、エンテロウイルス、アデノウイルス、呼吸器多核体ウイルス(RSV)、パラインフルエンザウイルス、流行性耳下腺炎ウイルス、レジオネラ・ニューモフィラ(*Legionella pneumophila*)、シュードモナス・エルジノーサ(*Pseudomonas aeruginosa*)、パークホルデリア・セパシア(*Burkholderia cepacia*)、マイコプラズマ・ニューモニエ(*Mycoplasma pneumoniae*)、マイコバクテリウム・ツベルクローシス(*Mycobacterium tuberculosis*)、クラミジア・ニューモニエ(*Chlamydia pneumoniae*)、マイコバクテリウム・アビウムイントラセルラーレ複合体(*Mycobacterium avium intracellulare complex*) (MAC)、カンジダ・アルビカンス(*Candida albicans*)、コクシジオイデス・イミチス(*Coccidioides immitis*)、ヒストプラズマ・カプスラーツム(*Histoplasma capsulatum*)、プラストミセス・デルマチチジス(*Blastomyces dermatitidis*)、クリプトコッカス・ネオフォルマンス(*Cryptococcus neoformans*)、およびアスペルギルス・フミガータス(*Aspergillus fumigatus*)に向けられた1つまたは複数のプローブを含む。特定の実施形態では、本発明によって提供されるパネルは、これらの生物のうちの1、2、3、4、5、6、7、8、9、10、15、20、25、30または33のそれぞれに対する1つまたは複数のプローブを含む。

【0077】

いくつかの実施形態では、パネルは呼吸器パネルであり、インフルエンザA(亜型H1、H3、H5およびH7を含む)、インフルエンザB、パラインフルエンザ(2型)、呼吸器多核体ウイルス、およびアデノウイルスを含む1つまたは複数の病原体に向けられた1つまたは複数のプローブを含有する。

【0078】

特定の実施形態では、パネルは呼吸器パネルであり、ストレプトコッカス・ニューモニエ(*Streptococcus pneumoniae*)、マイコプラズマ・ニューモニエ(*Mycoplasma pneumoniae*)、ヘモフィルス・インフルエンザエ(*Haemophilus influenzae*)、クラミドフィラ・ニューモニエ(*Chlamydophila pneumoniae*)、およびレジオネラ種、レジオネラ・ニューモフィラ(*Legionella pneumophila*)、SARSウイルス、H1N1、H5N1、グラム陰性桿菌、モラクセラ・カタラーリス(*Moraxella catarrhalis*)、スタフィロコッカス・アウレウス(*Staphylococcus aureus*)、結核、および呼吸器多核体ウイルス(RSV)を含む1つまたは複数の病原体に向けられた1つまたは複数のプローブを含有する。特定の実施形態では、本発明によって提供されるパネルは、これらの生物うちの1、2、3、4、5、6、7、8、9、10、11、12、13、または14のそれぞれに対する1つまたは複数のプローブを含む。

【0079】

いくつかの実施形態では、パネルは、ジフテリア、エプスタイン・バーウイルス(EBV)、シャーガス、HIV、ウエストナイルウイルス、マラリア、梅毒、デング熱、バベ

10

20

30

40

50

シア、異種指向性マウス白血病ウイルス関連ウイルス（X M R V）、B型肝炎、C型肝炎、ウイルス性出血熱（エボラおよびマールブルグウイルスを含む）のうちの1つまたは複数に向けられた1つまたは複数のプローブを含む血液パネルである。特定の実施形態では、本発明によって提供されるパネルは、これらの生物のうちの1、2、3、4、5、6、7、8、9、10、12、13、または14のそれぞれに対する1つまたは複数のプローブを含む。より特定の実施形態では、血液パネルは、H I V、B型肝炎、C型肝炎、およびトリパノソーマ・クルージ（*T r y p a n o s o m a c r u z i*）（シャーガス）のそれぞれに対する1つまたは複数のプローブを含む。さらなる実施形態では、血液パネルは、H I V、B型肝炎、C型肝炎、およびトリパノソーマ・クルージ（*T r y p a n o s o m a c r u z i*）（シャーガス）病原体、ならびにH L A、K i r、A B OおよびR h式血液マーカー座位などのヒト宿主ゲノム配列のそれぞれに向けられた1つまたは複数のプローブを含む。

10

【0080】

いくつかの実施形態では、パネルは、参照によって本明細書中に援用される米国特許出願公開第2009/0291854号明細書の段落26および27で開示されるものを含む1つまたは複数の病原体に向けられた1つまたは複数のプローブを含有する血液パネルである。特定の実施形態では、本発明によって提供されるパネルは、これらの生物のうちの1、2、3、4、5、6、7、8、9、10、15、20、25、または30のそれぞれに対する1つまたは複数のプローブを含む。

20

【0081】

いくつかの実施形態では、パネルは敗血症パネルであり、大腸菌（*E . c o l i*）、クレブシエラ、プロテウス、エンテロバクター種、シュードモナス・エルジノーサ（*P s e u d o m o n a s a e r u g i n o s a*）、ナイセリア・メンингイティディス（*N e i s s e r i a m e n i n g i t i d i s*）およびバクテロイデスのような主にグラム陰性の細菌、ならびにスタフィロコッカス・アウレウス（*S t a p h y l o c o c c u s a u r e u s*）、ストレプトコッカス・ニューモニエ（*S t r e p t o c o c c u s p n e u m o n i a e*）および他の連鎖球菌のような一般的なグラム陽性細菌を含む1つまたは複数の病原体に向けられた1つまたは複数のプローブを含む。特定の実施形態では、本発明によって提供されるパネルは、これらの生物のうちの1、2、3、4、5、6、7、8、9、または10のそれぞれに対する1つまたは複数のプローブを含む。

30

【0082】

いくつかの実施形態では、パネルは水、土壤、または農業パネルであり、例えば、G . ランブリア（*G . l a m b l i a*）、クリプトスピリジウム、サルモネラ、赤痢菌、カンピロバクター、カンジダ、大腸菌（*E . c o l i*）、エルシニア、エロモナス、または他の小さい寄生生物に向けられた1つまたは複数のプローブを含む。特定の実施形態では、パネルは、水および/または土壤中の一般的な汚染物質であるジアルジアおよび/またはクリプトスピリジウムに対する1つまたは複数のプローブを含む。特定の実施形態では、本発明によって提供されるパネルは、これらの生物のうちの1、2、3、4、5、6、7、8、9、10、または11のそれぞれに対する1つまたは複数のプローブを含む。

40

【0083】

いくつかの実施形態では、パネルは食料品または農業パネルであり、エシェリキア・コリ（*E s c h e r i c h i a c o l i*）、サルモネラ、シゲラ・ソネイ（*S h i g e l l a s o n n e i*）、カンピロバクター、リステリア（例えば、リステリア・モノサイトゲネス（*L i s t e r i a m o n o c y t o g e n e s*））、エルシニア・エンテロコリチカ（*Y e r s i n i a e n t e r o c o l i t i c a*）、エルシニア・シュードツベルクロシス（*Y e r s i n i a p s e u d o t u b e r c u l o s i s*）、ビブリオ・コレラ（*V i b r i o c h o l e r a*）、およびクロストリジウム（例えば、*C . b o t u l i n u m*）のうちの1つまたは複数に向けられた1つまたは複数のプローブを含む。特定の実施形態では、食料品または農業パネルは、エシェリキア・コリ（*E s c h e r i c h i a c o l i*）O 1 5 7 : H 7、腸管出血性のエシェリ

50

キア・コリ (*Escherichia coli*) (EHEC)、腸内毒素原性エシェリキア・コリ (*Escherichia coli*) (ETEC)、腸管侵入性エシェリキア・コリ (*Escherichia coli*) (EIEC)、腸管病原性エシェリキア・コリ (*Escherichia coli*) (EPEC)、サルモネラ、リストリア、エルシニア、カンピロバクター、クロストリジウム種、およびブドウ球菌種に向けられた1つまたは複数のプライマーを含む。特定の実施形態では、農業または食料品パネルは、キシレラ・ファスティディオーサ (*Xylella fastidiosa*) およびキントモナス・アクソノポディス (*Xanthomonas axonopodis*) などの一般的な柑橘類汚染物質に対する1つまたは複数のプローブを含有する。特定の実施形態では、本発明によって提供されるパネルは、これらの生物のうちの1、2、3、4、5、6、7、8、9、10、15、20、またはそれ以上のそれぞれに対する1つまたは複数のプローブを含む。10

【0084】

真菌パネルは、いくつかの実施形態では、参照によって本明細書中に援用される米国特許出願公開第2010/0129821号明細書の段落162および180ならびに表1および2に記載される1つまたは複数の真菌に向けられた少なくとも1つのプローブを含む。特定の実施形態では、本発明によって提供されるパネルは、これらの生物のうちの1、2、3、4、5、6、7、8、9、10、15、20、または30のそれぞれに対する1つまたは複数のプローブを含む。特定の実施形態では、真菌パネルは、アスペルギルスおよび/またはカンジダ・アルビカンス (*Candida albicans*) に向けられた1つまたは複数のプローブを含む。20

【0085】

いくつかの実施形態では、本発明によって提供されるパネルは、本明細書中で記載されるような複数の病原体に向けられたプローブ、ならびにHLA、KIR、ABOおよびRh式血液マーカー座位などの特定のヒトゲノム配列に向けられたプローブを含み、同じサンプルにおける遺伝子型同定および病原体検出を可能にする。

【0086】

いくつかの実施形態では、パネルは、被験者の遺伝子型同定のための被験者パネルである。特定の実施形態では、被験者パネルは、少なくとも1、2、3、4、5、6、7、8、9、10、20、40、80、100、200、400、800、1000、5000、または10000の被験者座位のためのプローブを含む。特定の実施形態では、パネルは、哺乳類被験者に対するものである。より特定の実施形態では、哺乳類はヒトである。いくつかの実施形態では、パネルは、遺伝性の遺伝子異常、および/または疾患のリスクの増大に関連する遺伝子型を検出するための出生前または新生児パネルである。特定の実施形態では、パネルは、キラー細胞免疫グロブリン様受容体 (KIR) 座位タイピングのため、そして、サイトカインSNP、例えば、以下のSNPのうちの1つまたは複数を検出するためのプローブを含む：-174におけるIL-6: C/G、-308におけるTNF-: G/A、-238におけるG/A、-1082におけるIL-10: G/A、-819C/T、-592におけるC/A。いくつかの実施形態では、パネルは遺伝子型HLAマーカーに対するプローブを含み、特定の実施形態では、クラスI (A-H) およびクラスIIのHLAマーカーのそれぞれに対する少なくとも1つのプローブを含む。他の実施形態では、パネルは、米国特許出願公開第2010/0137426号明細書の段落25、57、および58、米国特許出願公開第2009/0305284号明細書の段落6および7、米国特許出願公開第2010/0144836号明細書の段落27に記載される遺伝子のうちの1つまたは複数、米国特許出願公開第2010/0143949号明細書の表1に記載されるマーカーのいずれか、あるいは米国特許出願公開第2010/0093558号明細書の段落14の遺伝子のいずれかに向けられたプローブを含む（これらは全て、参照によって本明細書中に援用される）。いくつかの実施形態では、パネルは、機能獲得型「癌遺伝子」 (ABL1、BCL1、BCL2、BCL6、CBFA2、CBL、CSF1R、ERBA、ERBB、EBRB2、ETS1、ETS1、ETV64050

、 F G R 、 F O S 、 F Y N 、 H C R 、 H R A S 、 J U N 、 K R A S 、 L C K 、 L Y N 、 M D M 2 、 M L L 、 M M T V - P y V T 、 M M T V n e u 、 M Y B 、 M Y C 、 M Y C L 1 、 M Y C N 、 N R A S 、 P I M 1 、 P M L 、 R E T 、 S R C 、 T A L 1 、 T C L 3 、 および Y E S など) 、 および / または機能喪失型の腫瘍抑制遺伝子 (A P C 、 B R C A 1 、 B R C A 2 、 M A D H 4 、 M C C 、 N F 1 、 N F 2 、 R B 1 、 P 5 3 、 および W T 1 など) に向けられたプローブを含む。いくつかの実施形態では、パネルは、 H L A 、 K i r およびサイトカイン遺伝子座に向けられたプローブを含む。特定の実施形態では、本発明によって提供されるパネルは、これらのマーカーのうちの 1 、 2 、 3 、 4 、 5 、 6 、 7 、 8 、 9 、 1 0 、 1 5 、 2 0 、 3 0 、 またはそれ以上のそれぞれに対する 1 つまたは複数のプローブを含む。

10

【 0 0 8 7 】

本発明によって提供される追加のパネルは、ウイルス、細菌、古細菌、原生動物、および真核生物、ならびに組み合わせに向けられたプローブを含む。特定の実施形態では、パネルは、約 1 、 2 、 3 、 4 、 5 、 6 、 7 、 8 、 9 、 1 0 、 1 5 、 2 0 、 2 5 、 3 0 または 3 5 のウイルス、約 1 、 2 、 3 、 4 、 5 、 6 、 7 、 8 、 9 、 1 0 、 1 5 、 2 0 、 2 5 、 3 0 または 3 5 の細菌、および約 1 、 2 、 3 、 4 、 5 、 6 、 7 、 8 、 9 、 1 0 、 1 5 、 2 0 、 2 5 、 3 0 または 3 5 の真核生物のそれぞれに対する少なくとも 1 つのプローブを含有する。特定の実施形態では、真核生物に向けられたパネル中のプローブは、少なくとも 1 、 2 、 3 、 4 、 5 、 6 、 7 、 8 、 9 、 または 1 0 の真菌に対するプローブを含む。特定の実施形態では、パネルはさらに、 1 、 2 、 3 、 4 、 5 、 6 、 7 、 8 、 9 、 または 1 0 の古細菌のそれぞれに対する少なくとも 1 つのプローブを含むことができる。

20

【 0 0 8 8 】

本発明のパネルを用いて検出することができる例示的なウイルス分類群には、アデノウイルス科、アロヘルペスウイルス科、アネロウイルス、アレナウイルス科、アルテリウイルス科、アスコウイルス科、アスファウイルス科、アストロウイルス科、バキュロウイルス科、バルナウイルス科、ベニウイルス、ビコウダウイルス科 (B i c a u d a v i r i d a e) 、ビルナウイルス科、ボルナウイルス科、プロモウイルス科、ブニヤウイルス科、カリシウイルス科、カウドウイルス目、カリモウイルス科、チェラウイルス (C h e r a v i r u s) 、クリソウイルス科、サーコウイルス科、クロステロウイルス科、コモウイルス科、コロナウイルス科、コルチコウイルス科、シストウイルス科、デルタウイルス、ジシストロウイルス科、エンドルナウイルス、フィロウイルス科、フラビウイルス科、フレキシウイルス科、フロウイルス (F u r o v i r u s) 、フセロウイルス科、ジェミニウイルス科、グローブロウイルス科 (G l o b u l o v i r i d a e) 、ヘパドナウイルス科、ヘペウイルス科、ヘルペスウイルス目、ヘルペスウイルス科、ホルデイウイルス、ハイポウイルス科、イデオウイルス、イフラウイルス (I f l a v i r u s) 、イノウイルス科、イリドウイルス科、レビウイルス科、リボスリクスウイルス科、ルテオウイルス科、マラコヘルペスウイルス科、マルナウイルス科 (M a r n a v i r i d a e) 、ミクロウイルス科、ミミウイルス科、モノネガウイルス目、ミオウイルス科、ナノウイルス科、ナルナウイルス科、ニドウイルス目、ニマウイルス科、ノダウイルス科、オフィオウイルス、オルトミクソウイルス科、ウルミアウイルス、パピローマウイルス科、パラミクソウイルス科、パルティティウイルス科、パルボウイルス科、ペクルウイルス、フィコドナウイルス科、ピコルナウイルス目、ピコルナウイルス科、プラズマウイルス科、ポドウイルス科、ポリドナウイルス科、ポリオーマウイルス科、ポモウイルス、ポティウイルス科、ポックスウイルス科、レオウイルス科、レトロウイルス科、ラブドウイルス科、ロニウイルス科、ルディウイルス科、サドワウイルス、サルタープロウイルス (S a l t e r p r o v i r u s) 、セキウイルス科、シホウイルス科、ソベモウイルス、テクティウイルス科、テヌイウイルス、テトラウイルス科、トバモウイルス、トブラウイルス、トガウイルス科、トンブスウイルス科、トティウイルス科、ティモウイルス科、およびアンプラウイルスが含まれる。非 D N A および / または一本鎖ウイルスは、当業者に既知の手段によって、例え逆転写によって、本発明で使用するために容易に適応されるであろう。特

30

40

50

定の実施形態では、本発明の混合物は、少なくとも1、2、4、6、8、10、15、20、30、50、100、150、200、250、300、または400種類のウイルスを検出するために1つまたは複数のプローブを含む。

【0089】

本発明によって提供されるパネルを用いて検出することができる例示的な細菌の種類には、ファーミキューテス門（例えば、バチルス目、ラクトバチルス目、クロストリジウム）、バクテロイデス門／緑色硫黄細菌門、アクチンバクテリア（*Actinobacteria*）、シアノバクテリア門、スピロヘータ目、クラミジア、アルファプロテオバクテリア（例えば、根粒菌、リケッチア）、ベータプロテオバクテリア（例えば、ボルデテラ、ナイセリア、バークホルデリア）、ガムマプロテオバクテリア（例えば、パスツレラ、キサントモナス、シュードモナス、腸内細菌、ビブリオ）、ならびにイプシロンおよびデルタプロテオバクテリアが含まれる。特定の実施形態では、本発明の混合物は、少なくとも1、2、4、6、8、10、15、20、30、50、100、150、200、250、300、または400種類の細菌を検出するために、1つまたは複数のプローブを含む。

10

【0090】

本発明によって提供されるパネルを用いて検出することができる例示的な古細菌の種類には、サーモコッカス目、サーモプラズマ目、メタノサルシナ目、メタノミクロビウム目、メタノコッカス目、メタノバクテリウム目、メタノピュルス目、ハロバクテリウム目、アーケオグロブス目、ナノアーケオタ門、およびクレンアーケオタ門（例えば、サーモプロテウス目、スルフォロブス目、およびデスルフロコッカス目）が含まれる。特定の実施形態では、本発明の混合物は、少なくとも1、2、4、6、8、10、15、20、30、50、00、150、200、250、300、または400種類の古細菌を検出するために、1つまたは複数のプローブを含む。

20

【0091】

本発明によって提供されるパネルを用いて検出することができる例示的な真核生物には、線形動物門、吸虫綱、ディプロモナス目、アピコンプレックス門、エンタメオビダエ（*Entamoebidae*）、キネトプラスト目、ディクティオステリダ（*Dictyostelia*）、ストラメノパイル、真菌（例えば、微胞子虫門、バシドマイコタ（*Basidiomycota*）、接合菌門、および子囊菌門（例えば、シゾサッカロミケス綱、サッカロミケス亜門、およびチャワンタケ亜門））が含まれる。特定の実施形態では、本発明の混合物は、少なくとも1、2、4、6、8、10、15、20、30、50、100、150、200、250、300、または400種類の真核生物を検出するために、1つまたは複数のプローブを含む。

30

【0092】

3 本発明の例示的な方法

3.1 プローブ設計

本発明によって提供されるプローブおよび混合物は、以下の実施例および本出願の一般的な教示によって、当業者により製造することができる。プローブの設計プロセス（プローブ設計「パイプライン」とも呼ばれる）は、入力として、プローブが設計され得るゲノムDNA配列のセットおよび標的生物の特定の株のセットを探る。ゲノムDNA配列は、全ゲノム、特定の遺伝子、または1つまたは複数の株におけるゲノム座標であり得る。あるいは、パイプラインは、入力として、ゲノム、遺伝子、または座標のセットを探り、いくつかの基準に基づいて標的に対する領域のセットを選択するであろう。パイプラインは、入力ゲノム、遺伝子、または相同プローブ配列セットにおける標的領域の座標と、既知のゲノムより大きいセットとの間で異なる領域などの基準を使用することができる。

40

【0093】

特定の実施形態では、対象の生物の標的ゲノムの配列が提供され、標的ゲノム内の長さn（n-mer）を有する連続したヌクレオチドの全ての可能なストリングが列挙され（本明細書では、標的ゲノムの「スライス」とも呼ばれる）、ここで、nは18、20、2

50

2、24、26、28、30、32、34、36、38、40、45、50、55、60、65、70、80、90、100、110、120またはそれ以上である。特定の実施形態では、nは18～50、18～36、20～32、または22～28ヌクレオチドである。さらに特定の実施形態では、nは18～26ヌクレオチドである。より特定の実施形態では、nは22～28、例えば、25ヌクレオチドである。いくつかの実施形態では、長さnのゲノムセグメントは、ほぼ1～nの間であるオフセットを有する。特定の実施形態では、オフセットは1である。

【0094】

特定の実施形態では、列挙されたn-merは、そのゲノム位置を同定するために注釈付きである。いくつかの実施形態では、n-merは、より迅速なスクリーニングを容易にするために、ゲノムの注釈を用いずにストリングに転換される。

10

【0095】

パイプラインは、n-merの適合性に従って、連結側のプローブ相同性領域（連結側ホーマー）として、そして伸長側プローブ相同性領域（伸長側ホーマー）として、それぞれのn-merに対する第1のスコアを生成し得る。n-merに対するスコアは、融解温度、一般的な配列組成、特定の位置における配列組成、およびn-merがそれ自体であるいは骨格配列とヘアピンを形成する傾向などの特徴に基づくことができる。

【0096】

パイプラインはn-merをフィルタリングして、実質的に同じまたは正確に同じ配列のものを除去することができる（すなわち、「複製（duplicata）スクリーン」）。連結側ホーマー候補のセットを生成するために、長さx（ここで、xは、長さnのゲノムセグメントの列挙において使用される最小のnである（上記のとおり））の同一の接尾語を有するn-merが考慮され、最高スコアを有するものが保持され得るが、ここで、スコアは、上記のように、連結側ホーマーとしてのn-merの適合性に基づいている。伸長側ホーマー候補のセットを生成するために、長さxの同一の接頭語を有するn-merが考慮され、最高スコアを有するものが保持され得る。

20

【0097】

いくつかの実施形態では、n-merのスコアリングは、相同プローブ配列として使用するのに適していないn-merを除去するための一連のスクリーンとして実施することができる。スクリーンは、複製および実質的に複製の配列を除去し、指定のT_m範囲の外側の配列を除去し（「T_mスクリーン」、例えば50～72の外側）、多過ぎる反復ヌクレオチドを有するストリングを有する配列を除去し（「反復スクリーン」、例えば、4つ以上の連続した同一ヌクレオチド）、そしてセルフハイブリッド形成する可能性がある配列を除去する（「ヘアピンスクリーン」、例えば、自己二量体化またはヘアピンを形成）ことを含む。これらのスクリーンは、相同プローブ配列について本出願で記載されるパラメータのいずれかを適用させるように調整することができる。スクリーンは、例えば、以下の表の実施形態のいずれかによって、任意の順序で実施することができる。

30

第1のスクリーン	第2のスクリーン	第3のスクリーン	第4のスクリーン
複製スクリーン	T _m スクリーン	反復スクリーン	ヘアピンスクリーン
複製スクリーン	T _m スクリーン	ヘアピンスクリーン	反復スクリーン
複製スクリーン	反復スクリーン	T _m スクリーン	ヘアピンスクリーン
複製スクリーン	反復スクリーン	ヘアピンスクリーン	T _m スクリーン
複製スクリーン	ヘアピンスクリーン	T _m スクリーン	反復スクリーン
複製スクリーン	ヘアピンスクリーン	反復スクリーン	T _m スクリーン
T _m スクリーン	複製スクリーン	反復スクリーン	ヘアピンスクリーン
T _m スクリーン	複製スクリーン	ヘアピンスクリーン	反復スクリーン
T _m スクリーン	反復スクリーン	複製スクリーン	ヘアピンスクリーン
T _m スクリーン	反復スクリーン	ヘアピンスクリーン	複製スクリーン
T _m スクリーン	ヘアピンスクリーン	反復スクリーン	複製スクリーン
T _m スクリーン	ヘアピンスクリーン	複製スクリーン	反復スクリーン
反復スクリーン	ヘアピンスクリーン	T _m スクリーン	複製スクリーン
反復スクリーン	ヘアピンスクリーン	複製スクリーン	T _m スクリーン
反復スクリーン	T _m スクリーン	ヘアピンスクリーン	複製スクリーン
反復スクリーン	T _m スクリーン	複製スクリーン	ヘアピンスクリーン
反復スクリーン	複製スクリーン	T _m スクリーン	ヘアピンスクリーン
反復スクリーン	複製スクリーン	ヘアピンスクリーン	T _m スクリーン
ヘアピンスクリーン	複製スクリーン	T _m スクリーン	反復スクリーン
ヘアピンスクリーン	複製スクリーン	反復スクリーン	T _m スクリーン
ヘアピンスクリーン	T _m スクリーン	複製スクリーン	反復スクリーン
ヘアピンスクリーン	T _m スクリーン	反復スクリーン	複製スクリーン
ヘアピンスクリーン	反復スクリーン	T _m スクリーン	複製スクリーン
ヘアピンスクリーン	反復スクリーン	複製スクリーン	T _m スクリーン

10

20

30

40

50

【0098】

候補ホーマー（またはそのサブセット、ここで、サブセットは上記のように生成されたスコアに基づいて選択され得る）は、標的生物の様々な株からのゲノムのセットに対して、そして既知のゲノムの一般的なデータベースに対してアライメントされる。1) ホーマーが一致する株の数と、2) ホーマーに隣接する配列決定される期待伸長領域内の、これらの株の間の単一ヌクレオチド多型（S N P）の数（すなわち、配列決定された伸長産物の期待読取り長さを仮定して、ホーマーが明らかにすることを期待されるS N Pの数）とを考慮に入れる第2のスコアが各ホーマーに割り当てられる。

【0099】

第2の（またはスクリーニングされた）n - m e rは、例えば、被験者のゲノム（生物サンプルの場合）および対象の生物以外の生物（ウイルス、細菌、古細菌、真菌、および他の真核生物を含む）の配列決定されたゲノムを含むゲノムの排除セット内のゲノム中の配列と特異的にハイブリッド形成するものを排除するためにフィルタリングされる。特定の実施形態では、ゲノムの排除セットは、標的生物以外の共生生物、非病原性生物、および病原性生物を含む。特定の実施形態では、スクリーニングされたn - m e rは、排除セット内の任意の配列に対して、18、19、20、21、22、23、24、25、26、27、28、29、30、35、40、または45ヌクレオチドのウインドウ内に1、2、3、4、5、6、7、8、9、または10未満のミスマッチを含有する場合には排除される。特定の実施形態では、スクリーニングされたn - m e rは、少なくとも22ヌクレオチド（例えば、25ヌクレオチド）のウインドウ内に少なくとも19または20の一一致を含有する場合には除去される。候補n - m e rは、配列比較のために当該技術分野において知られている任意の手段によって、排除セットに対してスクリーニングすることができる。特定の実施形態では、候補n - m e rは、Megablastによって、排除セ

ットに対してスクリーニングされる。いくつかの実施形態では、スクリーニングされた n -mer は、ゲノム注釈（標的生物のゲノム中のこれらの位置など）を含有するようにフォーマットされ、他の実施形態では、これらはさらに、ゲノム注釈のないストリングとしてスクリーニングされる。

【0100】

特定の実施形態では、スクリーニングされた n -mer は、これらが少なくとも 1 つの追加のハイブリッド形成ゲノム中の配列と特異的にハイブリッド形成することを保証するためにさらにスクリーニングされる。いくつかの実施形態では、追加のハイブリッド形成ゲノムは、標的生物の追加の配列決定されたゲノムである。特定の実施形態では、追加のハイブリッド形成ゲノムは密接に関連しているが、例えば、同じ属または血清型に属する別個の種である。いくつかの実施形態では、スクリーニングされた n -mer は、ゲノムの排除セットと特異的にハイブリッド形成するものを排除するためのスクリーニングの前に、これらが追加のハイブリッド形成ゲノムと特異的にハイブリッド形成することを保証するためにスクリーニングされ、他の実施形態では、これらは後でスクリーニングされる。特定の実施形態では、スクリーニングされた n -mer は、ゲノムの排除セット内の配列と特異的にハイブリッド形成する配列を排除するためにスクリーニングされる前に、まず、これらが少なくとも 1 つの追加のハイブリッド形成ゲノムと特異的にハイブリッド形成することを保証するためにスクリーニングされる。

10

【0101】

いくつかの実施形態では、スクリーニングされた n -mer は、さらに、標的生物のゲノムにおいて特定の反復閾値未満、例えば、20、19、18、17、16、15、10、9、8、7、6、5、4、3、または 2 回未満などでこれらが標的生物のゲノム中で発生することを保証するためにスクリーニングされる。特定の実施形態では、スクリーニングされた n -mer は、標的生物のゲノムにおいて正確に 1 回発生する。

20

【0102】

スクリーニングされた n -mer が、特異的ハイブリダイゼーションの所望のパターン（すなわち、標的生物のゲノムと特異的にハイブリッド形成し、排除セットと特異的にハイブリッド形成しない）を保証するためにさらにスクリーニングされたら、連結側ホーマーおよび伸長側ホーマー候補は、候補プローブへ構築され得る。候補ホーマーの対は、人による事前選択またはコンピュータによる方法により選択される所定の対象の領域を捕獲するように選択され得る。他の実施形態では、候補相同プローブ配列の対は、所定長さの領域、例えば、少なくとも 2、3、4、5、6、7、8、9、10、15、20、30、40、50、60、80、100、125、150、200、250、300、350、400、500、600、700、800、900、1000、1200、1400、1600、1800、または 2000 ヌクレオチドの領域を捕獲するように選択される。いくつかの実施形態では、ホーマー対は、特定の標的生物株に対して決定される最大伸長距離の範囲内である。

30

【0103】

候補プローブのためのスコアは、1) プローブの結合が期待される株の各対の間で観察される SNP または挿入欠失（挿入または欠失またはこれらの組み合わせ）の数を、選択された最大値に至るまで計算し、2) (1) からの値の合計を生成して、プローブが明らかにし得る SNP または挿入欠失の総数をもたらし、3) プローブが機能する確率の推定値を、(2) からの合計に掛け合わせることによって生成することができる。このプロトコルはプローブの最終スコアである。プローブが機能する確率は、以下の何れかを考慮し得る：

40

- i) 連結ホーマーの配列
- ii) 伸長ホーマーの配列
- iii) プローブ骨格の配列、
- iv) 2 つのホーマー間の伸長領域の配列、
- v) 2 つのホーマー T_m 、

50

v i) プローブがそれ自体でヘアピンを形成する傾向、
 v i i) 伸長領域の配列組成、
 v i i i) 伸長領域の特定の部分、n - m e r、またはこれらの組み合わせの配列組成
 、および
 i x) 伸長領域の長さ。

【0104】

あるいは、プローブのスコアは、別の特定のゲノムセットを排除しながら、特定のゲノムセットまたは単一のゲノムのみとハイブリッド形成する、あるいはハイブリッド形成するのが好ましいプローブに対してスコアがより高いように生成され得る。

【0105】

いくつかの実施形態では、候補プローブのスコアは全ての対象の株の間で観察されるS N Pの合計を含まないが、代わりに、観察されるS N Pの数および特に選択された値のより小さい数の合計を含む。

【0106】

いくつかの実施形態では、プローブは、最終プローブのセット（「出力セット」）に順次付加される。上記のように計算された最高の候補プローブスコアを有するプローブをまず選択することができる。その時点で、全ての残存する候補プローブのスコアは、予め選択されたプローブにより区別されない株の間のS N Pを明らかにするプローブがより高くスコア化され、予め選択されたプローブにより区別される株の間を区別するS N Pを明らかにするプローブがより低くスコア化されるように再計算される。いくつかの実施形態では、残存する候補プローブのスコアは、これらが出力セットのために既に選択されたプローブとクロスハイブリッド形成する傾向を反映するように更新され得る。

【0107】

全ての可能なプローブのサブセットであり得るスコア化されたプローブのセットがあれば、株AおよびB（ここで、Aは株S1、S2、S3などのセット中にあり、Bは別の株のセット中にある）の全ての対が、少なくともある最小数のS N P、挿入欠失、またはその両方によって区別されることが期待されるまで、プローブスコアが減少する順にプローブを選択することによって、プローブは、最終プローブ出力セット中への含有について選択され得る。

【0108】

いくつかの実施形態では、全ての可能なプローブのサブセットであり得るスコア化されたプローブがあれば、1) 最高スコアを有するプローブを選択し、2) 既に選択されたプローブによって明らかにされたS N Pまたは挿入欠失の数を、まだ考慮中のプローブによって明らかにされた数から差し引くことにより、残存するプローブのスコアを再計算することによって、プローブは、最終プローブ出力セット中への含有について選択され得る。このようにして、既に選択された全てのプローブがあれば、プローブのスコアは、どのくらいの新しい情報をプローブが提供するかを反映するように更新され得る。

【0109】

ホーマーのプローブへの構築は、検出可能な部分およびプライマーなどの骨格配列の挿入を含み得る。

【0110】

特定の実施形態では、構築されたプローブの混合物は、二次構造を形成する可能性のある配列、または混合物中の他のプローブと特異的にハイブリッド形成する配列を排除するために、さらにスクリーニングされる。

【0111】

選択されたプローブのセットがあれば、プローブ選択ソフトウェアは、特定の標的生物株セットの中でプローブが明らかにするS N Pまたは挿入欠失の数に基づいた評価を提供することができる。ソフトウェアは、この情報を2 D グリッドの画像として表示することができ、一方の軸は株または種であり、他方の軸は特定のプローブの伸長領域における位置であり、そのグリッドエントリーの色は、その位置におけるその株 / 種の遺伝子型を示

10

20

30

40

50

す。ソフトウェアは、この情報をツリーとして表示することができ、ここで、ツリー中の各ノードはプローブに相当する。ノードからのエッジのセットは、そのプローブおよびツリー中の全ての上位(ancestor)プローブによって観察されるSNPまたは挿入欠失に従って区別することができないゲノムのセットに相当し得る。

【0112】

選択されたプローブのセットがあれば、ソフトウェアは、各プローブがハイブリッド形成すると期待される株の数に基づいた評価も提供することができる。ソフトウェアは、この情報を2Dグリッドの画像として表示することができ、一方の軸はゲノムであり、他方の軸はプローブであり、交差点における色はプローブがゲノムとハイブリッド形成し得るかどうかを示すか、あるいは色は、ハイブリダイゼーションの確率または可能性を示すことができる。

10

【0113】

さらなる実施形態では、プローブは、これらが株のセット間でいくつのSNPを明らかにするかに基づかないが、標的座位のリストに基づいて選択することができ、それぞれの座位は、単一のゲノム中の単一のヌクレオチドである。標的座位のセットは、1つまたは複数の基準ゲノム中の座位の塩基セットから誘導され、全ての関連ゲノム中の標的座位の完全セットは、基準ゲノムを互いのゲノムとアライメントさせることによって、塩基セットから誘導され得る。この方法は、例えば、薬物耐性突然変異が病原体の基準株において記載されており、その病原体の株または単離ゲノムのセットにおいてこれらの突然変異を検出し得るプローブが設計される場合に適用可能である。

20

【0114】

標的座位のリストに基づいてプローブを選択するこのような方法では、n-merrは上記のように生成され得る。これらの方では、プローブが機能する確率も上記のように計算され得る。しかしながら、このような方法では、プローブがランク付けおよび/または選択される最終スコアは、通常、プローブが機能する確率と、プローブの伸長領域（または、伸長領域の期待される配列決定読み取り）に含まれ得る標的座位の数との積に基づいている。従って、プローブは、多数の対象の株に対して、情報的価値のある産物（産物が標的座位を含有することを意味する）を生成すると期待される場合に高くスコア化され、多くの株において産物を生成しない場合、あるいはその産物が対象の座位を含有しない場合に低くスコア化され得る。

30

【0115】

いくつかの実施形態では、本明細書において記載される方法のいずれかによって生成される最終プローブは、相同プローブ配列（プローブアーム）があるゲノムセットの何れかに対してもはや完全に一致しないように修飾されてもよい。このゲノムセットは、プローブが設計されるゲノムセットであってもなくてもよく、そしてプローブがスコア化されるゲノムセットであってもなくてもよい。このような実施形態では、プローブをスコア化するために使用されるパラメータは、不完全な一致を相殺するように変更され得る。例えば、本方法では、通常よりも高い融解温度を有するプローブアームが選択されていてもよく、そしてプローブアームおよびゲノム間の不完全な一致の融解温度が通常の範囲内であるように、プローブアーム中のどの1つまたは複数のヌクレオチドに修飾するかが選択されていてもよい。

40

【0116】

特定の実施形態では、上記の方法は、少なくとも10、9、8、7、6、5、4、3、または2メガベースの標的ゲノムにおいて、シングルコアPentium Xeon 2.5ghzプロセッサを用いて16、14、12、10、8、6、もしくは4日間、または72、48、36、24、12、10、8、6、もしくは4時間かかる。

【0117】

一般に、プローブは、上記のように特定の標的生物に対して作成される。特定の実施形態では、複数の生物に向けられたプローブを含む混合物、例えばパネルは、プローブクロスハイブリダイゼーションを最小限にするか排除するため、例えば、混合物中の1つまた

50

は複数の相同プローブ配列またはプローブ骨格配列と特異的にハイブリッド形成するプローブを排除するために、パネルにより検出される各標的生物に対する候補プローブを互いにスクリーニングすることによって、例えばペアワイズ比較によって編集される。

【0118】

図7は、本発明によって提供されるプローブおよび混合物の製造方法の例示的な実行のフローチャートである。図7は、例えば、例えば標的ゲノムの提供100、およびn - m e rのセットへのスライシングの実施100を示す。n - m e rは、一連のスクリーン250（例えば、ヘアピン（253）、T_m（254）、反復（252）および複製（251）スクリーン）を含むプロセス200によってスクリーニングされる。次にn - m e rは、排除セット200および1つまたは複数の追加のハイブリッド形成ゲノム300に対する特異的ハイブリダイゼーションの所望のパターンについて、プロセス300によってスクリーニングされ、ここで、排除セット200および追加のハイブリッド形成ゲノム300はデータベースから得られる。例えば、プロセスは、少なくとも1つの追加のハイブリッド形成ゲノムに対するハイブリダイゼーションについてのフィルタリング330、2未満の反復閾値（例えば、1つの標的ゲノムにつき1ヒット）についてのフィルタリング340、被験者（例えば、ヒト）ゲノムに対するフィルタリング350、および排除セットに対するフィルタリング360を含んでいてもよい。スクリーニングされたn - m e rは、注釈がなければ、ゲノム中のその位置を決定するために標的ゲノムに注釈が付けられてもよい（370）。プローブは、フィルタ425によって対象の領域を捕獲するように対をフィルタリング（420）するプロセス400において構築され、例えば、フィルタ425-1によって、指定の対象の領域の長さを有するように、そして骨格配列40を含むようにフィルタリングされる。プローブは、二次構造を排除するためにフィルタリングされる（450）。プローブの混合物（例えば、パネル）はプロセス500によって作成され、フィルタリング（550）されて、混合物中の他のプローブ50に対する特異的ハイブリダイゼーションが排除される。本出願の教示に従って当業者により実験的検証600が実施されてもよい。

【0119】

上記で同定されるコンポーネントのそれぞれのうちの1つだけが上記図面に示されているが、当業者は、これらのコンポーネントがいずれもいくつでも提供され得ることを理解するであろう。さらに、当業者は、開示されるシステムのそれぞれの1つまたは複数のコンポーネントが、図中に示される別のコンポーネントに結合され得る、または取り込まれ得ることを認識するであろう。図中に示されるコンポーネントの1つまたは複数は、1つまたは複数のコンピューティングシステムにおいてソフトウェアで実行され得る。例えば、これらは、プロセッサにより実行されたときにコンピュータに方法のステップを実施させるコンピュータ可読指令の1つまたは複数のコンピュータユニットを含み得る1つまたは複数のアプリケーションを含むことができる。コンピュータ可読指令は、メモリまたはディスクなどのコンピュータ可読媒体上に保存され得る。このような媒体は、通常、非一時的な保存を提供する。あるいは、図中に示されるコンポーネントの1つまたは複数は、例えば、専用コンピュータまたは汎用コンピュータなどのハードウェアコンポーネントまたはハードウェアおよびソフトウェアの組み合わせであってもよい。コンピュータまたはコンピュータシステムは、内部または外部データベースも含むことができる。コンピュータまたはコンピュータシステムのコンポーネントは、ローカルバスインターフェースによって接続されていてもよい。

【0120】

当業者は、上記の段階が別個のソフトウェアモジュールで具体化されてもよいことを理解するであろう。開示されるコンポーネントは別々のユニットとして上記で説明されているが、当業者は、1つまたは複数のユニットにより提供される機能性が結合されてもよいことを認識するであろう。当業者により理解されるように、ユニットの1つまたは複数は任意的なものであり、特定の実施形態における実行から除外されてもよい。

【0121】

10

20

30

30

40

50

3.1.1 ホーマーおよび構築プローブのスコアリングのための例示的なアルゴリズム
上記の方法を含むプローブ設計の方法は、ホーマーのスコアリングおよび完成プローブのスコアリングのための方法を含むことができ、ここで、スコアはプローブが機能する確率に相当する。

【0122】

ホーマーおよびプローブのスコアリングアルゴリズムの中核は融解温度に基づくことができる。ある温度において二本鎖型で存在し得る核酸分子の集団の割合を説明するために、一般に、ロジスティック関数が使用される。Tが実験温度であり、T_mが核酸の融解温度であり、そしてsが二本鎖型から解離型への移行の勾配を説明するパラメータであるとすると、

$$p(T, s) = 1 / (1 + e^{-(T_m - T) / s})$$

は、二本鎖型で存在する集団の割合である（図8においてT_mの関数で示される）。いくつかの実施形態では、分子反転プローブが、標的配列を上手く増幅する見込みが高いことを反映するスコアを有するために、いくつかのことが発生しなければならない：

- 1) プローブの開始アームは、標的核酸とハイブリッド形成しなければならない、
- 2) ポリメラーゼは、伸長を開始しなければならない、
- 3) プローブの連結アームは、標的核酸とハイブリッド形成しなければならない、
- 4) 伸長は、伸長アームと連結アームの間の錠型配列全体を交差しなければならない、
そして
- 5) リガーゼは、伸長産物を連結アームに連結させなければならない。

【0123】

いくつかの実施形態では、上記の事象（1）および（3）は、プローブアームの融解温度に基づいたロジスティック関数を用いて説明することができる。事象（2）および（5）は、開始および連結部位を直接包囲するヌクレオチドに関して説明することができる（例えば、それぞれ、プローブアームの端部の2つの核酸および伸長ロジスティック関数領域の端部の2つの核酸によって説明することができる）。事象（4）は、伸長領域のヌクレオチド組成によって説明される。

【0124】

事象（1）および（3）は同一の式およびパラメータを用いて計算されてもよいし、あるいは別々に計算されてもよい。T_mは、プローブアームの融解温度であるとされてもよい。プローブアームがハイブリッド形成する確率は、

$$P_{hyb\ on\ Target} = (p(T, s) / (p(T, s) + \text{sum}_{other}(p_{other}(T, s))))^* p(T, s)$$

で説明することができ、式中、sum_{other}(p_{other}(T, s))は、プローブアームのゲノムへの意図されない一致またはオフターゲットの一致の融解温度に対するロジスティック関数の合計である。従って、モデルは、全ての部位にわたるハイブリダイゼーションに対する意図される部位へのハイブリダイゼーションの比率に、正しい部位において利用可能である場合にプローブアームがハイブリッド形成する確率を乗じたものとして、プローブアームがハイブリッド形成する確率を説明することができる。

【0125】

ゲノムへのプローブアームの各一致のための融解温度（オントーゲットの一致およびいくらかのオフターゲット、すなわち不完全な一致）は、プローブアームとオフターゲット結合部位との間のミスマッチ、ハイブリダイゼーション混合物中のプローブ核酸の濃度、およびハイブリダイゼーション混合物中の様々なイオン（例えば、Na⁺、Mg⁺⁺、K⁺、Tris）の濃度を考慮に入れることができる標準的な融解温度計算機を用いて計算することができる。

【0126】

モデルは、オフターゲットの一致の合計が、プローブアーム配列のゲノム配列への不正確なアライメントによって決定されるオフターゲットの一致と、プローブアームのT_mによって予測されるオフターゲットの一致の一般的なセットとの両方を含むようにさらに伸

10

20

30

40

50

長され得る。例えば、予測されるオフターゲットの一致のセットの合計は、30から $T_m - k$ （ここで、 $k = 10$ である）までの t （プローブアームの融解温度）のそれぞれの値において、予測されるオフターゲットの一致の数が

$$a^{\wedge}(T_m - t)$$

（式中、 a は、1.4の値を有する定数である）に等しくなるように生成され得る。 t のそれぞれの値において、プローブアームのゲノムまたはゲノムセットへのオフターゲットの一致または不完全な一致の数は、上記の式に従って予測される。オフターゲットの一致の数は、 t が減少するにつれて指数関数的に増大すると推定される。すなわち、オフターゲットの一致の数は、オントーゲットの一致およびオフターゲットの一致（または一致の類）間の融解温度の差が増大するにつれて指数関数的に増大し得る。これは、プローブアームと、ゲノム中のオフターゲット部位との間の一致がより短くなるので、期待される挙動であり得る。従って、融解温度は低下し、このような一致の数はより大きくなり得る。特定の融解温度における読み取りカウントによって決定されるような、プローブの効率に対する融解温度の効果は、連結および伸長プローブアーム（ホーマー）のそれぞれに対して、図9および10においてそれぞれ示される（図10の「開始ホーマーは」伸長プローブアームを指し、両方の図中の上側の円の弧は、その値を中心とする T_m のピン（bin）に対する平均配列読み取りカウントを示し、両方の図中の中央の円の弧は〔すなわち、底部の円のフラットな線ではない〕はサンプル標準偏差を示す）。

10

【0127】

事象（4）の成功した伸長の確率は、伸長領域におけるジヌクレオチド配列を横切る伸長確率の積で説明され得る。各ジヌクレオチドには、ポリメラーゼがそれを上手く取り込む確率を割り当てることができ、伸長領域を交差するポリメラーゼの確率は、伸長領域を横切るこれらの確率の積であり得る。

20

【0128】

MI P（分子反転プローブ）産物配列決定読み取りの公的なデータベースは、例えば、Porreca et al. Nat Methods. Nov; 4(1): 931-6 (2007)による“Multiplex amplification of large sets of human exons”、およびDeng et al., Nat Biotechnol. 27(4): 353-60 (2009)による“Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogrammingを含む、上記のモデルのパラメータを知るために使用することができる。

30

【0129】

3.2 プローブ捕獲および検出

本発明は、試験サンプル中の1つまたは複数の対象の生物の存在の検出方法を提供する。特定の実施形態では、方法は、上記のように、捕獲反応において上記のプローブを含む混合物を上記の試験サンプルのいずれかと接触させるステップを含む。特定の実施形態では、プローブを含む混合物は、ポリメラーゼ酵素およびヌクレオチド三リン酸（NTP）と共に、試験サンプルから抽出された核酸と接触され、混合物中の少なくとも1つの相同プローブ配列のポリメラーゼ依存性伸長によって、少なくとも1つの対象の領域を捕獲する。特定の実施形態では、相同プローブ配列のポリメラーゼ依存性伸長の後に、伸長された（すなわち、ポリメラーゼによって）相同プローブ配列の末端が他の相同プローブ配列の末端に連結され、対象の生物のゲノムからの対象の領域を含有する環状化プローブが生成される。いくつかの実施形態では、連結反応は、標的アームが標的とハイブリッド形成される間に発生する。他の実施形態では、標的アームは、他のプローブ分子に対するトランク連結よりも自己連結を好む反応条件下の溶液（例えば、希釈連結溶液）中で、標的から解離されて連結される。実例については、図2(A)または図2(C)を参照されたい。

40

【0130】

50

図2(C)は、本発明によって提供される方法の1つの特定の実施形態を図解する。簡単に言うと、対象の生物中の標的配列に対するプローブのハイブリダイゼーションの後、ポリメラーゼ媒介性の標的-配列指向性の、3'相同プローブ配列に対するヌクレオチドの付加が行われ、プローブの5'相同プローブ配列における妨害物のために終端する。連結反応は、末端の3'ヌクレオチドをアームH2の5'ヌクレオチドに結合させる。

【0131】

サンプルは、一本鎖DNAを消化するためにエンドヌクレアーゼで処理される。プローブ骨格に相補的なプライマーは、配列決定のためにMIPをdsDNAに増幅する。サンプル反応産物または増幅反応産物の多重化のために、この段階の増幅プライマーはサンプル特異的ヌクレオチドバーコード配列を含有することができ、例えば、これらはアダプタマーブライマーである。独特のプライマーバーコード分子配列は、そのため、各試験サンプルを同定する。例えば、100プローブのパネルは、50の個々の試験サンプルと接触される。配列読取りにおいて検出される相同プローブ配列は、対象の生物、例えば、特定の病原体または株を同定する。各試験サンプルの増幅反応は、1つの独特のプローブセットを用いて行われる。増幅プライマー内の各バーコードは、患者に対する識別子の機能を果たすために使用することができ、例えば、バーコードを含有する。従って、50対の増幅プライマー(増幅反応産物のそれぞれに対して1つ)および1つの100プローブのパネル(例えば、100の対象の生物に対する)が、50サンプルの多重アッセイのために必要とされる。

【0132】

図2(A)は、代替の実施形態を図解する。いくつかの実施形態では、各試験サンプルは、独特のプローブセット、例えばパネルと接触される。各試験サンプルについての増幅反応産物がプールされる。各試験サンプルは、独特的プローブセットと接触されるので、相同プローブ配列および捕獲配列は、標的生物および試験サンプルの両方を同定する。いくつかの実施形態では、プローブ認識配列をさらに含む従来のプライマー対(すなわち、相同プローブ配列を含む)は、増幅アーチファクトを低減するために少ないサイクル数(10未満)を用いて、対象の領域を増幅するためにサンプル核酸と接触される。次に、従来のプライマー対増幅産物のプローブ認識配列に向けられたプローブが適用される。ポリメラーゼ伸長および連結は、従来のプライマー対の相同プローブ配列および介在する対象の領域を捕獲する。独特のバーコード化プローブ配列は、サンプル(例えば、患者)の多重化を可能にする。配列読取りは、相同プローブ配列(対象の生物を同定する)およびバーコード(サンプル、例えば患者に関連する)を含むであろう。100プローブパネルおよび50試験サンプルの例では、対象の生物はそれぞれ、対象の生物、例えば病原体を同定する一対の相同プローブ配列を有する。試験サンプルは、独特的プローブセットと接触されるであろう。プローブ骨格内の各バーコードは、サンプル識別子の機能を果たすために使用され得る。従って、この説明的な実施形態では、それぞれに100プローブを有する50セットのプローブが使用される。

【0133】

本発明によって提供される方法で使用するためのポリメラーゼには、Taqポリメラーゼ(Lawyer et al., J. Biol. Chem., 264:6427-6437(1989)、Genbank受入:P19821)、Lawyer et al., PCR Meth. Appl., 2:275-287(1993)に記載される5'3'ヌクレアーゼ欠損「Stuffel」断片を含む)、PHUSION(商標)高忠実度(high fidelity)組換えポリメラーゼ(NEB)、およびパイロコッカス・フリオサス(Pyrococcus furiosus)(Pfu)ポリメラーゼ(例えば、米国特許第5,545,552号明細書を参照)、ならびにTopoTaqおよびPfuC2などのらせん-ヘアピン-らせんドメインを含むポリメラーゼ(Pavlov et al., PNAS, 99:13510-15(2002))が含まれる。より特定の実施形態では、ポリメラーゼは、TaqポリメラーゼのStuffel断片(3'5'フルーフリーディング活性がさらに欠如している)などの5'3'ヌクレアーゼ

10

20

30

40

50

欠損である。5'，3'エキソヌクレアーゼ活性が欠如したポリメラーゼは、例えば、スクリーニングまたは合理的設計の方法に基づいて、当該技術分野において知られている手段によって生成することができる。例えば、ポリメラーゼ変異体は、TaqのStoffe1断片への1つまたは複数のポリメラーゼの配列アライメントに基づいて、および/または、解かれたポリメラーゼ構造に配列を「通すこと（threading）」によって設計することができる（例えば、MMDB ID 56530、81884および81885）。

【0134】

特定の実施形態では、本発明の方法において使用するためのポリメラーゼは、Pfu、T4DNAポリメラーゼ、またはT7DNAポリメラーゼなどの非置換ポリメラーゼである。他の実施形態では、本発明によって提供される方法で使用するためのポリメラーゼは等温增幅に適切なポリメラーゼであり、捕獲および/または增幅反応は、例えば、金属イオン濃度を調節することによって、ならびに/あるいは、特定のポリメラーゼおよび/または追加の酵素、例えば、ヘリカーゼまたはニッキング酵素（プライマー生成RCRおよびEXPARなど）を用いることによって等温的に実施される。例えば、米国特許第6,566,103号明細書、Murakami et al., Nucl. Acid. Res., 37(3) e19 (2009)、Tan et al., Biochemistry, 47: 9987-99 (2008)、Vincent et al., EMBO Rep., 5(8): 795-800 (2004)を参照されたい。等温增幅において使用するためのポリメラーゼとしては、例えば、Bst、Bsuおよびphi 29 DNAポリメラーゼ、および大腸菌（E. coli）DNAポリメラーゼIが挙げられる。

10

20

30

30

【0135】

他の実施形態では、プローブの混合物は、上記のように、捕獲反応において、試験サンプルから抽出された核酸、リガーゼ酵素、およびn-merオリゴヌクレオチドのプールと接触される。実例については、図2(B)を参照されたい。特定の実施形態では、n-merオリゴヌクレオチドは、少なくとも2、3、4、5、6、7、8、9、10、12、14、16、18、20、22、24または25ヌクレオチドの長さである。より特定の実施形態では、これらはランダムな六量体である。他の実施形態では、これらは、相同プローブ配列とハイブリッド形成する第1および第2の標的配列間の対象の領域の長さのポリヌクレオチドである。いくつかの実施形態では、n-merオリゴヌクレオチドは、1、2、3、4、5、6、7、8、9、もしくは10のロッド核酸（LNA）、または10、20、30、40、50、60、70、80、90、もしくは100%のLNAを含有する。

40

【0136】

リガーゼ酵素は、n-merオリゴヌクレオチドを本発明によって提供されるプローブに連結させて、対象の生物からの対象の領域を含有する環状化プローブを生成する。プローブ骨格に相補的なプライマーは、配列決定のために、プローブをdsDNAに増幅する。いくつかの実施形態では、例えば多重化のために、増幅プライマーはアダプタマープライマーであり、サンプル識別バーコード配列を含有する。独特のバーコード配列は、そのため、各試験サンプルを多重に同定する。各病原体は、配列読取り内の相同プローブ配列および連結されたn-merの独特的組み合わせによって同定される。より特定の実施形態では、n-merオリゴヌクレオチドは、1つまたは複数（例えば、1、2、3、4、5、6、または7）のロッド核酸を含む7-merであり、相同プローブ配列は10または12塩基であり、7塩基の対象の領域によって分離される標的配列と特異的にハイブリッド形成する。

40

【0137】

本発明の方法において使用するためのリガーゼは、T4、T7、および熱安定性リガーゼ、例えば、Taqリガーゼ（Takahashi et al., J. Biol. Chem., 259: 10041-47 (1984)、および国際公開第91/17239号パンフレットに記載される）、およびAMPLIGASE（商標）などを含む。

50

【0138】

特定の他の実施形態では、本発明によって提供される従来のPCRプライマーの対（従来のプライマー対）を含む混合物は、対象の生物中の2つの標的領域間の対象の領域を増幅するために、サンプル核酸と接触される。特定の実施形態では、限定された数の増幅ステップが実施される。特定の実施形態では、25、20、15、10、9、8、7、6、5、4、3、または2回未満の増幅サイクルが実施される。特定の実施形態では、従来のプライマー対の混合物は、対象の領域を増幅するために、試験サンプルから抽出された核酸、ポリメラーゼ、およびヌクレオチド三リン酸と接触される。この方法の実例は図3に示される。従来のプライマー対の多数の組み合わせを使用して、同じサンプルチューブ内で、あるいはブーリングのために別々に、反応を多重化することができる、いくつかの実施形態では、従来のプライマー対中の普遍的プローブ認識配列（例えば、バーコード）に結合するプライマーは、ヌクレオチドバーコード、および次世代DNA配列決定技術プライマーのための認識部位を導入する。

10

【0139】

本発明の一部として、従来のプライマー対は、様々な追加の方法において使用することができる。例えば、いくつかの実施形態では、従来のプライマー対は、少なくとも1つの標的核酸を含有する疑いがあるサンプル核酸と接触され得る。特定の実施形態では、対象の領域をサンプル核酸から直接増幅するためにPCRが使用され得る。他の実施形態では、捕獲反応産物、例えば、1つまたは複数の環状化プローブを増幅するために、従来のプライマー対が使用され得る。他の実施形態では、対象の領域を含有する疑いがあるサンプル核酸は、従来のプライマー対を用いて増幅され、次に、環状化捕獲のために本発明によって提供されるプローブと接触される。いくつかの実施形態では、従来のプライマー対は、サンプル核酸、およびビオチン化ヌクレオチドなどの修飾ヌクレオチドと接触される。ビオチン化ヌクレオチドなどの修飾ヌクレオチドを用いるいくつかの実施形態では、次に、得られた捕獲または増幅反応産物は、例えば、ステプトアビジン（steptravidin）基質によるアフィニティ捕獲によって、次の処理、例えば、本発明によって提供されるプローブを用いる環状化捕獲のために単離することができる。さらなる実施形態では、単一の従来のプライマーは、サンプル核酸中の対象の領域の線状増幅のために使用され、次に、環状化捕獲のために、本発明によって提供されるプローブと接触され得る、他の実施形態では、5'ビオチン部分を含有する単一の従来のプライマーは標的配列を増幅するために使用されてから、例えば、本発明によって提供される特異的な従来のプライマー対を用いる直接配列決定によって、あるいはランダムな六量体プライミングによって、配列決定のためにストレプトアビジン捕獲を用いてサンプルから濃縮されてもよいし、あるいは、本発明によって提供されるプローブを用いる環状化捕獲のために使用されてもよい。

20

30

30

【0140】

特定の実施形態では、捕獲反応を含む方法は、さらに、捕獲反応産物を1つまたは複数のエキソヌクレアーゼと接触させて、線状核酸を除去するステップを含む。特定の実施形態では、エキソヌクレアーゼは、exo I、exo I I I、exo V I I、およびexo Vのうちの少なくとも1つを含む。さらに特定の組み合わせでは、エキソヌクレアーゼは、100:1、50:1、25:1、10:1、5:1、2:1、1:1、1:2、1:5、1:10、1:25、1:50、または1:100（単位対単位）までのエキソヌクレアーゼIおよびエキソヌクレアーゼI I Iの混合物である。

40

【0141】

特定の実施形態では、本発明の方法はさらに、捕獲反応産物を増幅反応において増幅するステップを含む。核酸を増幅する多数の方法は当該技術分野において知られており、ポリメラーゼ連鎖反応（例えば、米国特許第4,683,195号明細書および同第4,683,202号明細書、ならびにMcPherson and Moller, PCR (the basics), Taylor & Francis; 2 edition (March 30, 2006)を参照）、OLA（オリゴヌクレオチド連結増幅）（例えば

50

、米国特許第5,185,243号明細書、同第5,679,524号明細書、および同第5,573,907号明細書を参照)、ローリング・サークル増幅(「RCA」、Banner et al., Nuc. Acids Res., 26:5073-78(1998)、Barany, PNAS, 88:189-93(1991)、およびLizardi et al., Nat. Genet. 19:225-32(1998)に記載)、およびストランド置換増幅(SDA、米国特許第5,455,166号明細書および同第5,130,238号明細書に記載)を含む。特定の実施形態では、増幅は、RCAなどの線状増幅である。より特定の実施形態では、捕獲反応産物(例えば、環状化プローブ)はRCAにおける鑄型として使用され、長い線状の反復ssDNA産物を生成する。いくつかの実施形態では、RCA反応は、サンプルを、ビオチン化ヌクレオチド、LNAヌクレオチド、またはIsodCもしくはIsodGなどの人工的な塩基対、または脱塩基フラン(dSpacerなど)などの修飾ヌクレオチドと接触させて、アフィニティ濃縮および精製を容易にすることを含むことができる。特定の実施形態では、線状反復ssDNAを含む増幅反応産物は、本発明によって提供される従来のプライマーと接触されて、長さ2、3、4、5、6、7、10、15、20、30、40、50、75、100、500ヌクレオチドを有する二本鎖DNAの短い伸長を生成することができる。特定の実施形態では、伸長の長さは、このポリメラーゼの延長の最適温度における伸長ステップの時間(例えば、37、42、45、68、72、74を含む温度で5、10、15、20、40、60秒)によって調節され得る。他の実施形態では、伸長の長さは、さらなる延長が防止されたヌクレオチド類似体、例えば、ジデオキシシトシン(dideoxyCytosine)、または3'修飾、例えばビオチン、またはアミノ基で終結されるは炭素スペーサーを有するヌクレオチドを反応中に混合することによって調節される。追加の特定の実施形態では、プライマーは、線状反復ssDNAのRCA増幅反応産物と接触され、単一のPCRサイクルのためのポリメラーゼによって伸長されて、RCA産物の反復単位に対する相補的配列を含有する短い一本鎖DNAを生成する。より特定の実施形態では、線状反復ssDNAのRCA増幅反応産物と接触されたプライマーは、制限酵素切断部位を含むdsDNA領域を生成する。従って、特定の実施形態では、プライマーが線状反復ssDNAのRCA増幅反応産物とハイブリッド形成して二本鎖DNA領域を形成する場合、増幅反応産物は、制限酵素と接触されてより短い断片を生成する。

【0142】

特定の実施形態では、増幅反応はアダプタマーブライマーを使用する。いくつかの実施形態では、増幅反応は、サンプル特異的プライマー、すなわちサンプルを同定するプローブ中に存在する配列とハイブリッド形成するプライマーを使用する。特定の実施形態では、増幅アーチファクトを回避するために、少數の増幅サイクル、例えば、25、20、15、10、9、8、7、6、5回未満のサイクルが使用される。

【0143】

特定の実施形態では、本発明によって提供される方法は、サンプル核酸、捕獲反応産物または増幅反応産物を、ビオチン分子などの捕獲されるように設計された部分、およびサンプル核酸、捕獲反応産物、または増幅反応産物とハイブリッド形成することができる核酸配列を含む二次捕獲オリゴヌクレオチド捕獲プローブと接触させるステップを含むことができる。このようなオリゴヌクレオチド(ビオチン化オリゴヌクレオチドなど)は、アフィニティ精製を用いてその標的核酸を濃縮するために使用され得る。いくつかの実施形態では、ビオチン化オリゴヌクレオチドは、捕獲された配列(すなわち、それは対象の領域に対して相補的である)、相同プローブ配列、または骨格配列(例えば、バーコード配列など)と特異的にハイブリッド形成し得る。特定の実施形態では、ビオチン化プローブは、好熱性または中温性ポリメラーゼを用いて、サンプル核酸、捕獲反応産物または増幅反応産物上で伸長され得る。より特定の実施形態では、方法は、ビオチン:ストレプトアビジン相互作用を用いる特定の捕獲反応産物の濃縮のために、捕獲反応産物をビオチン化オリゴヌクレオチドと接触させることを含む。

【0144】

10

20

30

30

40

50

本発明の方法によって捕獲される配列は、例えば、アレイハイブリダイゼーションまたは直接配列決定を含む手段によって検出することができる。いくつかの実施形態では、捕獲された配列は、増幅することなく配列決定によって検出され得る。多数の配列決定方法は当該技術分野において知られており、本発明の方法において使用することができ、例えば、米国特許第6,946,249号明細書およびMetzker, Nat. Rev. News, Genetics, 11:31-46(2010)、Ansorge, Nat. Biotechnol., 25(4):195-203(2009)、Shendure and Ji, Nat. Biotechnol., 26(10):1135-45(2008)、Shendure et al., Nat. Rev. Genet. 5:335-44(2004)において概説される。いくつかの実施形態では、配列決定方は、DNAポリメラーゼまたはDNAリガーゼのいずれかの特異性に依存しており、例えば、ピロシーケンス、塩基伸長配列決定(単一の塩基の段階的伸長)、合成による多塩基配列決定(例えば、末端標識ヌクレオチドによる配列決定を含む)、およびゆらぎ(wobble)配列決定(連結に基づく)が含まれる。伸長配列決定は、例えば、米国特許第5,302,509号明細書で開示される。末端リン酸標識ヌクレオチドおよびその使用方法の例示的な実施形態は、例えば、米国特許第7,361,466号明細書、米国特許出願公開第2007/0141598号明細書(2007年6月21日に公開)、およびEid et al., Science, 323:133-138(2009)に記載されている。リガーゼに基づく配列決定方法は、例えば、米国特許第5,750,341号明細書、PCT公報国際公開第06/073504号パンフレット、およびShendure et al., Science, 309:1728-1732(2005)に開示されている。特定の実施形態では、本発明によって提供される方法で使用される配列決定技術は、Sanger配列決定、微量電気泳動配列決定、ナノポア配列決定、ハイブリダイゼーションによる配列決定(例えば、アレイベースの配列決定)、単一分子のリアルタイム観察、および環状アレイ配列決定を含み、例えば、ピロシーケンス(例えば、454 SEQUENCING(登録商標)、例えば、Margulies et al., Nature, 437:376-380(2005)を参照)、ILLUMINA(登録商標)またはSOLLEXA(登録商標)配列決定(例えば、Turcatti et al., Nucl. Acids Res., 36:e25(2008)を参照、また、米国特許第7,598,035号明細書、同第7,282,370号明細書、同第7,232,656号明細書、および同第7,115,400号明細書も参照)、ポロニー配列決定(例えば、SOLID(商標)、Shendure et al. 2005を参照)、および合成による配列決定(例えば、HELIКОS(登録商標)、例えば、Harris et al., Science, 320:106-109(2008)を参照)が含まれる。

【0145】

特定の実施形態では、捕獲プローブは、特定の配列決定技術による配列決定のための処理を容易にする配列を含有し、例えば、合成による配列決定のためのアンカー部位、配列決定反応の開始のためのプライマー部位、または特定の増幅産物の配列決定のためにオリゴヌクレオチドアダプターの連結を改善するための開裂を可能にする制限酵素部位としての役割を果たすことができる配列を含有する。いくつかの実施形態では、環状化捕獲プローブは、捕獲プローブのポリメラーゼ媒介性の伸長を刺激して、元の環状プローブの少なくとも1つ~100万またはそれ以上の鎖状体化コピーを含む、環状化プローブの配列に相補的な配列を生成するオリゴヌクレオチドによって接触される。

【0146】

本発明によって提供される混合物および方法は、任意の適切な検出手段(上記のものを含むがこれらに限定されない)と共に使用するために容易に適合させることができる。ILLUMINA(登録商標)またはSOLLEXA(登録商標)配列決定を用いる特定の実施形態では、より短い相同プローブ配列が、本発明によって提供されるプローブ、および従来のプライマー対において使用され得る。より特定の実施形態では、相同プローブ配列は、約8、9、10、11、12、13、14、15、16、17、18、19、また

10

20

30

40

50

は20塩基であろう。より特定の実施形態では、プローブまたは従来のプライマー対の標的配列間の対象の領域は、約2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、または50塩基である。さらにより特定の実施形態では、本発明によって提供されるプローブは、ポリメラーゼ依存性の合成および連結によって、あるいは約2、3、4、5、6、7、8、9、10、11、12、13、14、15、20、25、30、35、40、45、または50塩基のn-merオリゴヌクレオチドの連結によって環状化され得る。さらにより特定の実施形態では、対象の領域は約7塩基であり、相同プローブ配列は10または12塩基である。さらなる実施形態では、ロックド核酸を含む7-merオリゴヌクレオチドが本発明によって提供されるプローブによって連結され、さらにより特定の実施形態では、7-merオリゴヌクレオチドは、少なくとも1、2、3、4、5、6、または7のロックド核酸(LNA)を含む。

10

【0147】

他の実施形態では、捕獲または增幅反応産物は、例えば、Binaladen et al, PLoS One. 2(2):e197(2007)に開示されるような合成によるエマルジョン滴(emulsion droplet)配列決定によって配列決定することができる。特定の実施形態では、エマルジョンPCRおよび合成による配列決定のために捕獲されたDNAのエマルジョンを容易にするために、捕獲産物はRCRにより増幅されて、単一DNA分子内により高いコピー数の捕獲産物を生成し得る。例えば、Drmanac et al, Science 327(5961):78-81(2010)を参考されたい。

20

【0148】

特定の実施形態では、異なるサンプルを含有する捕獲反応産物および/または增幅反応産物は、検出の前に結合される。特定の実施形態では、捕獲および/または增幅反応産物は、検出の前に組み合わせ的にプールされ、例えば、個々の捕獲反応産物および/または增幅反応産物のM×Nアレイは行および列によってプールされ、プールが検出される。行および列プールからの結果は、次に、デコンボリューションされて、個々のサンプルの結果を提供することができる。より高次元のアレイおよびプールが同様に使用されてもよい。他の実施形態では、捕獲反応産物および/または增幅反応産物は、識別バーコード配列を含有する。特定の実施形態では、增幅プライマーは、サンプル特異的バーコード配列を含有する。従って、捕獲反応産物および/または增幅反応産物のプール中に含有される配列のサンプル源は、そのバーコード配列によって同定される。

30

【0149】

本発明によって提供される方法は、特定の標的増幅産物または増幅産物セットなどの、捕獲反応産物または増幅反応産物中の特定の核酸を直接検出することを含んでいてもよい。従って、いくつかの実施形態では、本発明の混合物は、特殊化したプローブセットを含み、例えば、5' 3'エキソヌクレアーゼ活性を有するDNAポリメラーゼによって放出される検出可能なレポーターおよび失活剤部分を含有するハイブリッド形成可能なプローブを使用するTaqman(商標)(米国特許第5,538,848号明細書)、反対の末端にレポーターおよび消光部分を有するヘアピンプローブを使用する分子ビーコン(米国特許第5,925,517号明細書)、蛍光ドナーおよびアクセプター部分をそれぞれ有する一対の隣接するプライマーを使用する蛍光共鳴エネルギー移動(FRET)プライマー(米国特許第6,174,670号明細書)、および、標的に結合されたときにだけ蛍光を発する単一の短いプローブであるLIGHTUP(商標)(米国特許第6,329,144号明細書)が含まれる。同様に、SCORPION(商標)(米国特許第6,326,145号明細書)およびSIMPLEPROBES(商標)(米国特許第6,635,427号明細書)は、単一のレポーター/色素プローブを使用する。増幅産物検出プローブは、使用される特定の検出モダリティに従って、そして上記の特許において考察されるように設計される。特定の実施形態では、特定の捕獲反応産物または増幅反応産物を検出するための定量的なリアルタイムPCRアッセイは、ILLUMINA(登録商標

40

50

) E C O リアルタイム P C R システム (商標) において実施され得る。

【 0 1 5 0 】

特定の実施形態では、本発明の方法は、試験サンプル中の対象の生物の濃度を推定するために、サンプル内部較正核酸 (S I C) の使用を含む。これは、対象の生物からの配列の頻度を S I C の既知の濃度に対して較正することによって行われ、試験サンプル中の対象の生物の推定濃度が提供される。より特定の実施形態では、対象の生物の推定濃度は、病状および / またはおそらく臨床診断に関連する基準濃度の対象の生物のデータベースと比較される。

【 0 1 5 1 】

いくつかの実施形態では、本発明の方法はさらに、結果をフォーマットして医師の意思決定を通知するステップを含む。「結果」は標的生物の検出の成果を指し、例えば、バイナリ (例えば、+ / -) 検出および濃度の推定値を含み、特に、捕獲反応産物または増幅反応産物の配列決定の結果に基づくことができる。特定の実施形態では、フォーマッティングは、場合により統計的信頼区間を含む試験サンプル中の生物濃度の推定値を提示することを含む。より特定の実施形態では、フォーマッティングはさらに、結果のカラーコーディングを含む。特定の実施形態では、フォーマッティングは、例えば、入院、生菌治療 (p r o b i o t i c t r e a t m e n t) 、抗生物質治療、および化学療法を含む、治療的介入のための勧告を含む。いくつかの実施形態では、フォーマッティングは、以下のうちの 1 つまたは複数を含む：査読された医学文献および経験的に定義されたサンプル結果のデータベース統計への言及。結果の例示的なフォーマットは図 6 に示される。

10

20

30

40

50

【 0 1 5 2 】

図 1 1 は、特に、配列決定結果の処理、分析、および出力のための方法の例示的な実施形態のフローチャートである。

【 0 1 5 3 】

3.3 配列分析

生の配列データの変換は 3 段階で生じることができ、すなわち、(1) 生の装置データの処理およびアライメントされた配列決定読み取りへの変換、(2) 読取りデータの統計的解釈、および(3) 出力の提供およびアーカイブへの保存である。

【 0 1 5 4 】

生の装置読み出しからの生データの、病原体ゲノム中の位置に関連する配列情報への処理は、少なくとも 2 つの以下のステップを含み得る：

1. 配列読み出し (「読み取り」) およびアライメントの前またはアライメント中のいずれかの関連の品質スコアファイルを統合するステップであって、配列決定プラットフォームが、エラーを捕獲し、読み取り長さを有する配列の低下を同定するために品質スコアを作成するステップ、

2. 病原体ゲノムの読み取りをアライメント / 位置付けするステップ。

【 0 1 5 5 】

いくつかの実施形態では、統計的分析および解釈は次に、全てのゲノムに対する全ての統計的に優位なヒットの説明に入り、場合により、病原体の耐性座位または独特の識別子などの対象の領域によるヒットを細分類する。

【 0 1 5 6 】

サンプル内に存在する生物の定量的な分析をもたらすために、配列決定装置からの生 F A S T Q データの処理、および基準ゲノムに対する定量化を示す例示的なワークフローは、図 1 2 に示される。

【 0 1 5 7 】

次世代配列決定読み取りから得られる配列の例示的なアライメントは図 1 4 に示される。個々に示されるように、配列決定読み取りは、プローブアーム領域を通してほぼ完全に一致する標的ゲノム D N A とアライメントされ得る。ポリメラーゼ伸長領域におけるアライメントは、この領域内の配列の変化を明らかにでき、これらの増幅産物配列を異なる株に割り当てることが可能になる。

【0158】

サンプル中の株を同定するための、基準株のデータベースに対する配列読取りアライメントの使用の略図は図15に示される。いくつかの読取りは、1つまたは複数の株の間で共通の領域に位置付けることができる。この略図では、ほとんどの読取りは株A、B、CおよびDにアライメントし、共通である。対照的に、他の読取りは、特定の株（例えば、株Dにのみアライメントする読取りのサブセット）に独特であり得る。いくつかの実施形態では、サンプル中に存在するそれぞれの独特的病原体の割合の定量的な推定値を提供するために、定量的モデルが使用されて、共通の読取りおよび独特的読取りの分布を予測する。

【0159】

10

いくつかの実施形態では、図16に示されるように、正確な多型モデリングおよび次世代配列決定による検出が実施される。3'プローブアーム、ポリメラーゼ伸長部位（矢印）、およびポリメラーゼ伸長領域の一部が上部に示される。下側のプロットは、配列読取りに沿って各ヌクレオチドにおいて、期待される標的配列と、配列読取りとの間で観察されるミスマッチを示す。ポリメラーゼ伸長領域を横切るミスマッチの頻度のモデリングは、バックグラウンドの配列決定エラーおよびノイズの結果ではない多型の正確な同定を可能にし得る。

【0160】

20

統計的分析は、一般に、全ての病原体に対するヒット密度などの簡単な要約統計量を含み、ここで、ヒット密度は、配列のウインドウ内のヒット数を高品質の読取り数によって除したものである。これは、病原体配列中の配列座標によって、あるいは「対象の領域」IDおよびその中心からの距離の組み合わせによって記録することができる。さらに、分類方法論を用いて、サンプルの病原体への正確な割り当てを提供することができる。利用可能なツールボックスは、最大尤度およびベイズアプローチ、線形判別に基づく方法論、ならびにニューラルネットワークアプローチを含む。このアプローチは、このようなアプローチのいずれか1つまたは組み合わせを用いることができる。同様または関連の問題において立証済みの実績がある既知の方法は、隠れマルコフモデル（HMM）、Parzen Windows、多変数回帰（LOESS回帰を含む）、およびサポートベクターマシン（SVM）である。いくつかの実施形態では、開示される方法は、最大の特異性および感度を達成するために基準データセットに対して評価されるこれらのアプローチのうちの1つまたは複数を用いる。最終的な分析は、本発明のシステムにおいて多数のサンプルを実行することと、「最も信頼できる（gold standard）」基準とに依存する。このことから、次に、これらのデータの特徴、アッセイを調べ、固定分析アルゴリズムを実行することができる。これらのアルゴリズムは正確には固定されているのではなく、入力データに適合される。この従来の分析は、本発明のシステムのライフサイクルに対して、数回実行される。上記のように実施される統計的解釈は、パワフルなコンピュータサービスにおける従来の分析に依存している。初期分析は分析および解釈についてのアルゴリズム方法を生成し、次に、これを本発明のシステムに配備することができる。

30

【0161】

30

従って、いくつかの実施形態では、プローブセットを用いる捕獲反応に従う配列決定およびその後の分析の目標は、サンプル中にそのDNAが存在する生物または株のセットを決定することである。いくつかの実施形態では、さらなる目標は、サンプル中のこれらの生物または株の相対量を決定することである。

40

【0162】

分析方法は、配列決定読取り中のエラーの確率のためのモデル、および生物の関連株の間で生じる突然変異のためのモデルに依存し得る。これらのモデルの最も簡単な型は、等しい確率を有するように全てのエラーまたは変化を処理する。ここで、その確率はデータから得られてもよいし、研究者の最良の推測に基づいて選択されてもよい。いくつかの実施形態では、より進歩したモデルは、同じ機械、サンプル調製、および分析ソフトウェアを用いて、既知の鋳型材料の配列決定データセットから異なるタイプのエラーの確率を学

50

習することができる。その他の進歩したモデルは、遺伝子もしくはゲノムの公的なデータベース、遺伝子もしくはゲノムの私的なデータベース、または配列決定読取りの構築されていない集合もしくは部分的に構築された集合からの既知の株のセットに基づいて、突然変異の確率を学習することができる。

【0163】

既知のゲノムのデータベースと、反応で使用されるプローブのセットに基づいて、期待される読取り配列のセットが計算される。期待される読取り配列はそれぞれ、1つのプローブおよび1つのゲノムから得ることができ、従って、期待される読取り配列の数はゲノムの数とプローブの数の積であり得る。

【0164】

反応から配列決定読取りのセット（または読み取りの対）が得られると、読取りは、期待される読取りのセットに対してアライメントされ得る。配列決定エラーのためのモデルを用いて、方法は、期待されるそれぞれの産物から得られる読取り（または読み取りの対）確率を計算し得る。次に、方法は、選択される最低の確率、例えば、1、.01、または.001よりも高度に読取りがアライメントされる、期待される全ての産物からの生物／株の結合として、サンプル中に存在し得る全ての生物または株のセットを算出する。

【0165】

いくつかの実施形態では、分析方法はさらに、

1) 期待される読取りのそれぞれにアライメントする各読取りの確率と、

2) サンプル中の各生物または株を観察する事前確率（このタイプの確率について、各生物または株は同様に確からしい）と、

3) 存在し得る生物または株の数の事前確率（このタイプの確率の最も簡単な形では、生物または株のそれぞれの数は同様に確からしいかもしだれない。別の形では、生物または株の数の確率は、ディリクレ分布を可能にし得る。）と

が与えられれば、割合または存在量が、観察される配列決定読取りのセットの実際の発生の確率を最小限にするように、各生物または株の相対的な割合または存在量を決定する。

【0166】

いくつかの実施形態では、分析方法は、「混合物モデル」によって、生物の相対的な割合または存在量を決定する。いくつかの実施形態では、モデルにおける隠れ変数は、生物または株の割合または存在量、および期待される読取りへの配列決定読取りの割り当てである（ここで、観察される読取りはそれぞれ、単一の期待される読取りに割り当てられる）。Expectation-Maximization、Gibbs Sampling、およびMetropolis-Hastingsを含む様々な方法を使用してこれらの隠れ変数の値を見出すことができ、隠れ変数および隠れ変数におけるプライア（priors）が与えられれば、データの確率が最大にされる。

【0167】

さらなる実施形態では、方法は、突然変異の確率を用いることにより、既知の生物の未知の株も混合物モデルに取り込む。このような実施形態では、未知の株のゲノムは、既知のすべてのゲノムに対して1つまたは複数のミスマッチを含有する、観察された読取りに基づいて生成される。これまで未知のゲノムは、既知のゲノムと同じ確率で混合物に付加され得る。

【0168】

いくつかの実施形態は、多重試験も補正する。どれか1つの技術に関して限定されずに、目的は、偽陽性および偽陰性を排除することである。FPRおよびFDR（false discovery rate）は、任意のシステムに適合することができるので、最も有望な補正の1つである。いくつかの実施形態では、追加の事例が試験されるにつれて時間とともに閾値は更新される。

【0169】

例示的な実施形態は、サンプルを（1）有意なヒット、（2）不確定なヒット、（3）ヒットの欠如または病原体が見当たらない、あるいは（4）低サンプル品質またはデータ

10

20

30

40

50

エラーに分類する。

【0170】

結果の出力は、(1)会社のサーバーに対して、(2)例えば、病院システム、電子医療記録(EMR)システム、または他のHL7またはxm1対応の保存システムに寄託するため、現存する健康記録のフレームワークで使用するために、xm1およびHL7フォーマットに対して、ならびに/または(3)医師にとって使いやすいグラフィックおよびテキストフォーマット、例えば、グラフ、表、概要テキスト、および可能な注釈付、基準情報にリンクするウェブフォーマットに対して、並行して生じることができる。出力フォーマットは任意であり、例えば、簡単なテキスト、スプレッドシートデータ、バイナリデータオブジェクト、暗号化および/または圧縮ファイルである。完全な記録は、独特の識別子により診断テストにリンクされたこれらの全てまたはいくつかを含むことができる。これらはコヒーレントオブジェクトに構築されてもよいし、あるいは独特の識別子の検索により構築されてもよい。

10

【0171】

図9は、配列決定データの分析およびフォーマッティングを実行するためのシステムアーキテクチャの例示的な実施形態の略図である。このシステムアーキテクチャは、配列決定分析(サーバー)、統計的測定の計算(計算)および出力またはディスプレイ機能(インターフェース)の分離を含む。このようなアーキテクチャの多数の実施形態が存在する。どの特定の物理的な実現に限定されることなく、好ましい実施形態は、これらの主要な構成要素を分析ワークフローおよびアーキテクチャに含む。

20

【0172】

3.4 例示的なプロトコル

プローブ、捕獲反応産物、および増幅反応産物の製造および使用方法は当該技術分野において知られており、本発明において使用され得る。例示的な方法は、例えば、Deng et al. 2009、およびLi et al., Genome Res., 19(9) 1606-15 (2009)に開示されている。

20

【0173】

例えば、本発明の混合物は捕獲反応(捕獲反応産物を形成するため)、増幅反応(増幅反応産物を形成するため)、および捕獲および/または増幅反応産物の配列決定のために、本質的にこれらの参考文献に記載されるように処理することができる。これらおよび他の参考文献において開示される方法は単に例示的ものであり、決して本発明を限定するものではない。例えば、Dengらは、Qiagen DNeasyカラムを用いて線維芽細胞、iPSまたはhES細胞の凍結したペレットからゲノムDNAを抽出し、これらをZymo DNA Methylation Gold Kit (Zymo Research)により重亜硫酸塩で変換した。本発明の方法では、例えば、DNAのメチル化を研究するために重硫酸変換が使用され得るが、必ずしも必須ではない。Dengらは、パドロック(padlock)プローブ(60nM)および200ngの重亜硫酸塩されたゲノムDNAを結合させ、10μlの1×Ampligase Buffer (Epicentre)中で混合し、95℃で10分間変性させてから、55℃で18時間ハイブリッド形成させ、その後、1μlのギャップ充填混合物(1×Ampligase緩衝液中、200μMのdNTP、2UのAmpliTaq Stoffel Fragment (ABI)および0.5単位のAmpligase (Epicentre))を反応に添加した。環状化のために、反応を55℃で4時間の後、95℃で1分間および55℃で4時間を5サイクル、インキュベートした。環状化の後に線状DNAを消化するために、2μlのエキソヌクレアーゼ混合物(10U/μlのエキソヌクレアーゼIおよび100U/μlのエキソヌクレアーゼII、USBを含有する)を反応に添加し、反応を37℃で2時間インキュベートしてから、95℃で5分間不活性化した。

30

【0174】

捕獲された配列を増幅するために、Dengらは、200nMのAmpF6.2-SOLプライマー、200nMのAmpR6.2-SOLプライマー、0.4×Sybr Gr

40

50

een I および 50 μ l の i Proof High - Fidelity Master Mix (Bio - Rad) を有する 100 μ l の反応中の PCR によって、98 で 30 秒間、8 サイクルの 8 で 10 秒間、58 で 20 秒間、72 で 20 秒間、14 サイクルの 98 で 10 秒間、72 で 20 秒間および 72 で 3 分間、10 - μ l の環状化産物を増幅した。6 % の PAGE (6 % の TBE ゲル、In Vitro) により、期待されるサイズ範囲 (344 ~ 394 bp) の増幅産物を精製した。

【0175】

次に、Deng からは、精製 PCR 産物を同じ鋳型 DNA において等モル比で 4 つのプローブセットと共にプールし、4 - μ l の鋳型 (10 ~ 15 ng / μ l)、200 μ M の dNTP、20 μ M の dUTP、200 nM の Amp F6.3 プライマー、200 nM mpAmp R6.3 プライマー、0.4 \times Sybr Green I および 200 μ l の 2 \times Taq Master Mix (NEB) を有する 4 \times 100 μ l の反応中で、94 で 3 分間、8 サイクルの 94 で 45 秒間、55 で 45 秒間、72 で 45 秒間および 72 で 3 分間、これらを再増幅した。Deng からは、PCR 増幅産物を Qiaquick カラムにより精製し、1 \times NEB 緩衝液中 4、MmeI : 約 3.6 nmole の精製 PCR 増幅産物、16 単位の MmeI (2 U / μ l、NEB)、100 μ M の SAM により 37 で 1 時間これらを消化した。Deng からは、消化物を再度カラム精製し、3 U の USeR 酵素 (1 U / μ l) により 37 で 2 時間消化してから、10 単位の S1 ヌクレアーゼ (10 U / μ l、In Vitro) により、1 \times S1 ヌクレアーゼ緩衝液中、37 で 10 分間消化した。Deng からは、断片化 DNA をカラムにより精製し、2.5 μ l の 10 \times 緩衝液、2.5 μ l の dNTP 混合物 (それぞれ、2.5 mM)、2.5 μ l の ATP (10 mM)、1 μ l の末端修復酵素混合物 (Epicentre)、および 15 μ l の DNA を含有する 25 - μ l の反応中、25 で 45 分間、DNA を末端修復した。約 100 ~ 500 ng の末端修復 DNA を、1 μ l の Quick Ligase を有する 30 μ l の 1 \times Quick Ligase 緩衝液 (NEB) 中、25 で 15 分間 60 μ M の Solexa 配列決定アダプターにより消化した。Deng からは、6 % PAGE により 150 ~ 175 bp のサイズの連結産物をサイズ選択し、5 μ l の鋳型、200 nM の Solexa PCR プライマー、0.8 \times Sybr Green I および 50 μ l の i Proof High - Fidelity Master Mix (Bio - Rad) を有する 100 μ l の反応中、98 で 30 秒間、12 サイクルの 98 で 10 秒間、65 で 20 秒間、72 で 20 および 72 で 3 分間、PCR によりこれらを増幅した。Deng からは、Qiaquick PCR 精製カラムにより PCR 増幅産物を精製し、Illumina Genome Analyzer においてこれらを配列決定した。

【0176】

Li からは、以下の方法を使用した。Li からは、15 μ l の反応中で 1 \times Ampligase 緩衝液 (Epicentre)、500 ng (0.25 amol) のゲノム DNA (例えば、試験サンプル DNA)、および 48 ng (1.32 pmol) のプローブ (各プローブ対 g DNA のモル比 = 100 : 1、他の比率の場合はそれに応じて数値が変化する) を混合し、95 で 10 分間、0.1 / 秒で 60 まで勾配させて変性してから、60 で 24 時間ハイブリッド形成した。次に、2 μ L のギャップ充填および密封混合物 (Ampligase 保存緩衝液 [Epicentre] 中、5.4 μ M の dNTP [100 \times 、1 \times 、10 \times 、1000 \times 、および 10,000 \times の場合はそれに応じて数値が変化する]、2 単位の Taq Stoffel 断片 [Applied Biosystems]、および 2.5 単位の Ampligase [Epicentre]) を添加し、反応を 60 で 15 分間、1 時間、1 日間、2 日間、または 5 日間インキュベートした。また Li からは、反応のサイクリングも試みた：60 で 1 日後、我々は 10 サイクルの 95 で 2 分の後、60 で 2 時間を適用した。線状 DNA を除去するために、Li からはインキュベーション温度を 37 に下げ、すぐに 2 μ L のエキソヌクレアーゼ I (20 単位 / μ L) および 2 μ L のエキソヌクレアーゼ III (200 単位 / μ L) (いずれも U.S.B から) を添加し、37 で 2 時間の後、94 で 5 分間、反応をインキュベートした。

【0177】

次に、Liらは、50 μLの2×iQ SYBR Green supermix (Bio-Rad)、10 μLの環鑄型（上から）、および40 pmolの順方向および逆方向プライマーのそれぞれ (IDT) を有する2つの100-μLのPCR反応によって環を増幅した。PCRプログラムは、96で3分、3サイクルの95で30秒間、60で30秒間、および72で30秒間、ならびに10サイクルの95で30分間、72で1分間、および72で5分間であった。所望のPCR産物をゲル精製し、定量化した。各サンプルについて、Liらは、Lumina Genome Analyzerバージョン1および更新バージョン2の両方によって、カスタムプライマーにより10~20 fmolのDNAを配列決定した。

10

【0178】

上記の記載は実例のために提示された。これは包括的ではなく、本発明を、開示される詳細な形態または実施形態に限定しない。本発明の修正および適応は、本明細書の考察および開示される実施形態の実施から当業者には明らかであろう。例えば、記載される実行は、ソフトウェア、ハードウェア、またはハードウェアおよびソフトウェアの組み合わせにおいて実行され得る。ハードウェアの例には、パーソナルコンピュータ、サーバー、ラップトップ、メインフレーム、およびマイクロプロセッサなどの計算または処理システムが含まれる。さらに、当業者は、図面に示されるレコードおよびフィールドが追加のフィールドを有していてもより少ないフィールドを有していてもよく、図面に示されるものは異なるフィールドを構成し得ることを認識するであろう。本明細書および実施例は単に例示的なものであると考えられることが意図され、本発明の真の範囲および趣旨は以下の特許請求の範囲によって示される。

20

【0179】

「約」、「少なくとも」、「未満」、および「よりも多い」などの本出願においていくつかのパラメータを説明する数値範囲の全てについて、記載は必ずしも列挙される値によって拘束される範囲を包含しなくてもよいことは理解されるべきである。従って、例えば、少なくとも1、2、3、4、または5という記載は、特に1~2、1~3、1~4、1~5、2~3、2~4、2~5、3~4、3~5、および4~5などの範囲も説明する。

【0180】

本明細書中で引用される全ての特許、出願、または非特許文献および基準配列情報などのその他の参考文献については、全ての目的および列挙される問題のために参照によってその全体が本明細書中に援用されることが理解されるべきである。参照によって本明細書中に援用される文献と本出願との間に矛盾が存在する場合には、本出願が支配するであろう。本出願で開示される基準遺伝子配列に関連するGeneIDまたは受入番号などの全ての情報（例えば、ゲノム座位、ゲノム配列、機能注釈、対立遺伝子変異体、および基準mRNA（例えば、エクソン境界を含む）およびタンパク質配列（保存ドメイン構造など）を含む）は、参照によってその全体が本明細書中に援用される。

30

【実施例】

【0181】

実施例1：プローブ生成プロセス

40

多数の異なる病原性生物（細菌、ウイルス、真菌および他の生物など）を同時に検出および同定することができる多重診断アッセイにおいて使用することができるDNAオリゴヌクレオチドプローブの設計のための方法が、本明細書において提供される。これは、所与の生物に対して瞬時に非常に特異的であり、臨床的対象の特定の領域を捕獲することができ、そして同一プール内の他の生物の核酸または他のプローブのいずれともクロスハイブリッド形成しないであろうプローブのプールの生成によって達成される。全ゲノム（またはゲノムの群）または特定の対象の領域（例えば、薬物耐性を付与する突然変異、薬物感受性、毒性、病原性、ヒトの伝播性の増大、および診断または臨床関連の他の特徴などの特定の特徴を反映する領域）のいずれかから、DNA（またはRNA）の候補相同性領域が選択される。これらの相同性領域は、特定の生物、株、亜株または血清型を同定する

50

ために使用することができる。

【0182】

DNAの特定の短い領域（通常、数千塩基以下の長さである）を予め選択することに限定されている現存のプライマー設計方法とは対照的に、本発明に従って、全ゲノムまたはゲノム群から始めることによってプライマーを設計した。これにより、特異性、T_m、および他のプローブの特徴に対する特定の基準を満たす核酸配列の最も広い可能な範囲から、最適な候補プローブの同定および検証が可能になる。

【0183】

通常、本方法によって提供されるプローブは、標的生物のゲノムの領域を捕獲するように設計された2つの相同プローブ配列（本明細書では、「ホーマー」とも呼ばれる）を含む。プローブの相同プローブ配列が、特定の標的とハイブリッド形成する場合、ギャップが充填されて環状産物が生成され、これは次に、配列決定するかあるいはアレイとハイブリッド形成させて、最終結果を得ることが可能である。プローブ「骨格」は2つの相同プローブ配列を接続させ、種々のリンカー、DNAバーコード、增幅部位、および／または制限部位を含む。構築された構造は、完成プローブである。本発明によって提供される例示的なプローブの概略図は図1に示される。

【0184】

この実施例は、本明細書において記載されるような、2つの一般的な病原体：ストレプトコッカス・ニューモニエ（*Streptococcus pneumoniae*）およびサルモネラ・エンテリカ（*Salmonella enterica*）に対して極めて特異的な捕獲プローブの生成を説明する。

【0185】

ストレプトコッカス・ニューモニエ（*Streptococcus pneumoniae*）について、標的ゲノム（gi 221230948 ref NC_011900.1）ストレプトコッカス・ニューモニエ（*Streptococcus pneumoniae*）ATCC700669、完全ゲノム）を、以下の表1に示される10の追加のS.ニューモニエ（*S. pneumoniae*）ゲノムと共に、NCBIからダウンロードした。

表1：追加のストレプトコッカス・ニューモニエ（*Streptococcus pneumoniae*）標的ゲノム

10

20

30

40

標的ゲノム
gi 194172857 ref NC_003028.3 ストレプトコッカス・ニューモニエ TIGR4
gi 15902044 ref NC_003098.1 ストレプトコッカス・ニューモニエ R6
gi 116515308 ref NC_008533.1 ストレプトコッカス・ニューモニエ D39
gi 169832377 ref NC_010380.1 ストレプトコッカス・ニューモニエ・ハンガリー 19A-6
gi 182682970 ref NC_010582.1 ストレプトコッカス・ニューモニエ CGSP14
gi 194396645 ref NC_011072.1 ストレプトコッカス・ニューモニエ G54
gi 225853611 ref NC_012466.1 ストレプトコッカス・ニューモニエ JJA
gi 225855735 ref NC_012467.1 ストレプトコッカス・ニューモニエ P1031
gi 225857809 ref NC_012468.1 ストレプトコッカス・ニューモニエ 70585
gi 225860012 ref NC_012469.1 ストレプトコッカス・ニューモニエ・タイワン 19F-14)

【0186】

サルモネラ・エンテリカ（*Salmonella enterica*）について、gi 29140543 ref NC_004631.1 サルモネラ・エンテリカ亜種エンテリカ血清型ティフィ（*Salmonella enterica* subsp. *enterica* serovar *Typhi*）株 Ty 2 の完全ゲノムを、最初の单一初期標的ゲ

50

ノムとしてダウンロードした。さらに、表2に示される14のS.エンテリカ(S. enterica)ゲノムをダウンロードした。

表2: 追加のサルモネラ・エンテリカ(Salmonella enteric)標的ゲノム

標的ゲノム		
gi 161501984 ref NC_010067.1	サルモネラ・エンテリカ亜種アリゾナエ血清型	
gi 16758993 ref NC_003198.1	サルモネラ・エンテリカ亜種エンテリカ血清型ティフィ株	CT18
gi 161612313 ref NC_010102.1	サルモネラ・エンテリカ亜種エンテリカ血清型パラティフィB株	SPB7
gi 56412276 ref NC_006511.1	サルモネラ・エンテリカ亜種エンテリカ血清型パラティフィA株	ATCC 9150
gi 62178570 ref NC_006905.1	サルモネラ・エンテリカ亜種エンテリカ血清型コレレスイス株	SC-B67
gi 194442203 ref NC_011080.1	サルモネラ・エンテリカ亜種エンテリカ血清型ニューポート株	SL254
gi 194733902 ref NC_011094.1	サルモネラ・エンテリカ亜種エンテリカ血清型シュワルツエンゲルント株	
CVM19633		
gi 198241740 ref NC_011205.1	サルモネラ・エンテリカ亜種エンテリカ血清型ダブリン株	CT_02021853
gi 197247352 ref NC_011149.1	サルモネラ・エンテリカ亜種エンテリカ血清型アゴナ株	SL483
gi 194447306 ref NC_011083.1	サルモネラ・エンテリカ亜種エンテリカ血清型ハイデルベルグ株	SL476
gi 224581838 ref NC_012125.1	サルモネラ・エンテリカ亜種エンテリカ血清型パラティフィC株	RKS4594
gi 207855516 ref NC_011294.1	サルモネラ・エンテリカ亜種エンテリカ血清型エンテリティデス株	P125109
gi 205351346 ref NC_011274.1	サルモネラ・エンテリカ亜種エンテリカ血清型ガリナルム株	287/91
gi 197361212 ref NC_011147.1	サルモネラ・エンテリカ亜種エンテリカ血清型パラティフィA株	
AKU_12601)		

10

20

【0187】

次に、初期標的ゲノムをDNAの全ての可能な25-塩基ストリング(25-mer)にスライスした。S.ニューモニエ(S. pneumoniae)の例では、初期標的ゲノムは、約2,253,000塩基の長さであり、それぞれ25塩基の2,221,290ストリングを含有するファイルを作成した。S.エンテリカ(S. enterica)の例では、このファイルは4,791,936のストリングを含有した。

【0188】

次に、25-merストリングのリストに一連のフィルタを適用したが、これは、FASTAファイルまたは他のフォーマットの場合よりも著しく速い。全ての複製配列および多過ぎる単一の反復(5以上)を有するいかなる配列も排除した。S.エンテリカ(S. enterica)の場合、これらの初期フィルタを適用した後4,295,818の候補配列が残存した。

【0189】

次に、セルフハイブリッド形成する可能性のあるプローブを同定するために非常に大きい候補セットの大規模の高速処理を可能するため、DNAのインシリコのストリング表示に基づいて、ヘアピンを形成する可能性のある(すなわち、セルフハイブリッド形成する可能性のある)全ての配列を排除した。ヘアピン/二量体化検索は、自己相補的であり得るオリゴヌクレオチド内の領域を探す。プローブ中のN個の塩基のセットが、プローブからD塩基離れた距離で同じプローブ中のN個の相補的な塩基と一致することを必要とする検索基準を確立した。これらの実行において、プローブ配列に由来する長さNを有する全ての可能な候補部分配列(subsequence)の逆補体をまず構築する、Rubyプログラミング言語で作成したスクリプトを用いた。次にスクリプトは、正確な一致についてプローブを検索し、一致が見出され、第1の配列の終わりと第2の配列の初めとの距離がD塩基よりも遠く離れている場合に、ヘアピンを報告する。検索およびマッチングは、この状況で結果を非常に迅速に送達することができるストリング処理機能を用いて配列のアレイおよび/またはハッシュにおいて実施される。この例では、Nは3よりも大きく7未満であり、Dは5よりも大きい。

【0190】

S.ニューモニエ(S. pneumoniae)からの候補25-merに対して、正確に13であるグアニジンおよびシトシン塩基の合計を有することに基づいて、約59の

30

40

50

T_m により 25-mer を同定した。S. エンテリカ (S. enterica) については、標的 T_m の選択は、以下に考察されるように、後の段階で実施した。このより早い段階でのこのスクリーンの実施は実質的に効率を高めることが後で分かった。

【0191】

これらのフィルタを適用した後、サルモネラ・エンテリカ (Salmonella enterica) から 1, 175, 631 の候補配列が残った。次のステップのために、ストリングファイルを FASTA - フォーマット化ファイルに変換した。

【0192】

次に、NCBI の MegabLAST バージョン 2.2.10 (他に記載されない限り、実施例における BLAST に対するいかなる参照 [すなわち、blast、blast ed、BLASTed など] も、MegabLAST を指す) を用いて、全ての候補 25-mer を、S. ニューモニエ (S. pneumoniae) および S. エンテリカ (S. enterica) についてそれぞれ表 1 および 2 に記載される同一生物の全ての標的ゲノムと比較した。その標的生物のゲノムの全てにおいて正確な一致を持たないかなる候補 25-mer も廃棄した。S. エンテリカ (S. enterica) の場合、42, 907 の候補 25-mer がこのステップの後に残った。次に、各標的ゲノムに対する各 25-mer のヒットの数を決定し、この例では、ゲノムにおいて正確に 1 回発生したものだけを保持した。

10

【0193】

ヒトゲノムとのハイブリダイゼーションを回避するために、候補 25-mer を、個々の染色体によって NCBI からダウンロードしたヒトゲノムに対して BLAST した。この研究で使用した配列は表 3 に示される。ヒトゲノム中の配列と 20 の連続した塩基のうち 19 を共有する候補 25-mer を廃棄した。サルモネラ・エンテリカ (Salmonella enterica) の場合、このステップの後に 42, 485 の候補 25-mer が残った。

20

表 3：ハイブリッド形成プローブのスクリーニングのためのヒトゲノム配列

ゲノム配列	
gi 89161185 ref NC_000001.9 NC_000001	ホモ・サピエンス(Homo sapiens)染色体1
gi 89161199 ref NC_000002.10 NC_000002	ホモ・サピエンス染色体 2
gi 89161205 ref NC_000003.10 NC_000003	ホモ・サピエンス染色体 3
gi 89161207 ref NC_000004.10 NC_000004	ホモ・サピエンス染色体 4
gi 51511721 ref NC_000005.8 NC_000005	ホモ・サピエンス染色体 5
gi 89161210 ref NC_000006.10 NC_000006	ホモ・サピエンス染色体 6
gi 89161213 ref NC_000007.12 NC_000007	ホモ・サピエンス染色体 7
gi 51511724 ref NC_000008.9 NC_000008	ホモ・サピエンス染色体 8
gi 89161216 ref NC_000009.10 NC_000009	ホモ・サピエンス染色体 9
gi 89161187 ref NC_000010.9 NC_000010	ホモ・サピエンス染色体 10
gi 51511727 ref NC_000011.8 NC_000011	ホモ・サピエンス染色体 11
gi 89161190 ref NC_000012.10 NC_000012	ホモ・サピエンス染色体 12
gi 51511729 ref NC_000013.9 NC_000013	ホモ・サピエンス染色体 13
gi 51511730 ref NC_000014.7 NC_000014	ホモ・サピエンス染色体 14
gi 51511731 ref NC_000015.8 NC_000015	ホモ・サピエンス染色体 15
gi 51511732 ref NC_000016.8 NC_000016	ホモ・サピエンス染色体 16
gi 51511734 ref NC_000017.9 NC_000017	ホモ・サピエンス染色体 17
gi 51511735 ref NC_000018.8 NC_000018	ホモ・サピエンス染色体 18
gi 42406306 ref NC_000019.8 NC_000019	ホモ・サピエンス染色体 19
gi 51511747 ref NC_000020.9 NC_000020	ホモ・サピエンス染色体 20
gi 51511750 ref NC_000021.7 NC_000021	ホモ・サピエンス染色体 21
gi 89161203 ref NC_000022.9 NC_000022	ホモ・サピエンス染色体 22
gi 89161218 ref NC_000023.9 NC_000023	ホモ・サピエンス染色体 X
gi 89161220 ref NC_000024.8 NC_000024	ホモ・サピエンス染色体 Y

30

40

50

【0194】

ヒトゲノムとの類似性を有する 25 - m e r を排除した後、残存する 25 - m e r を、25, 991 の微生物および 3, 602 のウイルスゲノムの N C B I データベースに対して B L A S T した。これらのゲノムのいずれかにおける配列に対して 20 の連続塩基のうち少なくとも 19 を共有する 25 - m e r を排除した。このフィルタを適用した後、S. エンテリカ (S. enterica) に対する 2, 245 の候補 25 - m e r が残った。

【0195】

S. エンテリカ (S. enterica) に対して、約 59 の T_m の選択（正確に 13 であるグアニジンおよびシトシン塩基の合計を有する配列のみを選択することによる）をこの段階で実施し、1, 116 の候補 25 - m e r が残った。

10

【0196】

次に、各生物に対して残存した候補 25 - m e r を、その本来の標的ゲノムに対して B L A S T して、ゲノム中のその開始位置および停止位置（すなわち、そのゲノム座標）を決定した。この情報を用いて、固定距離によって分離された 25 - m e r の対を選択した。S. エンテリカ (S. enterica) の場合、正確に 100 塩基の標的長さ（第 1 の 25 - m e r のスタートから第 2 の 25 - m e r のエンドまで）に及ぶプローブ対を選択して、18 のこのような候補プローブ対が得られた。S. ニューモニエ (S. pneumoniae) の場合、100、200、300、400 および 500 塩基の長さを有する配列を標的とするために、全部で 58 のプローブを設計した。S. ニューモニエ (S. pneumoniae) のためのプローブ中に含有される 25 - m e r は、プローブのゲノム位置および標的長さを示す表 4 に示される。

20

【0197】

次に、一般的なリンカー

A G A T C G G A A G A G C G T C G T G T A G G G A A A G C T G A G C A A A T G T T A T C G A G G T C (配列番号 7) を用いて 25 - m e r 対を完成プローブに構築した。S. ニューモニエ (S. pneumoniae) のために構築されたプローブは表 5 に示される。S. エンテリカ (S. enterica) のために構築された相同プローブ配列の対は、相同プローブ配列の各対に対するゲノム位置情報を含む表 6 に示される。

【0198】

さらなる実施形態では、プローブ構築の前に、候補 25 - m e r は、混合物中の全ての他の候補 25 - m e r および / または構築プローブに対して B L A S T され、混合物中の他の任意の配列（例えば、相同プローブ配列、骨格、または構築プローブ）とクロスハイブリッド形成し得るものが排除される。一実施形態では、混合物中の他のプローブ配列（例えば、骨格または相同プローブ配列）中に含有される 20 の連続した塩基のうち 19 を含有する 25 - m e r が排除される。

30

【0199】

フィルタリングされたらすぐに、25 - m e r は、2 つの 25 - m e r および骨格（様々なリンカー、D N A バーコード、普遍的增幅プライマー、および必要に応じて他の配列を含むことができる）を含む候補プローブに構築される。次に、構築されたプローブは、可能なクロスハイブリダイゼーションのための代替または追加のスクリーンとして、プール中の全ての他の構築プローブに対して B L A S T され得る。ヘアピンおよび / またはセルフハイブリダイゼーションのための最終分析が実施される。次に、検証された構築プローブは、有用なプローブのデータベースに追加される。プローブまたはプローブ混合物（例えば、プローブパネル）の生成プロセスにおける例示的な実行のフローチャートは図 7 に示される。

40

表4: S. ニューモニエ (S. pneumonia) 標的プローブのための 25-mer 配列

プローブ ID	H1 pos.	H2 pos.	標的 スタート	標的 エンド	標的 長さ	H1 (伸長アーミー)	H2 (連結アーミー)
>strep.pneumo-01	645-669	720-744	645	744	100	TATGGAGGACCAAGGCCCTGGTAAGA (配列番号 8)	GCGCGTGTAAATATCCCTGCCG (配列番号 9)
>strep.pneumo-02	673097-673121	673172-673196	673097	673196	100	GGTGTGCGCAACCTGTTCTGTTT (配列番号 10)	GCGGCTCGTCAAAATCTTGACCTTC (配列番号 11)
>strep.pneumo-03	707096-707120	707171-707195	707096	707195	100	CAGCCTGGTTACCCAGTTTACTG (配列番号 12)	GGTGAGAACGAGAACGAGAACCGTGC (配列番号 13)
>strep.pneumo-04	720981-721005	721056-721080	720981	721080	100	AATTCAATCGGGTACCCCTGTGGAAG (配列番号 14)	ATTGGGGATCGTGTCCAGCCTTGG (配列番号 15)
>strep.pneumo-05	767921-767945	767996-768020	767921	768020	100	AGGTGTCAATGCCATGGTGGTGAA (配列番号 16)	CACACCTGATGTGGTACACGTGATG (配列番号 17)
>strep.pneumo-06	777532-777556	777607-777631	777532	777631	100	CGACGGGATTATCGGTGGCTTTAC (配列番号 18)	TTGTCAGGGTGGCAGAAGATACTCG (配列番号 19)
>strep.pneumo-07	865658-865682	865733-865757	865658	865757	100	CTTCAGCGTTGCTGCCAGTAA (配列番号 20)	CAACACGACGAATCAGTCACTGGC (配列番号 21)
>strep.pneumo-08	963949-963973	964024-964048	963949	964048	100	CCTAGTGAGATTGTCGGTGACTTGC (配列番号 22)	GAATTAGCCAAAGTTTGAGCGTCCGG (配列番号 23)
>strep.pneumo-09	1313943-1313967	1314018-1314042	1313943	1314042	100	GCCCCACCTTACCCATAGAAATGGTC (配列番号 24)	CAAGTCTAACGACATCCTGCTCCGG (配列番号 25)
>strep.pneumo-10	1348377-1348401	1348452-1348476	1348377	1348476	100	GGCCACACATACTCATCAAGGGTTGAC (配列番号 26)	ATTCAAGTGGGCTACTTCCCTGTCCGC (配列番号 27)
>strep.pneumo-11	14221943-14221967	1422018-1422042	1421943	1422042	100	CATCCCTCGCTAGCAATTGAGCTAG (配列番号 28)	TGGCCTGAGGATAGAAAACCAATCCC (配列番号 29)
>strep.pneumo-12	1471291-1471315	1471366-1471390	1471291	1471390	100	GATTCTTCTGTCGCCAGAAGGCCAAGC (配列番号 30)	TTACTCTCATCGCATTAGCCGAGC (配列番号 31)
>strep.pneumo-13	1528931-1528955	1529006-1529030	1528931	1529030	100	AATGCCACACTACGGGTGGTGTGCCAC (配列番号 32)	CTTGGCAGAAATCGGCTCAATCAAGG (配列番号 33)
>strep.pneumo-14	1553359-1553383	1553284-1553383	1553284	1553383	100	GCCGCCAAAGAACACCAGCATCTA (配列番号 34)	ACACAGAAAGGGGGTTAATAGGG (配列番号 35)
>strep.pneumo-15	1665069-1665093	1665144-1665168	1665069	1665168	100	CGTGCCCTGTTGGAAAGGCCAATTGA (配列番号 36)	CGATACCTTGTCCCCATAGCTCCACT (配列番号 37)
>strep.pneumo-16	1780734-1780809-	1780833-1780833	1780734	1780833	100	TTGACCTCAGCGATTACCTGCAAGC (配列番号 38)	GGCTGGATTTCGTCAGCTTCATCT (配列番号 39)
>strep.pneumo-17	1822203-1822227	1822278-1822302	1822203	1822302	100	AGAGCTTCTTCATGAGTGGAGGCC (配列番号 40)	TAACGCTCCAATTCCGCAATCAGTCG (配列番号 41)
>strep.pneumo-18	1832185-1832209	1832260-1832284	1832185	1832284	100	GGCGCCCTTGGCCTGATTGATTA (配列番号 42)	CCAACCGTTCTCTCCAAAGCAAGCA (配列番号 43)

>strep.pneumo-19	1836264-1836288	1836339-1836363	1836264	1836363	100	CTTGGCTCAAAGTCATGCCATCTG (配列番号 44)	CTGTCACAAACGGGAACACGGGTATA (配列番号 45)
>strep.pneumo-20	1888158-1888182	1888233-1888257	1888158	1888257	100	CCGCTTCGAGCAAATTGCTCAAAGAC (配列番号 46)	GGTAAGAAACGAAACCTGAAGGCC (配列番号 47)
>strep.pneumo-21	1939796-1939820	1939871-1939895	1939796	1939895	100	ATAGCTGGACCATGAGGTTGACTG (配列番号 48)	ACTCTTGTGACTAGAGCACCGTGAG (配列番号 49)
>strep.pneumo-22	1960075-1960099	1960150-1960174	1960075	1960174	100	GGACGGGTTAAAGCGTGAGATTGGTG (配列番号 50)	TCAGGCCAAACCGTTCAAGACTCCTG (配列番号 51)
>strep.pneumo-23	1991584-1991608	1991659-1991683	1991584	1991683	100	CGTGGACGAGCTAGACACGAT (配列番号 52)	ACGTTCTAACCAAGCTTGACAGGCC (配列番号 53)
>strep.pneumo-24	1993533-1993557	1993608-1993632	1993533	1993632	100	CTACTTCTGCAGGCCAGTTCTGGATG (配列番号 54)	CGCACACGGCTCTGCAACATGTTCTTT (配列番号 55)
>strep.pneumo-25	2014591-2014615	2014666-2014690	2014591	2014690	100	CACCCGGGTCCTCATATAAGTTGG (配列番号 56)	TCCCACGAACTTAGCACCTGTTGC (配列番号 57)
>strep.pneumo-26	2040994-2041018	2041069-2041093	2040994	2041093	100	GCTGCGGCTCATTCAAATAGAG (配列番号 58)	AGAATGGCACGGTTGGAGAACGATGG (配列番号 59)
>strep.pneumo-27	2051649-2051673	2051724-2051748	2051649	2051748	100	CCTGAAGAAGGTAAGAGTCTCACCC (配列番号 60)	AAGGCAAGCCAAAGTCAGTATGGCTG (配列番号 61)
>strep.pneumo-28	2064289-2064313	2064364-2064388	2064289	2064388	100	AGTCAACTGACTGGCATCTACACCG (配列番号 62)	ATTCGGCCAAGGGGACACATTG (配列番号 63)
>strep.pneumo-29	2161108-2161132	2161183-2161207	2161108	2161207	100	GTGCGGTTGGAGATAACGCAAGTAA (配列番号 64)	GACACTATTGAACGACGTGCTGACG (配列番号 65)
>strep.pneumo-30	70613-70637	70788-70812	70613	70812	200	CATCGTTGGCGTATTTCGTCAGTACCC (配列番号 66)	TTCCATGGCCAACCAGCATAGCATCC (配列番号 67)
>strep.pneumo-31	459298-459322	459473-459497	459298	459497	200	CTGGGTGCTGAGGACAAGTACAAGGA (配列番号 68)	TTTCTCAAGTTCTGGGGAGGC (配列番号 69)
>strep.pneumo-32	891891-891915	892066-892090	891891	892090	200	GATTGGTCCAATAGTGCCTGATAACG (配列番号 70)	TTCCCTCTTCTGCCAGTCATGCTGG (配列番号 71)
>strep.pneumo-33	952083-952107	952258-952282	952083	952282	200	CCTTGCAAGTTGGTTCGAAACCAAGG (配列番号 72)	GGCATACGGTTGGATTTCGGTTGCA (配列番号 73)
>strep.pneumo-34	1077528-1077552	1077703-1077727	1077528	1077727	200	GAGGTCCAAACGATTCTCAACCTGC (配列番号 74)	GCTGAACGAACATTGGCCAGACTTG (配列番号 75)
>strep.pneumo-35	1079653	1079828	1079629	1079828	200	CTTGGCCCTGCTCTCTCGTTTCAAAC (配列番号 76)	AAAGGCAATGGACTCTTCCAAAGGCC (配列番号 77)
>strep.pneumo-36	1320102-1320126	1320277-1320301	1320102	1320301	200	TATCGGTTGGGTACGTTCAAGGTGCT (配列番号 78)	CAATTCCCTGTCCTCAGCTAGATCCG (配列番号 79)
>strep.pneumo-37	1377167-1377191	1377342-1377366	1377167	1377366	200	CTCCTGAATAGCAGACAGATAGGC (配列番号 80)	AAGACCAGAGCGAAATTCCGGTGTG (配列番号 81)
>strep.pneumo-38	1544171-1544195	1543996-1544020	1544195	1543996	200	CATCCATGAGACGAGCTAGGTGTC (配列番号 82)	AGTTTGACGGTTCTCAGGTACACGG (配列番号 83)
>strep.pneumo-39	1567063-1567087	1567238-1567262	1567063	1567262	200	TGAAGGGCTTGTAGGCCGTGAACG (配列番号 84)	TCCACTCTGGGGTTATCCGCATC (配列番号 85)

>strep.pneumo-40	1594512-1594536	1594687-1594711	1594512	1594711	200	CTGCCATGCCACTAGTAGCACAAA (配列番号 86)	GCCATCTCCACGATCATTGAGGCTA (配列番号 87)
>strep.pneumo-41	18387870-1837894	1838045-1838069	1837870	1838069	200	AGTCGGCTCAAACCTGTTAACGCCACC (配列番号 88)	AAACGGGTGATGGAGTGGCCAGCAT (配列番号 89)
>strep.pneumo-42	1904806-1905005	1904981-1905005	1904806	1905005	200	GTGCCCACTCTATCGCTTCTCTAG (配列番号 90)	GTCCGAACACTGCTTGGCTTGTGAGG (配列番号 91)
>strep.pneumo-43	1943489-1943513	1943664-1943688	1943489	1943688	200	TCGTACTGGCAGGGTGTATGATGT (配列番号 92)	CAAAGGAAGGCCCTGTAAGCGGTCTCTG (配列番号 93)
>strep.pneumo-44	2061201-2061225	2061376-2061400	2061201	2061400	200	ACCAAACCTTCAAGAACGGGAGCCA (配列番号 94)	TAGCAGTCATGGTGCCTGGTT (配列番号 95)
>strep.pneumo-45	2179622-2179646	2179797-2179821	2179622	2179821	200	TTCCAGCGAGCTGGTCAAAATTGAC (配列番号 96)	TGATGGCTTGGATGACTTTGCGAGC (配列番号 97)
>strep.pneumo-46	626697-626721	626972-626996	626697	626996	300	CCACCAAGATAATTGACGGGAAAGC (配列番号 98)	GTTGAGGCAACGAAGGGGGTACTT (配列番号 99)
>strep.pneumo-47	1120572-1120596	1120847-1120871	1120572	1120871	300	CAACCTGACGTCCACCTGCATAAAGA (配列番号 100)	CCGTGAGTAGGAAATTCCATCATCAG (配列番号 101)
>strep.pneumo-48	1153293-1153317	1153568-1153592	1153293	1153592	300	GTATCCTCTATCGTTGGGGAGGA (配列番号 102)	GTTCACACTGGGACTGGTCAAACACC (配列番号 103)
>strep.pneumo-49	1309537-1309561	1309812-1309836	1309537	1309836	300	TAGACCGCGGACTGAGTTGTTGCA (配列番号 104)	CTATCCACACCAACGCCCTTATGGAA (配列番号 105)
>strep.pneumo-50	1434430-1434454	1434705-1434729	1434430	1434729	300	GTTCCTTGGGTTCATCTGTTCCACC (配列番号 106)	AAGTAACCACCTGCTGAGAGCAAGG (配列番号 107)
>strep.pneumo-51	1437830-1437854	1438105-1438129	1437830	1438129	300	GGAGCAGGGTGCCTGACACTCTTCAT (配列番号 108)	CACCTCCGGCATAGCTCTTCCCTTCT (配列番号 109)
>strep.pneumo-52	1006724-1006748	1007099-1007123	1006724	1007123	400	CGTCCCTCTAAAGAAGCAAGCCGT (配列番号 110)	GATTTCACCCACAAACTTCCCTCGGG (配列番号 111)
>strep.pneumo-53	2102469-2102493	2102844-2102868	2102469	2102868	400	TCAGCTGCATTGGATCTGCTCCAC (配列番号 112)	TCATTTCACCCATCTGGCCGAG (配列番号 113)
>strep.pneumo-54	347420-347444	347795-347819	347420	347819	400	CTGTATCGAGTCACATGGCCAGCA (配列番号 114)	AAGGACGAGCAATATCCTCTATGCC (配列番号 115)
>strep.pneumo-55	162037-162061	162512-162536	162037	162536	500	CCATTAGGATTCCAGGTCCCATTC (配列番号 116)	CGCAAACACTCGATAATTGAGCTGGAGG (配列番号 117)
>strep.pneumo-56	879373-879397	879848-879872	879373	879872	500	GAGTACACTCCAGATGTAACGGCTC (配列番号 118)	TCGGTGGTGGAGGATTCAAGCTCAA (配列番号 119)
>strep.pneumo-57	993493-993517	993968-993992	993493	993992	500	ACCTGCAGGTTGATGAAGGAGATCG (配列番号 120)	CAATCTCTGGCTTGGACAGGCCA (配列番号 121)
>strep.pneumo-58	1119326-1119350	1119801-1119825	1119326	1119825	500	CACGGAGACTCTTGACACTAGACTC (配列番号 122)	AGGGCACCAAGAAAGGCCCTCAAAGG (配列番号 123)

表5：ストレプトコッカス・ニューモニエ (Streptococcus pneumoniae) のために構築されたプローブ配列

プローブID	構築プローブ
>strep.pneumo -01	GGCGGTGTTAAATATCCCTGCCGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTATGGAGGACCGGCCTTGGTAAGA (配列番号 124)
>strep.pneumo -02	GC GGCTGTCAAATCTTGTACATGGCTCGTCAGATCGGAAGAGGGCTGTTGCGCAACCTGTTCTGTT (配列番号 125)
>strep.pneumo -03	GGT GAGAACGAAAGACAAAGAACCGGTCAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCAGCCTGGTTACCCA GTTCTTACTG (配列番号 126)
>strep.pneumo -04	ATTGTTGGATCGTGTCCAGCCCTGGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCAAATCGGGTGCACCGTGGATGAC CCTGTTGGAAAG (配列番号 127)
>strep.pneumo -05	CACACCTGATGTGGTACACGTGATGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCAATGGGTCAATGCCAT GCGGGGTGAA (配列番号 128)
>strep.pneumo -06	TTGTCAGGTGGCAGAAAGATACTCGAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCCTAGTGAGATTGTC GTGGCTTAC (配列番号 129)
>strep.pneumo -07	CAACACGACGAATCAGTCTCACTGGCAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCCTAGTGAGATTGTC TCGCCAGTAA (配列番号 130)
>strep.pneumo -08	GAATTAGCCAAAGTTGAGCGTCCGGAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCATCTCGCTAGCAA CGTGACTTGC (配列番号 131)
>strep.pneumo -09	CAAGCTTAAGACATCCCTGCTCCGTGAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCACATACTCATCAAGGGTTGAC (配列番号 132)
>strep.pneumo -10	TTGGCTTGTAGAAACCAATCCAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCATCTCGCTAGCAA TTGCAAGCTAG (配列番号 134)
>strep.pneumo -11	ATTCAAGTGGCTACTTCTGTGCGAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCATCTCGCTAGCAA CAAGGGTGCAC (配列番号 133)
>strep.pneumo -12	TTACTCTCATCCGCATTAGCCGACGAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCATCTCGCTAGCAA AAGCCAAGC (配列番号 135)
>strep.pneumo -13	CTTGGCAGAAATCGGCTCAATCAAGGGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCATGCCACACTACGG TGTGTTGTCAC (配列番号 136)
>strep.pneumo -14	ACCAACAGAAAGGGGGTTAATAGGGAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCGGCAAAGAAAGAC ACCAGCACTCA (配列番号 137)
>strep.pneumo -15	CGATACCTTGTCCCCATAGCTCCACTAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCGTGGCAAAGGATTA AGGCAATTGA (配列番号 138)
>strep.pneumo -16	GGCTGGATTGCTCCAGCTTCATCTAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCTTGCACCTCAGCGATTA CCTGCAAGC (配列番号 139)
>strep.pneumo -17	TAACGCTCCAAATCCGCAATCAGTCGAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCAAGGCTTCTTCATGA GTGGAGGCC (配列番号 140)
>strep.pneumo -18	CCAACCGTTCTCTCCAAAGCAAGGAAAGATCGGAAGAGCTGAGCAAATGTTATCGAGGTCCGGCCCTTGAGCCTGATTGATTA (配列番号 141)

>strep.pneumo -19	CTGTCACAACGGGATAAAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGCCTTGGCTCAAAGTCATGCTCCATCTG (配列番号142)
>strep.pneumo -20	GGTAAGAACAGAACCTGAAAGAGCCAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGCTGGACGCAATTGCTCAAAGAC (配列番号143)
>strep.pneumo -21	ACTCTTGTGACTAGAGCACCGTGAAGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGCTGGACGCAATTAGTTGACT (配列番号144)
>strep.pneumo -22	TCAGCCAAACCGTTCAAGAACACTCCTGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGCTGGACGGTAAAGCGTGAGATTGTC (配列番号145)
>strep.pneumo -23	ACGTCTAACCAAGCTTACAGCCAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGCTGGACGAGTCAGATGAGACAGAT (配列番号146)
>strep.pneumo -24	CGCCACGGGCTCTGCAACATGTTCTTAACTGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGCTGGCTCCATTGTTCTGGATG (配列番号147)
>strep.pneumo -25	TCCCCGAATCTTAGCCACCTGTTGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGCTGGCTCCACCCGGGTCTCTCAATAAGTGG (配列番号148)
>strep.pneumo -26	AGAAATGGCACGCTGGAGAACCGATGGAGATCGGAAGAGCGTCGTAGGGAAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGCTGGCTCCATTCAAAATAGAG (配列番号149)
>strep.pneumo -27	AAGGCCAAGGCCAAGTCAGTATGGCTGAGATCGGAAGAGCGTCGTAGGGAAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCAACTGACTGGCAGTCTCACCC (配列番号150)
>strep.pneumo -28	ATTTCGGCCAAAAGGGAGCCACATTGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCAACTGACTGGCATACTACACCG (配列番号151)
>strep.pneumo -29	GACACTATTGAACGACGCTGACTGAGATCGGAAGAGCGTCGTAGGGAAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGGGAGATACTGAAGTAA (配列番号152)
>strep.pneumo -30	TTCCATGGCAAACCGCATAGCCATCCAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCATCGTTGGGTATTCGTCAGTAC (配列番号153)
>strep.pneumo -31	TTTCCTCAAGTTCTCGGGAGGCCAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGGGACAGTAAAGGA (配列番号154)
>strep.pneumo -32	TTCCCTCTTCTGCCAGTCTATGCTGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTAGGGCCATAGTGCCCCGATAAC (配列番号155)
>strep.pneumo -33	GGCATACGGTTGGATTTCGGTTGCAAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCCTGGAGTTGGTC (配列番号156)
>strep.pneumo -34	GCTGAACGAACATTGGCAGACTTGAAGATCGGAAGAGCGTCGTAGGGAAAGAGCTGAGCAAATGTTATCGAGGTCCCTGGCTCAAACGATTCTCAACCTGC (配列番号157)
>strep.pneumo -35	AAAGGCAATGGACACTCTTCCAAGGCCAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCCTGGCTCAAAC (配列番号158)
>strep.pneumo -36	CAATTCCCTGTCTCAGCTAGATCCGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTATCGGGTACCTGAGGTGCT (配列番号159)
>strep.pneumo -37	AAGACCAGGCCAAATTCCGTTGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCCTGAATAGCAGACAGATGGG (配列番号160)
>strep.pneumo -38	AGTTGACGGTTCTCAGGTACACGGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCATCCATGAGACGAGTCATGGTGTCT (配列番号161)
>strep.pneumo -39	TCCACTCTGGTGGTTATCGGCATCAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTGAAGGGCTGTGATTAGCCGTGAAC (配列番号162)

>strep.pneumo	GCCATCTCCACGATCATTGAGGCTAAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCCTGCCATGCCACTAG
-40	TAGCACCaaa (配列番号163)
>strep.pneumo	AAACGGTGTGGAGTCCAGCATAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCAACTGTT
-41	TAACGCCACC (配列番号 164)
>strep.pneumo	GTCCGAACTAGCTTGCTGTTGAGGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTCTATCG
-42	CTTCTCTAG (配列番号 165)
>strep.pneumo	CAAAGGAAGCCTGTAAGCGTGTCTGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTGTAGGGCAGGT
-43	GTCAATGATGT (配列番号 166)
>strep.pneumo	TAGCAGTCATAGGTGCCCTGGTTAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCAACAAACCTTCAGAA
-44	GCGGAGGCCA (配列番号167)
>strep.pneumo	TGATGGCTTGATGACTTTGGAGCAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTTCCAGCGAGCTGCG
-45	TCAAATTGAC (配列番号168)
>strep.pneumo	GTTGAGGCCAACGAAAGGAGGGTACTTAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCACCAAGATAATTGA
-46	CGGGCAAAAGC (配列番号169)
>strep.pneumo	CCGTGAGTACGAATTCCATCAGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTAGACGTCAC
-47	CTGCATAAAGA (配列番号 170)
>strep.pneumo	GTTCACCTGGGACTCTGGTCAAAACACAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTATCGTTT
-48	GGGGAGGA (配列番号171)
>strep.pneumo	CTATCCACACCACCGCTTATGGAAAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTAGACCGCCACTGAG
-49	TTCGTTTGCA (配列番号 172)
>strep.pneumo	AAGTAACCACCTGCTGAGAGCAAGGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTCTGGGTTCAT
-50	CTGTTCCACC (配列番号 173)
>strep.pneumo	CACCTCCGCATAGCTCTTCTCTAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTGGAGCAGGTGCTGAC
-51	ACTTCTTCAT (配列番号 174)
>strep.pneumo	GATTTCACCAACCAACTTCTCGGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCGTCCTAAAGAA
-52	GCAAGCCGT (配列番号175)
>strep.pneumo	TCATTCAACCTTCATCTGGCCGAGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCTGCACTGGAT
-53	CTGCTCCAC (配列番号176)
>strep.pneumo	AAGGACGAGCATATCCCTATGCCAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCCTGTATCGAGTCACA
-54	TGGTCCAGCA (配列番号177)
>strep.pneumo	CGCAAACACTCGATAATGAGCTGGAGGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCCATAGGATCCAG
-55	GTCCCCATTGC (配列番号178)
>strep.pneumo	TCGGTGGTGGAGATTCAAGACTCAAGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCACTCCAGAT
-56	GTAACGGGCTC (配列番号179)
>strep.pneumo	CAATCTCTGGCTTGGACGAGCCAAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCAACTGAGGTTGATG
-57	AACGAGATCG (配列番号 180)
>strep.pneumo	AGGGCACCAAGAAAAGGCCTCAAAAGGAGATCGGAAGAGCGTCGTAGGGAAAGCTGAGCAAATGTTATCGAGGTCCACGGAGACTCTTGA
-58	CACTAGACTC (配列番号181)

表6：サルモネラ・エンテリカ (Salmonella enterica) のために構築された相同プローブ配列の対

>sal-918813 sal-729167 163786-163885 100 GCGGTAATAGGGTGAACGTTATGGG (配列番号 182) TACCCAACCGTTCACAGGTGGAAAG (配列番号 183)	10
>sal-91537 sal-495107 163787-163886 100 CGGTAATAGGGTGAACGTTATGGC (配列番号 184) ACCCAACCGTTCACAGGTGGAAAGT (配列番号 185)	
>sal-1023952 sal-888277 163814-163913 100 GTTCCAGCGTTGCGTTGATGCTTC (配列番号 186) TGAAATTCCGCTTCGCGGGACCAA (配列番号 187)	
>sal-591159 sal-1123128 163815-163914 100 TTCCAGCGTTGCGTTGATGCTTCG (配列番号 188) GAAATTCCGCTTCGCGGGACCAA (配列番号 189)	
>sal-244766 sal-1039899 164829-164928 100 TGATGCCGGTATTGCGCTTGGCGAT (配列番号 190) ATGCGCGATTATCCGATATTGGC (配列番号 191)	20
>sal-379412 sal-999649 164841-164940 100 TCGCTTGGCGATGCGGTACAACCTT (配列番号 192) CCCGATATTGGCTGGATATCGATG (配列番号 193)	
>sal-643175 sal-704852 164981-165080 100 GCCTTGCCATCGTTTACGCGTGA (配列番号 194) GATCGTAGCATCCCCCTGCATACCTT (配列番号 195)	30
>sal-231120 sal-422707 164982-165081 100 CCTTGCCATCGTTTACGCGTGAG (配列番号 196) ATCGTAGCATCCCCCTGCATACCTT (配列番号 197)	
>sal-1053463 sal-69659 165054-165153 100 AATTACGCCGGAAGCCGCGTTAATG (配列番号 198) TGCCTTGCCATCGTTTACGCGTG (配列番号 199)	
>sal-492477 sal-239882 165083-165182 100 ATGATAAAGCGTTGCGTCTCTCCGC (配列番号 200) CGGGCTATATCGGTGGGAGTTGTT (配列番号 201)	40
>sal-239882 sal-596706 165157-165256 100 GAAGACTTACAGGATGGCGGTTGT (配列番号 202) ATGATAAAGCGTTGCGTCTCTCCGC (配列番号 203)	
>sal-120080 sal-428037 2922400-2922499 100 CGAATGGCAGGACTCGCTACTGAA (配列番号 204)	

ACGGGCAATGCACAAATCAAAGCGG (配列番号 205)	
>sal-662112 sal-1072150 2922404-2922503 100	
TGGCAGGACTCGCTTACTGAAGATG (配列番号 206)	
GCAATGCACAAATCAAAGCGGCGGT (配列番号 207)	
>sal-1071952 sal-10611 2939265-2939364 100	10
AAACTTCGGTGCAGGGTTAGGCAT (配列番号 208)	
TGGCAGAGCGAGTGACTATCTGAAG (配列番号 209)	
>sal-241367 sal-804215 4263827-4263926 100	
GGCTTGGCAACGTAGGCTTCTCA (配列番号 210)	
ATGCACATCACACCGTCCTGACCAA (配列番号 211)	
>sal-8740 sal-671757 4265448-4265547 100	
GCAGGCGATCTTAATCATCTGCGG (配列番号 212)	
AAACGCTTCGCGTTGGCGAGGTTA (配列番号 213)	
>sal-33849 sal-322827 4265449-4265548 100	20
CAGGCGATCTTAATCATCTGCGGG (配列番号 214)	
AACGCTTCGCGTTGGCGAGGTTAA (配列番号 215)	
>sal-848714 sal-549807 4265674-4265773 100	
CTGCAGATCGTCCCAGTCGGATTAA (配列番号 216)	
GTATGCAGATCGTCAGGATGGCCAA (配列番号 217)	

(表 6 中の配列の上側の見出しには、相同プローブ配列の識別子がそれぞれの順番で示された後、ゲノム標的座標と、第1の 25-mer のスタートから第2の 25-mer のエンドまでの標的配列の長さが示される)

30

【 0 2 0 0 】

実施例 2：M. ツベルクローシス (M. tuberculosis) 特異的プローブの生成

特異的なプローブは、本質的に、S. ニューモニエ (S. pneumoniae) について実施例 1 で説明される通りに製造した。簡単に言うと、標的ゲノム (gi|57116681|NC_000962.2) マイコバクテリウム・ツベルクローシス (Mycobacterium tuberculosis) H37Rv、完全ゲノム) を 25-mer にスライスし、これを 40% の CG 含量を有する (その後、固定 T_m) ようにフィルタリングして、実施例 1 に記載されるように、複製配列、二次構造を有する配列、および同一ヌクレオチドの 4 つよりも多い連続した反復を有する配列を排除した。また、表 7 の M. ツベルクローシス (M. tuberculosis) ゲノムと特異的にハイブリッド形成する配列を選択するように 25-mer をスクリーニングした。

40

表 7 : M. ツベルクローシス (M. tuberculosis) の追加の標的ゲノム

標的ゲノム
gi 50953765 NC_002755.2 マイコバクテリウム・ツベルクローシス CDC1551
gi 148659757 NC_009525.1 マイコバクテリウム・ツベルクローシス H37Ra
gi 148821191 NC_009565.1 マイコバクテリウム・ツベルクローシス F11
gi 253796915 NC_012943.1 マイコバクテリウム・ツベルクローシス KZN 1435

【 0 2 0 1 】

10

実施例 1 の場合と同様にヒトゲノムに対して 25 - m e r をスクリーニングして、ヒト DNA と特異的にハイブリッド形成する可能性のあるものを排除した。実施例 1 と同じ微生物およびウイルスゲノムの N C B I データベースに対して特異的にハイブリッド形成しないようにプローブ配列をスクリーニングした。100 ヌクレオチドの長さの標的領域を捕獲するように、25 - m e r を対でプローブに構築した。M. ツベルクローシス (M. tuberculosis) プローブ配列対およびそのゲノム位置は表 8 に記載される。

表 8 : M. ツベルクローシス (M. tuberculosis) のために構築された相同プローブ配列の対

### mtb-gc10-5778 mtb-gc10-13476 1697202-1697301 100 >mtb-gc10-5778 ATCAGCGTCTCACGTATCTTTGAT (配列番号 218) >mtb-gc10-13476 GCTCGTTTGATCCGATTTCTGTTT (配列番号 219)	20
### mtb-gc10-10249 mtb-gc10-21740 1697207-1697306 100 >mtb-gc10-10249 CGTCTCACGTATCTTTGATGGAAA (配列番号 220) >mtb-gc10-21740 TTTGATCCGATTTCTGTTCGCCA (配列番号 221)	
### mtb-gc10-14718 mtb-gc10-21512 1697208-1697307 100 >mtb-gc10-14718 GTCTCACGTATCTTTGATGGAAAC (配列番号 222) >mtb-gc10-21512 TTTGATCCGATTTCTGTTCGCCA (配列番号 223)	30
### mtb-gc10-18048 mtb-gc10-20799 1697209-1697308 100 >mtb-gc10-18048 TCTCACGTATCTTTGATGGAAACG (配列番号 224) >mtb-gc10-20799 TTGATCCGATTTCTGTTCGCCAAT (配列番号 225)	

mtb-gc10-13476 mtb-gc10-9738 1697276-1697375 100
 >mtb-gc10-13476
 GCTCGTTTGATCCGATTCTGTTT (配列番号 226)
 >mtb-gc10-9738
 CGACGAATGCAATCAGGTCAAAATA (配列番号 227)

mtb-gc10-5979 mtb-gc10-3490 1697348-1697447 100
 >mtb-gc10-5979
 ATCGACGAATGCAATCAGGTCAAAA (配列番号 228)
 >mtb-gc10-3490
 ACGCGGTGTCTCCAATTAGAATAA (配列番号 229)

mtb-gc10-9738 mtb-gc10-13364 1697350-1697449 100
 >mtb-gc10-9738
 CGACGAATGCAATCAGGTCAAAATA (配列番号 230)
 >mtb-gc10-13364
 GCGGTGTCTCCAATTAGAATAACA (配列番号 231)

mtb-gc10-1167 mtb-gc10-18133 1697421-1697520 100
 >mtb-gc10-1167
 AACGCGGTGTCTCCAATTAGAATA (配列番号 232)
 >mtb-gc10-18133
 TCTGCGACATATTCAATATGGTGCT (配列番号 233)

mtb-gc10-2966 mtb-gc10-6093 1697501-1697600 100
 >mtb-gc10-2966
 ACATATTCAATATGGTGCTCGGGAA (配列番号 234)
 >mtb-gc10-6093
 ATCGTCTCCTGTGAGATAATTGCAT (配列番号 235)

mtb-gc10-10988 mtb-gc10-9385 1697583-1697682 100
 >mtb-gc10-10988
 CTGTGAGATAATTGCATCCGATCAT (配列番号 236)
 >mtb-gc10-9385
 CCGTTTCTGGTTTGATGATGAT (配列番号 237)

mtb-gc10-15828 mtb-gc10-14219 1697591-1697690 100
 >mtb-gc10-15828
 TAATTGCATCCGATCATATAGGGCT (配列番号 238)
 >mtb-gc10-14219
 GGTTTTCTGGTTTGATGATCAAATCCG (配列番号 239)

mtb-gc10-7551 mtb-gc10-12444 2632341-2632440 100
 >mtb-gc10-7551
 CAAAACTTGATATGACCGATCTCAC (配列番号 240)
 >mtb-gc10-12444
 GATATCGCGCTATCGGTAAACTAAT (配列番号 241)

mtb-gc10-8929 mtb-gc10-2100 3487428-3487527 100
 >mtb-gc10-8929
 CATTACCTCTATCACTTCGGCTAA (配列番号 242)
 >mtb-gc10-2100
 AATCCGAACGAACACATAGCATTG (配列番号 243)

10

20

30

40

mtb-gc10-17338 mtb-gc10-13891 4056910-4057009 100 >mtb-gc10-17338 TCATGTTGATAAGGCGACGAAAC (配列番号 244) >mtb-gc10-13891 GGCCTTATCTAAACCACTGAAGTTT (配列番号 245)
mtb-gc10-8689 mtb-gc10-13874 4062276-4062375 100 >mtb-gc10-8689 CATCCTTATAGGAACATCACAGACT (配列番号 246) >mtb-gc10-13874 GGCATTCCGTAGCTTTGAAATTTC (配列番号 247)
mtb-gc10-17547 mtb-gc10-8941 4062278-4062377 100 >mtb-gc10-17547 TCCTTATAGGAACATCACAGACTTC (配列番号 248) >mtb-gc10-8941 CATTTCGTAGCTTTGAAATTCCCC (配列番号 249)
mtb-gc10-9500 mtb-gc10-7386 4062279-4062378 100 >mtb-gc10-9500 CCTTATAGGAACATCACAGACTTCA (配列番号 250) >mtb-gc10-7386 ATTTCGTAGCTTTGAAATTCCCC (配列番号 251)
mtb-gc10-11046 mtb-gc10-21368 4062280-4062379 100 >mtb-gc10-11046 CTTATAGGAACATCACAGACTTCAC (配列番号 252) >mtb-gc10-21368 TTTCCGTAGCTTTGAAATTCCCC (配列番号 253)

(表 8 中の配列の上側の見出しには、相同プローブ配列の識別子がそれぞれの順番で示された後、ゲノム標的座標と、第 1 の 25-mer のスタートから第 2 の 25-mer のエンドまでの標的配列の長さとが示される)

10

20

30

40

50

【0 2 0 2】

さらに、M. ツベルクローシス (M. tuberculosis) 感染の主要な第 1 線治療の 2 つであるリファンピシンおよびイソニニアジドへの耐性を付与する突然変異が発生していることが示される遺伝子に焦点を合わせて、M. ツベルクローシス (M. tuberculosis) ゲノムの特定の領域に対してプローブ配列を生成した。

【0 2 0 3】

これらのプローブは特異性について実施例 1 に記載されるようにスクリーニングしたが、この場合は、特定の T_m に限定しなかった。特に、リファンピシン耐性突然変異が集中される M. ツベルクローシス (M. tuberculosis) rpoB 遺伝子の特定の 81 塩基領域を捕獲するように設計した。この領域を捕獲するように設計された 2 つのプローブ配列対は次の通りである：

> m t b - H 3 7 R v - r p o B - p r - 0 1 - H 1 : G G T C G C C G C G A T C A A
G G A G T T C T T C (配列番号 254)
> m t b - H 3 7 R v - r p o B - p r - 0 1 - H 2 : C A T C G A A A C G C C G T A
C C G C A A G G T G (配列番号 255)
> m t b - H 3 7 R v - r p o B - p r - 0 2 - H 1 : G T T C A T C G A A A C G C C
G T A C C G C A A G (配列番号 256)
> m t b - H 3 7 R v - r p o B - p r - 0 2 - H 2 : A C C C A G G A C G T G G A G
G C G A T C A C A C (配列番号 257)。

【0 2 0 4】

イソニアジド耐性突然変異が発生するM.ツベルクローシス (M. tuberculosis) in hA 遺伝子に特異的なプローブを同様に同定した。この領域を捕獲するよう に設計された一対のプローブ配列は次の通りである：

>mtb-37rv-inha-pr-01-H1 : TCGAACTCGACG T G C A
A A A C G A G G A (配列番号 258)

>mtb-37rv-inha-pr-01-H2 : G G C G T A T T C G T A T G C T
T C G A T G G C C (配列番号 259)

【0205】

実施例3：C.ディフィシル (C. difficile) 毒素A遺伝子に向けられたプローブの生成

10

クロストリジウム・ディフィシル (Clostridium difficile) の 毒素A遺伝子に特異的なプローブは、本質的に、S.ニューモニエ (S. pneumoniae) について実施例1で説明される通りに製造した。簡単に言うと、標的病原体 (クロストリジウム・ディフィシル (Clostridium difficile) 630) の標的領域 (g i 1 1 5 2 4 9 0 0 3 : 7 9 5 8 4 3 - 8 0 3 9 7 5 クロストリジウム・ディフィシル (Clostridium difficile) 630 - tcdA 遺伝子) を実施例1で説明されるように 25-mer にスライスし、複製配列、二次構造を有する配列、または同一ヌクレオチドの4つよりも多い連続した反復を有する配列を排除した。この場合、固定CGまたは固定Tmについてはスクリーニングしなかった。また、以下のC.ディフィシル (C. difficile) 毒素A遺伝子配列と特異的にハイブリッド形成するように、プローブ配列をスクリーニングした：g i 2 6 0 6 8 1 7 6 9 : 7 1 8 4 7 4 - 7 2 6 6 0 6 クロストリジウム・ディフィシル (Clostridium difficile) CD196、完全ゲノム；g i 2 6 0 6 8 5 3 7 5 : 7 1 5 9 9 5 - 7 2 4 1 2 7 クロストリジウム・ディフィシル (Clostridium difficile) R20291、tcdA 遺伝子；およびg i 1 4 4 9 2 5 g b M 3 0 3 0 7 . 1 C L O T O X A C D C.ディフィシル (C. difficile) 毒素A遺伝子、完全c ds。実施例1の場合と同様にヒトゲノムに対して25-merをスクリーニングして、ヒトDNAとクロスハイブリッド形成する可能性のあるものを排除した。実施例1と同じ微生物およびウイルスゲノムのNCBIデータベースに対して特異的にハイブリッド形成しないようにプローブ配列をスクリーニングした。100~200ヌクレオチドの長さの標的領域を捕獲するようにプローブ配列対を構築した。クロストリジウム・ディフィシル (Clostridium difficile) 毒素Aプローブのための対は、各プローブ配列対に対するゲノム位置情報を含む以下の表11に記載される。

20

30

表 9 : C. ディフィシル (C. difficile) のために構築されたプローブ配列

>cdif-toxA-1.L50 pos1467-1566 CTCGCTCCACAATAAGTTAAGTGG (配列番号 260)	10
ATTCAGCTACCGCAGAAAACCTCTAT (配列番号 261)	
>cdif-toxA-1.L120 pos1467-1566 CTCGCTCCACAATAAGTTAAGTGG (配列番号 262)	
ATTCAGCTACCGCAGAAAACCTCTAT (配列番号 263)	
>cdif-toxA-2.L50 pos8185-8284 TGATGGAGTAAAAGCCCCTGGGATA (配列番号 264)	
CTTTATGCCTGATACTGCTATGGCT (配列番号 265)	
>cdif-toxA-2.L120 pos8185-8284 TGATGGAGTAAAAGCCCCTGGGATA (配列番号 266)	
CTTTATGCCTGATACTGCTATGGCT (配列番号 267)	
>cdif-toxA-3.L100 pos3114-3263 ATAACAGAGGGGATACCTATTGTAT (配列番号 268)	20
CCTCAGTTAACGGTTCAACTTTATGC (配列番号 269)	
>cdif-toxA-3.L170 pos3114-3263 ATAACAGAGGGGATACCTATTGTAT (配列番号 270)	
CCTCAGTTAACGGTTCAACTTTATGC (配列番号 271)	
>cdif-toxA-4.L150 pos1528-1727 ATAAATAGTCTATGGAGCTTGATC (配列番号 272)	
TTTTATGCCAGAACGCTCGCTCCACA (配列番号 273)	30
>cdif-toxA-4.L250 pos1528-1727 ATAAATAGTCTATGGAGCTTGATC (配列番号 274)	
TTTTATGCCAGAACGCTCGCTCCACA (配列番号 275)	

【 0 2 0 6 】

実施例 4 : H I V 中の薬物耐性突然変異を検出するためのプローブの生成

この実施例は、H I V - 1 の存在を検出し、薬物耐性突然変異を検出し得るプローブの選択方法を提供する。H I V R T 、プロテアーゼ、融合、およびインテグラーゼ遺伝子における 65 の薬物耐性座位のリストをまず作成した。これらの座位は、スタンフォード大学の H I V D r u g R e s i s t i r a n c e D a t a b a s e および以下のウェブサイトの表から得た：

http://hivdb.stanford.edu/cgi-bin/NRTIResiNote.cgi
 http://hivdb.stanford.edu/cgi-bin/NNRTIResiNote.cgi
 http://hivdb.stanford.edu/cgi-bin/PIResiNote.cgi
 http://hivdb.stanford.edu/cgi-bin/FIResiNote.cgi

40

50

http://hivdb.stanford.edu/cgi-bin/INIResiNote.cgi。

【0207】

また、1522のHIVゲノム配列のセットは、NCBIからダウンロードした。BioPerl module Bio::Tools::dpAlignを用いて、1522のゲノム配列のそれぞれにおける各耐性突然変異の位置を決定した。それぞれのゲノムについて、3つのフレーム全ておよび両方のオリエンテーションに対して、各遺伝子をアライメントさせて、最良のアライメントを決定した。次に、耐性突然変異の位置をセンサス配列からゲノム配列へ位置付けた。

【0208】

プローブ設計パイプラインへの入力として、1522のHIVゲノム配列のうちの100をランダムに選択した。候補プローブ配列（プローブアーム）のセットを生成するために、20～30の長さを有し、100の入力配列のいずれかにおける任意の50塩基の耐性突然変異内で発生する全てのn-merのリストを作成した。これらのn-merは、耐性突然変異の少なくとも1つを明らかにできる配列決定読取りを生成し得る候補プローブ配列である場合に選択した。3を超える長さを有するホモポリマー（例えはプローブのマイクロアレイ合成中に使用され得る酵素に関連する制限部位）を含有するn-merである場合に、複製をn-merのリストから除去した。候補プローブ配列をさらにフィルタリングして、100の入力HIV株のうちの20またはそれ以上において存在するものだけを保持した。

【0209】

次に、プローブ設計ソフトウェアは、各n-merに対して、連結側プローブアームおよび伸長側プローブアームとしてのその望ましさを説明する2つのスコアを生成した。スコアは本明細書において記載されるように生成し、望ましいプローブアームの融解温度の分布を通常よりも2度高くなるように選択した。各候補プローブアームがスコア化されたらすぐに、長さ20の共通の接頭語を共有するセットから最良候補が選択され、ここで、最良候補は、連結側プローブアームとしてのスコアおよび開始側プローブアームとしてのスコアの最高合計によって同定した。不十分にスコア化された候補プローブアーム（すなわち、機能する期待確率が.25未満であったもの）をさらなる考察から捨て去った。このプロセスにより、様々な長さ（20～30ヌクレオチド）を有する候補プローブアームを調べて、最良の融解温度および他の特徴を有するものを見出すという目標が達成された。

【0210】

次に、残存するプローブアームをそれぞれ、ヒトゲノム配列（Genome Reference Consortiumによって生成された2009年2月のヒト基準配列[GRC37/hg19]、http://genome.ucsc.edu/cgi-bin/hgGatewayで入手可能）と、短い読取りアライメントプログラムBowtie（http://bowtie-bio.sourceforge.net/index.shtmlで入手可能）を使用する、米国特許第6,252,059号明細書で提示される配列との、2つの排除データベースに対してアライメントさせた。1または0個のミスマッチでいずれかのデータベースと一致するいかなる候補プローブアームも廃棄した。次に、残存する候補プローブアームを、Bowtieを用いて100のHIV標的ゲノムとアライメントさせた。

【0211】

次に、プローブ捕獲領域によって包含される耐性突然変異部位の標的リストを準備した。リストは、各株に位置付けされるように全ての既知の耐性突然変異に対して1つのエンタリーを含有する（すなわち、65*100=6500エンタリー）。次に、少なくとも2つのプローブアームの配列決定読取りがリスト上の各エンタリー（すなわち、各株における各突然変異部位）を含むようにプローブアームを選択するように、プローブアーム選

10

20

30

40

50

択プロセスを設計した。

【0212】

各候補プロープアームに対して、プロープアームが連結側プロープアームおよび開始側プロープアームとして使用される場合に、プロープアームの配列読取りによって包含され得る6500のリストにおける耐性突然変異部位の数を決定した。これは、各ゲノムに対する候補プロープアームのBowieアライメントを調べ、プロープアームの位置の固定距離(50塩基)内の耐性突然変異部位の数をカウントすることによって行った。このステップは、候補プロープアームが良好に一致するHIV株の数を考慮に入れる。

【0213】

100のHIV標的株を任意の順序で処理して、その株において互いに85~250塩基の範囲内で発生する候補プロープアーム配列に基づいて、各株に対する候補完成プロープ(すなわち、完成プロープに構築するためのプロープアーム配列の対)を生成した。プロープが機能する期待確率が.5よりも大きい場合にだけ、各候補プロープを保持した。次に、このプロープからの配列決定読取りによって包含され得る耐性突然変異のリスト(6500の中から)を完成した。これは、包含範囲リスト(coverage list)を表す。この計算は、プロープを形成するために結合される2つの候補プロープアームからのリストを結合し、候補プロープアームが300塩基以内であり、そのゲノムにおける正しいオリエンテーションにある場合にだけ、ゲノムのエントリーが保持される。

【0214】

各プロープに対する包含範囲リストの合計に基づいて候補プロープをソートし、最高の合計を有するプロープ、すなわち、最大数の耐性突然変異を包含するプロープを選択した。

【0215】

2つのプロープによって既に包含されている耐性突然変異を反映するように、残存する候補プロープに対する包含範囲リストを更新した。包含されていないどの耐性突然変異も包含しないプロープを考察から除去した。

【0216】

このプロープ選択プロセスの実施において、プロープが全く残らないか、あるいは全ての耐性突然変異が2つのプロープによって包含されていれば、プロセスは中止され得る。プロープが残っていれば、候補リストは各プロープに対する包含範囲リストの合計に基づいて再度ソートされ、最高の合計を有するプロープ、すなわち、最大数の耐性突然変異を包含するリストからのプロープが選択され得る。

【0217】

いくつかの場合には、選択された全てのプロープのプロープアームに突然変異を導入した。突然変異は、プロープアームが1522のHIVゲノムのいずれかと19塩基対よりも多く一致しなくなるまで骨格側から開始して、捕獲側に向かって機能させて、プロープアーム内の各位置におけるバリエーションを試みることによって発生させた。プロープアームにおけるこのような全てのバリエーションの融解温度を計算し、1.5度に最も近い融解温度の低下を起こすバリエーション(Melting 5.0.3(<http://www.ebi.ac.uk/compneur-srv/melting/melting5-doc/melting.html>で入手可能))によって計算されるように、オリジナルプロープアームおよび突然変異プロープアームの不完全な二本鎖に基づく)を新しいプロープアームとして保持した。従って、初期パラメータにおける所望の融解温度を上昇させ、ミスマッチを有するより低い融解温度の達成を試みることによって、最終プロープアームは、実験条件下で非突然変異プロープと同様に挙動し得る。

【0218】

次に、Bowieにより1522の全てのHIVゲノムに対して突然変異プロープアームをアライメントさせ、1522のうちのいくつが少なくとも1つのプロープによって捕獲され得るか、そして1522の株を横切る65の耐性突然変異のうちのいくつが捕獲されるかを決定した(理論上は1522*65、すなわち全部で98930の座位が存在

10

20

30

40

50

するが、全ての耐性突然変異が全ての株に対して位置付けされ得るわけではないので 86, 905 の座位が同定可能であった）。この分析に基づいて、標的株のセットを増加させ、323 株に対してプロセスを繰り返した。オリジナルの 100 株に、最初のラウンドにおいて少数のプローブでしか捕獲されないかどのプローブでも捕獲されなかつた新しい 223 株を加えたものを使用した。初期パラメータに対する唯一の変化は、最初の 20 ではなく、7 以上の株において見出される候補プローブアームを保持したことであった。

【0219】

プローブ設計プロセスの最終ステップは、467 の予備的なプローブ配列をフィルタリングして、プール中の他のプローブとクロスハイブリッド形成またはクロスプライム (cross-prime) し得るプローブを除去することであった。このフィルタリングはプローブの互いに対するアライメントおよびそれ自体のアライメントに基づき、その後、アライメントされた領域において融解温度を計算して、実験条件下で二本鎖が形成する尤度を決定した。このフィルタリングにより、ヘアピンを形成する可能性があるために 34 プローブが除去され、他のプローブとクロスプライムする可能性があるために 56 プローブが除去されて、376 プローブが残された。これらの 376 のプローブは、1522 株のうちの 1384 に対して少なくとも 1 つのプローブを含有する。いくつかのプローブは 2000 を超える株を捕獲するが、多くの捕獲はただ 1 つまたは数個である。これは通常、多数の株において耐性突然変異を捕獲したプローブがまず選択され、1 つまたはいくつかの株に特異的なプローブが最後に選択されるような、プローブが選択される順序を反映する。

10

20

【0220】

実施例 5：HPV の株を区別するプローブの生成

この実施例は、ヒト乳頭腫ウイルスの配列決定された 288 株（137 の別個の種類からなり、いくつかの種類は多数の分離株または株を有する）の公的に利用可能なゲノムを検出および区別し得るプローブの選択方法を提供する。プローブ選択プロセスの目標は、これらのプローブによって捕獲される対象の領域からの配列読取りが、株の任意の対の間で少なくとも 7 つの SNP または小さい挿入欠失を明らかにし得るようにプローブを選択することである。

【0221】

プローブ設計パイプラインは、288 株全てからの長さ 18 ~ 26 の全ての n-mer のリストを生成することにより開始される。次に、3 よりも大きい長さを有するホモポリマーストレッチを含有するか、あるいは特定の制限酵素部位を含有する n-mer を廃棄した（特定の酵素は、マイクロアレイ上で合成されたプローブを処理するために使用されるので、プローブ配列中のこのような部位は、いくつかの実施形態では、全てのプローブが全ての可能な合成オプションと適合性であることを保証することが可能でないかもしれない）。次に、HIV 特異的 n-mer について実施例 4 で記載されたように、連結側プローブアームおよび開始側プローブアームとしてのその望ましさに従って、残存する 9, 825, 946 の n-mer のそれぞれをスコア化した。実施例 4 の場合と同様に、18-塩基接頭語が与えられた最高スコアリングのプローブを保持した。方法はさらに、プローブをフィルタリングして、ヒトゲノムに対して完全または 1 塩基対ミスマッチを有するものを除去し、プローブ選択で使用するために 715, 533 が残った。

30

40

【0222】

各軸に沿って 288 の HPV 株のそれぞれを用いて四角いマトリックスを構築した（しかし、マトリックスの上半分のみを使用して、各ペアワイズ結果を四角いマトリックス中に 1 回だけ示す）。マトリックス中の各エントリーは、選択されるプローブからの期待される読取りとともに方法が包含しようとする SNP または小さい挿入欠失の数を示した。従って、このマトリックスは所望の SNP のマトリックスであり、すなわち、マトリックスは、株の任意の対の間で完成プローブセットがいくつの差異を明らかにするように選択されるかを示した。この場合、全てのエントリーを 7 に設定した（または「初期化した」）。他のプローブ設計タスクはマトリックスを異なって初期化し得る。例えば、2 つの株

50

が臨床的に同一であると考察されると、マトリックスは、これらの株に対してゼロエントリーを有するかもしれない、これらを区別する必要がないことが示される。特定の株がより高い包含範囲を必要とする場合には、これらの株に対応するエントリーはより高い値を含有し得る。

【0223】

各n-merのプローブアームとしての有用性を決定するために、プローブ選択方法を用いて、n-merによって株の間でいくつのSNPが明らかにされるかを決定した。従って、Bowtieを用いて288株のセットに対してn-merをアライメントさせ、各n-merのアライメントにおいて1つのミスマッチを可能にした。各n-merおよびn-merがアライメントされる株の各対（順序依存性の形で）について、このn-merが連結側プローブアームとして使用される場合には、n-merの下流側の2つの領域のアライメントを実施して、各領域を通して配列決定読取りから観察され得るSNPおよび小さい挿入欠失の数を決定した。アライメントにおいて使用される隣接領域の長さは期待される配列決定読取り長さに依存し、この場合には、50塩基の隣接領域が使用される。n-merの50塩基上流側のアライメントも実施して、n-merを開始側プローブアームとして使用する場合に検出され得るSNPおよび小さい挿入欠失の数を決定した。従って、各n-merに対して、株の対の間で観察される差異の2つのマトリックスを計算した。一方は、連結側プローブアームとしてのn-merに対するマトリックスであり、他方は、開始側プローブアームとしてのn-merに対するマトリックスである。1つのn-merに対するアライメントの一例は以下に示されるが、ここで、アスタリスクはその位置における100%の同一性を示し、株は左側に示される。

FM955841	AGTTGTTGCAACAGCATTGCGACTATATCTGGGTTA (配列番号 276)
M32305	AGCTGTTGCAACAGCATTGTGACTATATATGGGTCC (配列番号 277)
FM955838	AGTTATTGCAACAGCATTGTGACTATATTGGGATTA (配列番号 278)
D90252	AGCTGTTGCAACAGCATTGTGACTATATCTGGGTCC (配列番号 279)
M22961	AGCTATTGCAACAGCATTGTGACTATATCTGGGTCC (配列番号 280)
NC_001531	AGCTATTGCAACAGCATTGTGACTATATCTGGGTCC (配列番号 281)

*** * ***** * *** *

【0224】

このn-merは、株FM955841とM32305との間にはSNPが3つあり、M22961とNC_001531との間ではなく、FM955838とD90252との間には6つあることを明らかにした。

【0225】

一対のn-merを含有するプローブを構築するために、全ての288HPV株を任意の順序で処理し、互いに300塩基の範囲内にあるn-merを結合することによって、各株に対してプローブを生成した。以下の値(1)および(2)に基づいて、各候補プローブをスコア化した：

(1) プローブが機能する確率、および
 (2) プローブにより株間で明らかにされ得るSNPまたは小さい挿入欠失の期待数（プローブにより株間で明らかにされ得るSNPまたは小さい挿入欠失の期待数は、2つのプローブアームに対して観察されるSNP / 挿入欠失マトリックスを合計することによって得た。プローブが機能しない（例えば、プローブアームが遠く離れすぎているか、または誤ったストランドオリエンテーションにある）株に相当する値をゼロに設定した。さらに、マトリックス中の最大値を、3または標的マトリックス中の対応するエントリーの値の小さい方の値に設定した。最終的なプローブの数は、このマトリックス中の全てのエントリーにわたる合計であった。）。

プローブの最終スコアは、値(1)と(2)の積であった。

【0226】

10

20

30

40

50

次に、最高スコアを有するプローブを選択し、次に、プローブの観察されたS N P /挿入欠失マトリックス値を、所望の標的マトリックスから差し引いた（結果における負の値をゼロに設定した）。次に、残存するプローブに対するスコアを更新した。残存するプローブは選択されたプローブによって既に包含されている株の間の差異を検出し得るので、スコアはこのプロセスの間に低下するだけであり得る。このようにして（すなわち、プローブを選択し、残存する候補プローブを再スコア化する）、標的マトリックスが全てゼロを含有する（選択されたプローブが、株の各対の間で、少なくとも7つのS N Pまたは挿入欠失を明らかに得ることを意味する）まで、あるいは、残存する候補プローブがどれも非ゼロスコアを有さなくなる（残存する候補プローブがどれも、まだ検出されていない株の間の差異を明らかにしないことを意味する）まで、プローブ選択を続けた。

10

【0227】

この反復プローブ選択プロセスにより、548のプローブを選択した。実施例4と同様に、ヘアピン、クロスプライミング（cross-priming）、およびクロスハイブリダイゼーションについてのプローブのフィルタリングにより、346のプローブが残った。

【0228】

これらの346のプローブおよび高リスクH P V株のセット（H P V 16、18、31、33、35、39、45、51、52、56、58、59）を用いてH P V株の検出のシミュレーションが実施される場合、73のプローブが産物を生成することが期待された。図17は、シミュレーションにおいてどのプローブ（x軸）がどの株（y軸）に対して機能したかのマトリックスを示し、白色ブロックは期待された産物を示し、そして黒色ブロックはプローブがその株から産物を生成しなかったことを示す。

20

【0229】

実施例6：臨床サンプルにおけるH P V株の検出

図18は、標的H P V株ゲノムに対する20の特異的H P Vプローブの群のための標的マトリックスを示す。プローブはプロットのx軸を横切って表され、株はy軸に沿って表される。白色領域は、表示される対応する株のゲノムに結合することが予想されるプローブを示し、黒色領域は、対応する株に結合することが予想されないプローブを示す。

【0230】

図19は、標的マトリックスを示す。27の特異的H P Vプローブのそれぞれによって同定されるS N Pの数および種類を示すように拡張される。異なるグレースケールシェーディングは、T、C、G、またはAのそれぞれに対する任意の特定の塩基変化、あるいは挿入欠失の存在G r a y = I n d e 1を示し、黒色は、その位置でその株からの読み取りがないことを示す。個々のプローブはx軸に沿って示され、各プローブは、1つよりも多いS N Pを捕獲する場合には、1つの列または多数の列に分割される。

30

【0231】

本明細書において記載されるような方法を用いて、図20に示されるゲル中の特定のサンプルに対するレーン番号によって示されるように、H P V 16指向性プローブ（N C 0 0 1 5 2 6 _ 4 0 0 5、N C 0 0 1 5 2 6 _ 3 9 9 9、またはN C 0 0 1 5 2 6 _ 7 2 9 9）またはH P V 18指向性プローブ（A Y 2 6 2 2 8 2 _ 7 1 7 4、A Y 2 6 2 2 8 2 _ 3 3 0 9、またはA Y 2 6 2 2 8 2 _ 1 4 5 0）を、H P V 16および18のいずれかを含有する臨床サンプル（T h i n P r e p）からのDNAと結合させた。ハイブリダイゼーションならびにそれに続くギャップ充填ポリメラーゼ伸長および連結（環状化捕獲）の後、P C Rを実施して、環状化プローブを検出した。いくつかのサンプル（レーン1～3および11～13で示される）において、期待サイズ（250 n t）のP C R增幅産物を検出した。H P V 16指向性プローブはH P V 16を検出し、H P V 18指向性プローブはH P V 18を検出したが、H P V 16は検出しなかった。

40

【0232】

図21は、上記の図20に対応するサンプル中で生成される增幅産物のS a n g e r配列決定のアライメントの一例を示す。H P V 16およびH P V 18基準ゲノムに対して配

50

列をアライメントさせ、ポリメラーゼ伸長領域によって捕獲された配列を示した。

【0233】

実施例7：臨床サンプルにおける細菌DNAの検出

本明細書に記載されるような環状化捕獲において単一の *S. saprophyticus* (*S. saprophyticus*) 指向性プローブを用いて、尿路感染症 (UTI) の患者からの臨床サンプル中で *Staphylococcus* · *Saprophyticus* (*Staphylococcus saprophyticus*) ゲノムDNAを検出した (図22A)。また、単一プローブ (「193」プローブ) または *Meca* 遺伝子領域に向けられたプローブを含むプールされたプローブ混合物 (「A11 *Meca* プローブプール」) のいずれかを用いて、細菌臨床分離株中においても *S. saprophyticus* (*S. saprophyticus*) DNAを検出した (図22B) (全てのサンプルにおいて、期待されるサイズのバンドが見られる；臨床分離株は NY356、GA15、および CA105) で示される)。

【0234】

基準 *Staphylococcus* · *Saprophyticus* (*Staphylococcus saprophyticus*) 株からのゲノムDNAとの、PCR増幅された環状化プローブの観察された配列決定読取りのアライメントにおいて観察されるように、順方向および逆方向の Sanger 配列決定によって、図22Aの *Staphylococcus* · *Saprophyticus* (*Staphylococcus saprophyticus*) 指向性プローブを用いる標的gDNAのポリメラーゼ伸長および捕獲が示された。

【0235】

また、Sanger 配列決定によって、*Staphylococcus* · *Aureus* (*Staphylococcus aureus*) 指向性プローブと結合させたときに、ゲノム *Staphylococcus* · *Aureus* (*Staphylococcus aureus*) 配列との、PCR増幅された環状化プローブの観察された配列決定読取りのアライメントにおいて示されるように、*Staphylococcus* · *Aureus* (*Staphylococcus aureus*) 標的gDNAのポリメラーゼ伸長および捕獲も示された (図23)。

【0236】

実施例8：臨床サンプルにおけるウイルスDNAの検出

また、培養インフルエンザウイルスから単離したRNAから逆転写されたcDNAは、5つの個々の分子反転プローブを用いて検出し、通常の Sanger (N) または次世代配列決定 (T、テールドプライマー) のための増幅は、図24に示される (プローブは 198、256、292、293、および 462 で示される； *S. sap* は *Staphylococcus* · *Saprophyticus* ゲノムDNA対照を示す)。

【0237】

実施例9：臨床UTIサンプルにおける細菌DNAの多重検出

尿路感染症における役割を有する可能性のある生物に向けられた 60 の完成プローブのプールを、3nM の全核酸濃度で調製した (等モル比の各プローブを含有する)。

【0238】

プローブプールを、約 4 μl の 33 の個々の臨床尿路感染症 (UTI) サンプルおよび 4 つの対照サンプルと 24 時間ハイブリッド形成させた。1マイクロリットルあたり 0.1 pg ~ 100 ng の可変量の dsDNAを含有するように picogreen により各臨床サンプルを定量化した。

【0239】

ポリメラーゼギャップ充填、連結、および消化反応を実施し、プローブの普遍的骨格とハイブリッド形成する 3' 位置と、 *Illumina* フローセル (*Illumina Inc.*, *San Diego*, CA) とのハイブリダイゼーションのために必要とされるアダプター配列を含有する 5' テールとを含有する普遍的プライマーにより、任意の環状化産物を増幅した。非ハイブリッド形成 6 - ヌクレオチドバーコード挿入断片を含有す

10

20

20

30

40

50

る個々の 3' プライマーを用いて、独特の DNA 配列タグにより個々の臨床サンプルからの増幅産物を標識化して、このサンプルからの配列読取りの部分配列の同定を可能にした。

【 0 2 4 0 】

期待されるサイズの増幅産物は、2 % アガロースゲルにおいて分解した後に切除した。配列決定のための調製において過剰のアガロースおよび塩から増幅産物を精製した。37 サンプルのそれぞれを 6 - ヌクレオチドバーコードによりバーコード化することにより、全てのサンプルを Illumina G A I I 装置における単一の配列決定ランへ多重化した。この分析中に含まれていない追加のサンプル（および異なるバーコード）により、これらのサンプルをさらに多重化した。配列決定ランは、およそ 3300 万の読取りを生じた。

10

【 0 2 4 1 】

60 の U T I プローブのためのプローブアームをゲノムおよび部分ゲノムの大きい集合とアライメントさせた。各プローブに対する各一致について、左側プローブアーム、伸長領域、右側プローブアーム、および 6 - ヌクレオチドバーコードと右側プローブアームとの間の 21 - ヌクレオチドの骨格配列からなる「期待される読取り」を構築した。これらの 10,886 の期待される読取りから Bowtie データベースを構築した。

20

【 0 2 4 2 】

読取りをアライメントするために、 Illumina 塩基呼び出しソフトウェアによって生成される F A S T Q ファイルをまず別々のファイル（各バーコードに対して 1 つ）に分割した。各バーコード（読取りの最初の 6 つのヌクレオチド）を既知のすべてのバーコードと比較した。読取りのバーコード部分が、任意の他のバーコードに対する一致よりも優れているバーコードに対して单一の一一致を有していれば、読取りをバーコードに割り当てる。バーコードに対する一致の品質は、配列決定読取りおよび期待されるバーコードがミスマッチする位置における塩基品質の合計であり、従って、高品質の一致は少ない合計（理想的にはゼロ）を有し、読取りからバーコードへのマッチングは、配列決定読取りの品質の説明となる。

20

【 0 2 4 3 】

実験において使用される 37 のバーコードのそれぞれは、1 つのバーコード当たり 11,245 ~ 4,874,885 の読取りの範囲で、少なくとも 1 つの読取りをもたらした。各バーコードに対する読取りは、コマンドラインオプション「 - p 8 - q - - t r i m 5 6 - - s o l e x a 1 . 3 - q u a l s - e 2 0 0 - - b e s t - s t r a t a - m 2 0 - k 2 0 」を有する Bowtie バージョン 0.12.7 を用いて、プローブデータベースに対して別々にアライメントさせた。従って、 Bowtie アライナー（ a l i g n e r ）は、期待される読取りに対する配列決定読取りのヒットだけを戻したが、これは、最良の一一致品質を有した（すなわち、いくつかの期待される読取りが同じ数のミスマッチを有する配列決定読取りと一致すれば、読取りは両方とも出力に含有された）。しかしながら、もう 1 つのミスマッチを有する別の期待される読取りは、その一致が、最良品質を有するものほど良好ではなかったので含有されていないであろう。さらなる詳細については、 Bowtie の「 - - b e s t - s t r a t a 」のドキュメンテーションを参照されたい。それぞれの bowtie アライメントは分析スクリプトに送り込まれた。各読取りに対して、スクリプトは、読取りがもっともらしく得られる株のセットを決定した（すなわち、読取りが最良の品質で一致する、期待される読取りに対応する株のセット）。この株のセットは、 G e n b a n k 受入番号のセット、例えば、「 A C L E 0 1 0 0 0 0 8 0 、 G G 6 6 8 5 7 8 、 N C _ 0 1 0 5 5 4 」で書くこともできるし、あるいはこれらの受入番号に相当する株のセットで書くこともできる。例えば、「 A C L E 0 1 0 0 0 0 8 0 、 G G 6 6 8 5 7 8 、 N C _ 0 1 0 5 5 4 」は、3 つのプロテウス・ミラビリス（ P r o t e u s m i r a b i l i s ）株であった。異なる読取りは、プロテウス・ミラビリス（ P r o t e u s m i r a b i l i s ）およびプロテウス・ペンネリ（ P r o t e u s p e n n e r i ）の両方を含む「 A B V P 0 1 0 0 0 0 2 5 、 A C L E 0 1 0 0 0 0 8 0 50

0、GG661996、GG668578、NC_010554」からの期待される読み取りに対して、同様に十分に位置付けすることができる。例えば、分析スクリプトは以下を報告し得る：

236 - プロテウス・ミラビリス (Proteus mirabilis) (ACLE 01000080、GG668578、NC_010554)

1 - プロテウス・ペンネリ (Proteus penneri)、プロテウス・ミラビリス (Proteus mirabilis) (ABVP01000025、ACLE 01000080、GG661996、GG668578、NC_010554)。

これにより、236の読み取りがP.ミラビリス (P. mirabilis) から期待される産物に位置し、そして1つの読み取りがP.ミラビリス (P. mirabilis) またはP.ペンネリ (P. penneri) から期待される産物に位置することが示される。従って、これらの結果は、第2のラインからの単一の読み取りが、P.ペンネリ (P. penneri) による同時感染というよりも、実際にP.ミラビリス (P. mirabilis) に由来した可能性がより高いので、P.ミラビリス (P. mirabilis) の存在を示すと解釈した。

【0244】

37の異なるサンプルからの結果は、様々な異なる生物による感染を示す。例えば、分析スクリプトにより、サンプル 7について以下が報告された：

2 - アグレガチバクター・アフロフィルス (Aggregatibacter aphrophilus)、プロテウス・ペンネリ (Proteus penneri)、プロテウス・ミラビリス (Proteus mirabilis) (ABVP01000025、ACLE 01000080、GG661996、GG668578、NC_010554、NC_012913)

324 - カンジダ・アルビカンス (Candida albicans) (AJ251858)

6 - クレブシエラ・ニューモニエ (Klebsiella pneumoniae) (ACZD01000012、EU682505、GG703525、NC_009648、NC_011283、NC_012731)

30109 - クレブシエラ・ニューモニエ (Klebsiella pneumoniae) (ACZD01000012、EU682505、GG703525、NC_009648、NC_012731)

5 - クレブシエラ・ニューモニエ (Klebsiella pneumoniae) (ACZD01000013、EU682505、GG703525、NC_009648、NC_012731)

7 - クレブシエラ・ニューモニエ (Klebsiella pneumoniae)、エシェリキア・コリ (Escherichia coli) (ACZD01000012、EU682505、GG703525、NC_009648、NC_010378、NC_012731、NC_013503)

2 - クレブシエラ・ニューモニエ (Klebsiella pneumoniae)、エシェリキア・コリ (Escherichia coli)、クレブシエラ・バリイコラ (Klebsiella variicola) (ACZD01000012、EU682505、GG703525、NC_009648、NC_010378、NC_01283、NC_012731、NC_013503、NC_013850)

30 - クレブシエラ・ニューモニエ (Klebsiella pneumoniae)、エシェリキア・コリ (Escherichia coli)、クレブシエラ・バリイコラ (Klebsiella variicola)、シトロバクター・コセリ (Citrobacter koseri) (ACZD01000012、EU682505、GG703525、NC_009648、NC_009792、NC_010378、NC_011283、NC_012731、NC_013503、NC_013850)

4 - クレブシエラ・ニューモニエ (Klebsiella pneumoniae)、

10

20

30

40

50

クレブシエラ・バリイコーラ (*Klebsiella variicola*) (ACZD 01000012, EU682505, GG703525, NC_009648, NC_011283, NC_012731, NC_013850)

656 - クレブシエラ・ニューモニエ (*Klebsiella pneumoniae*)、クレブシエラ・バリイコーラ (*Klebsiella variicola*) (ACZD 01000013, EU682505, GG703525, NC_009648, NC_011283, NC_012731, NC_013850)

2 - ラクトバチルス・ヘルベティカス (*Lactobacillus helveticus*)、ラクトバチルス・デルブリュッキ (*Lactobacillus delbrueckii*) (ACLM01000017, AEAT01000083, CP000156, CP002429, GG700753, NC_008054, NC_008529, NC_010080, NC_014727)

549 - プロテウス・ミラビリス (*Proteus mirabilis*) (ACLE 01000080, GG668578, NC_010554)

27 - プロテウス・ペンネリ (*Proteus penneri*)、プロテウス・ミラビリス (*Proteus mirabilis*) (ABVP01000025, ACLE 01000080, GG661996, GG668578, NC_010554)

7 - プロビデンシア・レットゲリ (*Providencia rettgeri*)、プロビデンシア・アルカリファシエンス (*Providencia alcalifaciens*)、プロテウス・ペンネリ (*Proteus penneri*)、プロテウス・ミラビリス (*Proteus mirabilis*)、プロビデンシア・ルスティジアニ (*Providencia rustigianii*) (ABVP01000025, ABXV02000043, ABXW01000004, ACCI02000067, ACLE 01000080, GG661996, GG668578, GG703820, GG705265, NC_010554)

76 - スタフィロコッカス・サプロフィチカス (*Staphylococcus saprophyticus*) (AF144088, AP008934, NC_007350)

310 - ウレアプラズマ・パルバム (*Ureaplasma parvum*) (CP00942, NC_002162, NC_010503)

25 - ウレアプラズマ・ウレアリチカム (*Ureaplasma urealyticum*) (CP001184, NC_011374)

5 - ウレアプラズマ・ウレアリチカム (*Ureaplasma urealyticum*)、ウレアプラズマ・パルバム (*Ureaplasma parvum*) (CP000942, CP001184, NC_002162, NC_010503, NC_011374)。

【0245】

この分析報告における読み取りの大部分は、尿路感染症の既知の一般的な原因であるクレブシエラ・ニューモニエ (*Klebsiella pneumoniae*) に由来した。またデータは、カンジダ・アルビカンス (*Candida albicans*) およびウレアプラズマ・パルバム (*Ureaplasma parvum*) を含む他の既知の尿路感染源 (*infectant*) の低レベルの存在も示す。

【0246】

カンジダ・アルビカンス (*Candida albicans*) ゲノムDNAのサンプルの結果は、恐らく、DNAを生成するために使用される細胞培養物の低汚染 (. 1 % 未満、読み取りカウントに基づく)、あるいは他のサンプルからの読み取りがこのサンプルのためのバーコードを含有するように見せる配列決定エラーのいずれかによる、C. アルビカンス (*C. albicans*) からの293, 384の読み取り、ならびにクレブシエラおよびプロテウスからの数百の読み取りを示した。

【0247】

10

20

30

40

50

この配列決定ランから尿路感染症サンプルの4つにおいて検出された異なる感染性種の割合は図25に示される。異なる主要な感染症は、プロテウス、クレブシエラ、およびウレアプラズマ感染症であると同定された。

【0248】

実施例10：環状化捕獲反応方法

環状化捕獲プロトコルは、特定のプローブおよび標的DNAサンプルに対して、最適なPCRサイクル数を決定するために種々のPCRサイクル数を用いて実施され得る（図25（i））。

【0249】

またプロトコルは、ギャップ充填および連結のために種々の長さの時間を用いて実施され得る。場合によっては、ギャップ充填は、わずか15分のインキュベーションの後に完了する（図25（ii））。

10

【0250】

プローブハイブリダイゼーションは、特定のプローブに対して最適ハイブリダイゼーション温度を決定するために、若干異なる温度で実施され得る。例えば、72または68のいずれかにおいて、10分という短い時間のハイブリダイゼーションの後に、実質的に環状化された産物が生成される（図25（iii）、インキュベーション時間（分）は各レーンに対して示される）。

【0251】

本明細書は、本明細書内で引用される参考文献の教示を鑑みて最も十分に理解される。本明細書内の実施形態は本発明の実施形態の実例を提供するものであって、本発明の範囲を限定すると解釈され得てはならない。当業者は、多くの他の実施形態が本発明によって包含されることを容易に認識する。本開示において引用される全ての刊行物、特許出願、および特許は、参照によってその全体が援用される。参照によって援用される資料が本明細書と矛盾するまたは一致しない範囲では、本明細書がこのようないかなる資料にも優先するであろう。本明細書における任意の参考文献の引用は、このような参考文献が本発明の先行技術であることの承認ではない。

20

【0252】

他に記載されない限り、特許請求の範囲を含む本明細書中で使用される成分、反応条件などの量を表す全ての数は、全ての場合において、「約」という用語で修飾されていると理解されるべきである。従って、反することが他に記載されない限り、数値パラメータは近似値であり、本発明が獲得しようとする所望の特性に応じて異なり得る。最低でも、そして特許請求の範囲に対する均等論の適用を制限する試みとしてではなく、各数値パラメータは、有効数字の数および通常のラウンディング（rounding）アプローチを鑑みて解釈されるべきである。本明細書における異なる量の有効数字を有する一連の数の記述は、より小さい有効数字が与えられた数がより大きい有効数字が与えられた数と同じ精度を有することを暗示すると解釈され得てはならない。

30

【0253】

「a」または「an」という語の使用は、特許請求の範囲および/または本明細書において「comprising」という用語と併用される場合、「1つ」を意味することができるが、「1つまたは複数」、「少なくとも1つ」、および「1つまたは1つも多い」という意味とも矛盾しない。特許請求の範囲における「or」という用語の使用は、代替のみを指すことが明確に示されない限り「および/または」を意味するために使用される、すなわち、代替は互いに排他的であるが、本開示は、代替のみ、そして「および/または」を指すという定義を指示する。

40

【0254】

他に記載されない限り、一連の要素に先行する「少なくとも」という用語は、一連の要素の全てを指すと理解されるべきである。当業者は、本明細書中に記載される特定の実施形態に対する多数の均等物を認識するか、あるいは単なるルーチン実験を用いて確かめることができるであろう。このような均等物は、特許請求の範囲によって包含されることが

50

意図される。

【0255】

他に定義されない限り、本明細書において使用される全ての技術および科学用語は、本発明が属する分野の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。本明細書中に記載されるものと類似または均等の任意の方法および材料を本発明の実施または試験において使用することができる。

【0256】

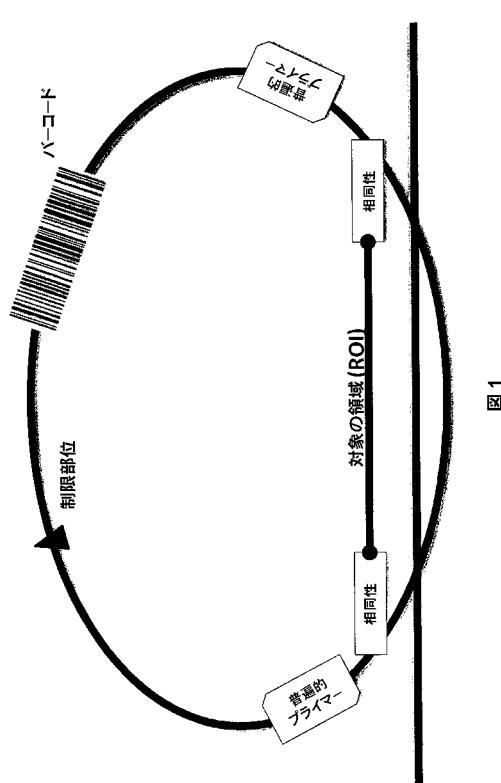
本明細書において考査される刊行物は、本出願の出願日よりも前のその開示についてのみ提供される。本明細書中のものはどれも、本発明が従来の発明によるこのような刊行物に先行する権利がないという承認として解釈されてはならない。さらに、提供される刊行物の日付は実際の公開日とは異なる可能性があり、独立して確認される必要があることもある。

【0257】

本発明の他の実施形態は、本明細書の考査および本明細書中に開示される実施形態の実行から当業者に明らかであろう。本明細書および実施例は単に例示的なものであると考えられることが意図され、本発明の真の範囲および趣旨は以下の特許請求の範囲によって示される。

10

【図1】



【図2】

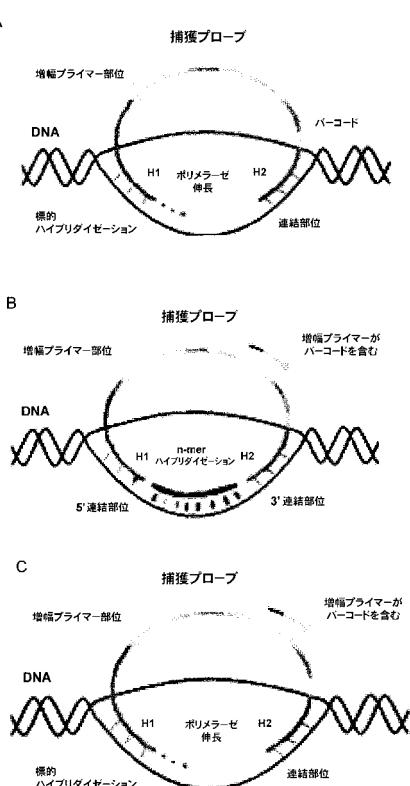
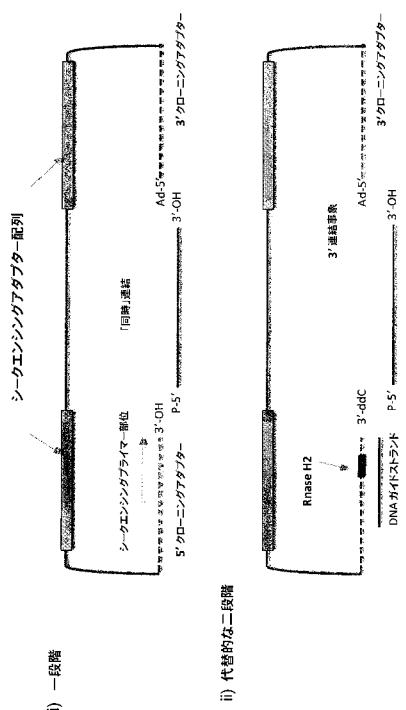
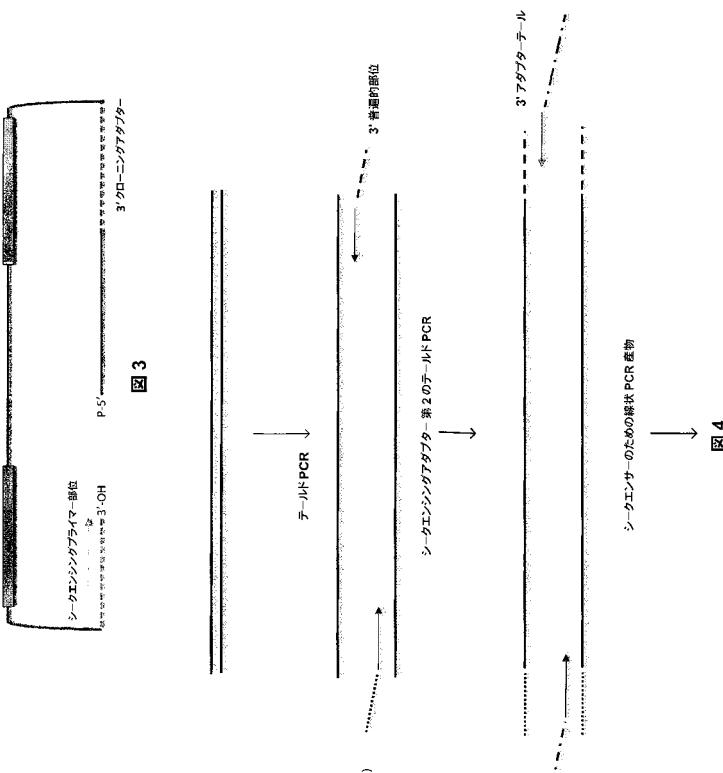


図2

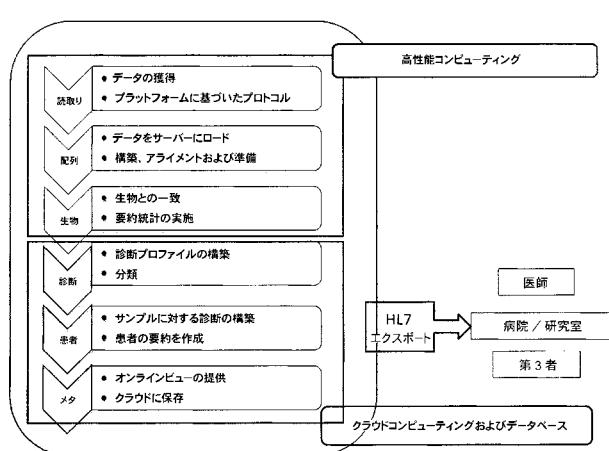
【図3】



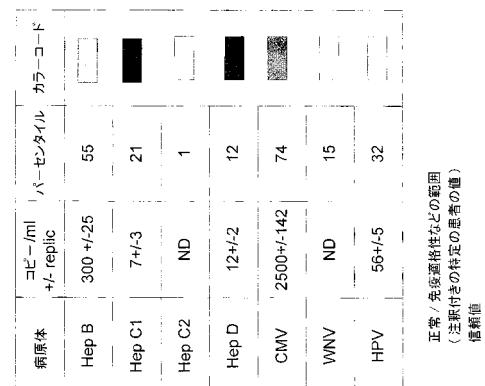
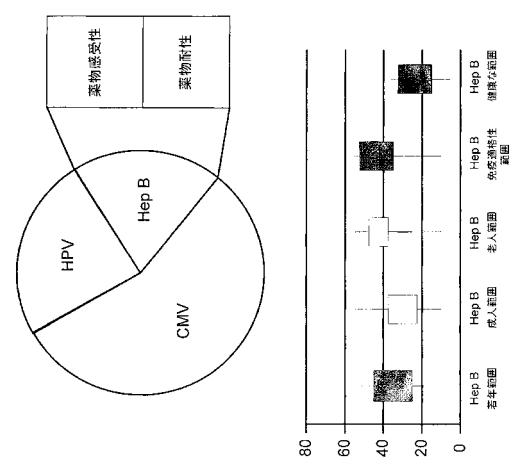
【図4】



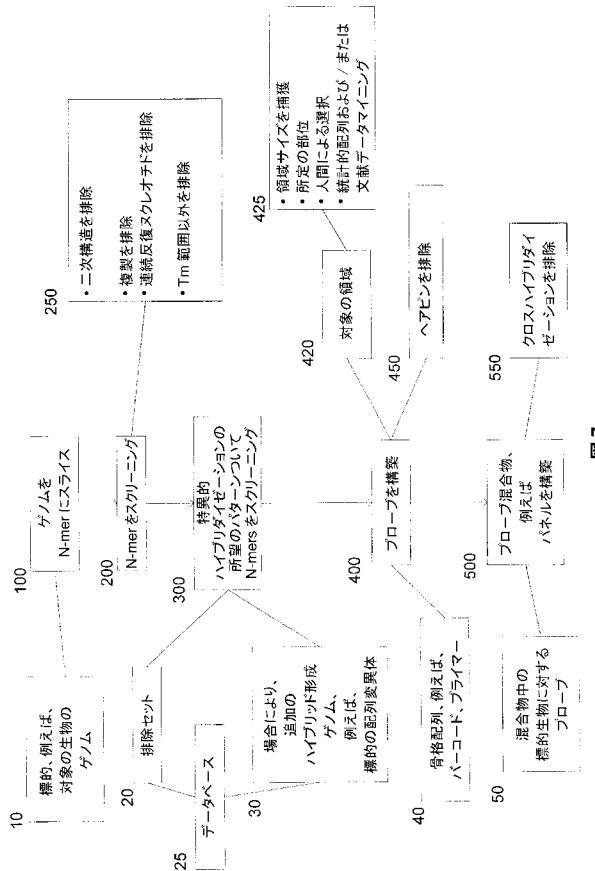
【図5】



【図6】



【図7】



【図8】

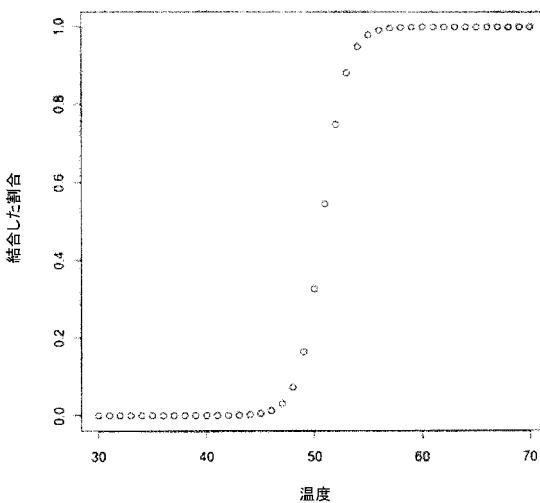


図8

【図9】

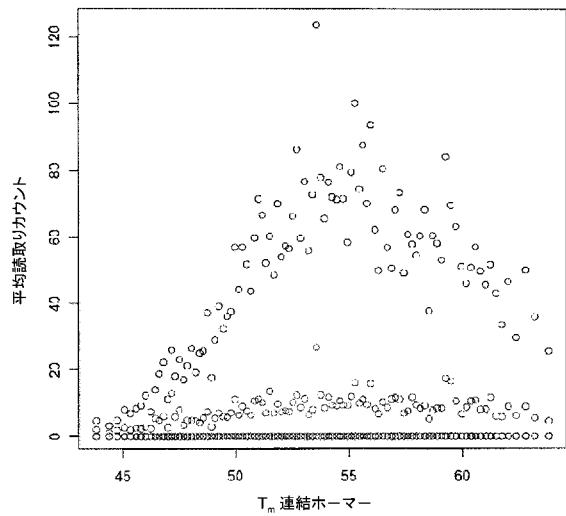


図9

【図10】

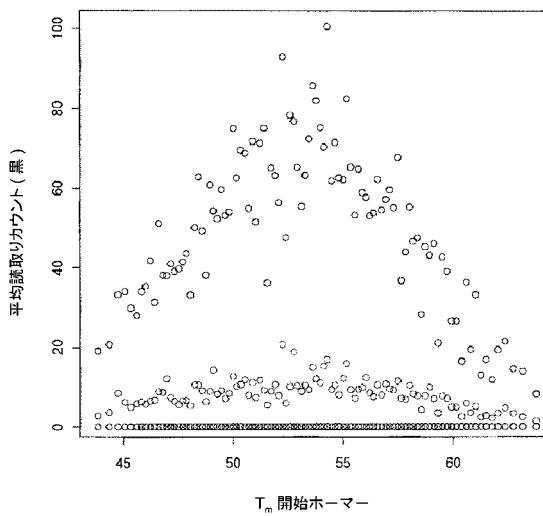
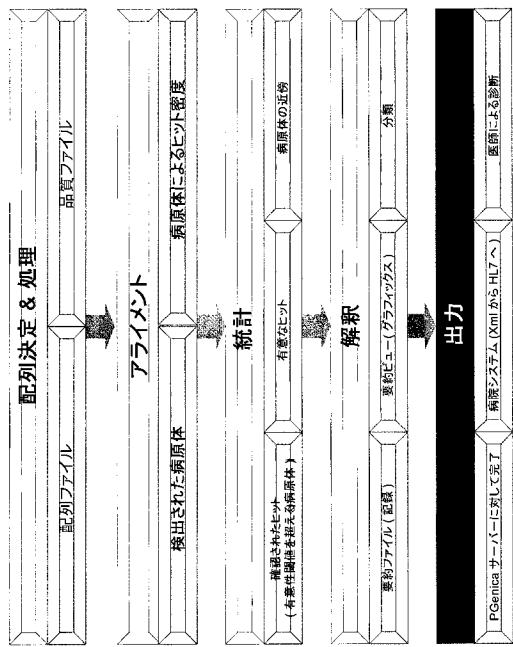
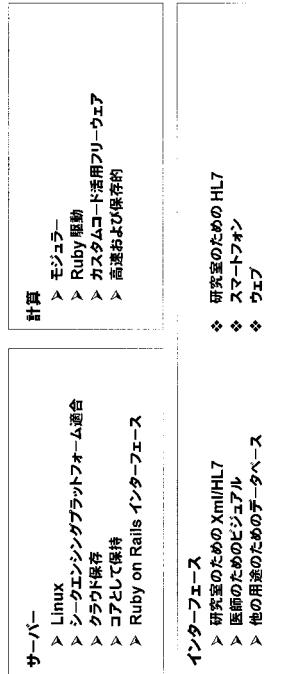


図10

【 図 1 1 】



【 図 1 2 】

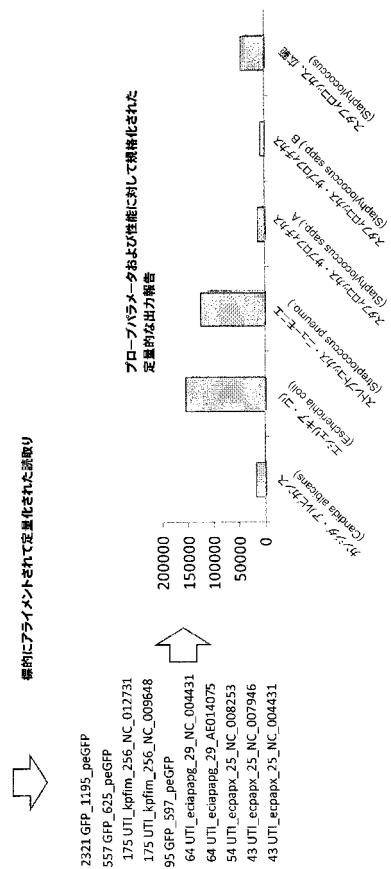


11

12

【 义 1 3 A 】

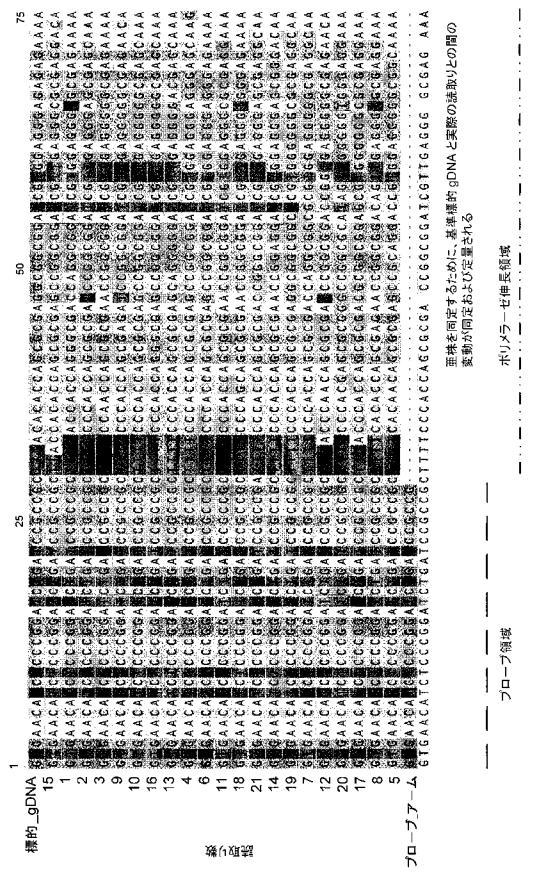
【図 1 3 B】



13A

13B
EX

【図 14】



【図 15】

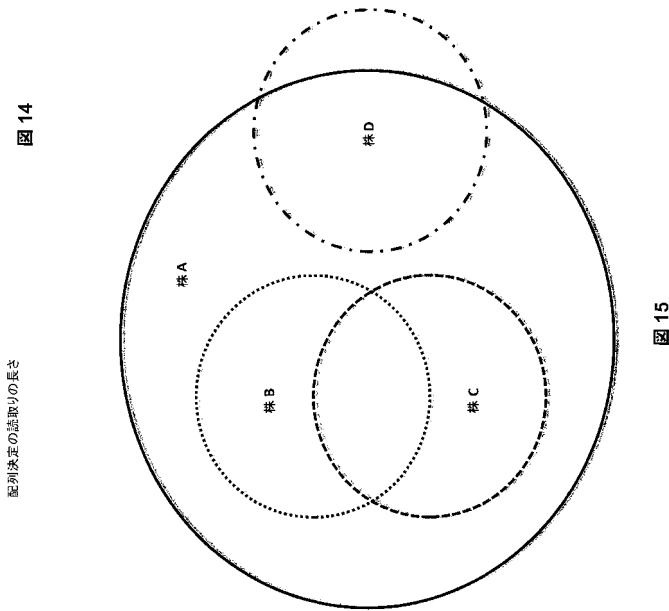
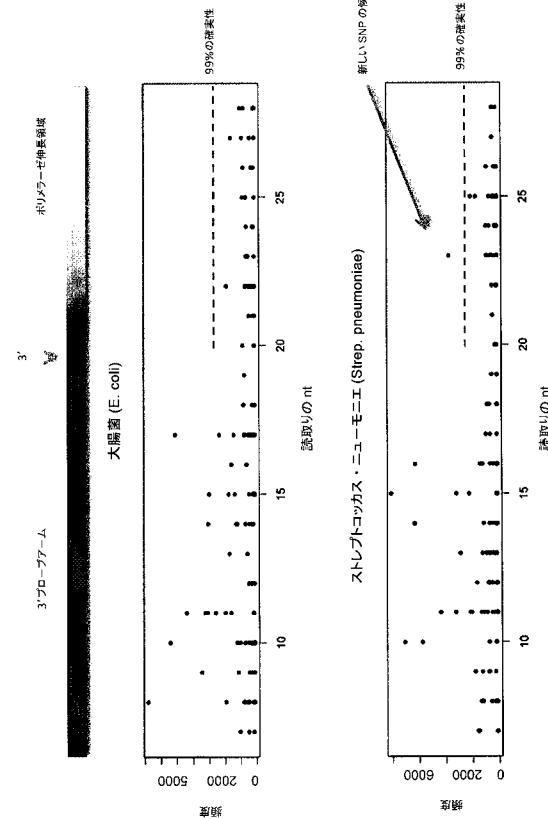


図 15

【図 16】



【図 17】

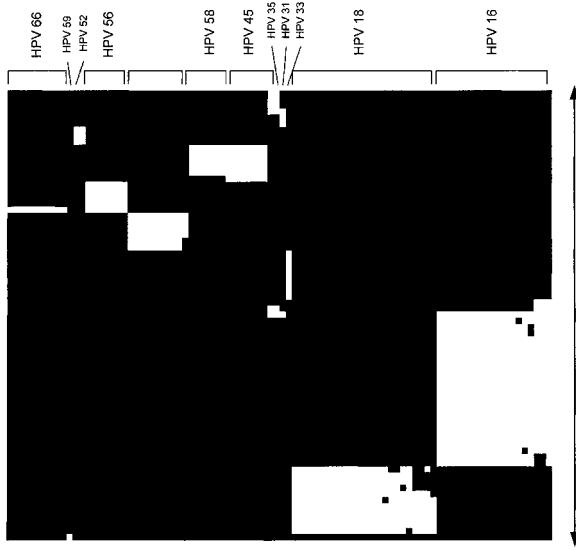


図 17

【図 18】

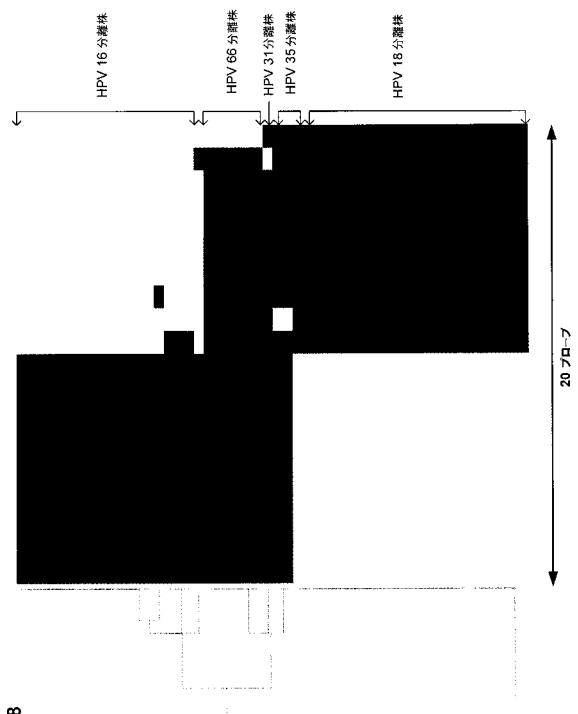


図 18

【図 19】

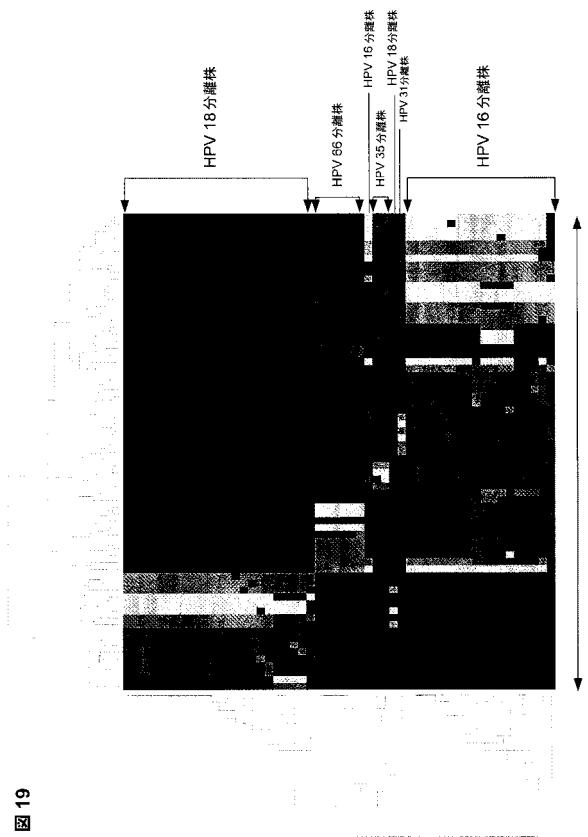


図 19

【図 20】

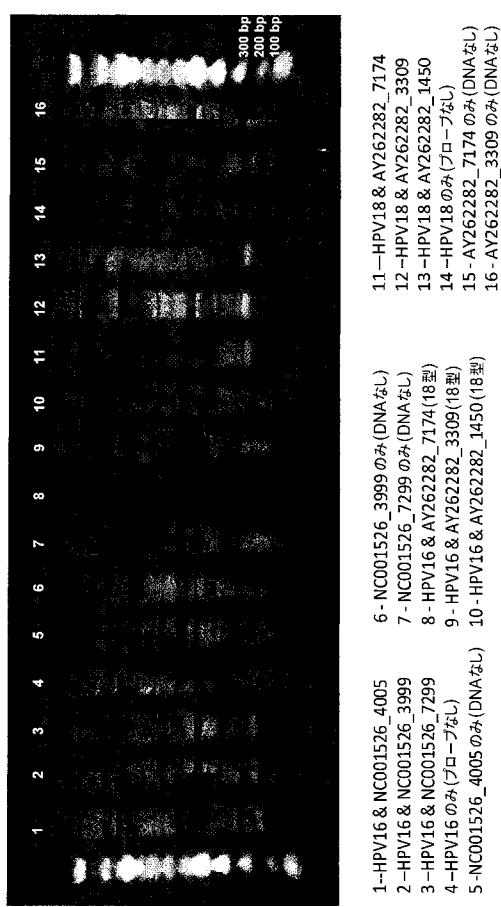


図 20

【図 21】

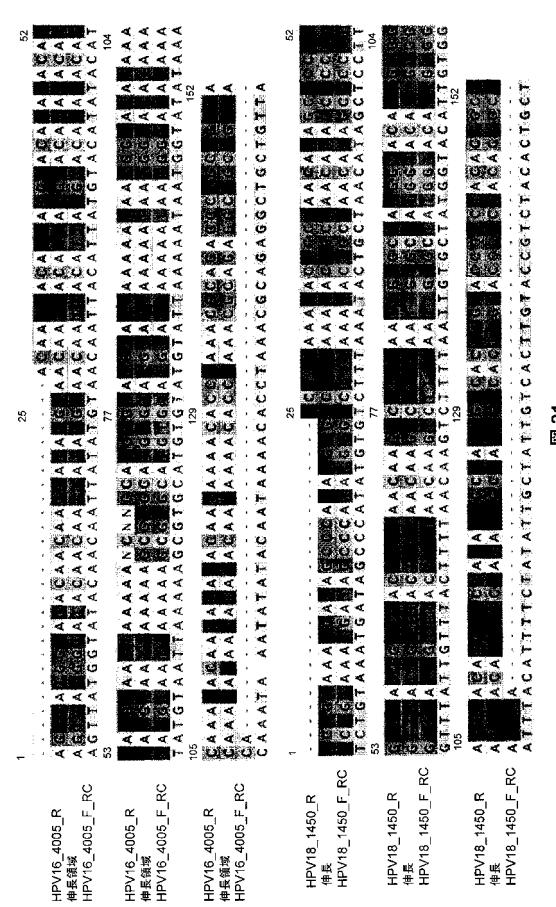
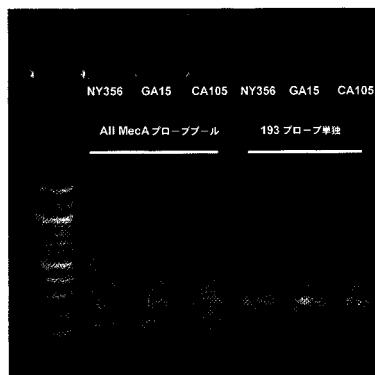


図 21

【 図 2 2 】

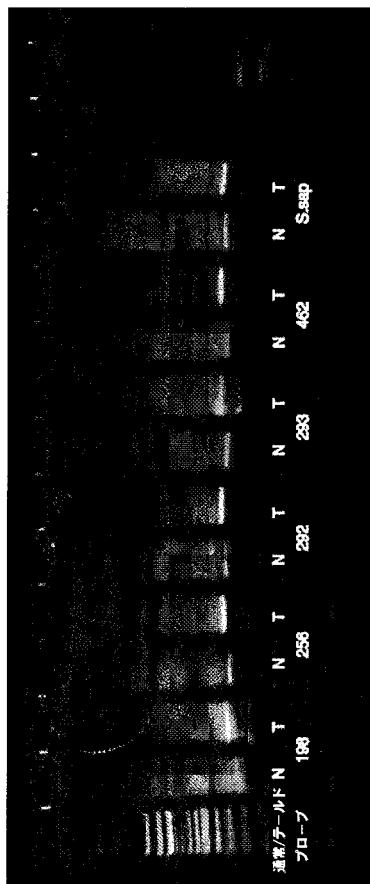


A



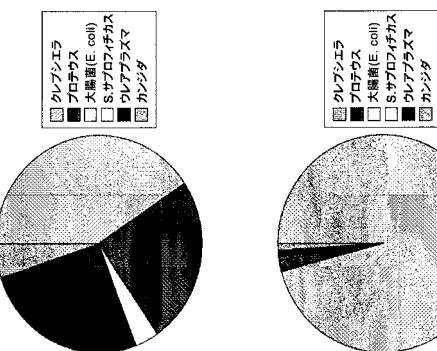
22

【 図 2 4 】



24

【 図 25 】



14

Plasma Type	Percentage
クリオジン	約45%
プロトウス	約30%
S.サルコチニカス	約15%
大腸菌	約10%

Category	Percentage
クラッシュエラ	30%
ブローラス	20%
大腸菌(E. coli)	25%
S.サルコバクテラス	10%
ウツララズマ	5%
カンジダ	10%

114

74

38

23

卷之三

【図 26】

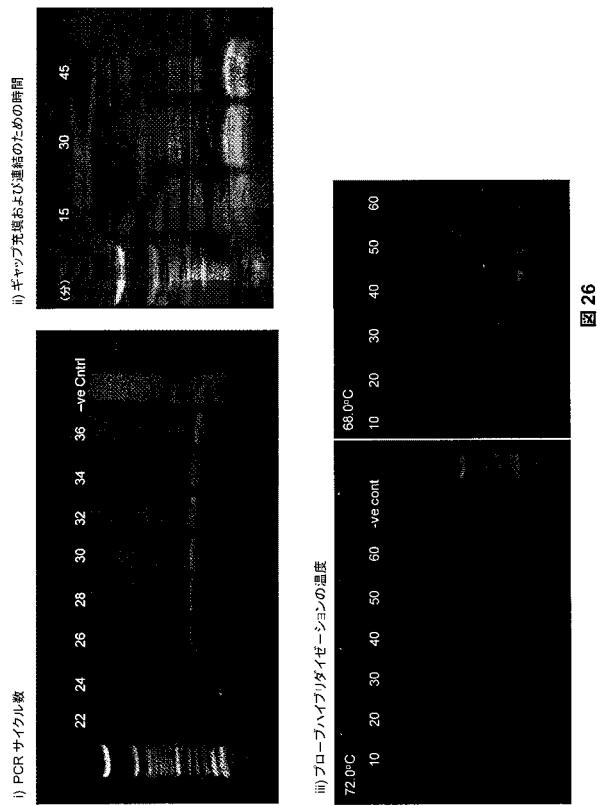


図 26

【国際調査報告】

INTERNATIONAL SEARCH REPORT		International application No. PCT/US2011/040106
A. CLASSIFICATION OF SUBJECT MATTER		
C12Q 1/68(2006.01)i, C12N 15/11(2006.01)i, G01N 33/53(2006.01)i		
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)		
C12Q 1/68		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Korean utility models and applications for utility models		
Japanese utility models and applications for utility models		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)		
PubMed		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2007/0264642 A1 (WILLIS, T. D. et al.) 15 November 2007 See claims 48-69, 80; figures 1-5.	1-36,38,65-71 ,75-77
A	LI, J. B. et al. Multiplex padlock targeted sequencing reveals human hypermutable CpG variations. <i>Genome Research</i> . September 2009, Vol. 19, No. 9, pages 1606-15. See abstract; figures 1, 2.	1-36,38,65-71 ,75-77
A	US 2007/0292861 A1 (THOMPSON, A.) 20 December 2007 See claims 1-10; figures 1-5; paragraphs [0018]-[0048].	1-36,38,65-71 ,75-77
A	DENG, J. et al. Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. <i>Nature Biotechnology</i> . April 2009, Vol. 27, No. 4, pages 353-360. See abstract; figures 1, 2.	1-36,38,65-71 ,75-77
A	US 2008/0026393 A1 (MINDRINOS, M. et al.) 31 January 2008 See claims 1-4, 8, 9, 13-18; figures 1-4; abstract.	1-36,38,65-71 ,75-77
<input type="checkbox"/> Further documents are listed in the continuation of Box C.		<input checked="" type="checkbox"/> See patent family annex.
<p>* Special categories of cited documents:</p> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier application or patent but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> <p>"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</p> <p>"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</p> <p>"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</p> <p>"&" document member of the same patent family</p>		
Date of the actual completion of the international search	Date of mailing of the international search report	
15 FEBRUARY 2012 (15.02.2012)	17 FEBRUARY 2012 (17.02.2012)	
Name and mailing address of the ISA/KR  Korean Intellectual Property Office Government Complex-Daejeon, 189 Cheongsa-ro, Seo-gu, Daejeon 302-701, Republic of Korea Facsimile No. 82-42-472-7140	Authorized officer Lee Hyun Ji Telephone No. 82-42-481-8745	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2011/040106

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.: 40,42-46,48,49,51-57,59,61,62,74,79-81 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:
Claims 40, 42-46, 48, 49, 51-57, 59, 61, 62, 74, 79-81 referring to claim 39, 41, 47, 49, 50, 58, 60, 64, or 73 are unclear because claims 39, 41, 47, 50, 58, 60, 64, and 73 do not comply with PCT Rule 6.4(a).
3. Claims Nos.: 37,39,41,47,50,58,60,63,64,72,73,78,82-84 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 an additional fee, this Authority did not invite payment of any additional fee.
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.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No.

PCT/US2011/040106

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2007-0264642 A1	15.11.2007	AU 2002-246612 B2 CA 2426824 A1 EP 1366192 A2 EP 1366192 B1 EP 1366192 B8 JP 04-287652 B2 JP 2004-528016 A US 2004-0101835 A1 US 2005-0026180 A1 US 2006-0252060 A1 US 2006-0275789 A1 US 2007-0178479 A1 US 2010-0330619 A1 US 6858412 B2 US 7700323 B2 US 7993880 B2 WO 02-057491 A2 WO 02-057491 A3	01.11.2007 25.07.2002 03.12.2003 12.12.2007 29.10.2008 03.04.2009 16.09.2004 27.05.2004 03.02.2005 09.11.2006 07.12.2006 02.08.2007 30.12.2010 22.02.2005 20.04.2010 09.08.2011 25.07.2002 25.07.2002
US 2007-0292861 A1	20.12.2007	EP 1756307 A1 US 2009-0156424 A1 US 2011-0118137 A1 US 7575867 B2 US 7618780 B2 WO 2005-113804 A1	28.02.2007 18.06.2009 19.05.2011 18.08.2009 17.11.2009 01.12.2005
US 2008-0026393 A1	31.01.2008	US 2011-086393 A1 US 7897747 B2	14.04.2011 01.03.2011

フロントページの続き

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

(特許庁注：以下のものは登録商標)

1. PENTIUM

2. Linux

(72)発明者 クム, ヨッヒエン

アメリカ合衆国 カリフォルニア州 94061, レッドウッドシティ, パークストリート 2
24

(72)発明者 ロルフ, フィリップ, アレクサンダー

アメリカ合衆国 マサチューセッツ州 02458, ニュートン, プレイステッドロード 34,
#2

F ターム(参考) 4B024 AA11 CA01 HA08 HA12

4B063 QA17 QA18 QA19 QQ02 QQ08 QQ42 QR08 QR32 QR42 QR50
QR55 QR62 QR72 QR77 QR82 QS03 QS25 QS28 QS34 QS36
QS39 QX02