

(19) 日本国特許庁(JP)

(12) 特許公報(B2)

(11) 特許番号

特許第4579577号
(P4579577)

(45) 発行日 平成22年11月10日(2010.11.10)

(24) 登録日 平成22年9月3日(2010.9.3)

(51) Int. Cl.		F I	
C 1 2 M	1/00	(2006.01)	C 1 2 M 1/00 Z N A A
C 1 2 Q	1/68	(2006.01)	C 1 2 Q 1/68 A
G O 1 N	33/53	(2006.01)	G O 1 N 33/53 M
G O 1 N	37/00	(2006.01)	G O 1 N 37/00 1 O 2
C 1 2 N	15/09	(2006.01)	C 1 2 N 15/00 A

請求項の数 6 (全 31 頁)

(21) 出願番号 特願2004-149387 (P2004-149387)
 (22) 出願日 平成16年5月19日(2004.5.19)
 (65) 公開番号 特開2005-3676 (P2005-3676A)
 (43) 公開日 平成17年1月6日(2005.1.6)
 審査請求日 平成19年5月21日(2007.5.21)
 (31) 優先権主張番号 特願2003-140793 (P2003-140793)
 (32) 優先日 平成15年5月19日(2003.5.19)
 (33) 優先権主張国 日本国(JP)

(73) 特許権者 000001007
 キヤノン株式会社
 東京都大田区下丸子3丁目30番2号
 (74) 代理人 100076428
 弁理士 大塚 康德
 (74) 代理人 100112508
 弁理士 高柳 司郎
 (74) 代理人 100115071
 弁理士 大塚 康弘
 (74) 代理人 100116894
 弁理士 木村 秀二
 (74) 代理人 100130409
 弁理士 下山 治
 (74) 代理人 100134175
 弁理士 永川 行光

最終頁に続く

(54) 【発明の名称】 情報処理装置および情報処理方法ならびに記憶媒体、プログラム

(57) 【特許請求の範囲】

【請求項1】

生物種の核酸配列の一部と相補的な核酸であるプローブを配置したDNAマイクロアレイを用いて、所定の検体をハイブリダイゼーション反応させた結果得られたDNAマイクロアレイ上の各プローブのシグナル強度に関する情報を処理する情報処理装置であって、

前記DNAマイクロアレイは、各生物種に特異的な複数種類のプローブがそれぞれ独立したスポットとして配置されており、

前記情報処理装置は、

既知の生物種が含まれる複数の検体をハイブリダイゼーション反応させた結果得られた、各検体ごとの、前記各プローブに対するシグナル強度を成分とするベクトルデータである第1の情報を保持する保持手段と、

前記所定の検体をハイブリダイゼーション反応させた結果得られた各プローブに対するシグナル強度を成分とするベクトルデータである第2の情報を取得する取得手段と、

前記保持手段に保持された前記第1の情報から、所定の生物種が含まれる検体をハイブリダイゼーション反応させた結果得られたベクトルデータ群と、該所定の生物種が含まれない検体をハイブリダイゼーション反応させた結果得られたベクトルデータ群とを選択し、該選択したベクトルデータ群の共分散行列を算出することにより生成されるベクトルフィルタであって、該所定の生物種が含まれているか否かを判定するのに寄与するプローブにおけるシグナル強度を抽出するためのベクトルフィルタと、

前記第1の情報のベクトルデータを前記ベクトルフィルタを用いてフィルタリングする

ことにより得られたベクトルデータと、前記第2の情報のベクトルデータを前記ベクトルフィルタを用いてフィルタリングすることにより得られたベクトルデータとを比較することにより、前記所定の検体に、前記所定の生物種が含まれているか否かを判定する判定手段と

を備えることを特徴とする情報処理装置。

【請求項2】

前記判定手段は分類木であり、該分類木は、2分岐分類木アンサンブルアルゴリズムを用いて作成されており、かつ、該分類木の各ノードにおける判定関数は、該各ノードに存在する学習データの中からランダムに選択されたカテゴリーの異なる2つの学習パターンのどちらに近いかを示す関数で定義されていることを特徴とする請求項1に記載の情報処理装置。

10

【請求項3】

生物種の核酸配列の一部と相補的な核酸であるプローブを配置したDNAマイクロアレイを用いて、所定の検体をハイブリダイゼーション反応させた結果得られたDNAマイクロアレイ上の各プローブのシグナル強度に関する情報を処理する情報処理方法であって、
前記DNAマイクロアレイは、各生物種に特異的な複数種類のプローブがそれぞれ独立したスポットとして配置されており、

前記情報処理方法は、

既知の生物種が含まれる複数の検体をハイブリダイゼーション反応させた結果得られた、各検体ごとの、前記各プローブに対するシグナル強度を成分とするベクトルデータである第1の情報を保持手段に保持する保持工程と、

20

前記所定の検体をハイブリダイゼーション反応させた結果得られた各プローブに対するシグナル強度を成分とするベクトルデータである第2の情報を取得する取得工程と、

前記第1の情報のベクトルデータをベクトルフィルタを用いてフィルタリングすることにより得られたベクトルデータと、前記第2の情報のベクトルデータを前記ベクトルフィルタを用いてフィルタリングすることにより得られたベクトルデータとを比較することにより、前記所定の検体に、前記所定の生物種が含まれているか否かを判定する判定工程と、を備え、

前記ベクトルフィルタは、

前記保持手段に保持された前記第1の情報から、所定の生物種が含まれる検体をハイブリダイゼーション反応させた結果得られたベクトルデータ群と、該所定の生物種が含まれない検体をハイブリダイゼーション反応させた結果得られたベクトルデータ群とを選択し、該選択したベクトルデータ群の共分散行列を算出することにより生成され、該所定の生物種が含まれているか否かを判定するのに寄与するプローブにおけるシグナル強度を抽出するように構成されていることを特徴とする情報処理方法。

30

【請求項4】

前記判定工程は分類木により処理され、該分類木は、2分岐分類木アンサンブルアルゴリズムを用いて作成されており、かつ、該分類木の各ノードにおける判定関数は、該各ノードに存在する学習データの中からランダムに選択されたカテゴリーの異なる2つの学習パターンのどちらに近いかを示す関数で定義されていることを特徴とする請求項3に記載の情報処理方法。

40

【請求項5】

請求項3または4のいずれか1項に記載の情報処理方法をコンピュータに実行させるための制御プログラム。

【請求項6】

請求項3または4のいずれか1項に記載の情報処理方法をコンピュータに実行させるための制御プログラムを格納した記録媒体。

【発明の詳細な説明】

【技術分野】

【0001】

50

本発明はいわゆるDNAマイクロアレイを用いた核酸配列解析に関するものであり、特に、菌などの微生物の種類を判定する技術に関するものである。

【背景技術】

【0002】

従来より、感染症を発症した患者の原因菌を判定する技術として“培養法”と呼ばれる技術がある。この方法は、患者から採取した血液に含まれる菌を特定の培地中で培養し、増殖した菌を観察することによって、感染症の原因菌を特定する。

【0003】

この技術の欠点は、原因菌の判定に数日の時間を必要とし、患者への治療方針を決める前に原因菌を判定することが、ほぼ不可能であるという点にある。つまり、本来は抗生物質などの治療薬を患者に投与する前に原因菌を特定しておくことが望ましいが、数日後の判定結果を待っている間は、患者の病状が悪化し、手遅れになってしまう可能性が高い。このため、原因菌の判定を行う前に、複数の可能性に対応した薬剤の投与をしなければならず、結果的に患者に対して薬剤の効果と引き換えに副作用のリスクを負わせることとなっていた。

【0004】

このような問題を解決するための1つの方法として、原因菌のDNA解析によって感染症の原因菌を判定する方法がある。この方法は、原因菌内ゲノムのある特定の部分を、例えばPCR(Polymerase Chain Reaction)法やLAMP法などの生化学的方法を用いて増幅し、その増幅した核酸配列を読み取ることによって原因菌の判定を行うものであり、かかる方法によれば、短時間での判定が可能であり、しかも原因菌の種類によることなく複数の原因菌について判定ができる。

【0005】

また、上記問題を解決する他の方法として、例えば、米国特許第6040138号明細書には、増幅された対象生物由来の核酸断片を、いわゆるプローブと呼ばれる核酸断片とハイブリダイゼーション反応させることにより、対象生物由来の核酸断片の量を定量する方法が開示されている。

【0006】

同明細書(米国特許第6040138号明細書)によれば、高密度に集積されたDNAマイクロアレイを用いることにより、1つの対象核酸断片に対して、多数のプローブを設定することが可能となり、その結果、対象核酸断片の配列を読み取るのと同じぐらいの情報を短時間で得ることができる。

【特許文献1】米国特許第6040138号公報

【発明の開示】

【発明が解決しようとする課題】

【0007】

しかしながら、上記従来技術に示した方法のうち、原因菌のDNA解析によって感染症の原因菌を判定する方法は、核酸配列の読み取りに非常に高い技術スキルとコストがかかるという問題がある。PCR法やLAMP法などのDNA増幅手法自体は簡単な手順で実施できるが、核酸配列の読み取りを可能にするためには、増幅された核酸断片について高い純度が要求される。また、シーケンサーと呼ばれる“核酸配列の読み取り装置”は高価であり、読み取りのための操作はPCR法やLAMP法によるDNA増幅に比べて遥かに複雑で、通常の検査技師のスキルでは実施が困難な場合が多い。

【0008】

一方、ハイブリダイゼーション反応による方法(米国特許第6040138号明細書に記載の方法)は、シーケンサーに比べて必要なスキルもコストも低く押さえられるものの、ハイブリダイゼーション反応の結果に基づいて生物種を判定するためのプロセスが必要となってくる。ここで生物種を判定するにあたっては、いわゆる“相同性検索”という方法が一般に用いられるが、かかる方法は塩基配列の似た原因菌を区別して判定することが困難であるという問題がある。これは、“相同性検索”がハイブリダイゼーション反応

10

20

30

40

50

の結果に基づいて各原因菌の存在確率を導き出すことにより結果的に生物種を特定する手法だからである。

【 0 0 0 9 】

一般に、相同性検索において塩基配列の似た原因菌を区別して判定するためには、DNAマイクロアレイでのハイブリダイゼーション反応において、「異なる種類の核酸断片に対して設定した異なるプローブは、それぞれ独立である」ということが大前提となってくる。例えば、遺伝子Aに対応するプローブが10個あった場合、そのプローブに遺伝子B由来の核酸断片は反応しないことが前提である。そして、かかる前提があってはじめて、例えば、前記10個のプローブのハイブリダイゼーション反応の結果得られたシグナル強度（例えば、蛍光強度）を平均することで得られた遺伝子Aの推定量に基づいて原因菌の有無（存在確率）が正しく判定されることとなる。

10

【 0 0 1 0 】

しかしながら、異なる生物種に由来する核酸断片であっても、それらが非常に似ている場合、プローブ同士が独立であるという上記前提は現実的ではなく、例えば、上記の遺伝子Aに対して設計したプローブであっても遺伝子Bがハイブリダイゼーション反応してしまうことがある（かかる現象のことを“クロスハイブリダイゼーション”という）。このため、「ある対象核酸断片に対応する複数のプローブの平均操作で得られた代表値を使って、その対象核酸断片の存在確率を導く」という上記手法は、類似した塩基配列を複数有する検体における原因菌の特定に際しては現実的ではなく、かかる手法により得られた判定精度は信頼性に欠けるという問題があった。

20

【 0 0 1 1 】

本発明は、上記課題を鑑みてなされたものであり、互いに類似する塩基配列を有する生物種が検体中に複数存在する場合において、簡易かつ安価に、しかも短時間で精度よく生物種を判定することを目的とする。

【課題を解決するための手段】

【 0 0 1 2 】

上記の目的を達成するために本発明に係る情報処理装置は以下のような構成を備える。即ち、

生物種の核酸配列の一部と相補的な核酸であるプローブを配置したDNAマイクロアレイを用いて、所定の検体をハイブリダイゼーション反応させた結果得られたDNAマイクロアレイ上の各プローブのシグナル強度に関する情報を処理する情報処理装置であって、

30

前記DNAマイクロアレイは、各生物種に特異的な複数種類のプローブがそれぞれ独立したスポットとして配置されており、

前記情報処理装置は、

既知の生物種が含まれる複数の検体をハイブリダイゼーション反応させた結果得られた、各検体ごとの、前記各プローブに対するシグナル強度を成分とするベクトルデータである第1の情報を保持する保持手段と、

前記所定の検体をハイブリダイゼーション反応させた結果得られた各プローブに対するシグナル強度を成分とするベクトルデータである第2の情報を取得する取得手段と、

前記保持手段に保持された前記第1の情報から、所定の生物種が含まれる検体をハイブリダイゼーション反応させた結果得られたベクトルデータ群と、該所定の生物種が含まれない検体をハイブリダイゼーション反応させた結果得られたベクトルデータ群とを選択し、該選択したベクトルデータ群の共分散行列を算出することにより生成されるベクトルフィルタであって、該所定の生物種が含まれているか否かを判定するのに寄与するプローブにおけるシグナル強度を抽出するためのベクトルフィルタと、

40

前記第1の情報のベクトルデータを前記ベクトルフィルタを用いてフィルタリングすることにより得られたベクトルデータと、前記第2の情報のベクトルデータを前記ベクトルフィルタを用いてフィルタリングすることにより得られたベクトルデータとを比較することにより、前記所定の検体に、前記所定の生物種が含まれているか否かを判定する判定手段とを備える。

50

【発明の効果】

【0013】

本発明によれば、互いに類似する塩基配列を有する生物種が検体中に複数存在する場合において、簡易かつ安価に、しかも短時間で精度よく生物種を判定することが可能となる。

【発明を実施するための最良の形態】

【0014】

図1は、本発明にかかる情報処理方法を含む検査処理全体の流れを示した図である。同図に示すように、検査にあたっては、はじめにDNAマイクロアレイを用いて既知の検体のハイブリダイゼーション反応実験を行い(ステップS101)、その結果得られたDNAマイクロアレイの蛍光強度に代表されるシグナル強度に関する情報(スキャン画像)を基準サンプルの反応結果として格納しておき(ステップS102、S104)、次に未知のサンプルについて反応実験の結果得られたDNAマイクロアレイの蛍光強度に関する情報(スキャン画像)について、当該基準サンプルの反応結果に基づいて、生物種の判定処理を行う(ステップS102、S103)。以下に、ステップS101及びステップS103の処理について、順を追って詳細に説明する。

【0015】

1. ハイブリダイゼーション反応実験(ステップS101)に関する説明

[1-1. ハイブリダイゼーション反応実験の流れ]

はじめに図4を用いてDNAマイクロアレイを用いたハイブリダイゼーション反応実験の実験手順全般について説明する。

【0016】

401の“サンプル”とは対象としている核酸が含まれているはずの液体や個体などの検体である。例えば、感染症の原因菌を特定する場合、ヒト、家畜等の動物由来の血液、喀痰、胃液、膣分泌物、口腔内粘液等の体液、尿及び糞便のような排出物等細菌が存在すると思われるあらゆる物がサンプルとなる。また、食中毒、汚染の対象となる食品、飲料水及び温泉水のような環境中の水等、細菌による汚染が引き起こされる可能性のある媒体がサンプルとして用いられることもある。さらに、輸出入時における検疫等の動植物もその対象となる。

【0017】

サンプル401は、402に示す“生化学的増幅”方法を用いて増幅される。例えば、感染症の原因菌を特定する場合には、16S rRNA検出用に設計されたPCR反応用プライマーを用いてPCR法によって対象核酸を増幅したり、或いはPCR増幅物を元にさらにPCR反応等を行って調整したりする。なお、PCR以外のLAMP法などの増幅方法により調整してもよい。

【0018】

増幅されたサンプル、または元のサンプル401は、シグナル強度検出のために各種標識法により標識される(ラベル混入403)。本発明においてシグナル強度とは、適当な手段により適宜検出・測定可能な信号強度であって、蛍光、放射能、化学発光等が含まれるが、蛍光強度であるのが好ましい。このための標識物質としては、通常Cy3、Cy5、Rodaminなどの蛍光物質を用いるのが好ましい。また、生化学的増幅処理(402)において、標識分子が混入されることもある。

【0019】

そして、標識分子が付加された核酸を用いて、DNAマイクロアレイ404とハイブリダイゼーション反応(405)を行う(詳細は後述)。例えば感染症の原因菌の特定をする場合、DNAマイクロアレイ404としては、菌に特異的なプローブを基板に固定したものをを用いる。各菌のプローブの設計は、例えば16S rRNAをコーディングしているゲノム部分より、当該菌に対し非常に特異性が高く、かつそれぞれのプローブ塩基配列で“出来るだけ”ばらつきのないハイブリダイゼーション感度が期待できるように行われる。DNAマイクロアレイ404のプローブを固定する担体(基板)は、ガラス基板、プ

10

20

30

40

50

ラスチック基板、シリコンウェハー等の平面基板が考えられる。また、凹凸のある三次元構造体、ビーズのような球状のもの、棒状、紐状、糸状のもの等を用いてもよい。

【0020】

通常、前記基板は、プローブDNAが固定化できるように表面処理したものを使用する。特に、化学反応が可能となるように表面に官能基を導入したものは、ハイブリダイゼーション反応の過程でプローブを安定に結合させるため、再現性の点で好ましい形態といえる。

【0021】

また、固定化にあたっては、例えば、マレイミド基とチオール(-SH)基との組み合わせを用いる例が挙げられる。これは、核酸プローブの末端にチオール(-SH)基を結合させておき、固相表面がマレイミド基を有するように処理しておくことで、固相表面に供給された核酸プローブのチオール基と固相表面のマレイミド基とが反応して核酸プローブを固定化させるものである。

【0022】

マレイミド基の導入にあたっては、まず、ガラス基板にアミノシランカップリング剤を反応させる。そして、そのアミノ基とEMCS試薬(N-(6-Maleimidocaproyloxy)succinimide : Dojin社製)との反応によりマレイミド基を導入する。DNAへのSH基の導入は、DNA自動合成機上5'-Thiol-ModifierC6(Glen Research社製)を用いることにより行うことができる。

【0023】

なお、固定化に利用する官能基の組み合わせとしては、上記したチオール基とマレイミド基の組み合わせ以外にも、例えばエポキシ基(固相上)とアミノ基(核酸プローブ末端)の組み合わせ等が挙げられる。また、各種シランカップリング剤による表面処理も有効であり、該シランカップリング剤により導入された官能基と反応可能な官能基を導入したオリゴヌクレオチドを用いてもよい。さらに、官能基を有する樹脂をコーティングする方法も有効である。

【0024】

ハイブリダイゼーション反応したDNAマイクロアレイ404の表面は洗浄され、プローブと結合していない核酸が剥がされた後、(通常は)乾燥される。その後、DNAマイクロアレイの基板に励起光を照射することで蛍光量の測定が行われる(406)。また、励起光を照射した状態でスキャンすることで蛍光強度に比例したスキャン画像(407)が得られる。

【0025】

[1-2. ハイブリダイゼーション反応の概要]

次に、上述したハイブリダイゼーション反応(405)の概要について図3を用いて説明する。図3はDNAマイクロアレイ上でのハイブリダイゼーション反応の様子を示した図である。生体内ではほとんどの場合、DNAは2重らせん構造をしており、その2本鎖の間の結合は塩基間の水素結合で実現されている。一方、RNAは1本で存在する場合が多い。塩基の種類はDNAの場合はACGTの4種類、RNAの場合はACGUの4種類であり、それぞれ水素結合ができる塩基対はA-T(U)、G-Cのペアとなっている。

【0026】

一般にハイブリダイゼーション反応とは、1本鎖状態の核酸分子同士がその中にある部分塩基配列を介して部分的に結合する状態をいう。なお、本実施形態では、図3の上側の基板にくっついた核酸分子(プローブ301)の方が下側のサンプル中にある核酸分子(302)より短い場合を想定している。よって、サンプル中に存在する核酸分子がプローブの塩基配列を含む場合には、このハイブリダイゼーション反応はうまくいき、サンプル中のターゲット核酸分子はDNAマイクロアレイにトラップされることとなる。

【0027】

[1-3. DNAマイクロアレイの原理]

次に、図5により感染症の菌を特定するために用いられるDNAマイクロアレイの原理

10

20

30

40

50

を説明する。図5で示したDNAマイクロアレイ(500-1、500-2)は、黄色ブドウ球菌を特定する目的で作られたDNAマイクロアレイの一例である。

【0028】

同図左の列は、当該DNAマイクロアレイを用いた場合の黄色ブドウ球菌野生株由来の処理系列であり、右の列は大腸菌野生株由来の処理系列である。例えば、左は黄色ブドウ球菌に感染した患者の血液を処理する流れで、右は大腸菌に感染した患者の血液を処理する流れだと考えてよい。

【0029】

どちらも基本的には同じ処理を行う。つまり、まず初めに例えば菌感染患者の血液や、痰などからDNAを抽出する(501-1、501-2)。この際に、一般的には、患者の体細胞由来の人間のDNAも含まれている可能性がある。

10

【0030】

なお、抽出されたDNAが少ない場合には、PCR法などの方法で増幅を行う。この際に蛍光物質もしくは蛍光物質を結合させることができる物質を標識として混入させるのが一般的である(502-1、502-2)。

【0031】

増幅をしない場合は、抽出されたDNAを用いて、相補鎖を作りながら蛍光物質もしくは蛍光物質を結合させることができる物質を標識として混入させる(503-1、503-2)。あるいは、そのまま直接抽出されたDNAに蛍光物質もしくは蛍光物質を結合させることができる物質を標識として付加させる。

20

【0032】

通常、PCR増幅を行う場合、感染症の菌特定目的であれば、いわゆる16S rRNAといわれるリボゾームRNAを構成する塩基配列の部分を増幅するのが一般的である。この場合、左の黄色ブドウ球菌のPCRプライマーと右の大腸菌のPCRプライマーとはほとんど同じものを使うこととなる。より具体的には、どのような菌の16S rRNAをコーディングしている部分でも増幅させることができるプライマーセットを用いて、マルチプレックスPCRを行う。この場合、結果的に、図5における右と左のどちらのハイブリ溶液も複数の種類の塩基配列を含むものとなる。この理由は次の図で詳しく述べる。

【0033】

これに対して、より詳しい配列解析を行いたい場合には、例えば、黄色ブドウ球菌用のPCRプライマーセット、大腸菌用のPCRプライマーセットを別々に設定する。この場合、仮に菌のゲノムの特定部分だけを選択的に増幅するようにプライマーを設定すれば、ハイブリ溶液に含まれる塩基配列の種類は非常に限定される。それでも通常、自然界に存在する菌の株は数種類に及ぶので、ハイブリ溶液に存在する塩基配列の種類は1種類となることは稀である。

30

【0034】

黄色ブドウ球菌を判定する目的のために設計されたDNAマイクロアレイが正しく動作するならば、左のハイブリ溶液では、スポットがポジティブに反応し(500-1)、右のハイブリ溶液では、スポットがネガティブに反応する(500-2)。

【0035】

40

これと全く同じように、大腸菌の存在を判定する目的のために設計されたDNAマイクロアレイが正しく動作するならば、左のハイブリ溶液では、スポットがネガティブに反応し、右のハイブリ溶液では、スポットがポジティブに反応する。本実施形態では、いろいろな菌に対してそれぞれ特異的に反応する数種類のスポットを同時に並べたDNAマイクロアレイを用いて、感染菌の判定を行う。

【0036】

[1-4. 複数の塩基配列が存在する理由]

次に、図6を用いて図5のハイブリ溶液に複数の種類の塩基配列が存在する理由を説明する。通常、自然界に存在する菌は、突然変異を頻繁に起こす。その結果、淘汰を経て生き残った主要な数種類の株が同時に存在することがある。例えば、院内感染などで問題を

50

起こす菌株は、通常は薬剤耐性がないはずの菌が、突然変異を起こすことによって薬剤耐性を獲得することで出現する。薬剤耐性を獲得した結果、殺菌努力を行っている衛生的な環境でも旺盛な繁殖力を持つ菌が出現したりする。このように、自然界に存在する同一の菌の塩基配列は、数種類のパリエーションをもつものである。

【0037】

図6に示したのは、黄色ブドウ球菌におけるMu50とMW2の2種類の株のゲノム構造である。それぞれの株のゲノムの総塩基数はそれぞれ2,878,040と2,820,462と異なっている。更に16sのリボゾームRNAをコーディングしている部位は、Mu50の場合は順方向2箇所、逆方向3箇所の合計5箇所であるのに対して、MW2では順方向3箇所、逆方向3箇所の合計6箇所存在する。

10

【0038】

これらの16sリボゾームRNAの各部位の塩基配列はそれぞれ非常に似ているが、全く同じではない。つまり、調べようとしている菌感染患者の体の中に存在する菌の株の種類がたとえ1種類であったとしても、図5のような一般的な処理でハイブリ溶液を調整すると、複数種類の似通った塩基配列がハイブリ溶液の中に存在することになる。そして、この複数の似通った核酸に対してDNAマイクロアレイを用いた実験を行った場合でも、安定して生物種の判定ができるようにするのが本願における生物種類判定方法の主たる目的である。

【0039】

[1-5.ハイブリダイゼーション反応実験の実施例]

20

次に、感染症の原因菌の特定を目的として実際に行ったハイブリダイゼーション反応実験の具体的な実施例について以下に詳説する。なお、本発明にかかる生物種類判定方法は、以下に述べる感染症の原因菌特定を目的としたものに限ったものではなく、MHCなどの人間の体質判定や、癌などの疾病に関わるDNA、RNAの解析などに用いてもよい。

【0040】

[1-5-1.プローブDNAの準備]

エンテロバクタークロアカエ(Enterobacter cloacae)菌検出用プローブとして表1に示す核酸配列(I-n)(nは数字)を設計した。

【0041】

具体的には、16s rRNAをコーディングしているゲノム部分より、以下に示したプローブ塩基配列を選んだ。これらのプローブ塩基配列群は、当該菌に対し非常に特異性が高く、十分かつそれぞれのプローブ塩基配列で“出来るだけ”ばらつきのないハイブリダイゼーション感度が期待できるように設計されている。

30

【0042】

【表1】

I-1	CAGAgAgCTTgCTCTCgggTgA
I-2	gggAggAAggTgTTgTggTTAATAAC
I-3	ggTgTTgTggTTAATAACCACAgCAA
I-4	gCggTCTgTCAAgTCggATgTg
I-5	ATTCgAAACTggCAggCTAgAgTCT
I-6	TAACCACAgCAATTgACgTTACCCg
I-7	gCAATTgACgTTACCCgCAgAAgA

40

【0043】

表中に示したプローブは、DNAマイクロアレイに固定するための官能基として合成後

50

、定法に従って核酸の5'末端にチオール基を導入した。官能基の導入後、精製し、凍結乾燥した。凍結乾燥した内部標準用プローブは、-30℃の冷凍庫に保存した。

黄色ブドウ球菌(A-n)、表皮ブドウ球菌(B-n)、大腸菌(C-n)、肺炎桿菌(D-n)、緑膿菌(E-n)、セラチア菌(F-n)、肺炎連鎖球菌(G-n)、インフルエンザ菌(H-n)、及びエンテロコッカス・フェカリス菌(J-n)(nは数字)についても同様な手法により以下に示すプローブセット(表2-1~2-9)を設計した。

【0044】

【表2-1】

A-1	gAACCGCATggTTCAAAAgTgAAAgA
A-2	CACTTATAgATggATCCgCgCTgC
A-3	TgCACATCTTgACggTACCTAATCAg
A-4	CCCCTTAgTgCTgCAgCTAACg
A-5	AATACAAAgggCAgCgAAACCGC
A-6	CCggTggAgTAACCTTTTAggAgCT
A-7	TAACCTTTTAggAgCTAgCCgTCgA
A-8	TTTAggAgCTAgCCgTCgAAggT
A-9	TAgCCgTCgAAggTgggACAAAT

10

20

【0045】

【表2-2】

B-1	gAACAgACgAggAgCTTgCTCC
B-2	TAgTgAAAgACggTTTTgCTgTCACT
B-3	TAAgTAACTATgCACgTCTTgACggT
B-4	gACCCCTCTAgAgATAgAgTTTTCCC
B-5	AgTAACCATTTggAgCTAgCCgTC
B-6	gAgCTTgCTCCTCTgACgTTAgC
B-7	AgCCggTggAgTAACCATTTgg

30

40

【0046】

【表 2 - 3】

C-1	CTCTTgCCATCggATgTgCCCA
C-2	ATACCTTTgCTCATTgACgTTACCCg
C-3	TTTgCTCATTgACgTTACCCgCAg
C-4	ACTggCAAgCTTgAgTCTCgTAgA
C-5	ATACAAAgAgAAgCgACCTCgCg
C-6	CggACCTCATAAAgTgCgTCgTAgT
C-7	gCggggAggAAgggAgTAAgTTAAT

10

【 0 0 4 7 】

【表 2 - 4】

D-1	TAgCACAgAgAgCTTgCTCTCgg
D-2	TCATgCCATCAGATgTgCCCAgA
D-3	CggggAggAAggCgATAAggTTAAT
D-4	TTCgATTgACgTTACCCgCAGAAgA
D-5	ggTCTgTCAAgTCggATgTgAAATCC
D-6	gCAggCTAgAgTCTTgTAgAgggg

20

【 0 0 4 8 】

【表 2 - 5】

E-1	TgAgggAgAAAgtgggggATCTTC
E-2	TCAGATgAgCCTAggTCggATTAgC
E-3	gAgCTAgAgTACggTAgAgggTgg
E-4	gTACggTAgAgggTggTggAATTC
E-5	gACCACCTggACTgATACTgACAC
E-6	TggCCTTgACATgCTgAgAACTTTC
E-7	TTAgTTACCAgCACCTCgggTgg
E-8	TAgTCTAACCgCAAgggggACg

40

【 0 0 4 9 】

【表 2 - 6】

F-1	TAgCACAgggAgCTTgCTCCCT
F-2	AggTggTgAgCTTAATACgCTCATC
F-3	TCATCAATTgACgTACTCgCAgAAg
F-4	ACTgCATTgAAACTggCAAgCTAgA
F-5	TTATCCTTTgTTgCAgCTTCggCC
F-6	ACTTTCAgCgAggAggAAggTgg

10

【 0 0 5 0 】

【表 2 - 7】

G-1	AgTAgAACgCTgAAggAggAgCTTg
G-2	CTTgCATCACTACCAgATggACCTg
G-3	TgAgAgTggAAAgtTCACACTgTgAC
G-4	gCTgTggCTTAACCATAgTAggCTTT
G-5	AAgCggCTCTCTggCTTgTAACT
G-6	TAgACCCTTTCCggggTTTAgTgC
G-7	gACggCAAgCTAATCTCTTAAgCCA

20

【 0 0 5 1 】

【表 2 - 8】

H-1	gCTTgggAATCTggCTTATggAgg
H-2	TgCCATAggATgAgCCCAAgTgg
H-3	CTTgggAATgTACTgACgCTCATgTg
H-4	ggATTgggCTTAgAgCTTggTgC
H-5	TACAgAgggAAgCgAAgCTgCg
H-6	ggCgTTTACCACggTATgATTCATgA
H-7	AATgCCTACCAAgCCTgCgATCT
H-8	TATCggAAgATgAAAgtgCgggACT

40

【 0 0 5 2 】

【表 2 - 9】

J-1	TTCTTTCTCCCgAgTgCTTgCA
J-2	AACACgTgggTAACCTACCCATCAg
J-3	ATggCATAAgAgTgAAAaggCgCTT
J-4	gACCCgCggTgCATTAgCTAgT
J-5	ggACgTTAgTAACTgAACgTCCCCT
J-6	CTCAACCggggAgggTCATTgg
J-7	TTggAgggTTTCCgCCCTTCAg

10

【 0 0 5 3 】

[1-5-2. 検体増幅用 PCR Primer の準備]

原因菌検出の為に 16S rRNA 核酸 (標的核酸) 増幅用 PCR Primer として表 2 に示す核酸配列を設計した。

【 0 0 5 4 】

具体的には、16S rRNA をコーディングしているゲノム部分を特異的に増幅するプローブセット、つまり約 1500 塩基長の 16S rRNA コーディング領域の両端部分で、特異的な融解温度をできるだけ揃えたプライマーを設計した。なお、変異株や、ゲノム上に複数存在する 16S rRNA コーディング領域も同時に増幅できるように複数種類のプライマーを設計した。

20

【 0 0 5 5 】

【表 3】

	Primer No.	配列
Forward Primer	F-1	5' GCGGCGTGCCTAATACATGCAAG 3'
	F-2	5' GCGGCAGGCCTAACACATGCAAG 3'
	F-3	5' GCGGCAGGCTTAACACATGCAAG 3'
Reverse Primer	R-1	5' ATCCAGCCGCACCTTCCGATAC 3'
	R-2	5' ATCCAACCGCAGGTTCCCCTAC 3'
	R-3	5' ATCCAGCCGCAGGTTCCCCTAC 3'

30

【 0 0 5 6 】

表中に示したプライマーは、合成後、高速液体クロマトグラフィー (HPLC) により精製し、Forward Primer : 3 種、Reverse Primer : 3 種を混合し、それぞれのプライマー濃度が、最終濃度 10 pmol / μl となるように TE 緩衝液に溶解した。

[1-5-3. Enterobacter_cloacae Genome DNA (モデル検体) の抽出]

[1-5-3-1] 微生物の培養 & Genome DNA 抽出の前処理

まず、エンテロバクター クロアカエ (Enterobacter cloacae) 標準株を、定法に従って培養した。この微生物培養液を 1.5 ml 容量のマイクロチューブに 1.0 ml (OD600=0.7) 採取し、遠心分離で菌体を回収した (8500rpm、5min、4)。

40

50

【 0 0 5 7 】

上精を捨てた後、Enzyme Buffer (50mM Tris-HCl : p.H. 8.0、25mM EDTA) 3 0 0 μ l を加え、ミキサーを用いて再懸濁した。再懸濁した菌液は、再度、遠心分離で菌体を回収した (8500rpm、5min、4)。

【 0 0 5 8 】

上精を捨てた後、回収された菌体に、以下の酵素溶液を加え、ミキサーを用いて再懸濁した。

Lysozyme 50 μ l (20 mg/ml in Enzyme Buffer)

N-Acetylmuramidase SG 50 μ l (0.2 mg/ml in Enzyme Buffer)

次に、酵素溶液を加え再懸濁した菌液を、3 7 のインキュベーター内で3 0 分間静置し、細胞壁の溶解処理を行った。

【 0 0 5 9 】

[1-5-3-2] Genome DNA 抽出

以下に示す微生物の Genome DNA 抽出は、核酸精製キット (MagExtractor -Genome- : TOYOB0社製) を用いて行った。

【 0 0 6 0 】

具体的には、まず、前処理した微生物懸濁液に溶解・吸着液 7 5 0 μ l と磁性ビーズ 4 0 μ l を加え、チューブミキサーを用いて、1 0 分間激しく攪拌した (ステップ 1)。

【 0 0 6 1 】

次に、分離用スタンド (Magical Trapper) にマイクロチューブをセットし、3 0 秒間静置して磁性粒子をチューブの壁面に集め、スタンドにセットした状態のまま、上精を捨てた (ステップ 2)。

【 0 0 6 2 】

次に、洗浄液 9 0 0 μ l を加え、ミキサーで 5 s e c 程度攪拌して再懸濁を行った (ステップ 3)。

【 0 0 6 3 】

次に、分離用スタンド (Magical Trapper) にマイクロチューブをセットし、3 0 秒間静置して磁性粒子をチューブの壁面に集め、スタンドにセットした状態のまま、上精を捨てた (ステップ 4)。

【 0 0 6 4 】

ステップ 3、4 を繰り返して 2 度目の洗浄 (ステップ 5) を行った後、7 0 % エタノール 9 0 0 μ l を加え、ミキサーで 5 s e c 程度攪拌して再懸濁した (ステップ 6)。

【 0 0 6 5 】

次に、分離用スタンド (Magical Trapper) にマイクロチューブをセットし、3 0 秒間静置して磁性粒子をチューブの壁面に集め、スタンドにセットした状態のまま、上精を捨てた (ステップ 7)。

【 0 0 6 6 】

ステップ 6、7 を繰り返して 7 0 % エタノールによる 2 度目の洗浄 (ステップ 8) を行った後、回収された磁性粒子に純水 1 0 0 μ l を加え、チューブミキサーで 1 0 分間攪拌を行った。

【 0 0 6 7 】

次に分離用スタンド (Magical Trapper) にマイクロチューブをセットし、3 0 秒間静置して磁性粒子をチューブ壁面に集め、スタンドにセットした状態のまま、上精を新しいチューブに回収した。

【 0 0 6 8 】

[1-5-3-3] 回収した Genome DNA の検査

回収された微生物 (Enterobacter cloacae 株) の Genome DNA は、定法に従って、アガロース電気泳動と 2 6 0 / 2 8 0 n m の吸光度測定を行い、その品質 (低分子核酸の混入量、分解の程度) と回収量を検定した。

【 0 0 6 9 】

10

20

30

40

50

本実施例では、約10 μ gのGenome DNAが回収され、Genome DNAのデグラデーションやrRNAの混入は認められなかった。回収したGenome DNAは、最終濃度50ng/ μ lとなるようにTE緩衝液に溶解し、以下の実施例に使用した。

【0070】

[1-5-4. DNAマイクロアレイの作製]

[1-5-4-1] ガラス基板の洗浄

合成石英のガラス基板（サイズ：25mm \times 75mm \times 1mm、飯山特殊ガラス社製）を耐熱、耐アルカリのラックに入れ、所定の濃度に調製した超音波洗浄用の洗浄液に浸した。一晚洗浄液中で浸した後、20分間超音波洗浄を行った。続いて基板を取り出し、軽く純水ですすいだ後、超純水中で20分超音波洗浄をおこなった。次に80 $^{\circ}$ Cに加熱した1N水酸化ナトリウム水溶液中に10分間基板を浸した。再び純水洗浄と超純水洗浄を行い、DNAチップ用の石英ガラス基板を用意した。

【0071】

[1-5-4-2] 表面処理

シランカップリング剤KBM-603（信越シリコン社製）を、1%の濃度となるように純水中に溶解させ、2時間室温で攪拌した。続いて、先に洗浄したガラス基板をシランカップリング剤水溶液に浸し、20分間室温で放置した。ガラス基板を引き上げ、軽く純水で表面を洗浄した後、窒素ガスを基板の両面に吹き付けて乾燥させた。次に乾燥した基板を120 $^{\circ}$ Cに加熱したオーブン中で1時間バークし、カップリング剤処理を完結させ、基板表面にアミノ基を導入した。次いで同仁化学研究所社製のN-マレイミドカプロイロキシスクシイミド(N-(6-Maleimidocaproyloxy)succinimido)（以下EMCSと略す）を、ジメチルスルホキシドとエタノールの1：1混合溶媒中に最終濃度が0.3mg/mlとなるように溶解したEMCS溶液を用意した。バークの終了したガラス基板を放冷し、調製したEMCS溶液中に室温で2時間浸した。この処理により、シランカップリング剤によって表面に導入されたアミノ基とEMCSのスクシイミド基が反応し、ガラス基板表面にマレイミド基が導入された。EMCS溶液から引き上げたガラス基板を、先述のEMCSを溶解した混合溶媒を用いて洗浄し、さらにエタノールにより洗浄した後、窒素ガス雰囲気下で乾燥させた。

【0072】

[1-5-4-3] プローブDNA

本実施例で作製した微生物検出用プローブを純水に溶解し、それぞれ、最終濃度（インク溶解時）10 μ Mとなるように分注した後、凍結乾燥を行い、水分を除いた。

【0073】

[1-5-4-4] BJプリンタによるDNA吐出、および基板への結合

グリセリン7.5wt%、チオジグリコール7.5wt%、尿素7.5wt%、アセチレノールEH（川研ファインケミカル社製）1.0wt%を含む水溶液を用意した。続いて、先に用意した7種類のプローブ（表1）を上記の混合溶媒に規定濃度なるように溶解した。得られたDNA溶液をバブルジェット（登録商標）プリンタ（商品名：BJF-850 キヤノン社製）用インクタンクに充填し、印字ヘッドに装着した。

【0074】

なお、ここで用いたバブルジェットプリンタ（登録商標）は平板への印刷が可能なように改造を施したものである。またこのバブルジェットプリンタ（登録商標）は、所定のファイル作成方法に従って印字パターンを入力することにより、約5ピコリットルのDNA溶液を約120マイクロメートルピッチでスポッティングすることが可能となっている。

【0075】

続いて、この改造バブルジェットプリンタ（登録商標）を用いて、1枚のガラス基板に対して、印字操作を行い、アレイを作製した。印字が確実に行われていることを確認した後、30分間加湿チャンバー内に静置し、ガラス基板表面のマレイミド基と核酸プローブ末端のチオール基とを反応させた。

10

20

30

40

50

【 0 0 7 6 】

[1-5-4-5] 洗淨

30分間の反応後、100mMのNaClを含む10mMのリン酸緩衝液(pH7.0)により表面に残ったDNA溶液を洗い流し、ガラス基板表面に一本鎖DNAが固定したDNAマイクロアレイを得た。

【 0 0 7 7 】

[1-5-5. 検体の増幅と標識化 (PCR増幅 & 蛍光標識の取り込み)]

検体となる微生物DNAの増幅、および、標識化反応は以下に示すとおりである。

【 0 0 7 8 】

Premix PCR 試薬 (TAKARA ExTaq)	25 μ l	
Template Genome DNA	2 μ l (100ng)	
Forward Primer mix	2 μ l (20pmol/tube each)	
Reverse Primer mix	2 μ l (20pmol/tube each)	
Cy-3 dUTP (1mM)	2 μ l (2nmol/tube)	
H2O	17 μ l	
<hr/>		
Total	50 μ l	10
		20

【 0 0 7 9 】

上記組成の反応液を以下のプロトコールに従って、市販のサーマルサイクラーで増幅反応を行った。

【 0 0 8 0 】

95°C	10 min.		
92°C	45 sec.	← 35 Cycles →	
55°C	45 sec.		
72°C	45 sec.		
72°C	10 min.		30

【 0 0 8 1 】

反応終了後、精製用カラム (QIAGEN QIAquick PCR Purification Kit) を用いてプライマーを除去した後、増幅産物の定量を行い、標識化検体とした。

[1-5-6. ハイブリダイゼーション]

上記「1-5-4. DNAマイクロアレイの作製」で作製したDNAマイクロアレイと上記「1-5-5. 検体の増幅と標識化 (PCR増幅 & 蛍光標識の取り込み)」で作製した標識化検体を用いて検出反応を行った。

【 0 0 8 2 】

[1-5-6-1] DNAマイクロアレイのプロッキング

B S A (牛血清アルブミンFraction V: Sigma社製) を1wt%となるように100mM NaCl / 10mM Phosphate Bufferに溶解し、この溶液に「1-5-4. DNAマイクロアレイの作製」で作製したDNAマイクロアレイを室温で2時間浸し、プロッキングを行った。プロッキング終了後、0.1wt% SDS (ドデシル硫酸ナトリウム) を含む2xSSC溶液 (NaCl 300mM, Sodium Citrate (trisodium citrate dihydrate, C₆H₅Na₃ · 2H₂O) 30mM, p.H. 7.0) で洗淨を行った後、純水でリンスしてからスピンドライ装置で水切りを行った。

【 0 0 8 3 】

[1-5-6-2]ハイブリダイゼーション

水切りしたDNAマイクロアレイをハイブリダイゼーション装置 (Genomic Solutions Inc. Hybridization Station) にセットし、以下 ([1-5-6-3]、[1-5-6-4]) に示すハイブリダイゼーション溶液、条件でハイブリダイゼーション反応を行った。

【 0 0 8 4 】

[1-5-6-3]ハイブリダイゼーション溶液

6 x SSPE / 10% Form amide / Target (2nd PCR Products 全量)
(6xSSPE: NaCl 900mM、NaH₂PO₄ · H₂O 60mM、EDTA 6mM、p.H. 7.4)

[1-5-6-4]ハイブリダイゼーション条件

65 3min 92 2min 45 3hr Wash 2x SSC / 0.1% SDS at 25
Wash 2 x SSC at 20 (Rinse with H₂O : Manual) Spin dry (65
で3分、92度で2分、45 で3時間ハイブリダイゼーション反応させた後、2x SSC / 0.1% S
DS、25 で洗浄、2xSSC、20 で洗浄後、純水でリンスしスピンドライした。)

[1-5-7. 微生物の検出 (蛍光測定)]

ハイブリダイゼーション反応終了後のDNAマイクロアレイをDNAマイクロアレイ用蛍光検出装置 (Axon社製、GenePix 4000B) を用いて蛍光測定を行った。

【 0 0 8 5 】

以上の実施例の結果得られたスキャン画像の一例を図7に示す。なお、図7において、より蛍光強度の強いプローブは、より濃い色で示している。

【 0 0 8 6 】

701はDNAマイクロアレイに黄色ブドウ球菌のゲノムを含むサンプルを反応させることで得られたスキャン画像で、702は大腸菌のゲノムを含むサンプルを反応させることで得られたスキャン画像の一例である。

【 0 0 8 7 】

図の左に書いているアルファベットは、プローブ配列のアルファベットで、AからJまでそれぞれ、黄色ブドウ球菌 (A)、表皮ブドウ球菌 (B)、大腸菌 (C)、肺炎桿菌 (D)、緑膿菌 (E)、セラチア菌 (F)、肺炎連鎖球菌 (G)、インフルエンザ菌 (H)、エンテロバクター・クロアカエ菌 (I)、及びエンテロコッカス・フェカリス菌 (J) に特異的に結合するように設計されたプローブである。

【 0 0 8 8 】

2. 生物種判定処理 (ステップS103) に関する説明

次に、上記ステップS101で得られたスキャン画像を用いて行われる生物種類判定処理 (ステップS103) について以下に説明する。

【 0 0 8 9 】

[2-1. システム構成]

図2は本発明の一実施形態にかかる情報処理方法 (生物種類判定方法) を実現するための情報処理装置の構成を示すブロック図である。

【 0 0 9 0 】

生物種類判定方法は、外部記憶装置201、中央処理装置 (CPU) 202、メモリ203、入出力装置204から構成される装置において実現される。外部記憶装置201は、本実施形態にかかる生物種類判定方法を実現するプログラムや、ハイブリダイゼーション反応の結果得られたスキャン画像を保持する。また、本実施形態によって導かれた生物種類判定結果を保持する機能を持つ。中央処理装置 (CPU) 202は生物種類判定方法のプログラムを実行したり、すべての装置の制御を行ったりする。メモリ203は中央処理装置 (CPU) 202で処理されるプログラム、及びサブルーチンやデータを一時的に記録する。入出力装置204は、ユーザとのインタラクションを行う。また、プログラムの実行トリガーは入出力装置204を介してユーザが入力する。さらに、ユーザは入出力装置204を介して判定結果を見たり、プログラムのパラメータを設定したりすることができる。

10

20

30

40

50

【 0 0 9 1 】

[2 - 2 . 一般的な生物種判定処理の概要]

次に、本発明の情報処理方法について詳説する前に本方法の特徴を明確にすべく、ハイブリダイゼーション反応実験により得られたスキャン画像についての従来の判定処理方法の具体例を示し、その問題点について検討する。

【 0 0 9 2 】

[2-2-1. 相同性検索による判定処理の場合]

すでに、上記「従来技術」においても米国特許第 6 0 4 0 1 3 8 号明細書を挙げて述べたように、未知のサンプルの生物種を判定するために、相同性検索により、複数の原因菌の存在を判定する方法がある。

【 0 0 9 3 】

例えば、DNA マイクロアレイを用いた発現解析で行っていた解析手法、つまり、図 7 において、A 群の複数プローブから J 群の複数プローブまで、それぞれの群の複数プローブの平均蛍光強度を対象としている菌の蛍光強度とした上で、個々の菌の存在確率を $\{ (X \text{ 群の蛍光強度}) / (A \sim J \text{ 群の蛍光強度の総和}) \}$ (X は A ~ J のどれか) とする解析手法が挙げられる。かかる解析手法によれば、複数の原因菌が未知のサンプルに混入されていても、それぞれの菌の存在確率を適当に導き出すことができる。

【 0 0 9 4 】

しかしながら、すでに従来技術においても述べたように、かかる解析手法を用いた場合、塩基配列の似た原因菌の存在有無を厳密に区別することができない。一例を挙げると、理想的には、ハイブリダイゼーション反応の結果は、7 0 1 の A の行のプローブだけが蛍光強度が高くなり、かつ、7 0 2 の C の行のプローブだけが蛍光強度が高くなるのが望ましい(この 7 0 1 の理想的な結果は、図 5 に示した実験結果の例と同じである)。

【 0 0 9 5 】

ここで、全てのプローブが、図 5 に示したような理想的な性質を持っていれば、A 群のプローブから J 群のプローブまで、それぞれの群のプローブの平均蛍光強度を対象としている菌の強度とすることができ、その場合、1 つのサンプルの実験結果が A ~ J 群の蛍光強度の 1 0 個の値となるので、例えば、個々の菌の存在確率を $\{ (X \text{ 群の蛍光強度}) / (A \sim J \text{ 群の蛍光強度の総和}) \}$ (X は A ~ J のどれか) とすることが可能である。

【 0 0 9 6 】

しかしながら、図 7 に示すように、実際には理想通りにならず、いわゆる“クロスハイブリダイゼーション反応”がおこり、7 0 1 の場合は、A 以外の行のプローブも蛍光強度が強く、また、7 0 2 の場合は、C 以外の行のプローブも蛍光強度が強くなってしまふ。更に、7 0 2 の場合、C の行でも蛍光強度の弱いプローブが存在する場合もある。

【 0 0 9 7 】

このように、存在確率を求めるやり方は、従来の DNA マイクロアレイを用いた mRNA の定量分析などにおいては適した方法であったが、図 7 のような結果となる感染症の原因菌の判定においては適していない。

【 0 0 9 8 】

[2-2-2. その他の方法による場合]

ハイブリダイゼーション反応の結果を用いて生物種を判定する方法は上記相同性検索のほかにも幾つか考えられる。例えば、あらかじめ既知の生物種からなる基準サンプルについてハイブリダイゼーション反応させた結果得られたスキャン画像を格納しておき、当該既知の生物種の基準サンプルによるスキャン画像に基づいてパターン認識させることで、未知のサンプルの生物種を判定する方法等が挙げられる。

【 0 0 9 9 】

そこで、以下にパターン認識による生物種判定の実現可能性について検討してみる。ここでは、特に、基準サンプルと未知のサンプルのそれぞれについて、全プローブの蛍光強度をまとめて一つのベクトルとして表現し、両者のベクトルを用いて判定する方法について検討することとする。

10

20

30

40

50

【0100】

例えば、図7の例のような蛍光強度が得られたとすると、A～J群の合計72個のプロープをまとめて1つの72次元ベクトルとする。つまり、1つのサンプル（基準サンプル or 未知のサンプル）の実験結果から1つのベクトル（かかるベクトルを「統合ベクトル」と称す）が得られる。そして、複数の基準サンプルから得られた複数の統合ベクトルと未知のサンプルから得られた統合ベクトルとを比較することで、未知のサンプルがどの生物種に対応するかをパターン認識により判定することができる。

【0101】

図8A、Bに、一例として2次元ベクトル（つまり、プローブ数が2個（X、Y）のDNAマイクロアレイを用いて得られたスキャン画像に基づいて導き出された統合ベクトル）におけるパターン認識の例を示す（このときプローブXもプローブYも黄色ブドウ球菌に対して特異的であることを期待して設計されたプローブだとする）。

10

【0102】

図8A、Bでは、基準サンプルとして合計64個のサンプルが与えられ、その測定結果から64個の統合ベクトルが得られている（同図において、X軸、Y軸の値はそれぞれプローブX、プローブYの蛍光強度を示す（実際には、各サンプルの測定値を正規化したものである））。64個のうち、例えば黄色ブドウ球菌由来の統合ベクトルを黒の点、大腸菌由来の統合ベクトルを白の点として表す。なお、ここでは説明の便宜上プローブを2種類としたが、通常は図7のように多くの種類が存在し、測定結果は次元の高いベクトルとなる。

20

【0103】

そして、図8（B）に示す複数の基準サンプル由来の統合ベクトルの分布図に対して、分類木（図8（A））を用いてパターン認識をすることができる。

【0104】

分類木とは、階層的に基準サンプルの分布する特徴空間を分割する方法であり、例えば図8（B）に示した基準サンプル集合で分類木を作成すると、図8（A）に示すような構造のものが出来る。図8（A）に示した分類木の各ノードは図8Bの図の中で太い線で示した境界線を示し、全体で見ると特徴空間を7つの部分区間に分割することができる。それぞれの部分区間は、分類木のリーフノードに対応し、白丸、黒丸で示してある。結局、図8（A）、（B）に示した例の場合は、黄色ブドウ球菌の部分区間が4つ、大腸菌の部分区間が3つとなる。

30

【0105】

分類木を用いた判定方法は、未知のサンプルに由来する統合ベクトルが与えられた場合に、どの部分区間に属するかを判断し、その属した部分区間に対応する生物種を判定結果とする（図8（A）、（B）の場合、例えば、Xの値が0.5以下で、かつYの値が0.5以下であれば、黄色ブドウ球菌由来の統合ベクトルが分布する領域に属することとなる。また、Xの値が0.5以下で、Yの値が0.75以上であれば、大腸菌由来のベクトルが分布する領域に属することとなる）。

【0106】

そして、階層的に分類木を辿っていくことによって、未知サンプルに由来する統合ベクトルがどの部分区間に属するかを判定できるので、一般的には非常に高速に生物種を判定できるというメリットがある。

40

【0107】

以上の説明からわかるように、統合ベクトルを用いてパターン認識を行う場合、1生物種あたり少なくとも1サンプル以上の基準サンプルを用意しておく必要があり、基準サンプルの数が少ないと、その生物種の判定精度が低下するか、あるいは不可能となってしまう。もちろん、原因菌が1つのみ存在する場合には問題は生じないが、上述したように感染症の原因菌の判定の場合、原因菌が複数存在するのが普通であり、そのような場合には、正しい判定結果が得られない。これは、ハイブリダイゼーション反応を行う検体中に含まれる原因菌の組み合わせによって、統合ベクトルの向き、大きさが異なってくるため

50

ある。つまり、検体中に原因菌が複数存在する場合、導かれた統合ベクトルは複数の原因菌の結果を足し合わせたものに近くなってしまうため、このような統合ベクトルに基づいて上述のパターン認識を行っても、未知のサンプルがどの基準サンプルにも似ていないという判定結果となってしまう。

【 0 1 0 8 】

これに対処するには、1つの生物種に対して、様々な原因菌と組み合わせた基準サンプルを多数用意しておき、それぞれについてハイブリダイゼーション反応させた結果を蓄積しておく必要がある。しかしながら、あらゆる原因菌の組み合わせに対する基準サンプルを用意しておくことは現実的ではない。したがって、このようなパターン認識は、類似する塩基配列を有する原因菌が含まれる場合の生物種判定にそのまま適用することは適切でない。

10

【 0 1 0 9 】

[2 - 3 . 本願に基づく生物種判定処理の特徴]

上述のような背景のもと、本願に基づく生物種判定処理では、上記「2-2-1. 相同性検索の場合」の問題点を解決すべくパターン認識による判定方法を採用しつつ、パターン認識を用いた場合の問題点に関する上記検討を考慮し、あらゆる原因菌の組み合わせを用意しておかなくても、類似する塩基配列について精度よく判定できるようにした。具体的には、基準サンプルと未知サンプルとのベクトル対比にあたり、ベクトルフィルタ処理を付加した点に特徴がある。以下に詳細を説明する。

【 0 1 1 0 】

20

[2 - 4 . 本願に基づく生物種判定処理の流れ]

図9は、本実施形態にかかる生物種類判定方法の処理を説明するための機能ブロック図である。901は“基準サンプルに対するスキャン画像”で、ターゲットとしている生物種由来の核酸断片を含む基準サンプルをハイブリダイゼーション反応させた結果得られたものである。通常この核酸断片には蛍光物質などの標識分子が付加されていて、DNAマイクロアレイとのハイブリダイゼーション反応の強度を簡単に測定できるようになっている。

【 0 1 1 1 】

902はハイブリダイゼーション反応数値化部であり、前記DNAマイクロアレイと前記基準サンプルとのハイブリダイゼーション反応の強度を数値化する処理を行う。903は測定結果ベクトル化部であり、前述のハイブリダイゼーション反応数値化部で得られたDNAマイクロアレイ上の各プローブに対する測定値を統計処理した後で、n次元のベクトルとして表現し直す処理を行う。904はベクトル正規化部であり、生成されたベクトルを正規化する。ベクトル化され正規化されたデータ(基準ベクトルデータ)は、基準ベクトルデータ収集部905に蓄えられる。

30

【 0 1 1 2 】

909は主成分分析部であり、基準ベクトルデータ収集部905に蓄えられた基準ベクトルデータについて主成分分析を行う。主成分分析部909における主成分分析の結果は、ベクトルフィルタ部905においてフィルタ処理する際に用いられる。

【 0 1 1 3 】

40

907は“未知サンプルに対するスキャン画像”で、基準サンプルに対するスキャン画像と同様に、ハイブリダイゼーション反応数値化部902において数値化処理され、測定結果ベクトル化部903にてベクトル表現に変換され、ベクトル正規化部904にて正規化される(未知ベクトルデータ)。

【 0 1 1 4 】

生物種判定部908では、パターン認識により未知サンプルの生物種を判定するが、判定にあたっては、ベクトルフィルタ部905にてフィルタリングされたベクトルを用いる。つまり、未知サンプルに対するスキャン画像より得られた正規化された未知ベクトルデータに対してベクトルフィルタ部905にてフィルタされたベクトルデータと、基準ベクトルデータ収集部906に蓄えられた基準ベクトルデータに対してベクトルフィルタ部9

50

05にてフィルタされたベクトルデータとを対比し、未知のサンプルの生物種を判定する。

【0115】

[2-5. 処理の詳細な説明]

以下に図9の各部の処理の詳細について説明する。

【0116】

[2-5-1. ベクトル正規化処理]

ベクトル正規化部904におけるベクトル正規化処理とは、各サンプルごとに得られた蛍光強度に基づいて導かれたベクトルを正規化する処理をいう。

【0117】

例えば、DNAマイクロアレイ上に同一プローブの複数のスポットがある場合は、その蛍光強度の平均値をプローブの蛍光強度とするのが一般的である。

【0118】

図10に同じ種類のプローブが複数スポットあるDNAマイクロアレイの例を示す。図10のDNAマイクロアレイでは、20種類のプローブが4セット基板上に固定されており、合計80個のスポットが存在する。この場合、それぞれ4つある同じ種類のプローブの平均強度をそのプローブの測定値として、20次元のベクトルとしてまとめる。

【0119】

図10に示したDNAマイクロアレイの例では、左上のプローブがポジティブコントロールとなっている。例えば、ポジティブコントロールとして予めどのプローブとも干渉しない核酸をスポットしておいて、ハイブリダイゼーション反応する直前にそのプローブ塩基配列の相補鎖となるオリゴヌクレオチドに蛍光色素を付加した物質を加える。また、サンプル中に必ず含まれるような部分塩基配列を持ったプローブをポジティブコントロールとすることもできる。

【0120】

DNAマイクロアレイの実験においては、蛍光強度が全体として高くなったり低くなったりすることがある。このような場合に、前述したポジティブコントロールを用いることで、全てのプローブの蛍光強度を正規化することができる。また、DNAマイクロアレイの中で一番蛍光輝度の高いスポットの測定値を基準に、全てのプローブの蛍光強度を正規化する方法もある。

【0121】

[2-5-2. ベクトルフィルタ処理]

次に本発明の特徴であるベクトルフィルタ部905におけるベクトルフィルタ処理について説明する。はじめに、ベクトルフィルタ処理の概念について説明する。上述のようにハイブリダイゼーション反応の結果得られた統合ベクトルは、プローブの数により決定される多次元空間において、各サンプルごとに決定される。このとき、検体にどのような原因菌が含まれているかによって、当該多次元空間における統合ベクトルが大きく異なってくる。

【0122】

例えば、単一の原因菌しか含まれていない検体の反応結果から得られた統合ベクトルと、当該原因菌のほか、他の原因菌も含む検体の反応結果から得られた統合ベクトルとは、大きく異なるベクトルとなる。

【0123】

つまり、どのような組み合わせの原因菌を含んだ検体に対する反応結果を基準サンプルとするかが判定結果に大きく影響を及ぼすこととなる。このため、あらゆる組み合わせの原因菌を含むハイブリダイゼーション反応の結果得られた基準サンプルを用意する方法も考えられるが、これは現実的ではないことはすでに述べた。

【0124】

そこで、組み合わせの影響を極力排除した状態で判定を行うようにした点に本発明の特徴がある。具体的には、統合ベクトルにフィルタをかけることで、未知サンプルから所定

10

20

30

40

50

の原因菌のベクトル成分（特定ベクトル）を抽出する一方、基準サンプルの統合ベクトルについても同様に所定の原因菌の特定ベクトルを抽出する。そして、該特定ベクトル同士を比較しパターン認識することで、基準サンプルの数をおさえつつ、当該所定の原因菌の有無を精度よく判定することを可能にするものである。

【0125】

このように、ベクトルフィルタ処理を行うことで、統合ベクトルから、所定の原因菌以外のベクトル成分を排除した特定ベクトルを抽出することで、限られた基準サンプルを用意しておくだけで、正しい判定結果が得られるようになる。

【0126】

次にベクトルフィルタ部905における具体的な処理について説明する。ベクトルフィルタ部906の最も原始的なアルゴリズムを図11に示す。1101は黄色ブドウ球菌用のフィルタで、黒いスポットの係数が1で白いスポットの係数が0を意味する。例えば、図7のようなハイブリダイゼーション反応の結果が得られた場合、701の実験データと702の実験データ共に1101のフィルタを適用する。すると、結果的に7次元だったベクトルが9次元にフィルタされることとなる。

【0127】

同様に、1102は大腸菌用のフィルタを意味し、黒いスポットの係数が1で白いスポットの係数が0である。このフィルタを、例えば、図7のような実験データに適用すると7次元のベクトルが7次元にフィルタされることとなる。結局、図7のような2つの基準サンプルの実験結果が得られたとすると、図9の基準ベクトルデータ収集部906では、9次元のベクトルが2つ、7次元の特定ベクトルが2つ収集されることとなる。これに対して未知サンプルの実験データからは、9次元の特定ベクトル、7次元の特定ベクトルが1つずつ得られ、それぞれを前記の同じ次元の基準ベクトルと比較し、黄色ブドウ球菌、大腸菌のそれぞれの存在有無が推定される。このように図10に示した黄色ブドウ球菌のフィルタや大腸菌のフィルタなどのように、生物種毎にフィルタを用意しておいて、そのフィルタでベクトルをフィルタし、特定ベクトルを求めることによって、生物種毎の存在有無がパターン認識で求められる。

【0128】

[2-5-3. 主成分分析処理]

図10では、説明の便宜上、それぞれの生物種プロープに対応する測定値に1をかけて、それ以外のプロープに0をかける単純なフィルタの例を示したが、一般には、予め得られている知見に基づき、各プロープの測定値ごとに0から1の間の定数をかけることによって、フィルタを実現することができる。そこで、以下にフィルタの構成方法について説明する。

【0129】

一般に、ベクトル群の情報を圧縮する技術に、主成分分析がある（詳細な技術は、例えば書籍「回帰分析と主成分分析 統計解析プログラム講座 2 芳賀 敏郎、橋本 茂司（著）」出版社：日科技連出版社；ISBN:4817120118；(1980/05)を参照）。これは、複数の測定結果ベクトルから得られる共分散行列を固有値分解し、それぞれの固有値に対応する固有ベクトルを用いて測定結果ベクトルを主成分分解（スペクトル分解）する方法である。固有ベクトルは正規直交基底にできるので、主成分分解（スペクトル分解）はいわゆる直交座標変換となる。よって、基本的に主成分分解前のベクトルの次元と主成分分解後のベクトルの次元は変わらない。しかし主成分分解後、固有値の極端に小さい値に対応する成分を無視できる場合が多く、この場合、ベクトルの次元を小さくすることができる。そしてこれがフィルタとして機能する。

【0130】

以下に、生物種毎に主成分分析を行いフィルタを構成する主成分分析部909での処理について説明する。

【0131】

図12が図9の主成分分析部909における処理の流れを示すフローチャートである。

まずフィルタを作る生物種を選択する。感染症の原因菌判定の場合、例えば黄色ブドウ球菌や大腸菌などを指定する。次に、図9の基準ベクトルデータ収集部906で集められた基準サンプルの統合ベクトルのうち、ステップS1201で選ばれた生物種サンプルに由来するものだけを選択する。この時、通常は1201で選ばれた生物種に由来する基準サンプルの統合ベクトルデータは全て選択する。また、ステップS1201で選択された生物種以外の生物種に由来する基準サンプルの統合ベクトルデータを選択する。

【0132】

次にステップS1202とS1203で選択した統合ベクトルデータ群の共分散行列を求めて、固有値を計算し、固有値の小さい値に対する成分を無視する。これによって、ステップS1201で選択した生物種を判定するのに重要なプローブの測定値がフィルタされることになる。

10

【0133】

例えば、プローブがn個あるとすると、ステップS1202とS1203で選択した統合ベクトルデータ群から得られる共分散行列はn x nの対称非負行列になり、その固有値もn個存在する。固有値を大きい順に並べたものを λ_i ($i = 1, 2, \dots, n$) とすると、第m成分までの累積寄与率 (accumulated propotion) は、

【0134】

【数1】

$$\sum_{i=1}^m \lambda_i / \sum_{i=1}^n \lambda_i$$

20

【0135】

で計算される。

これが例えば80%以上の時点で、主成分分解 (スペクトル分解) を打ち切り、それより小さい固有値に対応する成分を無視する。なお、上述の例で80%という数字はユーザが与える任意の比率に設定してよい。また、ユーザに個々の固有ベクトルを固有値と共に表示して見せて、無視する主成分分解成分をユーザに選択させてもよい。

【0136】

30

この時、ステップS1201で選択された生物種以外の生物種に由来する基準サンプルの統合ベクトルデータを全て選択してもよいが、判定すべき生物種が多い場合はステップS1203で選択される統合ベクトルの数がステップS1202で選択される統合ベクトルの数より遥かに多い状況が生じる。するとその後の主成分分析の結果がステップS1203で選択された統合ベクトルに引きずられ、結果的に全ての生物種で主成分分析の結果がほとんど同じになってしまうことがある。これを回避するために、例えば、ステップS1203で選ばれる統合ベクトルの数をステップS1202で選ばれる統合ベクトルの数と同じ程度にするという工夫をする。

【0137】

例えば、ステップS1202で選択された統合ベクトルの数と同じ数だけステップS1201で選択された生物種以外の生物種に由来する基準サンプルの統合ベクトルデータの中からランダムに選択する。この時、色々な生物種の基準サンプルの統合ベクトルデータを選択するために、判定すべき生物種の数 N だとすると、例えば、生物種毎に収集された統合ベクトルの $1 / (N - 1)$ の数の基準サンプルの統合ベクトルをランダムに選択していく。するとステップS1203で選択される統合ベクトルの数はステップS1202で選択される統合ベクトルの数とほとんど同じになる。

40

【0138】

[2-5-4. 判定処理]

一般に、ベクトル同士の比較、分類は“パターン認識”と呼ばれる技術で行う。その詳しい技術内容は例えば、IEEE Transaction on Pattern Analysis and Machine Learning,

50

Vol. 22, No. 1, January 2000, pp.4-pp.37にある“Statistical Pattern Recognition: A Review” Anil K. Jain, Robert P.W. Duin, and Jianchan Mao. の論文にレビューされている。本発明の生物種類判定方法にはパターン認識の技術である、k-Nearest-Neighbor法、分類木、Support Vector Machine、ベイズ識別法、ブースティング法、ニューラルネットなどのいずれの方法についても適用できる。

【0139】

ここでは、k-Nearest-Neighbor法によるパターン認識と分類木を用いたパターン認識とについて説明する。k-Nearest-Neighbor法はパターン認識のアルゴリズムで最も原始的な方法である。k-Nearest-Neighbor法とは基準サンプルと未知のサンプルの特定ベクトル同士の距離を計算して、基本的に近い距離にある基準サンプルと同じ生物種を判定結果とする方法である。特定ベクトル同士の距離には、

10

【0140】

【数2】

$$\sqrt{\frac{\sum_{i=1}^n (x_i - y_i)^2}{n}} \quad (X \text{は未知サンプル、} Y \text{は基準サンプル由来のベクトル})$$

【0141】

の式で表されるユークリッド距離を用いるのが一般的である。また、

20

【0142】

【数3】

$$\frac{\sum_{i=1}^n |x_i - y_i|}{n}$$

【0143】

で示される絶対値距離を用いてもよい。k-Nearest-Neighbor法とは、距離の近い順に基準サンプルを並べて、最も近い方からK個の中で最も数の多い基準サンプルの生物種を判定結果とする方法である。特に1-Nearest-Neighbor法とは、未知のサンプル由来の特定ベクトルと最も近い基準サンプルの生物種を判定結果とする方法である。

30

次に分類木を用いたパターン認識について図13～図15を用いて説明する。はじめに図13を用いて分類木を用いたパターン認識の概要について説明する。図13に示すように、分類木を用いたパターン認識では、まず、学習パターン1301から複数の分類木1303を作成する分類木作成処理を実行する(1302)。

【0144】

この学習パターン1301はいわゆる「教師付きデータ」と呼ばれ、どのカテゴリーに属するパターンか予めわかっている。感染症原因菌判定の例で言えば、DNAチップのハイブリパターンとその菌の情報とがペアで含まれている。また、一般にこの分類木1303を作成する工程を学習フェーズと呼ぶ。

40

【0145】

次に属するカテゴリーの不明なパターン(未知パターン1304)を学習フェーズで作成した分類木1303を用いてパターンマッチングし(1305)、属するカテゴリーを推定する。感染症原因菌判定の例で言えば、DNAチップのハイブリパターンを元に、その菌の判定を行う。一般にこの未知パターン1304をパターンマッチングする工程をパターン認識フェーズと呼ぶ。

【0146】

本実施形態では、学習フェーズにおいて、同じ学習パターン1301から複数の分類木1303を作成する。学習フェーズにおいて作成された分類木1303がn個であった場

50

合、パターン認識フェーズにおいても、それぞれの分類木 1 3 0 3 に対応して n 個の認識結果が得られることとなる。そして最終的な認識結果 1 3 0 6 は、これら n 個の認識結果の多数決投票により行う。なお、作成されたそれぞれの分類木 1 3 0 3 が確率も含めた分類木であった場合には、n 個全ての確率付き認識結果をカテゴリー毎に足し合わせて、もっとも確率の高いカテゴリーを全体の認識結果 1 3 0 6 とする（このようなアルゴリズムは、一般にアンサンブルアルゴリズム (ensemble algorithm) と呼ばれる)。

【 0 1 4 7 】

次に図 1 4 を用いて上記分類木作成処理 (1 3 0 2) の詳細について説明する。図 1 4 は分類木作成処理 (1 3 0 2) の流れを示すフローチャートである。同図に示すように、分類木を作成するにあたっては、まず、初期設定としてステップ S 1 4 0 1 においてル
10
ートノードをカレントノードに設定する。ここでカレントノードとは、現在着目しているノードであり、学習パターンのサブセット (フルセットやヌルセットも有りうる) を含んでいる。また、ルートノードとは、分類木の一番親のノードであり、全ての学習パターンを含んでいる。

【 0 1 4 8 】

次にステップ S 1 4 0 2 においてカレントノードが学習パターンを含むかどうかを判定する。カレントノードがルートノードである場合には、常に学習パターンを含むが、分類木作成処理が進み分類木が細分化された結果、下位層のノードがカレントノードとなった場合には、当該カレントノードに学習パターンが含まれない場合も生じえる。もし含ま
20
ない場合は、ステップ S 1 4 0 8 に進み、カレントノードを NULL ノードとする。ここで NULL ノードとは、判定結果が不明なノードであり、未知のパターンが NULL ノードに落ちた場合は、その親のノードのカテゴリー存在確率をパターン認識結果とする。

【 0 1 4 9 】

カレントノードが NULL ノードに設定された場合には、当該ノードについてはそれ以上子ノードを生成する必要がないため、当該ノードの親ノードに戻る (つまり、当該ノードの親ノードをカレントノードに設定する (ステップ S 1 4 1 0))。

【 0 1 5 0 】

一方、カレントノードが学習パターンを含む場合 (NULL ノードでない場合)、ステップ S 1 4 0 3 においてリーフノードの条件を満たすかどうかを確認する。リーフノードとは子ノードを持たないノードをいい、確認にあたっては、例えばカレントノードに含
30
まれる学習パターンのエントロピーを求め、当該エントロピーがある閾値以下であればリーフノードと判定する。なお、このとき用いる閾値が 0 に設定されていたならば、リーフノードには単独のカテゴリーに属する学習パターンしか含まれないこととなる。

【 0 1 5 1 】

ステップ S 1 4 0 3 における確認の結果、リーフノードの条件を満たすと判定された場合には、ステップ S 1 4 0 9 に進みカレントノードをリーフノードとする。上述のようにリーフノードは子ノードを持たないノードであることから、当該ノードに対して子ノードが生成されることはないため、当該ノードの親ノードに戻る (つまり、当該ノードの親ノードをカレントノードに設定する (ステップ S 1 4 1 0))。

【 0 1 5 2 】

一方、カレントノードが NULL ノードでもリーフノードでもない場合には、ステップ S 1 4 0 4 においてノードの判定関数を決定し、子ノードを生成する (ステップ S 1 4 0 5)。そして、ステップ S 1 4 0 7 にて、それぞれの子ノードを順次カレントノードに設定し、ステップ S 1 4 0 2 からステップ S 1 4 0 5 及びステップ S 1 4 0 8 からステップ S 1 4 1 0 までの処理を繰り返し、更に深く分類木を作成していく。
40

【 0 1 5 3 】

なお、本実施形態では、分類木は 2 分岐分類木を用いることとし、ステップ S 1 4 0 5 にて生成される子ノードの数は常に 2 個である (このようなアルゴリズムを 2 分岐分類木アンサンブルアルゴリズムという)。かかるアルゴリズムをプログラムで実装した場合、カレントノードを子ノードに設定するのはループで実現され、2 分岐ということは、ルー
50

プの回数が2となる。

【0154】

ルートノード以下の全ての子ノードが展開されたら、つまり、最終的にルートノードの下の全てのノードの末端がNULLノードかリーフノードになったら（ステップS1406において「No」となったら）、ステップS1411に進み、親ノードがルートノードであるか否かを確認する。ルートノード以下の全ての子ノードが展開された時点において、カレントノードが分類木の下位層のノードに位置していた場合には、ステップS1411、ステップS1410、ステップS1406を繰り返すことで、上位層のノードに移動していき、カレントノードがルートノードの1つ下のノードまで移動した時点で、処理を完了する。

10

【0155】

次に図15を用いてノードにおいて判定関数を決定する工程（ステップS1404）の概要を説明する。図14で説明してきた分類木作成処理（1302）は、分類木作成アルゴリズムに普遍のものである。そして、個々の分類木作成処理において作成された分類木が高い認識率を得るためには、各ノードにおいて、どのようなアルゴリズムで判定関数を決定するかという点が重要となってくる。そこで、本実施形態ではランダムサンプリングを用いて判定関数を決定することとした。

【0156】

図15はランダムサンプリングを用いて判定関数を決定する処理の概要を説明するための図であり、1501はカレントノードに含まれる学習パターンの分布を示している。ここでは、説明の簡略化のためカテゴリーは白と黒で表現した2種類とし、カレントノードには白の学習パターンが7個、黒の学習パターンが5個、合計で12個の学習パターンが含まれているものとする。

20

【0157】

判定関数を決定するにあたっては、まず、カレントノードが含む全ての学習パターンから1つランダムに学習パターンを選び、次にその選ばれた学習パターンが属するカテゴリーとは異なる他のカテゴリーの学習パターンから1つランダムに学習パターンを選ぶ（1503が選択された学習パターン）。そして、この2つの学習パターンのどちらに近いかを示す関数を判定関数とする。

【0158】

2つの学習パターンのどちらに近いかを示す関数を求める際の指標となる「近さ」はユークリッド距離を用いるのが普通であるが、それに限ったことではなく、距離空間を構成できる距離尺度であれば何であってもよい。学習パターンがベクトルでユークリッド距離を採用すると、判定関数で仕切られる判定曲線1502は超平面となる。一般により複雑な距離尺度を用いると、より複雑な判定曲線が得られる。なお、図15の場合には、判定曲線1502の紙面右側が黒の学習パターンに近い領域であり、判定曲線1502の紙面左側が白の学習パターンに近い領域となる。

30

【0159】

なお、ランダムに選ぶ学習パターンの個数は2つに限られず、あるカテゴリーからm個、他のカテゴリーからm個選び、それぞれm個の平均パターンを求め、その2つのパターンからの距離を測ることで、判定関数を求めるようにしてもよい。

40

【0160】

このように、本実施形態では分類木作成処理にあたり、2分岐分類木アンサンブルアルゴリズムを用いるとともに、さらに、分類木の各ノードにおける判定関数を、各ノードに存在する学習データの中からランダムに選択された、互いにカテゴリーの異なる2つの学習パターンのどちらに近いかを示す関数でもって定義することにより、作成された分類木において高い認識率を実現することが可能となる。

【0161】

3.まとめ

以上の説明から明らかなように、本実施形態では、従来の培養法に替えて、DNAマイ

50

クロアレイによるハイブリダイゼーション反応を用いて判定を行うことにより、簡易かつ安価に、しかも短時間での生物種判定を実現した。

【0162】

また、ハイブリダイゼーション反応の結果として得られたスキャン画像を用いて生物種の判定を行うにあたっては、従来の存在確率を求める方法に替えて、基準サンプルとの対比に基づいてパターン認識を用いて判定することにより、互いに類似する塩基配列を有する生物種であっても判定できるようになった。

【0163】

このとき、基準サンプルとの対比において、基準サンプル上の全プローブの蛍光強度を統合した統合ベクトルを用いるのではなく、統合ベクトルから所定の原因菌成分を抽出して得られた特定ベクトルを用いるようにすることで、予め用意しておく基準サンプルの数が少なくても判定できるようになった。

【0164】

なお、統合ベクトルから特定ベクトルを抽出する際に用いられる原因菌ごとのベクトルフィルタは、所定の原因菌が含まれる複数の基準サンプルの統合ベクトルを主成分分析することで取得することとした。

【0165】

なお、本発明は、複数の機器（例えばホストコンピュータ、インタフェイス機器、リーダー、プリンタなど）から構成されるシステムに適用しても、一つの機器からなる装置（例えば、複写機、ファクシミリ装置など）に適用してもよい。

【0166】

また、本発明の目的は、前述した実施形態の機能を実現するソフトウェアのプログラムコードを記録した記憶媒体を、システムあるいは装置に供給し、そのシステムあるいは装置のコンピュータ（またはCPUやMPU）が記憶媒体に格納されたプログラムコードを読み出し実行することによっても、達成されることは言うまでもない。

【0167】

この場合、記憶媒体から読出されたプログラムコード自体が前述した実施形態の機能を実現することになり、そのプログラムコードを記憶した記憶媒体は本発明を構成することになる。

【0168】

プログラムコードを供給するための記憶媒体としては、例えば、フロッピー（登録商標）ディスク、ハードディスク、光ディスク、光磁気ディスク、CD-ROM、CD-R、磁気テープ、不揮発性のメモ리카ード、ROMなどを用いることができる。

【0169】

また、コンピュータが読出したプログラムコードを実行することにより、前述した実施形態の機能が実現されるだけでなく、そのプログラムコードの指示に基づき、コンピュータ上で稼働しているOS（オペレーティングシステム）などが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【0170】

さらに、記憶媒体から読出されたプログラムコードが、コンピュータに挿入された機能拡張ボードやコンピュータに接続された機能拡張ユニットに備わるメモリに書込まれた後、そのプログラムコードの指示に基づき、その機能拡張ボードや機能拡張ユニットに備わるCPUなどが実際の処理の一部または全部を行い、その処理によって前述した実施形態の機能が実現される場合も含まれることは言うまでもない。

【図面の簡単な説明】

【0171】

【図1】本発明の情報処理方法を含む検査全体の流れを示した図である。

【図2】本発明の一実施形態にかかる情報処理方法（生物種類判定方法）を実現するための情報処理装置の構成を示すブロック図である。

10

20

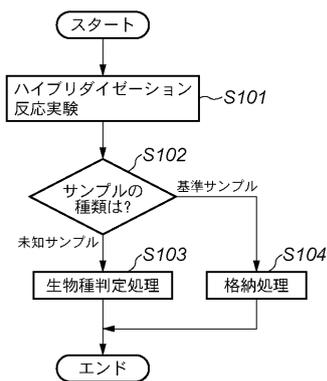
30

40

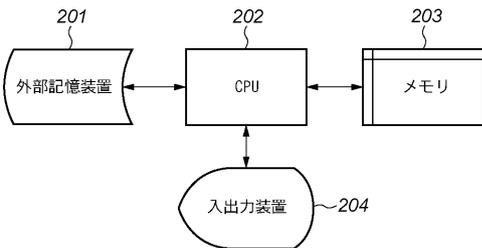
50

- 【図3】DNAマイクロアレイ上のハイブリダイゼーションの様子を示した図である。
- 【図4】DNAマイクロアレイを用いたハイブリダイゼーション反応実験の実験手順全般について説明するための図である。
- 【図5】感染症の菌を特定するDNAマイクロアレイの原理を説明するための図である。
- 【図6】ハイブリ溶液に複数の種類の塩基配列が存在する理由を説明するための図である。
- 【図7】ハイブリダイゼーション反応後の蛍光強度を表す画像の一例を示す図である。
- 【図8】ベクトルの分布図とそれを判別する分類木の一例を示す図である。
- 【図9】本発明の一実施形態にかかる生物種類判定方法の処理を説明するための機能ブロック図である。
- 【図10】同じ種類のプローブが複数スポットあるDNAマイクロアレイの例を示す図である。
- 【図11】本発明の一実施形態にかかる情報処理方法において用いられる原始的なベクトルフィルタの一例を示す図である。
- 【図12】主成分分析処理の流れを示すフローチャートである。
- 【図13】学習フェーズとパターン認識フェーズを説明する図である。
- 【図14】分類木作成処理の流れを示すフローチャートである。
- 【図15】ノード判定関数を決定するステップを説明するための図である。

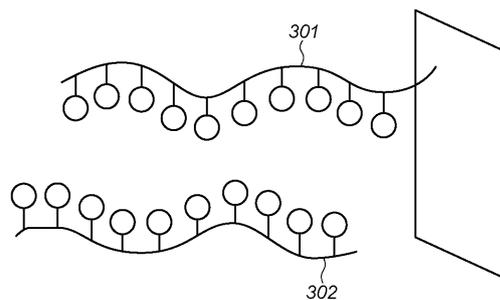
【図1】



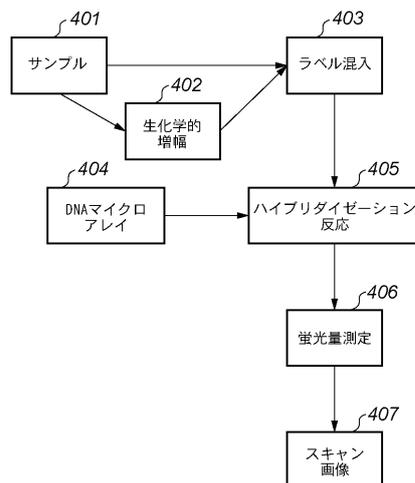
【図2】



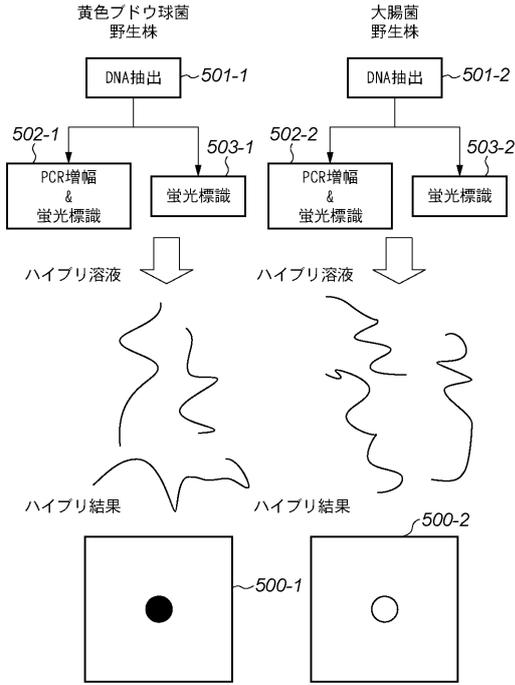
【図3】



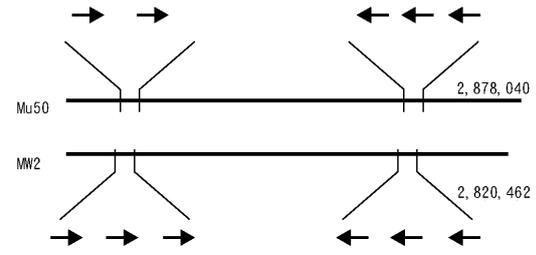
【図4】



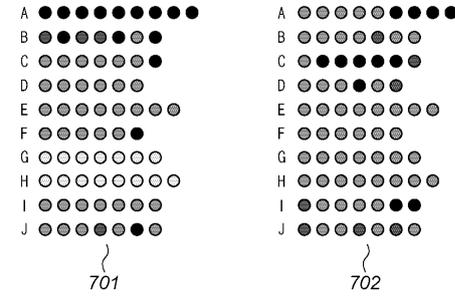
【図5】



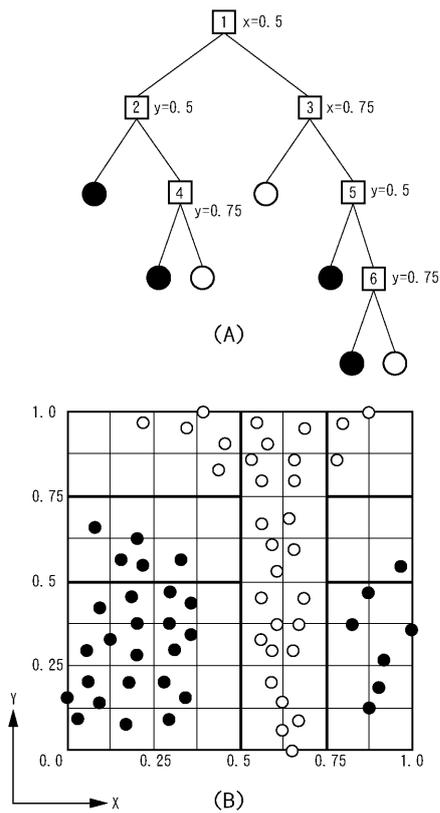
【図6】



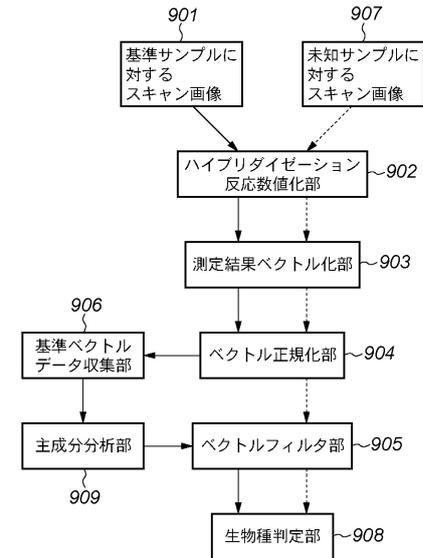
【図7】



【図8】

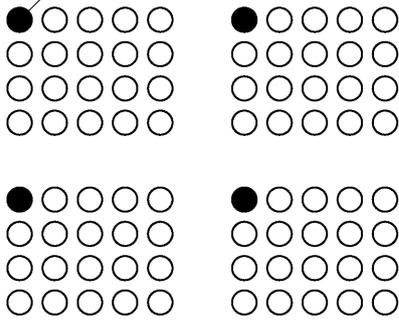


【図9】

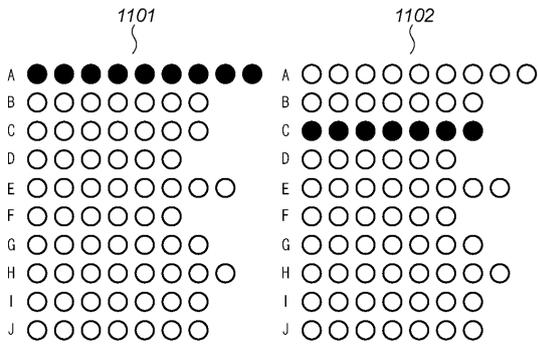


【図10】

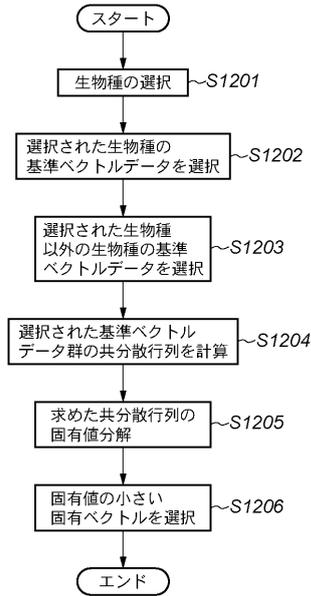
ポジティブコントロール



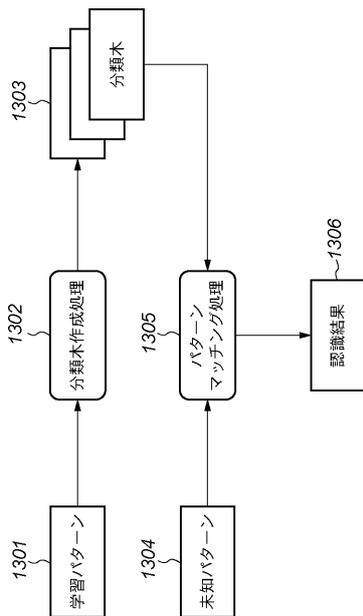
【図11】



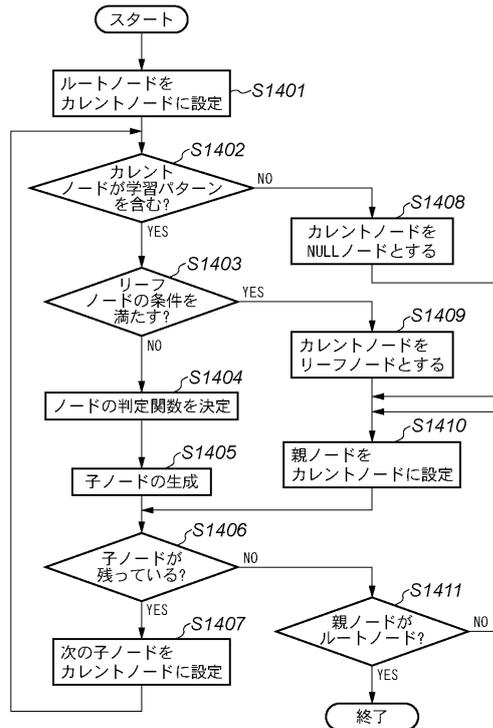
【図12】



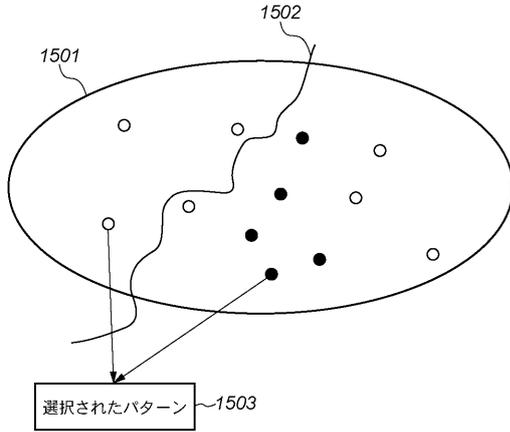
【図13】



【図14】



【図15】



【配列表】

0004579577000001.app

フロントページの続き

(72)発明者 吉井 裕人
東京都大田区下丸子3丁目30番2号 キヤノン株式会社内

審査官 小金井 悟

(56)参考文献 特表2000-504575(JP,A)
特開2002-357606(JP,A)
国際公開第03/012078(WO,A2)
Genome Res., 1998年 5月, Vol.8, No.5, pp.435-448
J. Clin. Microbiol., 1999年 1月, Vol.37, No.1, pp.49-55

(58)調査した分野(Int.Cl., DB名)

C12N 15/00 - 15/90

C12M 1/00

C12Q 1/68

G01N 33/53

G01N 37/00

JSTPlus/JMEDPlus/JST7580(JDreamII)

PubMed