

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6845256号
(P6845256)

(45) 発行日 令和3年3月17日(2021.3.17)

(24) 登録日 令和3年3月1日(2021.3.1)

(51) Int. Cl.		F I	
C 1 2 Q	1/68	(2018.01)	C 1 2 Q 1/68
C 1 2 Q	1/686	(2018.01)	C 1 2 Q 1/686 Z
C 1 2 M	1/00	(2006.01)	C 1 2 M 1/00 A
C 1 2 N	15/09	(2006.01)	C 1 2 N 15/09 Z

請求項の数 27 (全 24 頁)

(21) 出願番号 特願2018-557201 (P2018-557201)
 (86) (22) 出願日 平成29年1月23日 (2017.1.23)
 (65) 公表番号 特表2019-511242 (P2019-511242A)
 (43) 公表日 平成31年4月25日 (2019.4.25)
 (86) 国際出願番号 PCT/EP2017/051327
 (87) 国際公開番号 W02017/125606
 (87) 国際公開日 平成29年7月27日 (2017.7.27)
 審査請求日 令和1年10月30日 (2019.10.30)
 (31) 優先権主張番号 1650527
 (32) 優先日 平成28年1月22日 (2016.1.22)
 (33) 優先権主張国・地域又は機関
 フランス (FR)

(73) 特許権者 515031137
 ユニヴェルシテ ドュ モンペリエ
 UNIVERSITE DE MONTPE
 LLIER
 フランス共和国, エフー34090 モン
 ペリエ, リュ オーギュスト ブルソネ
 163

最終頁に続く

(54) 【発明の名称】 生物学的サンプルを分類する方法

(57) 【特許請求の範囲】

【請求項1】

測定用生物学的サンプルを分類する方法であって、
 前記方法が、

- 少なくとも1つの測定曲線と呼ばれる、前記測定用生物学的サンプルの少なくとも1つのDNA融解曲線の取得(9)；
- 前記少なくとも1つの測定曲線に由来するディスクリプタの分析による、異なる可能グループの中の所定グループに前記測定用生物学的サンプルが属していることの決定(10)；

を含み、

前記取得及び決定の各工程が技術的手段(101, 102)の使用を伴っており、
 前記ディスクリプタが、各測定曲線の二次導関数の1以上の点を含み、前記の点または前記の各点が特有の温度値に関する融解曲線の二次導関数の値を含むことを特徴とする、前記方法。

【請求項2】

前記ディスクリプタが、各測定曲線の二次導関数の少なくとも30個の点を含むことを特徴とする、請求項1に記載の方法。

【請求項3】

前記ディスクリプタが、各測定曲線の一次導関数の1以上の点、及び/又は各測定曲線の1以上の点を含むことを特徴とする、請求項1又は2に記載の方法。

【請求項 4】

前記ディスクリプタが、各測定曲線の 1 以上の百分位数を含むことを特徴とする、請求項 1 ~ 3 のいずれか一項に記載の方法。

【請求項 5】

前記決定 (1 0) が、ランダムフォレスト法による決定を含むことを特徴とする、請求項 1 ~ 4 のいずれか一項に記載の方法。

【請求項 6】

前記方法が、学習 (6) を含み、その学習 (6) が、
 - 異なる初期グループに属する異なる基準用生物学的サンプルから開始される、基準曲線と呼ばれる異なる DNA 融解曲線の取得 (1) ; それに続く、
 - 基準曲線から開始するディスクリプタの決定 (2 , 3) ; それに続く、
 - 前記ランダムフォレスト法によるいくつかの木の構築を含む、前記ランダムフォレスト法による森林の構築 (8) であって、各木の各ノードで検討される 1 つ又は複数の変数が、1 以上のディスクリプタを含み、各木の各葉が、異なる可能グループのうちの 1 つのグループにのみ相当する、前記構築 ;
 を含むことを特徴とする、請求項 5 に記載の方法。

10

【請求項 7】

前記ディスクリプタの決定 (2 , 3) が、
 - ディスクリプタの予備決定 (2) ; 次いで、
 - 特定の不要なディスクリプタの削除 (3) ;
 を含むことを特徴とする、請求項 6 に記載の方法

20

【請求項 8】

特定のディスクリプタの削除 (3) が、0 . 9 5 を超えるピアソン相関係数をペアで表示するディスクリプタの組毎に、1 つのディスクリプタだけを保留することを含むことを特徴とする、請求項 7 に記載の方法。

【請求項 9】

- 異なる基準曲線を取得した後におけるいくつかの基準曲線の識別であって、前記基準曲線が、あいまいなグループと呼ばれる同じ初期グループに相当し、かつ、いくつかのサブグループに分離されたプロファイルを有する、前記識別 ; 及び
 - いくつかの可能グループへのこのあいまいなグループの分離 (4) ;
 を含むことを特徴とする、請求項 1 ~ 8 のいずれか一項に記載の方法。

30

【請求項 10】

- 異なる基準曲線を取得した後におけるいくつかの基準曲線の識別であって、前記基準曲線が、併合されたグループと呼ばれるいくつかの初期グループに相当し、そして、1 つだけのグループ内で組み合わせられたプロファイルを有する、前記識別 ; 及び
 - 1 つだけの可能グループへのこれらの組み合わせられたグループの統一 (5) ;
 を含むことを特徴とする、請求項 1 ~ 9 のいずれか一項に記載の方法。

【請求項 11】

前記測定用生物学的サンプルが所定グループに属することを決定する工程の信頼インデックスの計算をさらに含むことを特徴とする、請求項 1 ~ 10 のいずれか一項に記載の方法。

40

【請求項 12】

前記信頼インデックスの計算が、
 - 前記所定グループに属する基準曲線間同士の平均近似性の分布についての計算 ;
 - 前記所定グループに属する基準曲線と前記の少なくとも 1 つの測定曲線との平均近似性についての計算 ; 及び
 - 前記所定グループに属し、そして前記の少なくとも 1 つの測定曲線と前記所定グループに属する前記基準曲線との平均近似性よりも小さいものである、前記所定グループに属する他の基準曲線への平均近似性を有する、基準曲線のレベルの計算 ;
 を含むことを特徴とする、請求項 11 に記載の方法。

50

【請求項 13】

前記測定用生物学的サンプルが所定グループに属することを決定する工程の後に、信頼インデックスの値の関数として、何らかの任意のグループに前記測定用生物学的サンプルを割り当てることを拒絶することをさらに含むことを特徴とする、請求項 11 又は 12 に記載の方法。

【請求項 14】

前記測定用生物学的サンプルの少なくとも 1 つの DNA 融解曲線の取得が、いくつかの標的 DNA 分子を標的とするいくつかのプライマーペアの同時存在下で得られた PCR の結果の少なくとも 1 つの融解曲線の取得を含むことを特徴とする、請求項 1 ~ 13 のいずれか一項に記載の方法。

10

【請求項 15】

測定用生物学的サンプルを分類するデバイス (100) であって、そのデバイスが、

- 少なくとも 1 つの測定曲線と呼ばれる測定用生物学的サンプルの少なくとも 1 つの DNA 融解曲線を取得 (9) するために配置および/またはプログラムされた手段 (101) ;

- 前記測定用生物学的サンプルが、前記少なくとも 1 つの測定曲線に由来するディスクリプタの分析によって、異なる可能グループの中から所定グループに属することを決定 (10) するためにプログラムされた手段 (102) ;

を含み、

20

前記ディスクリプタが、各測定曲線の二次導関数の 1 つ以上の点を含み、前記の点または前記の各点が特有の温度値に関する融解曲線の二次導関数の値を含むことを特徴とする、前記デバイス。

【請求項 16】

前記ディスクリプタが、各測定曲線の二次導関数の少なくとも 30 個の点を含むことを特徴とする、請求項 15 に記載のデバイス。

【請求項 17】

前記ディスクリプタが、各測定曲線の一次導関数の 1 以上の点、及び/又は各測定曲線の 1 つ以上の点を含むことを特徴とする、請求項 15 又は 16 に記載のデバイス。

【請求項 18】

前記ディスクリプタが、各測定曲線の 1 つ以上の百分位数を含むことを特徴とする、請求項 15 ~ 17 のいずれか一項に記載のデバイス。

30

【請求項 19】

決定 (10) するためにプログラムされた手段 (102) が、ランダムフォレスト法によって決定 (10) するためにプログラムされた手段 (102) を含むことを特徴とする、請求項 15 ~ 18 のいずれか一項に記載のデバイス。

【請求項 20】

前記デバイスが、学習のために配置および/またはプログラムされた手段 (101, 102) を含み、

- 基準曲線と呼ばれる異なる DNA 融解曲線を取得 (1) するために配置および/またはプログラムされ、そして、異なる初期グループに属する異なる基準用生物学的サンプルから出発する手段 (101) ; それに続く、

40

- 前記基準曲線から前記ディスクリプタを決定 (2, 3) するために配置および/またはプログラムされた手段 (102) ; それに続く、

- 前記ランダムフォレスト法によっていくつかの木を構築 (8) するために配置および/またはプログラムされた手段 (102) を含む、前記ランダムフォレスト法によって森林を構築するために配置および/またはプログラムされた手段 (102) であって、各木の各ノードで検討された 1 つまたは複数の変数が、1 つまたは複数のディスクリプタを含み、各木の各葉が、異なる可能グループの中から 1 つのグループのみに相当する、前記手段 (102) ;

50

を含むことを特徴とする、請求項 19 に記載のデバイス。

【請求項 21】

前記ディスクリプタを決定 (2, 3) するためにプログラムされた手段 (102) が、
 - ディスクリプタを予備決定 (2) するためにプログラムされた手段 (102) ;
 - その予備決定の後に、特定の不要なディスクリプタを削除 (3) するためにプログラムされた手段 (102) ;

を含むことを特徴とする、請求項 20 に記載のデバイス。

【請求項 22】

特定のディスクリプタを削除 (3) するためにプログラムされた手段 (102) が、0.95 より大きいピアソン相関係数をペアで表示するディスクリプタの組毎に、1つのディスクリプタだけを保留するためにプログラムされた手段 (102) を含むことを特徴とする、請求項 21 に記載のデバイス。

10

【請求項 23】

- あいまいなグループと呼ばれる同じ初期グループに相当し、更にいくつかのサブグループに分離されたプロファイルを有するいくつかの基準曲線を、異なる基準曲線の取得後に識別するためにプログラムされた手段 (102) ; 及び

- このあいまいなグループをいくつかの可能グループへ分離 (4) するためにプログラムされた手段 (102) ;

を含むことを特徴とする、請求項 15 ~ 22 のいずれか一項に記載のデバイス。

【請求項 24】

20

- 併合されたグループと呼ばれるいくつかの初期グループに相当し更に1つのグループに組み合わせられたプロファイルを有するいくつかの基準曲線を、異なる基準曲線の取得後に識別するためにプログラムされた手段 (102) ; 及び

- これらの併合されたグループを1つの可能グループに統一 (5) するためにプログラムされた手段 (102) ;

を含むことを特徴とする、請求項 15 ~ 23 のいずれか一項に記載のデバイス。

【請求項 25】

前記測定用生物学的サンプルが所定グループに属することを決定するための前記工程の信頼インデックスを計算するためにプログラムされた手段 (102) を更に含むことを特徴とする、請求項 15 ~ 24 のいずれか一項に記載のデバイス。

30

【請求項 26】

前記信頼インデックスを計算するためにプログラムされた手段 (102) が、

- 前記所定グループに属する基準曲線間同士の平均近似性の分布を計算するためにプログラムされた手段 (102) ;

- 前記所定グループに属する前記基準曲線と前記の少なくとも1つの測定曲線との平均近似性を計算するためにプログラムされた手段 (102) ; 及び

- 前記所定グループに属し、更に、前記の少なくとも1つの測定曲線と前記所定グループに属する前記基準曲線との平均近似性よりも小さいものである、前記所定グループに属する他の基準曲線への平均近似性を有する、基準曲線のレベルを計算するためにプログラムされた手段 (102) ;

40

を含むことを特徴とする、請求項 25 に記載のデバイス。

【請求項 27】

前記信頼インデックス値の関数として、何らかの任意のグループに前記測定用生物学的サンプルを割り当てることを、前記測定用生物学的サンプルが所定グループに属することを決定する工程の後に、拒絶するためにプログラムされた手段 (102) を更に含むことを特徴とする、請求項 25 又は 26 に記載のデバイス。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、生物学的サンプルを分類する方法、およびそれに関連するデバイスに関する

50

。

【0002】

このようなデバイスによって、ユーザは、いくつかの可能グループの中から生物学的サンプルを分類することができる。本発明の技術分野は、生物学的分類についての分野である。

【背景技術】

【0003】

公知の文献(WO2013/166373)には、細胞サンプルまたは被験体におけるIL-6/STAT3シグナル伝達経路の調節状態を決定する方法が記載されている。細胞サンプルまたは被験体におけるIL-6/STAT3シグナル伝達経路の調節状態は、
10
発現サインの16個の遺伝子のうちの1個または複数の発現レベルに基づいて分析することができる。バイオマーカー発現は、好ましくは、SYBR Green法を用いたRT-PCRによって決定され、そして、発現データは、ランダムフォレスト法を用いて、対照サンプルと分析及び比較される。選択された変数(このケースでは16個の遺伝子)の決定は、この問題に特有であり、そして新しい問題毎に手動で実施しなければならない。

【0004】

グループ内において生物学的サンプルを分類する方法については、特に可能グループの数が多い場合には、例えば、以下のいくつかの技術的問題が生じる可能性がある：

- 分類の細かさをどのように向上させるのか、および/または
- 既知のグループに属していないサンプルをどのように扱うことができるか、
- 変数を手動で選択することなく、どのように方法を自動化するのか。

20

【0005】

本発明の目的は、これらの問題の少なくとも1つを解決することである。

【発明の開示】

【0006】

本発明の目的は、測定用生物学的サンプルを分類する方法であって、その方法が、

- 少なくとも1つの測定曲線と呼ばれる、前記測定用生物学的サンプルの少なくとも1つのDNA融解曲線の取得(好ましくは、技術的反復とも呼ばれる、前記測定用生物学的サンプルのいくつかのDNA融解曲線の取得)(通常は、各測定曲線が異なる点(point)を含み、各点が温度の関数として前記測定用サンプルのDNAのレベルまたは変性量に比例する量に相当するか、またはそのレベルまたは変性量を代表する量に相当する) ;

30

- 前記生物学的サンプルが、前記の少なくとも1つの測定曲線に由来するディスクリプタ(descriptor)の分析による、異なる可能グループの中の所定グループに属することの決定 ;

を含み、

前記ディスクリプタが、

- 各測定曲線の一次導関数の1つ以上(好ましくは、いくつか)の点、および/または、各測定曲線の二次導関数の1つ以上(好ましくは、いくつか)の点 ; および/または
- 各測定曲線の一次導関数の1つ以上(好ましくは、いくつか)の点、及び各測定曲線の二次導関数の1つ以上(好ましくは、いくつか)の点 ; および/または
- 各測定曲線の1つ以上の点 ; および/または
- 各測定曲線の1つ以上の百分位数 ;

40

を含むことを特徴とする、前記方法

によって達成される。

【0007】

前記測定用生物学的サンプルの少なくとも1つのDNA融解曲線の取得は、(例えば、いくつかの病原体に相当する)いくつかの標的DNA分子を標的とするいくつかのプライマーペアの同時存在下で得られたPCRの結果の少なくとも1つの融解曲線の取得を含む

50

ことができる。これは、「マルチプレックス」条件と呼ばれる。この実施形態は、検索のスピードアップ、例えば、全く同一である生物学的サンプル中に一緒に存在することはめったにないいくつかの病原体に有用である。例えばサンプル中に一緒に存在するいくつかの病原体のまれなケースを同定する：通常、融解曲線は、存在する病原体と同じくらいの多くの変曲点を有する；それゆえ、異なる「所定グループ」は、これらの異なる病原体の存在の異なる組み合わせを含むことができる。

【0008】

前記決定は、ランダムフォレスト法による決定を含むことができる。本発明による方法は、学習を含むことができ、その学習は、以下を含む：

- 学習前に決定された異なる既知の初期グループに属する異なる基準用生物学的サンプルから出発して、基準曲線と呼ばれる、異なるDNA融解曲線の取得すること；それに続く、

- 前記基準曲線からディスクリプタを決定すること；それに続く、
 - ランダムフォレスト法によるいくつかの木の構築を含む、ランダムフォレスト法による森林の構築であって、各木が、各ノードにおいて、閾値に関連するディスクリプタの1つの使用を含み、そして、その各木が、前記基準曲線を2つのサブセットに分離することを可能にし、各木の各葉が、異なる可能グループの中の1つのグループにのみに相当する、前記構築。

【0009】

前記ディスクリプタの決定は、以下を含むことができる：

- ディスクリプタの予備的決定；それに続く、
 - 特定の不要なディスクリプタの削除。

【0010】

特定のディスクリプタの削除は、0.95より大きいピアソン相関係数をペアで表示するディスクリプタの組毎に、1つのディスクリプタだけの保留を含むことができる。

【0011】

本発明による方法は、以下を含むことができる：

- 異なる基準曲線の取得後に（および/または、好ましくは、木の構築前に）、あいまいなグループと呼ばれる、同じ初期グループに相当し、更にいくつかのサブグループに分離されたプロファイルを有する、いくつかの基準曲線の識別すること；
 - このあいまいなグループをいくつかの可能グループに分離すること。

【0012】

本発明による方法は、以下を含むことができる：

- 異なる基準曲線の取得後に（および/または、好ましくは、木の構築後に）、併合されたグループと呼ばれる、いくつかの初期グループに相当し、そして、1つのグループだけに組み合わされたプロファイルを有する、いくつかの基準曲線を識別すること；および
 - これらの併合されたグループを1つの可能グループだけに統一すること。

【0013】

本発明による方法は、前記測定用生物学的サンプルが所定グループに属することを決定するための工程の信頼インデックスの計算をさらに含むことができる。信頼インデックスの計算は、以下を含むことができる：

- 前記所定グループに属する基準曲線間同士の平均近似性の分布の計算；
 - 少なくとも1つの測定曲線と前記所定グループに属する基準曲線との平均近似性の計算；および

- 前記所定グループに属し、更に、前記の少なくとも1つの測定曲線と前記所定グループに属する前記の他の基準曲線との平均近似性よりも小さいものである、前記所定グループに属する他の基準曲線への平均近似性を有する、基準曲線のレベルの計算。

【0014】

本発明による方法は、前記測定用生物学的サンプルが所定グループに属することを決定

10

20

30

40

50

するための工程の後に、信頼インデックス値の関数として、何らかの任意のグループに前記測定用生物学的サンプルを割り当てることを拒絶することをさらに含むことができる。

【0015】

本発明のさらに別の観点によれば、測定用生物学的サンプルを分類するためのデバイスが提案され、このデバイスは、

- 少なくとも1つの測定曲線と呼ばれる、前記測定用生物学的サンプルの少なくとも1つのDNA融解曲線の取得（好ましくは、技術的反復とも呼ばれる、前記測定用生物学的サンプルのいくつかのDNA融解曲線の取得）（通常は、好ましくは各測定曲線が異なる点を含み、各点が、温度の関数として、前記測定用サンプルのDNAのレベルまたは変性量に比例する量に相当するか、またはそのレベルまたは変性量を代表する量に相当する）

10

- 前記生物学的サンプルが、前記の少なくとも1つの測定曲線に由来するディスクリプタの分析によって、異なる可能グループの中から所定グループに属することを決定するように配置および/またはプログラムされた手段；

を含み、

前記ディスクリプタが、

- 各測定曲線の一次導関数の1つ以上（好ましくは、いくつか）の点、および/または、各測定曲線の二次導関数の1つ以上（好ましくは、いくつか）の点；および/または
- 各測定曲線の一次導関数の1つ以上（好ましくは、いくつか）の点、及び各測定曲線の二次導関数の1つ以上（好ましくは、いくつか）の点；および/または
- 各測定曲線の1つ以上の点；および/または
- 各測定曲線の1つ以上の百分位数；

20

を含むことを特徴とする。

【0016】

前記決定するために配置および/またはプログラムされた手段は、好ましくは、ランダムフォレスト法により決定するために配置および/またはプログラムされた手段を含む。本発明による前記デバイスは、学習するために配置および/またはプログラムされた手段を含むことができ、その学習は、以下を含む：

- 学習前に決定された異なる既知の初期グループに属する異なる基準用生物学的サンプルから出発して、基準曲線と呼ばれる、異なるDNA融解曲線を取得するために配置および/またはプログラムされた手段；それに続く、

30

- 前記基準曲線からディスクリプタを決定するために配置および/またはプログラムされた手段；それに続く、

- 前記ランダムフォレスト法によって木を構築するために配置および/またはプログラムされた手段であって、その手段が、前記ランダムフォレスト法によっていくつかの木を構築するために配置および/またはプログラムされた手段を含み、各木が、各ノードにおいて、閾値に関連するディスクリプタの1つの使用を含み、そして、その各木が、前記基準曲線を2つのサブセットに分離することを可能にし、各木の各葉が、異なる可能グループの中の1つのグループにのみに相当する、前記手段。

【0017】

40

ディスクリプタを決定するために配置および/またはプログラムされた手段は、以下を含むことができる：

- ディスクリプタの予備的決定のために配置および/またはプログラムされた手段；
- 予備的決定の後に、特定の不要なディスクリプタを削除するために配置および/またはプログラムされた手段。

【0018】

特定のディスクリプタを削除するために配置および/またはプログラムされた手段は、0.95より大きいピアソン相関係数をペアで表示するディスクリプタの組毎に、1つのディスクリプタだけの保留を含むために配置および/またはプログラムされた手段を含むことができる。

50

【 0 0 1 9 】

本発明による前記デバイスは、以下を含むことができる：

- 異なる基準曲線の取得後に（および/または、好ましくは、木の構築前に）、あいまいなグループと呼ばれる、同じ初期グループに相当し、更にいくつかのサブグループに分離されたプロファイルを有する、いくつかの基準曲線の識別するために配置および/またはプログラムされた手段；及び
- このあいまいなグループをいくつかの可能グループに分離するために配置および/またはプログラムされた手段。

【 0 0 2 0 】

本発明による前記デバイスは、以下を含むことができる：

- 異なる基準曲線の取得後に（および/または、好ましくは、木の構築後に）、併合されたグループと呼ばれる、いくつかの初期グループに相当し、そして、1つのグループだけに組み合わされたプロファイルを有する、いくつかの基準曲線を識別するために配置および/またはプログラムされた手段；及び
- これらの併合されたグループを1つの可能グループだけに統一するために配置および/またはプログラムされた手段。

【 0 0 2 1 】

本発明による前記は、前記測定用生物学的サンプルが所定グループに属することを決定するための工程の信頼インデックスの計算をさらに含むことができる。前記信頼インデックスを計算するために配置および/またはプログラムされた手段は、好ましくは、以下を含む：

- 前記所定グループに属する基準曲線間同士の平均近似性の分布を計算するために配置および/またはプログラムされた手段；
- 少なくとも1つの測定曲線と前記所定グループに属する基準曲線との平均近似性を計算するために配置および/またはプログラムされた手段；および
- 前記所定グループに属し、更に、前記の少なくとも1つの測定曲線と前記所定グループに属する前記の他の基準曲線との平均近似性よりも小さいものである、前記所定グループに属する他の基準曲線への平均近似性を有する、基準曲線のレベルを計算するために配置および/またはプログラムされた手段。

【 0 0 2 2 】

本発明による前記デバイスは、前記測定用生物学的サンプルが所定グループに属することを決定するための工程の後に、信頼インデックス値の関数として、何らかの任意のグループに前記測定用生物学的サンプルを割り当てることを拒絶するために配置および/またはプログラムされた手段をさらに含むことができる。

【 図面の簡単な説明 】

【 0 0 2 3 】

本発明の他の利点および特徴は、決して限定的ではない実施および実施形態の詳細な説明、および以下の添付図面を読むことにより明らかになるであろう。

- 図1は、本発明による方法の好ましい実施形態のフローチャートである。
- 図2は、本発明による方法の実施を学ぶための、いくつかのいわゆる「基準」融解曲線12を示す。

- 図3は、以下を示す：

サンプルのいくつかのいわゆる標準化された「測定」融解曲線13（「技術的反復」（“*replicats techniques*”）（それが、いくつかの可能グループ中における所定グループに属することを決定しようとする）、

これらの測定曲線の百分位数（*percentile*）14、

これらの測定曲線の一次導関数15、

これらの測定曲線の二次導関数16。

- 図4は、いくつかのサンプル（生物学的反復（*replicats biologicals*））のいくつかの基準融解曲線（技術的反復）（すべてのサンプルは、初期グルー

10

20

30

40

50

「マイコバクテリウム (Mycobacterium fortuitum)」に属する)を示す。

- 図5は、初期グループ「M. avium」または「M. szulgai」に属するいくつかのサンプル(生物学的反復)のいくつかの基準融解曲線(技術的反復)を示す。

- 図6は、3種のディスクリプタ、およびと、2つの可能グループ(グループ(groupe)1およびグループ2)との場合において作り出されたランダムフォレストの木を示す。各ノード(nœud)における質問に対する肯定的な答えは、次の枝の左側を選択することに相当する。各ノードにおける質問に対する否定的な答えは、次の枝の右側を選択することに相当する。

- 図7は、本発明による信頼度指標についての原理を図式的に示す。

- 図8は、正しいグループに割り当てられた観測または正しくないグループに割り当てられた観測を識別するための、グループ内の近似性(proximité)に基づく、信頼度指標の使用に関連するROC曲線である。

- 図9は、本発明によるデバイス100の実施形態を模式的に示す。

【発明を実施するための形態】

【0024】

これらの実施形態は決して限定的ではないので、本発明の変形は、記載または図示された他の特徴とは別に(たとえ、下記の選択が、これらの他の特性を含むフレーズ内において孤立している場合であっても)、本明細書において記載または図示された特徴の選択のみを含むと考えることができる。但し、特性についてのこの選択が、技術的な優位性を与えるか、または最新技術に対する本発明を区別するのに十分である場合に限る。この選択は、構造的な詳細を有さないか、および/または構造的な詳細の一部のみ(但し、この部分だけで、技術的な利点を与えるか、または先行技術に対して本発明を区別するのに十分なものである)を有する少なくとも1つの特性(好ましくは、機能的な特性)を含む。

【0025】

したがって、本発明による方法の好ましい実施形態が、図1～図8を参照して説明される。本明細書の以降の部分では、用語「本発明による方法」は、この実施形態のみしか意味しておらず、これは決して限定的なものではない。

【0026】

この実施形態の目的は、異なる種を区別することができることにある。マイコバクテリウム属の異なる種の間での識別は、一例として選択されており、これは決して限定的なものではない。

【0027】

この実施形態において、「生物学的サンプル」は、生物学的材料を含有する任意のタイプのサンプルであるか、または生物学的材料を含有することができる任意のタイプのサンプルに相当する。好ましくは、それは、マイコバクテリアを含有することができるサンプル、および/またはマイコバクテリアのデオキシリボ核酸(または「DNA」)若しくは微量のDNAを含有することができるサンプルである。

【0028】

融解曲線を得る原理

この実施形態において、「高感度DNA融解」と呼ばれる分子生物学技術は、「HRM」(高感度融解: high resolution melting)とも呼ばれる。このHRM技術は、二本鎖DNAから出発して実施される。HRMによる分析を行う前に、目的の突然変異が位置することができるDNAの断片を、PCR(「ポリメラーゼ連鎖反応」)により増幅する。次いで、サンプルは、PCRによって標的化及び増幅されたDNA断片の多くのコピーを含む。

【0029】

次いで、HRM分析は、変性を引き起こすために、PCRによって増幅されたDNA断片を正確に制御が行われる加熱からなる。したがって、HRM分析中において、DNAの変性をモニタすることによって、標的DNA断片の特定の融解プロファイルを決定するこ

10

20

30

40

50

とが可能になる。

【0030】

「融解プロファイル」（「融解曲線」とも呼ばれる）は、温度の関数としての1つ（または各平均）のDNA分子の変性の進行に相当している。本発明が意味する範囲内では、融解曲線は、必ずしもグラフ曲線ではなく、温度の関数としての1つ（または各平均）のDNA分子の変性中における1つの（または平均して）DNA分子のこの変性の間のこの曲線のいくつかの点の値のリストまたは表であってもよい。

【0031】

PCR反応は、例えば、以下の3つの工程によって構成されるサイクルの繰り返しを含む：

- 90 ~ 99 の温度での5 ~ 30秒間の工程、
- それに続く、58 ~ 64 の温度での5 ~ 30秒間の工程、
- それに続く、70 ~ 74 の温度での5秒 ~ 1分間の工程。

【0032】

PCRサイクルを構成する3つの工程は、それぞれ、以下の工程に相当する：

- (90 ~ 99 での) DNAの変性、
- (58 ~ 64 での) DNAのプライマーへのハイブリダイゼーション、
- (70 ~ 74 での) プライマーから開始するDNAポリメラーゼによるDNAの伸長。

【0033】

このサイクルは、通常、40 ~ 50回、好ましくは45回繰り返される。

【0034】

前記PCR反応を行う前に、好ましくは95 で10分間、前記生物学的サンプル中に含まれるDNAを最初に変性させる工程が行われる。

【0035】

この最初の変性工程は、PCRサイクルの前に実施される加熱工程である。これは、特に二本鎖DNAを完全に脱ハイブリダイズすることによって、DNAの二次構造を破壊することによって、またはDNAポリメラーゼを活性化することによって、増幅反応中のマトリックスとして役立つサンプルのDNAを調製することを可能にする。

【0036】

前記PCR反応は、例えば、少なくとも以下：

- 生物学的サンプルに含まれるDNA、
 - 3mMの割合でのMgCl₂、
 - 0.4μMの割合でのプライマー、および
 - 少なくとも1ユニットのDNAポリメラーゼ
- を含む反応混合物を使用して行われる。

【0037】

前記PCR反応の後において、増幅産物の変性を生じさせ、かつ前記増幅産物の融解プロファイルを得るための、例えば、60 ~ 100（好ましくは、65 ~ 95）の間で徐々に加熱する工程が続いて行われる。

【0038】

徐々に加熱するこの段階は、制御された方法で実施されるサンプルの加熱に相当し、この間では、温度は、例えば0.2 / 秒の上昇のように、経時的に段階的に増加する

【0039】

前記増幅産物の変性は、通常、好ましくはLC Green、LC Green Plus、ResoLight、EvaGreen、ChromoflyおよびSYTO9から選択される蛍光マーカーを用いて、モニタリングされる。

【0040】

増幅および融解の工程は、高感度融解キット LightCycler（登録商標）480 master kit（Roche）を用いて行った。反応混合物は、最終容量1

10

20

30

40

50

0 μl中の、2X Master Mix、MgCl₂、センス、およびアンチセンスプライマー、ゲノムDNA、および水からなる。増幅手順は、最初の変性、それに続く45サイクルの変性、ハイブリダイゼーション、および伸長からなる。増幅後において、融解プログラムによって、95 °Cでの1分間の加熱、40 °Cでの1分間の冷却、それに続く65 ~ 95 °Cでの0.2 °C/sの上昇速度による温度上昇の追加、及び蛍光の連続測定が行われた。各反応は、LightCycler (登録商標) 480 system (Roche) を用いて、96ウェルプレートで3回行われた。各HRM分析は、DNAマトリックスが水に置き換えられたネガティブコントロールを含んでいる。

【0041】

本発明は、有利には、遺伝子のサブセットの発現を測定するよりも融解曲線を得ることが有利であることに留意されたい。

10

【0042】

学習

図1に示すように、本発明による方法の実施形態は学習6を含み、この学習 (apprentissage) 6は、以下を含む。

- 学習前に決定された異なる既知の「初期グループ」に属する異なる基準用生物学的サンプルから出発する、基準曲線と呼ばれる異なるDNA融解曲線の取得1；次に、
- 基準曲線からのディスクリプタの決定2，3；次に
- ランダムフォレスト法によるいくつかの木の構築を含む、ランダムフォレスト法による森林の構築8（ここで、各木は、各ノードにおいて、閾値に関連するディスクリプタのうちの一つの使用を含み、そして基準曲線を2つのサブセットに分離することを可能にし；各木 (arbre) の各葉 (feuille) は、異なる可能グループの内における1つだけのグループにのみ相当する。木の分類では、葉は、常に1つだけのグループに相当し、これは、一般に、この葉に「落下する」大多数のグループの観測である。純粋な葉の特徴は、この葉に「落下する」基準用サンプルの観測のすべてが、同じグループに属することである）。

20

【0043】

「可能グループ (groupes possibles)」(このグループ中で、未知の生物学的サンプルを分類することが求められる) は、学習工程中に使用される異なる基準用生物学的サンプルの異なる「初期グループ」からなり、場合により、(例えば、グループを分離するための少なくとも1つの工程、および/またはグループを合理化するための以下に記載されるようなグループを統一するための少なくとも1つの工程によって) 変形される。好ましくは、「可能グループ」は、異なる「初期グループ」の少なくとも一部を含む。

30

【0044】

この学習段階は、アプリケーションのタイプ毎に1回実施される(新しい基準用サンプルおよび/または新しいグループを含めるために、この段階を繰り返すオプションが存在する)。これは、(最終の)可能グループを定義して、決定ルールを構築するという目的があり、以下を有する：

- 学習工程の入力として：それぞれの事前に定められた初期グループに、1つ毎に、割り当てられた標準化された学習の融解プロファイルのライブラリ。各初期グループは、いくつかの生物学的反復によって表されなければならない。各生物学的反復は、いくつかの技術的反復によって表されなければならない。

40

- 学習工程の出力として：割り当てについてのルール決定

標準化された信号の読み込み：学習に役立つための異なる「基準」DNA融解曲線の取得工程1については、上述の融解曲線を得るためのプロトコールが使用され、そして、例えばLightCycler (登録商標) 480 (Roche) に関連するソフトウェアによって提供されるような標準化の方法を応用することによって、連続する6つの実験が、マイコバクテリウム19の異なる種(または、「初期グループ」)に相当する417 HRMプロファイル(すなわち、417の基準融解曲線)を作り出すことを可能にする異

50

なる日付で行われた。各種は、いくつかの生物学的反復（種毎に2～20個の生物学的反復）のいくつかの技術的反復により表される。「生物学的反復」は、全く同一である種における異なる個体に由来する異なる生物学的サンプルのために使用される用語である。全く同一である生物学的反復の「技術的反復」は、同じ生物学的サンプルから得られた異なる融解曲線に使用される用語である。ソフトウェア入力は、機械出力でソフトウェアによって標準化された後における融解プロファイルの座標を含むテキストファイルである。

【0045】

それぞれの種における生物学的反復の分布は、表1に示されており、異なる生物学的反復についての一組の技術的反復に関連する標準化された基準曲線12は、図2に示されている。

【0046】

【表1】

種	生物学的反復の数
“M. abscessus”	3
“M. avium”	20
“M. bohemicum”	4
“M. bovis”	4
“M. chelonae”	4
“M. flavescens”	3
“M. fortuitum”	5
“M. gastri”	4
“M. gordonae”	5
“M. haemophilum”	3
“M. immunogenicum”	3
“M. kansasii”	9
“M. marinum”	4
“M. phlei”	4
“M. phocaicum”	4
“M. pseudoschottsii”	3
“M. smegmatis”	6
“M. szulgai”	4
“M. wolinskyi”	3
合計	95

これらの19種は、19の最初のグループを形成する。

【0047】

ディスクリプタの決定：ディスクリプタがその後決定される。ディスクリプタの決定2, 3は、最初に、「基準」融解曲線D(T)（温度“T”の関数としてのDNAの変性「D」（通常、%または蛍光シグナルとして））からのディスクリプタの予備決定2を含み、これは以下を考慮に入れる：

- 標準化された各融解曲線の測定点（本ケースでは、180の点）：従って、180のディスクリプタが得られ、これによって、各融解曲線を特徴付けることが可能となる；
- 百分位数（0から100までの、101の分位点（quantile））：i番目の百分位数は、標準化された融解曲線の温度（ここにおいて、i%のDNAの変性が起こっている）に使用される用語である。したがって、101の分位点に相当する101のディスクリプタが得られ（測定点の間における補間によって得られ）、そして、このディスクリプタによって、各融解曲線を特徴付けることが可能となる。

【0048】

更に、ディスクリプタの予備決定2では、厳密な意味での融解曲線は、曲線をより正確

10

20

30

40

50

に記載することが可能な得られたデータで補完される。

標準化された各融解曲線についての一次（数値）導関数

$$\frac{dD(T)}{dT}$$

。したがって、180の点（すなわち、変性レベルについての180の値）で最初に確立された融解曲線については、178個の追加のディスクリプタが得られ、これによって、各融解曲線の一次導関数を特徴付けることが可能となる。

標準化された各融解曲線についての二次（数値）導関数

$$\frac{d^2D(T)}{dT^2}$$

10

。したがって、180点（すなわち、変性レベルの180の値）で最初に確立された融解曲線については、176個の追加のディスクリプタが得られ、これによって、各融解曲線の二次導関数を特徴付けることが可能となる。

【0049】

最後に、以下が得られる：

180 + 101 + 178 + 176 = 635 各融解曲線または技術的の反復を記載するためのディスクリプタ。

【0050】

20

ディスクリプタの決定2, 3は、以下を含む：

- 上述のディスクリプタの予備決定2、それに続く、
- 特定の不要なディスクリプタの削除3。

【0051】

データの不要性は、可能グループを学習することによって有害である。さて、融解曲線の連続する値またはその導関数の連続する値との間には、非常に強い相関がある。そのため、0.95を超えるピアソン相関係数をペアで表示するディスクリプタのセット毎に、1つのディスクリプタしか保留されない。したがって、特定のディスクリプタの削除3は、0.95より大きいピアソン相関係数をペアで表示する各組のディスクリプタ毎に、1つだけのディスクリプタの保留を含む。最後に、不要なディスクリプタを削除した後に、

30

（初期の635のうちの）208個のディスクリプタが保留され、これは、以下を含む：

- 融点曲線の38の点；
- 12の百分位数；
- 一次導関数の59の点；
- 二次導関数の99の点。

【0052】

選択された各ディスクリプタの位置は、図3の縦線に示されている。

【0053】

これは、本発明による方法の利点を明確に示している。（第1および第2の、特に第2の）融解曲線の導関数は、最終的に保留されているディスクリプタの大部分を含んでいるので、生物学的サンプルが所定の可能な群に属することを決定することを可能にする識別データにおいて、非常に豊富であると考えることができる。これは、融解プロファイル同士により微細な識別において反映される。

40

【0054】

グループの合理化：

本発明による方法の実施形態は、様々な学習の複雑さを伴う多数の問題または応用に適用することができる。多かれ少なかれ遺伝的に近いグループ同士を区別する必要があるかもしれない。したがって、最初のグループがすべて融解曲線によって区別できるかどうかを知ることは、演繹的には不可能である。こうした理由で、学習中において、「グループの合理化」の工程が挿入される。これは、区別可能なまたは区別できない初期のグループ

50

の周囲を定めることを可能にする。この工程は、2つの主要な知見の結果である：

- 特定の初期グループは、異なる種のサブグループによって構成される；
- 特定の初期グループは、融解曲線によって相互に区別されない。

【0055】

図2では、次のことが理解され得る：

1) 生物学的または技術的反復は、全く同一の初期グループ内において非常に異なるプロファイルを有することができる。この現象は、2つの初期グループ、特に、図4に示す初期グループ「*M. fortuitum*」に現れる。したがって、これらの2つの初期グループのそれぞれは、学習のためにいくつかの(この場合には、2つの)可能なサブグループに分割され、学習の作業を容易にすることができる；そして、

10

2) 異なる初期グループについての全ての生物学的または技術的反復は、十分にコンパクトにすることができる。

【0056】

したがって、上記のケース1)では、本発明による方法の実施形態(より正確には、学習6)は、以下を含む：

- 異なる基準曲線を取得した後(しかしながら、好ましくは、木の構築前)における、あいまいなグループ(例えば、初期グループ“*M. fortuitum*”(図4)、つまり、初期グループ「*M. kansasii*」)と呼ばれ、そしていくつかのサブグループに分離されているプロファイルを有している同じ初期グループに相当するいくつかの基準曲線の識別。この識別は、例えば、グループ内での隔たりに関する基準を使用して実施

20

され、このために閾値が定められる；そして、

- このあいまいなグループの、いくつかの可能グループ12aおよび12bへの分離4

【0057】

同様に、上記のケース2)では、本発明による方法の実施形態(より正確には、学習6)は、以下を含む：

- 異なる基準曲線の取得後(しかしながら、好ましくは、木の構築後；グループの溶解または統一には、木の最初の構築が必要な場合がある)における、1つだけのグループに組み合わされたプロファイルを有する、併合されたグループと呼ばれる幾つかの初期グループに相当するいくつかの基準曲線の識別。この識別は、例えば、学習のエラー率を用いて

30

実施される。すなわち、1つのグループにエラーが生じた場合には、十分に分類されていない観測は、観測と組み合わせることができ(この組み合わせのために、それらの観測は、最も近似している)、融合されたグループまたは統一されたグループを形成することができる；そして、

- これらの併合されたグループの、1つだけの可能グループへの統一5。

【0058】

本発明による方法のこの実施形態では、最初のグループ「*M. szulgai*」12cと「*M. avium*」12dとは、例えば非常に近似しているが、本発明による方法の分析の微細さのせいで、それらの近似性にもかかわらず最終的には統一されない。

【0059】

最後に、表2に列挙される以下の21の可能な最終グループが得られる：

40

【表 2】

ランダムフォレスト方法で最終的に使用される 可能グループのそれぞれの名前	
“M. abscessus”	
“M. avium”	
“M. bohemicum”	
“M. bovis”	
“M. chelonae”	
“M. flavescens”	
“M. fortuitum 1”	10
“M. fortuitum 2”	
“M. gastri”	
“M. gordonae”	
“M. haemophilum”	
“M. immunogenicum”	
“M. kansasii 1”	
“M. kansasii 2”	
“M. marinum”	
“M. phlei”	20
“M. phocaicum”	
“M. pseudoschottsi”	
“M. smegmatis”	
“M. szulgai”	
“M. wolinskyi”	

【0060】

「グループの合理化」の工程は、以下に説明するランダムな森林を構築した後に、さらに、反復的に行うことができる。最初に、パラメータの最適化後、交差検証(cross-validation)に適合したランダムフォレスト法が2つのブロックに応用される。次に、悪いグループに割り当てられた生物学的反復に基準が与えられる。これらの反復毎に、新しいグループが、作成され、そして、この悪い方に割り当てられた反復と、間違っ

て割り当てられたグループの最も近似する生物学的反復とを一緒に集める。したがって、「ハイブリッド」グループが、作製され、そして、二重タグ(double etiq

uette)を含む。学習サンプルのすべての生物学的反復が正しく割り当てられるまで、この手順が再び開始される。この工程の最後に、1つまたは複数の「タグ」を有する一定数のグループが得られる。

【0061】

もちろん、この工程は、ハイブリッドグループの作成を含むことができ、このハイブリッドグループは、いくつかの初期グループを含む。しかし、多数のグループを有する予測の状況では、可能性の数を大幅に削減する可能性を有することは、非常に価値がある。なおさら、この方法ではグループ全体が別のグループと融合することが強制されないために、理論は、生物学的反復の規模で行われる。したがって、初期のグループが、別のグループに近づく生物学的反復のサブセットと異種である場合には、2つの可能な最終グループが最終的に得られる：最終グループは、初期グループとハイブリッド最終グループとの反復のみを含む。

【0062】

学習方法のパラメータを予測し決定する方法の定義：

学習6は、最終的に、ランダムフォレスト法による森林の構築8を含む。

【0063】

このランダムフォレスト法の操作は、ここでは、本発明による技術的反復/生物学的反

復としてのデータの構造に応用される。技術的な反復によって、溶解プロファイルを得るための技術的なばらつき（非常に限定されたばらつき）を考慮に入れることが可能となる。生物学的なばらつきは、学習の中心にある。なぜならば、本発明による方法の実施形態が、実際の使用条件下で直面するばらつきを反映しているからである。これは、全く同一の可能グループの個体間で観測され得る配列の相違に関連している。

【0064】

したがって、 k 個の異なる可能グループ（この例では $k = 21$ の可能グループ、表 2 参照）を区別するために、周知のランダムフォレスト法が使用される（このランダムフォレスト法がよく知られた一般的な考慮事項に関する詳細については、参考文献 [2]、[3]、[4] 参照）。（木の分類に基づいた）この方法の原理は、各木について、 n 個の開始基準融解曲線（「観測」とも称される）（この例では、 $n = 417$ の基準融解曲線）のサブセット毎に使用して、そして、そして、木のノード毎に、 p 個の開始ディスクリプタのサブセット（「ばらつき」とも称される）（この例では、 $p = 208$ 個のディスクリプタ）を使用して、いくつかの分類木を構築することである。この方法は、2つのパラメータに依存する：

- n 木 (n arbore) 構築された木の数、すなわち、最終決定のルールに含まれる木の総数、および
- m 木 (m arbore) 各木の各ノードで調査されたディスクリプタの数 ($mtry \leq p$)、すなわち、各木の各ノードの学習中に利用可能な一組のディスクリプタからランダムに選択されたディスクリプタの数。

【0065】

これらの2つのパラメータは、学習データ（基準曲線）上の2つのブロックの交差検証によって最適化工程7の間に決定される。（学習6の一部を形成する）この工程7について、そして、交差検証の使用毎に、作業は、生物学的反復のスケールで、すなわち、交差検証の各工程において実施され、生物学的反復の技術的反復は、学習ブロックにすべて割り当てられるか、または、検証ブロックのすべてに割り当てられるかのいずれかである。この制約は、実際の学習条件を最も近似して模倣するという利点がある。選択されたパラメータは、学習/テストブロック内における100個のランダム分布で得られた「十分に分類される」の平均パーセンテージを最大にするパラメータである。したがって、ペア（ n 木、 m 木）の可能な値毎に、森林は、 n 個の観測の半分に基づいてランダムフォレスト法によって（複数の木を使用して）構築され、この森林は、 n 個の観測の残りの半分でテストされる（これによって、 k 個の可能グループのそれぞれに属しているかどうかは、それが、実際にはすでにわかっている）。次いで、ペア（ n 木、 m 木）の値が選択され、平均で最良の結果をもたらす森林が構築される（この理由としては、値のペア毎に、100個の森林を含む100個のランダムな分布が実行されるからである）。 n 木 = 1000個の木、ノード毎に m 木 = 10個の変数の最適数が得られる。

【0066】

次に、ランダムフォレストの n 木 = 1000の木は、 $n = 417$ 個の観測を用いて構築される。

各木を構築するために：

- 利用可能な $n = 417$ から100個の観測（または基準曲線）がランダムに採取され、
- 各木の各ノード17において、利用可能な $p = 208$ からランダムに m 木 = 10のディスクリプタが採取され、そして、採取された m 木 = 10のディスクリプタから、Giniインデックスと一緒に最適化するディスクリプタおよび関連する閾値が選択される。各岐の端18または純粋である「葉」18に必要な多くのノードが存在する（すなわち、そこに「落ちる」観測は、すべて同じ初期グループに属する。）

【0067】

図6は、3つのディスクリプタ、および（ $p = 3$ ）と2つの可能グループ（ $k =$

2、グループ1及びグループ2)とを有する場合における、ランダムフォレストの木の例を示している。この木が、図2における本発明による方法の実施形態の場合に相当しない場合(k=21の可能グループ、p=208のディスクリプタ、大きすぎて図示できない木)であっても、この木は、本発明による方法のこの実施形態において構築された木の原理を完全に示している。

【0068】

各ノード17は、通常、ディスクリプタに関連してもたらされる質問に相当する：このディスクリプタは、しきい値未満(または、それ以下)の値を有しているかどうか。

【0069】

例えば：

- 融解曲線の点の一部を形成するディスクリプタ(38個のディスクリプタ)毎に：このディスクリプタに特有の温度の値について、変成率(pourcentage de denaturation)(または、融解曲線の他の任意の単位)が所定の閾値未満(または、以下)の値を有するかどうか。

- 百分位数の一部を形成するディスクリプタ(12個のディスクリプタ)毎に：このディスクリプタに特有の変成率(または、融解曲線の他の任意の単位)の値について、温度が所定の閾値未満(または、以下)の値を有するかどうか。

- 一次導関数の点の一部を形成するディスクリプタ(59個のディスクリプタ)毎に：このディスクリプタに特有の温度の値について、融解曲線の一次導関数が所定の閾値未満(または、以下)の値を有するかどうか。

- 二次導関数の点の一部を形成するディスクリプタ(99個のディスクリプタ)毎に：このディスクリプタに特有の温度の値について、融解曲線の二次導関数が所定の閾値未満(または、以下)の値を有するかどうか。

各葉18は、k個の可能な最終グループのうちの1つに相当する。

【0070】

信頼インデックスの学習：

構築することによって、ランダムフォレスト法は、木の数を調べること(この木の数において、2つの観測が同じ葉の中に「落ちる」)によって、観測間同士の近似性を計算することを可能にする。この近似性は、予測の信頼指標を計算するために、従って、可能グループの1つに観測を割り当てることを場合により拒否するために使用される。

【0071】

したがって、ランダムフォレストの構築後、学習工程6の間において、全く同一の可能な同じグループに属する学習ライブラリの生物学的反復のすべてのペアについて、ペアの近似性の分布が計算される。生物学的反復間同士の近似性は、その技術的反復間同士で計算された近似性の最小値(いわゆる、完全リンク法)によって定められる。この分布は、カーネル法でスムーズにすることができる。この操作は、可能グループ毎に繰り返され、したがって、各グループに特有のグループ内での距離の分布が得られる。

【0072】

予測：未知の、いわゆる「測定用」生物学的サンプルについての分類

予測は、本発明による方法の実施形態の現在の工程である。(特に、グループの合理化の工程の後の)学習の出力として得られる可能な最終グループ(「クラス」とも呼ばれる)のうちの1つに対する割り当てを得るために、生物学的サンプルに決定ルールを応用することを目的としており、そして、この割り当ては、信頼指標と一致している。

したがって、結果は、以下を伴うことになる：

- 予測工程の入力として：全く同一の「測定用」生物学的サンプルから得られた融解プロファイル(技術的反復)のセット。

- 予測工程の出力として：「測定用」サンプルの信頼指標への割り当てに関する決定、または、場合により、可能グループのいずれかへの割り当ての拒絶。

【0073】

実際、本発明による方法の実施形態の目的は、融解曲線による「測定用」サンプルのデ

10

20

30

40

50

ィスクリプション (d e s c r i p t i o n) に基づいて、学習中に定められた k 個の可能な最終グループの 1 つにこの個体を割り当てること (監視方法)、および提案された決定に信頼指標を割り当てることのできるかどうかを決定することである。

【 0 0 7 4 】

したがって、本発明による方法の実施形態は、少なくとも 1 つの測定曲線と呼ばれる、測定用生物学的サンプルの DNA 融解の標準化曲線 (上記参照) 少なくとも 1 つの取得 (9) を含んでおり、各測定曲線は異なる点を含み、各点は、温度の関数としての測定用サンプルの DNA の変性についてのレベル (通常、%) 若しくは量に比例 (例えば、蛍光シグナル) する量に相当するか、または代表する量に相当する。この取得は、(実験室での) PCR および融解曲線それ自体、および / またはこの融解曲線のデータ (例えば、コンピュータデータ) の単純なダウンロードを実施することを含むことができる。

10

【 0 0 7 5 】

場合により、この融解曲線についての PCR は、いくつかの標的 DNA 分子を標的とするいくつかのプライマーペアを同時に存在させて行うことができる。続いて、「マルチプレックス」条件を言及する。

【 0 0 7 6 】

本発明による方法の実施形態は、更に、測定用生物学的サンプルが、k 個の異なる最終の可能グループ中において、所定グループに属するというを、学習段階中に構築された木の森林に基づくランダムフォレスト法によって決定 10 することを含む。この決定は、学習段階中に構築された木の森林に基づくランダムフォレスト法による、少なくとも 1 つの測定曲線から生じるディスクリプタの分析を含む。このディスクリプタは、以下を含む：

20

各測定曲線の 1 つ以上の点、および / または

各測定曲線の変性レベルの 101 の百分位数の一部または全部、および / または

各測定曲線の一次導関数の 1 つ以上の点 (通常、少なくとも 30 個の点)、および / または

各測定曲線の二次導関数の 1 つ以上の点 (通常、少なくとも 30 個の点)

【 0 0 7 7 】

測定用生物学的サンプルの技術的反復は、ランダムフォレストを独立して行うことになり、更に、可能グループが、技術的反復のそれぞれに割り当てられる。デフォルトによって、測定用生物学的サンプルは、技術的反復毎に予測されたグループのうちの多数のグループに割り当てられる。複数のグループの場合、信頼指標を使用して決定することができる。

30

ランダムフォレストは確率的方法 (いくつかの応用は、種々の結果を与えることができる) であるので、この方法は、生物学的サンプルの割り当てを予測するために数回 (この実施形態では、3 回) 応用される。

【 0 0 7 8 】

各ディスクリプタの位置は、図 3 の縦軸に示されている。

図 3 は、以下を示している：

- 左上には、全く同一の測定用サンプルの 3 つの測定融解曲線 13 (3 つの技術的反復) (それが、k 個の可能グループのうち 1 つに属するかを決定する) ；
- 右上には、3 つの技術的反復 13 に相当する百分位数を示す 3 つの曲線 14 ；
- 左下には、3 つの測定融解曲線 13 の 3 つの一次導関数 15 ；及び
- 右下には、3 つの測定融解曲線 13 の 3 つの二次導関数 16 。

40

【 0 0 7 9 】

信頼指標の計算：

本発明による方法の実施形態の品質は、初期学習ライブラリの品質によって決定される。生物学的なばらつきにおいてこれが豊かになればなるほど、学習は、ますます、新しいサンプルの多様性に対して正確で一般化することができるようになる。しかしながら、学習ライブラリの品質にかかわらず、新しいサンプルの予測中において、それに完全に異な

50

るサンプルに遭遇する可能性が依然としてある。この場合でも、従来の学習方法は、それに最も近い可能グループに新しいサンプルを割り当てることによって予測を供給しているであろう。本発明による方法の実施形態は、何らかの任意の可能グループに新しいサンプルを割り当てることを拒否することができなければならない。

【0080】

このために、本発明による方法の実施形態は、所定グループに測定用生物学的サンプルが属することを決定する工程の信頼指標の計算を含む。

【0081】

信頼指標を計算するこの工程には、2つの目的がある：

- 作成された予測の信頼性を定量化すること；
- 生物学的反復が、任意の可能グループに割り当てられていないことを予測することを可能にすること。

10

【0082】

ランダムフォレストには、観測間同士の近似性の測定値を提供するという利点がある。ランダムフォレスト法におけるこの周知の「近似性」の概念の詳細については、参考文献[3]および[4]を参照することができる。

【0083】

これらの測定は、信頼インデックスを提供するために使用される。実際、可能グループの観測に予測される観測（その観測が、可能グループに割り当てられている）が近い場合には、予測される観測が可能グループの観測（その観測が、可能グループに割り当てられている）から遠い場合よりも、分類の品質の方が潜在的に優れている。この原則は、信頼インデックスを構築するために使用される。

20

【0084】

学習データは、全く同一の可能グループの基準融解曲線の平均近似性の分布を計算するための基礎として採用された。次に、測定用生物学的サンプルが可能グループに割り当てられると、このグループの生物学的反復へのその平均近似性が計算され、そして、この可能グループの基準融解曲線の近似性と比較される。次いで、基準融解曲線のパーセンテージ（この基準融解曲線の近似性は、予測される測定融解曲線の近似性よりも小さい）を計算することが可能である。このパーセンテージは、予測されたグループに属するかの可能性の推定であり、信頼インデックスとして使用される。

30

【0085】

図7は、2つの可能グループ22, 23（基準融解曲線（x：バツ）32および33は、それぞれ、このグループに属する）を有するこのインデックスの原理を示している。

【0086】

その（または、各々の、または平均の）測定曲線13a（：三角形）が、

- グループ23よりもグループ22に近く、そして、
- グループ22の融解曲線32の間で観察される近似性と一致するような、グループ22の融解曲線32への近似性を有してしている場合には、曲線13aに相当するサンプルは、グループ22に属すると考えられ、そして、本発明による方法の実施形態は、決定されたグループが確かにグループ22であることを確認する。

40

【0087】

その（または、各々の、または平均の）測定曲線13b（：三角形）が、

- グループ23よりもグループ22に近いが、
- グループ22の融解曲線32の間で観察される近似性よりも非常に低い、グループ22の融解曲線32への近似性を有している場合には、曲線13bに相当するサンプルは、グループ22またはグループ23のいずれにも属さず、そして、好ましくは、本発明による方法の実施形態は、測定用生物学的サンプルを、所定グループ22に（場合により、何らかの任意のグループに）割り当てることを拒絶することを含む。

【0088】

学習段階において、全く同一のグループに属する学習ライブラリの生物学的反復のすべ

50

てのペアについての近似性の分布を計算した。

【0089】

予測工程において、任意の新しい観測（すなわち、任意の新しい「測定」融解曲線）について、グループの基準用生物学的反復へのその平均近似性（その平均近似性は、グループに割り当てられていた）が、同じ方法によって計算される。続いて、以前に得られた全体分布は、このグループに属する可能性の推定を計算するために使用される。

【0090】

新しい観測がこの工程を通過する場合には、この可能グループに属する可能性は、予測されたグループと同時にユーザに提供される。

【0091】

もちろん、この最後の工程を応用する可能性は、学習ライブラリの十分なサイズに依存する。

【0092】

図8は、この指標に関連するROC (Receiver Operating Characteristic) を示す ($AUC = 0.9988$ 、 AUC は、ROC曲線の面積、または「その曲線下の面積」である)。

【0093】

この図8は優れた品質のものである。実際、図7を参照すると、 0.14 に等しい閾値が信頼インデックスのために使用される場合には、(ループ19で表されている)この閾値以下の未満のインデックスを有する全ての観測は、分類されていない観測であり、1つのよく分類された観測だけがこの閾値を超えるインデックスを有する。このインデックスのおかげで、よく分類された観測は非常に満足に識別することができる。

【0094】

(ランダムフォレストの異なる応用の結果として、または異なる技術的反復の矛盾した結果として)いくつかの可能グループへの割り当ての場合、または信頼性の低いインデックスの場合には、生物学的反復に関して(森林の木のスベテから)少なくとも一回予測される可能グループのセットへの近似性を再度計算することができる。これらの可能グループのうちの一つが、 0.14 の閾値を超えるインデックスを示す場合には、この測定曲線は、最大インデックス値を有する可能グループにおそらく属するものとして、タグ付けすることができる。

【0095】

このルールを基準曲線に応用してその有効性をテストすると、融解曲線は、「回収 (recovered)」であり、この「回収」では、ほとんどが間違った種に割り当てられていたが、正しい種が少なくとも1回は予測されており、計算された信頼インデックスが 0.14 より大きかった。

【0096】

最後に、信頼インデックスは、測定融解曲線に同じ回数割り当てられた2つの可能グループの間で決定するために使用することができる。

したがって、要約すると、信頼インデックスの計算は、以下を含む：

- 所定グループに属する基準曲線の間における平均近似の分布を計算する工程、
- 所定グループに属する基準曲線と測定用生物学的サンプルの平均近似を計算する工程、
- 所定グループに属し、そして、(少なくとも一つの基準曲線と所定グループに属する基準曲線との平均近似より小さくなる)所定グループに属する他の基準曲線への平均近似性を有する、基準曲線のレベルを計算する工程。

【0097】

測定用生物学的サンプルが所定グループに属すると決定した工程の後に、本発明による方法の実施は、以下を含む：

(結果を提供する工程、通常は、スクリーン上に表示されるか、またはコンピュータメモリまたは電子メモリに格納される工程11として、)拒絶するか、または信頼インデックスの値の関数として何らかの任意の可能グループに測定用生物学的サンプルを割り当てな

10

20

30

40

50

い、より正確には、

- 信頼インデックスが閾値未満である場合には、測定用生物学的サンプルを何らかの任意の可能グループに割り当ててを拒否すること、または
- 信頼インデックスが閾値を上回る場合には、測定用生物学的サンプルを可能最終グループの1つに割り当ててを承認すること。

【0098】

ランダムフォレストの生の結果と信頼インデックスの使用とを組み合わせることにより、観測の95.74%が、正しく割り当てられる。観測の残りの4.26%に関して、それらは、本発明による方法の実施形態によって、それらの割り当てに関して疑わしいものとして明確に識別される。

10

【0099】

図3の曲線の場合には、サンプルは、グループ「アブセサス (abscessus)」に属するものとして識別される。

【0100】

本発明による方法のこの実施形態では、以下の各工程が、純粹に抽象的に、または純粹に知的に実行されるのではないが、(一般に、本発明による方法のこの実施形態の他のすべての工程と同様に、)技術的手段の使用を伴っている：

- 取得1；
- ディスクリプタ2, 3の決定；
- グループ4, 5の合理化；
- 学習6；
- パラメータの最適化7；
- ランダムフォレスト8の構築；
- 取得9；
- ディスクリプタの分析、ランダムフォレストの応用、いくつかの可能グループのうちからの所定グループの決定、信頼インデックスの計算、測定用生物学的サンプルを所定グループに割り当ててを確認する決定、および測定用生物学的サンプルを何らかの任意のグループへ割り当ててを拒否する工程10；
- 結果の提供または表示11。

20

【0101】

図9は、本発明による方法のこの実施形態を実施するために配置され、プログラムされた、本発明によるデバイス100の実施形態を模式的に示す。

30

【0102】

デバイス100は、以下の各工程を実施するために構成され、プログラム化された手段102を含む：

- ディスクリプタ2, 3の決定；
- グループ4, 5の合理化；
- パラメータの最適化7；
- ランダムフォレスト8の構築；
- ディスクリプタの分析、ランダムフォレストの応用、いくつかの可能グループのうちからの所定グループの決定、信頼インデックスの計算、および測定生体サンプルを所定グループに割り当ててを確認する決定、または生物学的サンプルを何らかの任意のグループに割り当ててを拒否する工程10；
- 結果の提供または表示11。

40

【0103】

デバイス100は、以下の工程を実現するように配置および/またはプログラムするための手段101および102を含む：

- 学習工程6、より正確には基準曲線を取得する工程1、
- 少なくとも1つの測定曲線を取得する工程9。

【0104】

50

手段102は、コンピュータ、および/または中央または計算ユニット、および/または(好ましくは、専用の)アナログ電子回路、および/または(好ましくは、専用の)デジタル電子回路、および/または(好ましくは、専用の)マイクロプロセッサ、および/またはソフトウェアを含む。これらの手段102は、好ましくは、結果を提供または表示する工程11のために、スクリーン若しくは印刷手段、またはデータをエクスポートする手段をさらに含む。

【0105】

手段101は、PCR機器を含み、および/または、変形例によれば、DNA融解曲線をロードおよび読み取りのために配置およびプログラムされたコンピュータ手段(USBポート、SDカードリーダー、コンピュータネットワークへの接続などと組み合わせたソフトウェア)を含む。したがって、これらの手段101は、手段102に接続されるか、または手段102の一部を形成する。

10

【0106】

もちろん、本発明は、上記の実施例に限定されず、本発明の範囲を逸脱することなく、これらの実施例に対して多くの調整を加えることができる。

【0107】

例えば、この概念の別の証明が、*Coxiella burnetii*、*Chlamydochloa spp*、*Neospora caninum*、*Toxoplasma gondii*、および*Anaplasma*を含めた他の微生物についても行われ、同様の成功を収めた。開発された方法によって、マルチプレックス条件下で(すなわち、上記の(5つの)病原体すべての標的DNA分子を標的とするすべてのプライマーペアを同時に存在させることでPCRを増幅させることを介して)異なる病原体を同定することを可能にすることは、実験によって証明されている。

20

【0108】

一般に、本発明は、任意の生物学的サンプル、特に、人間、動物、植物、ウイルス、細菌、古細菌、菌類、酵母、ウイロイド、真核生物、または原生動物に応用可能である。

【0109】

もちろん、本発明の種々の特性、形態、変形例、および実施形態は、互いに組み合わせることができる。

引用文献：

30

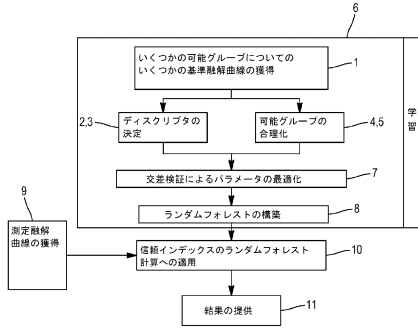
[1] : WO2013 / 166373

[2] : Random Forests , Leo Breiman, paru dans Machine learning , 45, 5-32, 2001

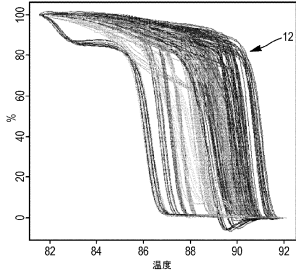
[3] : Classification and Regression by RandomForest , Andy Liaw and Matthew Wiener, R News, vol . 2/3, December 2002, ISSN 1609-3631

[4] : Manual on setting up, using, and understanding random Forests V3.1 , Leo Breiman, 2002

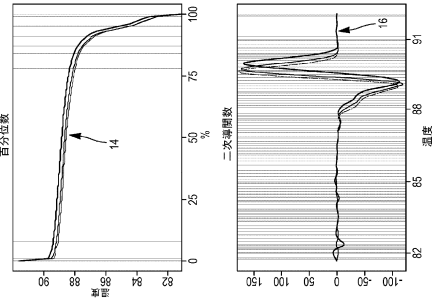
【図1】



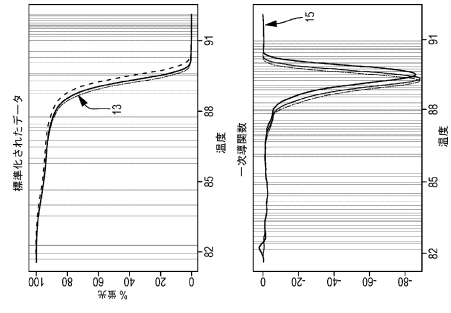
【図2】



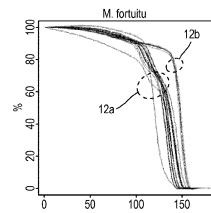
【図3】



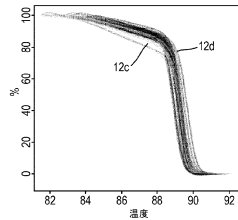
【図4】



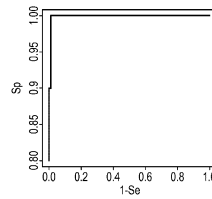
【図5】



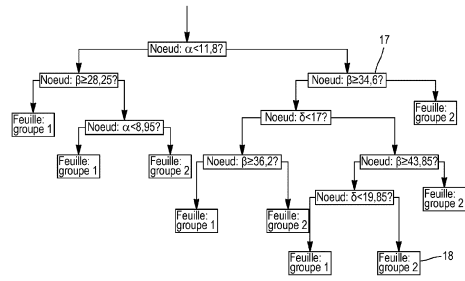
【図6】



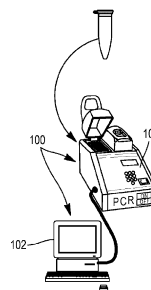
【図8】



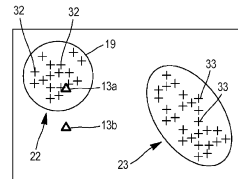
【図6】



【図9】



【図7】



フロントページの続き

(73)特許権者 518261283

アンスティテュ ドゥ ルシェルシュ プール ル デヴロップマン
INSTITUT DE RECHERCHE POUR LE DEVELOPPEMENT
フランス共和国, 13002 マルセイユ 2, プールヴァール ダンケルク 44, イムブル
ル セクスタン

(73)特許権者 509025832

サントル ナシオナル ドゥ ラ ルシェルシェ シアンティフィク
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
フランス共和国, エフ - 75016 パリ, リュ ミシェル - アンジュ 3

(74)代理人 100139594

弁理士 山口 健次郎

(74)代理人 100185915

弁理士 長山 弘典

(74)代理人 100090251

弁理士 森田 憲一

(72)発明者 アヴァール, ジャン - クリストフ

フランス共和国, 34980 モンフェリエ シュル レ, リュ デュ フール 5

(72)発明者 レーヌ, クリステル

フランス共和国, 34070 モンペリエ, リュ エドゥアール セット, 8, レジダンス エド
ゥアール セット バティマン A

審査官 田中 晴絵

(56)参考文献 特表2012-519002(JP, A)

米国特許出願公開第2013/0218476(US, A1)

特表2013-513387(JP, A)

国際公開第2014/120821(WO, A1)

特開2015-144606(JP, A)

The R Journal, [online], 2013年, インターネット<URL:https://www.researchgate.net/p
ublication/256121929_Surface_Melting_Curve_Analysis_with_R>

(58)調査した分野(Int.Cl., DB名)

C12Q 1/68 - 1/6897

C12M 1/00 - 3/10

C12N 15/00 - 15/90

JSTPlus/JMEDPlus/JST7580(JDreamIII)

CAplus/MEDLINE/EMBASE/BIOSIS/WPIDS(STN)