

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6327473号
(P6327473)

(45) 発行日 平成30年5月23日 (2018. 5. 23)

(24) 登録日 平成30年4月27日 (2018. 4. 27)

(51) Int. Cl.		F I	
G 0 6 F	19/22	(2011. 01)	G O 6 F
C 1 2 Q	1/68	(2018. 01)	C 1 2 Q
C 1 2 N	15/09	(2006. 01)	C 1 2 N
			C 1 2 Q
			19/22
			1/68
			15/00
			1/68
			Z N A
			A
			A
			Z

請求項の数 9 (全 23 頁)

(21) 出願番号	特願2014-558586 (P2014-558586)	(73) 特許権者	510192802
(86) (22) 出願日	平成26年1月22日 (2014. 1. 22)		国立研究開発法人国立国際医療研究センター
(86) 国際出願番号	PCT/JP2014/051208		東京都新宿区戸山1丁目2番地1号
(87) 国際公開番号	W02014/115747	(73) 特許権者	504300088
(87) 国際公開日	平成26年7月31日 (2014. 7. 31)		国立大学法人帯広畜産大学
審査請求日	平成29年1月18日 (2017. 1. 18)		北海道帯広市稲田町西2線1番地
(31) 優先権主張番号	特願2013-12546 (P2013-12546)	(73) 特許権者	000135036
(32) 優先日	平成25年1月25日 (2013. 1. 25)		ニプロ株式会社
(33) 優先権主張国	日本国 (JP)		大阪府大阪市北区本庄西3丁目9番3号
(31) 優先権主張番号	特願2013-136645 (P2013-136645)	(74) 代理人	100130111
(32) 優先日	平成25年6月28日 (2013. 6. 28)		弁理士 新保 斉
(33) 優先権主張国	日本国 (JP)	(72) 発明者	秋山 徹
			東京都新宿区戸山1-21-1 国立国際医療研究センター内

最終頁に続く

(54) 【発明の名称】 結核菌の菌株の同定方法及び遺伝子変異の検出方法

(57) 【特許請求の範囲】

【請求項 1】

コンピューターによるサンプル中の結核菌の菌株を同定する方法であって、

(1) コンピューターが入力されたサンプル中の結核菌のゲノム情報を記憶する工程と

、
(2) コンピューターが記憶している結核菌のコンセンサスゲノム情報に含まれる特定の菌株の遺伝子配列と、サンプル中の結核菌の遺伝子配列についてアライメント処理を行う工程と、

(3) サンプル中の結核菌のゲノム情報における特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致することで、特定の菌株であることを同定する工程と、

(4) サンプル中の結核菌の菌株の同定結果を表示する工程と、

を具備し、

前記結核菌のコンセンサスゲノム情報は、結核菌H37Rv株、結核菌KZN605株、結核菌RGT B423株、結核菌RGTB327株、結核菌 Erdman株、結核菌CTRI-2株、結核菌CDC1551株、結核菌CCDC5180株、結核菌CCDC5079株、結核菌KZN4207株、結核菌KZN1435株、結核菌F11株、結核菌H37Ra株、類縁菌CIPT 140010059株、類縁菌GM041182株、類縁菌BCG str. Mexico株、類縁菌BCG str. Moreau RDJ株、類縁菌BCG str. Tokyo 172株、類縁菌BCG Pasteur 117 3P2株の一部又は全部のゲノム情報に対して、それぞれの遺伝子配列の逆位若しくは転移を補正して、アライメント処理した遺伝子配列について、共通する遺伝子配列領域を共通

遺伝子配列領域として、菌株間で異なる遺伝子配列領域を菌株の特有の遺伝子配列領域としてコンセンサスゲノム情報に含ませることを特徴とするコンピューターによるサンプル中の結核菌の菌株を同定する方法。

【請求項 2】

コンピューターによるサンプル中の結核菌の菌株及び変異を同定する方法であって、

(1) コンピューターが入力されたサンプル中の結核菌のゲノム情報を記憶する工程と

(2) コンピューターが記憶している結核菌のコンセンサスゲノム情報に含まれる特定の菌株の遺伝子配列と、サンプル中の結核菌の遺伝子配列についてアライメント処理を行う工程と、

(3) サンプル中の結核菌のゲノム情報における特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致することで、特定の菌株であることを同定する工程と、

(4) 同定された菌株に対応する結核菌コンセンサスゲノム情報に含まれる遺伝子配列を基準遺伝子配列とし、基準遺伝子配列とサンプル中の結核菌の遺伝子配列とを比較処理して、該基準遺伝子配列と異なる遺伝子情報を遺伝子変異情報として同定する工程と、

(5) サンプル中の結核菌の菌株の同定結果、及び遺伝子変異情報の結果を表示する工程と、

を具備し、

前記結核菌のコンセンサスゲノム情報は、結核菌H37Rv株、結核菌KZN605株、結核菌RGT B423株、結核菌RGTB327株、結核菌 Erdman株、結核菌CTRI-2株、結核菌CDC1551株、結核菌CCDC5180株、結核菌CCDC5079株、結核菌KZN4207株、結核菌KZN1435株、結核菌F11株、結核菌H37Ra株、類縁菌CIPT 140010059株、類縁菌GM041182株、類縁菌BCG str. Mexico株、類縁菌BCG str. Moreau RDJ株、類縁菌BCG str. Tokyo 172株、類縁菌BCG Pasteur 117 3P2株の一部又は全部のゲノム情報に対して、それぞれの遺伝子配列の逆位若しくは転移を補正して、アライメント処理した遺伝子配列について、共通する遺伝子配列領域を共通遺伝子配列領域として、菌株間で異なる遺伝子配列領域を菌株の特有の遺伝子配列領域としてコンセンサスゲノム情報に含ませることを特徴とするコンピューターによるサンプル中の結核菌の菌株及びその遺伝子変異を同定する方法。

【請求項 3】

結核菌のコンセンサスゲノム情報を構築における遺伝子配列の逆位の補正が、結核菌KZN605株の遺伝子配列932051番目と932052番目の間、3479594番目と3459595番目の間、結核菌KZN1435株の遺伝子配列931985番目と931986番目の間、3479865番目と3479866番目の間、結核菌KZN4207株の遺伝子配列932007番目と932008番目の間、3476553番目と3476554番目の間の遺伝子配列の一部又は全部を対象とすることを特徴とする請求項 1 又は 2 のいずれかに記載のコンピューターによる同定方法。

【請求項 4】

結核菌のコンセンサスゲノム情報が遺伝子配列 1 であることを特徴とする請求項 1、2 又は 3 のいずれかに記載のコンピューターによる同定方法。

【請求項 5】

遺伝子変異情報の結果を表示する工程において、該遺伝子変異を含む遺伝子配列領域に関する情報をも表示できることを特徴とする、請求項 2、3 又は 4 のいずれかに記載のコンピューターによる同定方法。

【請求項 6】

前記の基準遺伝子配列と異なる遺伝子情報を遺伝子変異情報として同定する工程において、コンピューターが記憶している結核菌の薬剤応答変異遺伝子情報と比較処理して、該遺伝子変異情報が特定の薬剤応答変異遺伝子情報である場合にはこれを同定するものであって、遺伝子変異情報の結果を表示する工程において、遺伝子変異情報と薬剤応答変異遺伝子情報の結果を表示することを特徴とする、請求項 2、3、4 又は 5 のいずれかに記載

のコンピューターによる同定方法。

【請求項 7】

薬剤応答変異遺伝子を有する配列を含んだ結核菌のコンセンサスゲノム情報を用いることを特徴とする、請求項 2、3、4、又は 5 のいずれかに記載のコンピューターによる同定方法。

【請求項 8】

結核菌のコンセンサスゲノム情報を生産する方法であって、結核菌H37Rv株、結核菌KZN 605株、結核菌RGTB423株、結核菌RGTB327株、結核菌Erdman株、結核菌CTRI-2株、結核菌CDC1551株、結核菌CCDC5180株、結核菌CCDC5079株、結核菌KZN 4207株、結核菌KZN1435株、結核菌F11株、結核菌H37Ra株、類縁菌CIPT 140010059株、類縁菌GM041182株、類縁菌BCG str. Mexico株、類縁菌BCG str. Moreau RDJ株、類縁菌BCG str. Tokyo 172株、類縁菌BCG Pasteur 1173P2株の一部又は全部のゲノム情報に対して、それぞれの遺伝子配列の逆位若しくは転移を補正して、アライメント処理した遺伝子配列について、共通する遺伝子配列領域を共通遺伝子配列領域として、菌株間で異なる遺伝子配列領域を菌株の特有の遺伝子配列領域としてコンセンサスゲノム情報に含ませることを特徴とする、結核菌のコンセンサスゲノム情報を生産する方法。

【請求項 9】

サンプル中の結核菌の菌株及び変異を同定するコンピューターシステムであって、

(1) コンピューターが入力されたサンプル中の結核菌のゲノム情報を記憶する結核菌サンプル情報記憶部と、

(2) コンピューターが結核菌コンセンサスゲノム情報を記憶するコンセンサスゲノム情報記憶部と、

(3) サンプル中の結核菌のゲノム情報に含まれる遺伝子配列と結核菌コンセンサスゲノム情報に含まれる遺伝子配列についてアライメント処理するアライメント処理部と、

(4) サンプル中の結核菌の特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致することで、特定の菌株であることを同定する菌株同定処理部と、

(5) 前記菌株同定処理部で同定された特定菌株に対応する結核菌コンセンサスゲノム情報に含まれる特定菌株の基準遺伝子配列を同定する基準遺伝子配列同定処理部と、

(6) 前記基準遺伝子配列とサンプル中の結核菌の遺伝子配列とを比較処理して、前記基準遺伝子配列と異なる遺伝子情報を遺伝子変異情報として同定する変異遺伝子情報検出部と、

(7) サンプル中の結核菌の菌株の同定結果、及び遺伝子変異情報を表示する情報出力部と、

を具備し、

前記結核菌のコンセンサスゲノム情報は、結核菌H37Rv株、結核菌KZN605株、結核菌RGTB423株、結核菌RGTB327株、結核菌 Erdman株、結核菌CTRI-2株、結核菌CDC1551株、結核菌CCDC5180株、結核菌CCDC5079株、結核菌KZN4207株、結核菌KZN1435株、結核菌F11株、結核菌H37Ra株、類縁菌CIPT 140010059株、類縁菌GM041182株、類縁菌BCG str. Mexico株、類縁菌BCG str. Moreau RDJ株、類縁菌BCG str. Tokyo 172株、類縁菌BCG Pasteur 1173P2株の一部又は全部のゲノム情報に対して、それぞれの遺伝子配列の逆位若しくは転移を補正して、アライメント処理した遺伝子配列について、共通する遺伝子配列領域を共通遺伝子配列領域として、菌株間で異なる遺伝子配列領域を菌株の特有の遺伝子配列領域としてコンセンサスゲノム情報に含ませることを特徴とするサンプル中の結核菌の菌株及びその遺伝子変異を同定するコンピューターシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、結核菌の菌株の同定及び結核菌ゲノムの遺伝子変異の同定・検出方法、これらの方法に用いる結核菌のコンセンサスゲノム情報の製造方法に関する。

【背景技術】

【0002】

結核は単一の感染症中では年間の死亡数が最大の感染疾患である。WHOは、年間に約1千万人が結核に感染し、2.5百万人がこの疾患で死亡していると推定している。また、Mycobacterium tuberculosis（結核菌）の多剤耐性結核菌株の出現など事態は悪化している。結核の制御において、患者の迅速な診断と感染源の同定が重要である。M.tuberculosis株のタイピングが爆発的流行に関する研究に極めて有用であることは証明され、様々な疫学的問題の解決に応用されている。

【0003】

現在、結核菌の検出および同定は、患者から採取したサンプルを培養し、得られた培養菌を用いて、VNTR、スポリゴタイピングやIS6110-RFLPにより実施されている。VNTRは、結核菌の遺伝子中に別個に複数存在する反復領域のそれぞれの反復数を計数し、各領域の数を比較して菌株特定を行うものである。

10

【0004】

スポリゴタイピングは、結核菌ゲノムに特異的に存在する36bpのダイレクトリピートに挟まれたスペーサ配列を含む領域の解析を行うものである。スペーサ配列は菌株によりその数（最大で50個程度）と配列が異なっており、PCR法で同領域を増幅後、スペーサの配列を利用してハイブリダイゼーションを行う。その結果、たとえば、ある菌株は、スペーサ1と2と3と5を保有、他の菌株はスペーサ2と4を保有、などという方法で菌株を特定するものである。

20

【0005】

IS6110-RFLPは、結核菌ゲノム中に最大で10個程度存在するほぼ同一の配列を持つ部分を検出するものである。結核菌ゲノムをPvuIIという制限酵素で切断後、電気泳動にて各断片を分離してサザントランスファーを行い、IS6110の配列を検証する。菌株毎にIS6110の数と、それらが含まれる断片の長さが異なるため、菌株の特定に使用されている。

【0006】

しかしながら、そのような方法は、培養に時間を相当に要することや、それらの方法自体が煩雑であったりする。また、結核菌ゲノムの遺伝子配列情報を直接に利用するものではないため、その同定精度も問題となる。

30

【0007】

一方、世界的にゲノム解析が進むなかで、結核菌においてもその全ゲノム解析が終了している（非特許文献1を参照）。そのような中で、結核菌の全遺伝子をPCR法で用いて増幅し、このPCR産物を固着化したDNAチップを作成することも提案されている（非特許文献2および非特許文献3）。このようなチップを用いて、結核菌等のゲノム構造の違いを検出する方法である。

【0008】

例えば、結核菌に特異的なオリゴヌクレオチドを利用するプローブ固相化基体を用いる結核菌の同定方法が開示されている（特許文献1）。また、結核菌の同定のために特異的な配列を用いたタイピング法の発明が開示されている（特許文献2、特許文献3）。これらの方法などにより、現在、世界中の結核菌は概ね10系統に分類することが可能となっている。

40

【先行技術文献】

【特許文献】

【0009】

【特許文献1】特開2004-113196「特異的繰り返し遺伝子配列が多数含まれる結核菌特異的遺伝子群を用いたタイピング用DNAチップおよびその利用」

【特許文献2】特許第3631252号「直列変異反復配列オリゴタイピングによる結核菌群細菌の検出および鑑別」

【特許文献3】特許第3134940号「鳥型結核菌複合種の増幅および検出」

50

【非特許文献】

【0010】

【非特許文献1】Nature, 393: 537-544, (1998)

【非特許文献2】Science, 284: 1520-1523, (1999)

【非特許文献3】Proc Natl Acad Sci USA, 98: 7534-7539, (2001)

【発明の概要】

【発明が解決しようとする課題】

【0011】

そこで、本発明は、サンプル中の結核菌ゲノムの遺伝子配列情報から、結核菌のコンセンサスゲノム情報を用いてサンプル中の結核菌の菌株の同定や遺伝子変異を同定・検出する方法、該コンセンサスゲノム情報を生産する方法を提供することを課題とする。

【課題を解決するための手段】

【0012】

上記課題を解決するために、本発明のコンピューターによるサンプル中の結核菌の菌株を同定する方法は次の構成を備える。すなわち、コンピューターによるサンプル中の結核菌の菌株を同定する方法であって、コンピューターが入力されたサンプル中の結核菌のゲノム情報を記憶する工程と、コンピューターが記憶している結核菌のコンセンサスゲノム情報に含まれる特定の菌株の遺伝子配列と、サンプル中の結核菌の遺伝子配列についてアライメント (Alignment) 処理を行う工程と、サンプル中の結核菌のゲノム情報における特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致を検出することで、予め記憶された該特定の遺伝子配列に対応する菌株の種類に基づき特定の菌株であることを同定する工程と、サンプル中の結核菌の菌株の同定結果を表示する工程を具備することを特徴とする。ここで、サンプル中の結核菌のゲノム情報における特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致するとは、全部又は一部の一致であっても良い。なお、サンプル中の結核菌のゲノム情報における特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致することで菌株と同定する場合に、ある菌株を特徴付ける遺伝子配列は公知の情報を用いることもできる。

【0013】

また、本発明であるコンピューターによるサンプル中の結核菌の菌株及び変異を同定する方法は次の構成を備える。すなわち、コンピューターが入力されたサンプル中の結核菌のゲノム情報を記憶させる工程と、コンピューターが記憶している結核菌のコンセンサスゲノム情報に含まれる特定の菌株の遺伝子配列情報とサンプル中の結核菌の遺伝子配列情報とについてアライメント処理を行う工程と、サンプル中の結核菌のゲノム情報における特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致を検出することで、予め記憶された該特定の遺伝子配列に対応する菌株の種類に基づき特定の菌株であることを同定する工程と、同定された菌株に対応する結核菌コンセンサスゲノム情報に含まれる遺伝子配列を基準遺伝子配列とし、基準遺伝子配列とサンプル中の結核菌の遺伝子配列とを比較処理して、該基準遺伝子配列と異なる遺伝子情報を遺伝子変異情報として同定する工程と、サンプル中の結核菌の菌株の同定結果、及び遺伝子変異情報の結果を表示する工程とを具備することを特徴とする。

【0014】

結核菌のコンセンサスゲノム情報は、結核菌H37Rv株、結核菌KZN605株、結核菌RGTB423株、結核菌RGTB327株、結核菌Erdman株、結核菌CTRI-2株、結核菌CDC1551株、結核菌CCDC 5180株、結核菌CCDC5079株、結核菌KZN4207株、結核菌KZN1435株、結核菌F11株、結核菌H37Ra株、類縁菌CIPT 140010059株、類縁菌GM041182株、類縁菌BCG str. Mexico株、類縁菌BCG str. Moreau RDJ株、類縁菌BCG str. Tokyo 172株、類縁菌BCG Pasteur 1173P2株の一部又は全部のゲノム情報に対して、それぞれの遺伝子配列の逆位若しくは転移を補正して、アライメント処理した遺伝子配列について、共通する遺伝子配列領域を共通遺伝子

配列領域として、菌株間で異なる遺伝子配列領域を菌株の特有の遺伝子配列領域としてコンセンサスゲノム情報に含ませることを特徴としても良い。また、ここに挙げる結核菌の菌株、及び類縁菌の菌株の一部は、変更又は欠損させても良い。

【0015】

結核菌のコンセンサスゲノム情報を構築における遺伝子配列の逆位の補正が、結核菌KZN605株の遺伝子配列932051番目と932052番目の間、3479594番目と3459595番目の間、結核菌KZN1435株の遺伝子配列931985番目と931986番目の間、3479865番目と3479866番目の間、結核菌KZN4207株の遺伝子配列932007番目と932008番目の間、3476553番目と3476554番目の間の遺伝子配列を対象とすることを特徴とするのもであっても良い。ここに挙げた遺伝子配列の逆位の補正は、全て実施してもよく、1又は2以上の補正をした結核菌のコンセンサスゲノム情報としても良い。また、結核菌ゲノムの遺伝子配列の補正は、必要に応じて逆位の補正だけではなく、転移の補正をしても良い。

10

【0016】

結核菌のコンセンサスゲノム情報は、遺伝子配列1としても良い。また、サンプル中の結核菌の同定や遺伝子変異が検出できる範囲で、遺伝子配列1の遺伝子配列に含まれる1又は2以上の塩基について、置換、欠損又は付加されていても良い。

【0017】

遺伝子変異情報の結果を表示する工程は、遺伝子変異を含む遺伝子配列領域に関する情報をも表示できるものとしても良い。本発明のコンセンサスゲノムによりサンプル中の結核菌の菌株を同定し、その遺伝子配列と比較した場合にサンプル中の結核菌の遺伝子配列に変異がある場合には、その変異箇所を示すとともに、その変異が含まれる遺伝子領域に関する名称、機能、特定番号や詳細情報へのリンク情報などの情報を示すことができるものであっても良い。

20

【0018】

また、本発明のコンピューターによるサンプル中の結核菌の薬剤耐性若しくは薬剤感受性を示す薬剤応答変異遺伝子を同定する方法は次の構成を備える。前記の基準遺伝子配列と異なる遺伝子情報を遺伝子変異情報として同定する工程において、コンピューターが記憶している結核菌の薬剤応答変異遺伝子情報と比較処理して、該遺伝子変異情報が特定の薬剤応答変異遺伝子情報である場合にはこれを同定するものであって、遺伝子変異情報の結果を表示する工程において、遺伝子変異情報と薬剤応答変異遺伝子情報の結果を表示できるものであっても良い。つまり、サンプル中の結核菌ゲノムの遺伝子配列に変異がある場合に、これが特定の薬剤に対する耐性若しくは感受性に関与する遺伝子変異であるときに、これを同定して結果表示させるものである。すなわち、サンプル中の結核菌ゲノムと結核菌コンセンサスゲノムに含まれる遺伝子配列を比較させて、サンプル中の結核菌の遺伝子配列に変異がある場合に、その遺伝子変異情報と予めコンピューターが記憶している結核菌の菌株毎の薬剤応答変異遺伝子情報とを比較して、該遺伝子変異が薬剤応答変異遺伝子情報であると同定する。

30

【0019】

本発明のコンピューターによるサンプル中の結核菌の薬剤耐性若しくは薬剤感受性を示す薬剤応答変異遺伝子を同定する方法は次の構成としても良い。すなわち、薬剤応答変異遺伝子を有する配列を含んだ結核菌のコンセンサスゲノム情報を用いることを特徴とするものである。本発明の結核菌コンセンサスゲノムに薬剤応答変異遺伝子を有する結核菌株の遺伝子配列情報を加えることで、サンプル中の結核菌の遺伝子配列と比較して、直接に結核菌の薬剤応答変異株までをも同定することができる。すなわち、結核菌の菌株が薬剤応答変異の観点からも分類され同定できるものである。

40

【0020】

結核菌のコンセンサスゲノム情報の製造・作製にあつては、結核菌H37Rv株、結核菌KZN605株、結核菌RGTB423株、結核菌RGTB327株、結核菌Erdman株、結核菌CTRI-2株、結核菌CDC1551株、結核菌CCDC5180株、結核菌CCDC5079株、結核菌KZN4207株、結核菌KZN1435株、結核菌F11株、結核菌H37Ra株、類縁菌CIPT 140010059株、類縁菌GM041182株、類縁菌BCG

50

str. Mexico株、類縁菌BCG str. Moreau RDJ株、類縁菌BCG str. Tokyo 172株、類縁菌BCG Pasteur 1173P2株の一部又は全部のゲノム情報のうち、それぞれの遺伝子配列の逆位を補正若しくは転移して、アライメント処理した遺伝子配列について、共通する遺伝子配列領域を共通遺伝子配列領域として、菌株間で異なる遺伝子配列領域を菌株の特有の遺伝子配列領域としてコンセンサスゲノム情報に含ませることを特徴としても良い。また、ここに挙げた遺伝子配列の転移や逆位の補正は、全て又は一部を実施することによる結核菌のコンセンサスゲノム情報としても良い。

【0021】

上記の発明はコンピューターシステムとしてもよい。すなわち、コンピューターが入力されたサンプル中の結核菌のゲノム情報を記憶する結核菌サンプル情報記憶部と、コンピューターが結核菌コンセンサスゲノム情報を記憶するコンセンサスゲノム情報記憶部と、サンプル中の結核菌の遺伝子配列と結核菌コンセンサスゲノム情報に含まれる遺伝子配列についてアライメント処理するアライメント処理部と、サンプル中の結核菌のゲノム情報における特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致を検出することで、予め記憶された該特定の遺伝子配列に対応する菌株の種類に基づき特定の菌株であることを同定する菌株同定処理部と、前記菌株同定処理部で同定された特定菌株に対応する結核菌コンセンサスゲノム情報に含まれる特定菌株の基準遺伝子配列を同定する基準遺伝子配列同定処理部と、前記基準遺伝子配列とサンプル中の結核菌の遺伝子配列とを比較処理して、前記基準遺伝子配列と異なる遺伝子情報を同定する変異遺伝子情報検出部と、サンプル中の結核菌の菌株の同定結果、及び遺伝子変異情報を表示する情報出力部とを具備するサンプル中の結核菌の菌株及びその遺伝子変異を同定するコンピューターシステムとしてもよい。なお、請求項1乃至請求項8までの発明を、コンピューターシステムの発明としてもよい。

【発明の効果】

【0022】

本発明の結核菌コンセンサスゲノムを利用する結核菌の菌株及び変異を同定する方法によると、サンプル中の結核菌の遺伝子配列情報により結核菌のすべての菌株同定とSNPレベルのタイピング、遺伝子変異の検出が可能となり、菌株の分類などの疫学的解析の解像度、解析時間が格段に向上することができる。

【図面の簡単な説明】

【0023】

【図1】結核菌ゲノム中に存在する逆位を示す説明図

【図2a】結核菌ゲノム中に存在する逆位の補正処理を示す説明図（H37Rv vs ErdmanのHarrプロット）

【図2b】結核菌ゲノム中に存在する逆位の補正処理を示す説明図（H37Rv vs KZN1405（補正前）のHarrプロット）

【図2c】結核菌ゲノム中に存在する逆位の補正処理を示す説明図（H37Rv vs KZN 1405（補正後）のHarrプロット）

【図3】結核菌の全ゲノム規模でのアライメントの一部を示すコンピューター画像

【図4】結核菌コンセンサスゲノムとの比較による変異位置を示すリスト

【図5】結核菌コンセンサスゲノムを利用して得られた結核菌・類縁菌の系統図（無根系統樹）

【図6a】結核菌コンセンサスゲノムを利用して得られた結核菌・類縁菌の系統図（無根系統樹）

【図6b】結核菌H37Rv株を利用して得られた結核菌・類縁菌の系統図（有根系統樹）

【図6c】全結核菌株のゲノムを利用して得られた結核菌・類縁菌の系統図（有根系統樹）

【図7】VNT Rパターンの比較結果を示す図

【図8】スポリゴタイピングパターンの比較結果を示す図

【図 9】結核菌コンセンサスゲノムを利用する解析ソフトウェアのイメージ図

【図 10】結核菌コンセンサスゲノムを利用する薬剤応答変異遺伝子情報を同定した説明図

【発明を実施するための最良の形態】

【0024】

これまでのタイピング法などでは菌株識別の解像度が不十分で集団発生株間の差異などの解析は事実上不可能であった。しかし、本発明者は、次世代シーケンサー出現によりゲノムワイドで配列情報を得ることが可能となり、全ゲノムレベルの比較が容易に実施可能となったことを鑑みて、既存の全ゲノム情報が登録されている13株の結核菌、さらに医学的に重要な類縁菌6株の標準的な情報をすべて組み込んだ仮想的なコンセンサスゲノム配列を作製し、これを利用することでサンプル中の結核菌の菌株及び遺伝子変異を同定できるとの知見を得て、本発明に至った。本方法によりサンプル中の結核菌の遺伝子配列情報により結核菌のすべての菌株同定とSNPレベルのタイピング、遺伝子変異の検出が可能となり、菌株の分類などの疫学的解析の解像度、解析時間が格段に向上することができる。

10

【0025】

以下に、図面等を基に本発明を説明する。ここでは、実験例及び実施例の一部を挙げるが、本発明の実施形態は、本発明の趣旨から逸脱しない範囲で、適宜設計変更可能である。

【0026】

(結核菌コンセンサスゲノムを構成する結核菌及び類縁菌)

本発明の構成要素である結核菌のコンセンサスゲノム情報の構築について説明する。既存の全ゲノム情報が登録されている13株の結核菌、さらに医学的に重要な類縁菌6株のゲノムに含まれる遺伝子配列について準備をした。表1は、本発明における結核菌コンセンサスゲノムの構築に利用した結核菌の菌株を示したものであり、左から通し番号、菌株の名称とアクセッションナンバーを掲げている。表2は、本発明における結核菌コンセンサスゲノムの構築に利用した結核菌の類縁菌株を示したものであり、左から通し番号、菌株の名称とアクセッションナンバーを掲げている。ここで、医学的に重要な類縁菌6株とは、ヒトから抽出される菌株であり、結核菌ゲノムと近似しており、且つ結核菌コンセンサスゲノムの構築に利用できるとの理由等から検討し選択したものである。

20

30

【0027】

【表 1】

1	Mycobacterium tuberculosis H37Rv, complete genome	CP003248.1
2	Mycobacterium tuberculosis KZN 605, complete genome	CP001976.1
3	Mycobacterium tuberculosis RGTB423, complete genome	CP003234.1
4	Mycobacterium tuberculosis RGTB327, complete genome	CP003233.1
5	Mycobacterium tuberculosis str. Erdman = ATCC 35801 DNA, complete genome	AP012340.1
6	Mycobacterium tuberculosis CTIR-2, complete genome	CP002992.1
7	Mycobacterium tuberculosis CDC1551, complete genome	AE000516.2
8	Mycobacterium tuberculosis CCDC5180, complete genome	CP001642.1
9	Mycobacterium tuberculosis CCDC5079, complete genome	CP001641.1
10	Mycobacterium tuberculosis KZN 4207, complete genome	CP001662.1
11	Mycobacterium tuberculosis KZN 1435, complete genome	CP001658.1
12	Mycobacterium tuberculosis F11, complete genome	CP000717.1
13	Mycobacterium tuberculosis H37Ra, complete genome	CP000611.1

【0028】

【表 2】

1	Mycobacterium canettii CIPT 140010059 complete genome	HE572590.1
2	Mycobacterium africanum GM041182 complete genome	FR878060.1
3	Mycobacterium bovis BCG str. Mexico, complete genome	CP002095.1
4	Mycobacterium bovis BCG str. Moreau RDJ complete genome	AM412059.2
5	Mycobacterium bovis BCG str. Tokyo 172 DNA, complete genome	AP010918.1
6	Mycobacterium bovis BCG Pasteur 1173P2, complete genome	AM408590.1

【0029】

(遺伝子配列の逆位補正)

次に、それぞれの遺伝子配列について、その逆位の補正を行った。感染症を引き起こす結核菌などの細菌やウイルスは、ゲノム遺伝子の中で、転移や逆位を起こしやすい。ここで、逆位とは、染色体の一部が180°反転して、同位置に収まった場合を指す。転移とは染色体の一部が当初の位置から外れ、染色体の他の位置に移動した場合を指す。本発明者は、結核菌に逆位・転移が少ないことを見出し、この逆位を全て補正することで、結核菌のコンセンサスゲノムの構築を実現することができた。逆位・転移の補正を行うことで、各遺伝子配列における重複配列を増やすことができ、結核菌の菌株同定や遺伝子変異の検出のための最適なコンセンサスゲノムを構築することができる。

【0030】

結核菌ゲノムの遺伝子配列における逆位の検出手順について説明する。

ゲノムの先頭から、一定数のヌクレオチドの単位 (Windowという) ごとに相同性を比較し、相同性が極端に低くなった場合に、代替の場所がゲノムの他の場所にあるかどうかを探索することを繰り返す。ちなみに、これは遺伝子配列の転移の探索についても同様である。比較をする際に相補的な配列も探索することで逆位を検出する。パラメータの設定に

より、相同性が低い部分でも検出され、これらの遺伝子配列逆位の検出作業は一般的な遺伝子解析ソフト (<http://asap.ahabs.wisc.edu/mauve/index.php>) を用いることにより実施可能である。ただし、遺伝子配列の逆位の正確な塩基の位置を特定するためには目視による確認は必要である。

【0031】

本発明の実施例では、上述した遺伝子配列の逆位の検出手順により結核菌KZN605株、結核菌KZN1435株と結核菌KZN4207株の3株について逆位の検出を実施した。その後に、アライメントを行って、コンセンサスゲノムを構築したものである。結核菌のゲノムサイズは約4.4Mbpであるが、この3株は共にゲノムの約0.9Mbp付近と3.4Mbp付近で遺伝子の逆位が起きていたことを見出した(図1)。このうち、3.4Mbp付近の遺伝子の向きが入れ替わる部位にはIS6110という結核菌の中に数カ所以上存在する繰り返し配列が存在しているので、これを指標にその前後の、配列を逆位が発生していない株と比較することで、ゲノムの入れ替わっている0.9Mbp付近の前方転移位置を探索した。なお、この部分には3株とも、hypothetical proteinとしてannotateされている機能不明の同一蛋白質がコードされている。

【0032】

図1は、Genome rearrangement map(ゲノム再構成地図)と呼ばれる図であり、上下二つのゲノムが左から右に5'→3'と並べられている。上下それぞれで対応する遺伝子がある部分を線で結んでいる。結核菌は類似性が高いのでほとんどの部分が線分で結ばれるので、塗りつぶされるように表示されている。途中の三角形になっているように見える部分は、1,000,000bpより少し5'側の部分付近から、上下ゲノムの対応か所(3,400,000bp付近)が反転(逆位)になっていることを示す。逆位の補正位置に対応して、1,000,000付近から3,400,000付近が反転していることを示している。

【0033】

次に、結核菌KZN605株、結核菌KZN1435株と結核菌KZN4207株の3株について逆位の検出した後に、逆位している遺伝子配列を逆転させることで補正を行った。具体的には、結核菌KZN605株の遺伝子配列932051番目と932052番目の間、3479594番目と3459595番目の間、結核菌KZN1435の遺伝子配列931985番目と931986番目の間、3479865番目と3479866番目の間、結核菌KZN4207遺伝子配列932007番目と932008番目の間、3476553番目と3476554番目の間の遺伝子配列が逆位であったので、これを逆転させることで補正を行った。

【0034】

これらの結核菌ゲノムの逆位の補正の結果についてHarrプロットを用いて説明する。図2では、H37Rv vs Erdman (a)、H37Rv vs KZN1405(補正前)(b)、H37Rv vs KZN1405(補正後)(c)を示している。グラフの縦軸はH37Rvであり、横軸はErdman、KZN1405としており、それぞれ結核菌のゲノム内の位置を表し、0の位置が5'末端、それぞれ値が大きくなるほど3'に近づき、一番大きな値が、約4.4Mbpである。図2aのように、線分が左下から右上に繋がっている場合には塩基配列がだまかにそれぞれの位置で一致していることを示している。図2bは、中央部分に線分が左上から右下へ伸びており、当該部分が二つの菌間で逆位になっていることが分かる。この部分を検出して逆位を補正した後に同様の解析を行ったのが図2cであり、図2aと同様に線分が左下から右上に向かうようになって、アライメントが可能になる。

【0035】

(結核菌コンセンサスゲノム構築のためのアライメント工程)

このように結核菌ゲノムの遺伝子配列の逆位を補正した後に、この結核菌とその類縁菌のゲノムの遺伝子配列についてアライメント処理を行った(図3)。アライメント処理は、2以上の遺伝子配列についてコンピューターに入力し、それぞれを5'側から3'側となるようにコンピューターの画面上に表示させて、各遺伝子配列の共通する部分が最大となるように計算して、並び替える(アライメント)処理を行う。必要に応じて、各遺伝子配列の共通する部分が最大となるように、ギャップを入れることがある。この処理方法は既にいくつかの公知技術があり本発明にあってはそれらを利用している。図3は、結核菌の全

ゲノム規模でのアライメント処理の一部を示すコンピューター画像を示している。

【0036】

本発明において、各結核菌ゲノムの遺伝子配列の逆位を補正することで、アライメント処理により多くの重複した遺伝子部分（具体的には、約95%の重複部分）を得ることができ、これにより初めて結核菌コンセンサスゲノムを構築することができた。他方で、この逆位の補正処理をしないと、アライメント処理をしても、僅かに、約40%のみの相同性となる。

【0037】

このようにして得られた重複配列部分は、その遺伝子配列情報を取り出して、結核菌のコンセンサスゲノムの構成要素とする。他方で、重複していない遺伝子配列についても、それぞれについて結核菌を特徴付ける遺伝子配列として結核菌のコンセンサスゲノムの構成要素とする。すなわち、結核菌ゲノムの重複部分と複数の重複していない遺伝子配列が連結した人工的な遺伝子配列となり、自然界に存在する結核菌ゲノムと比較して、重複していない遺伝子配列が複数連結することから、その情報量は大きくなるものである。具体的には、自然界に存在する結核菌ゲノムの情報量は約4.4Mbpであるのに対して、本発明のコンセンサスゲノムの情報量は約4.9Mbpとなる。

【0038】

（結核菌コンセンサスゲノムの遺伝子配列とその利用）

配列番号1は、本発明における結核菌ゲノムのコンセンサスゲノムの遺伝子配列の実施例である。遺伝子の大きさは、約4.9Mbpである。本発明における結核菌ゲノムのコンセンサスゲノムは、拡張することもでき、たとえば、今後、新しい結核菌株や結核菌類縁菌株が発見された場合には、これらの菌株のゲノム遺伝子について、遺伝子配列にある逆位や転移の補正をした後に、アライメント処理をすることで本発明の結核菌ゲノムのコンセンサスゲノムに追加することができる。

【0039】

本発明において構築した結核菌コンセンサスゲノムは、サンプル中にある未知の結核菌株の同定をするほかに、その菌株のゲノムにある遺伝子変異についても検出することができる。まず、このサンプル中の結核菌のゲノム解析を行い（例えば、illumina社のMiSeq, Genome Analyzerシリーズ、HiSeqシリーズ、Lifetechnologies社のion proton、ion PGM, SoLid シリーズ、Roche社のGSシリーズ、PacBio社のPacBio RSシリーズなどの従来のサンガー法を基本原理とした塩基配列解析法とは異なる新たな塩基配列解析法を採用している次世代シーケンサー、さらに従来のサンガー法による遺伝子配列解析など）、得られた全ゲノムの遺伝子配列情報についてコンピューターに入力する。次に、本発明において構築した結核菌コンセンサスゲノムの遺伝子配列情報とアライメント処理を行う。このアライメント処理は、市販されているコンピューターソフトウェアを用いた。

【0040】

なお、結核菌ゲノムの遺伝子配列にある逆位を補正することでコンセンサスゲノムが構築されるが、サンプル中の菌株の遺伝子配列と結核菌コンセンサスゲノムの遺伝子配列とをアライメントすると、遺伝子配列にある逆位の補正により遺伝子配列の一致率が減少することもある。そこで、結核菌コンセンサスゲノムの中で、逆位を補正した遺伝子配列に補正処理したことを示す特定の情報を付加させて、サンプル中の菌株の遺伝子配列とアライメント処理する際に、一致確率を増やすように、逆位の補正を解除した遺伝子配列をも生成させて、逆位の補正の処理と未処理でのアライメント処理等ができるようにしても良い。

【0041】

本発明の結核菌コンセンサスゲノムには、表1及び表2に含まれる結核菌とその類縁菌を特徴付ける遺伝子配列が含まれているので、結核菌コンセンサスゲノムとアライメント処理することで、サンプル中にある未知の菌株について、菌株を同定することができる。すなわち、表1及び表2に含まれる結核菌とその類縁菌を特徴付ける遺伝子配列のいずれかの遺伝子配列と、サンプル中の未知の菌株の遺伝子配列と一致（部分一致も含む）する

遺伝子配列があることを検知することで、菌株を同定することができる。

【0042】

(結核菌コンセンサスゲノムの結核菌対応配列検出率)

表3は、本発明の結核菌コンセンサスゲノムの結核菌対応配列検出率を示す。この結核菌対応配列検出率によって、本発明の結核菌コンセンサスゲノムが高度に標準化した結核菌ゲノムであることを示す。複数の結核菌の菌株が含まれる検体を試料として、結核菌標準株である結核菌H37Rv株のゲノム情報と本発明の結核菌コンセンサスゲノムを用いて、試料中の結核菌群の結核菌対応遺伝子配列の検出率を測定した。加熱殺菌した2種類の結核菌培養物の検体を用い、NEB社の遺伝子増幅キットにより遺伝子増幅して、QIAGEN社の遺伝子精製キットにより精製した遺伝子を遺伝子解析装置MiSeq (Illumina社) にて遺伝子配列の解析を行った。得られた遺伝子配列データは、仮想的に全ゲノムの遺伝子配列を反映していると考えられ、試料中の結核菌対応遺伝子配列の検出率を測定することができる。遺伝子配列の解析の結果、2種類の結核菌培養物の検体において、それぞれ約200万塩基のデータを得ることができた。この遺伝子配列情報をそれぞれ本発明の結核菌コンセンサスゲノムと結核菌H37Rv株のゲノム情報にマッピングして、結核菌対応遺伝子配列の検出率を測定した。いずれの検体においても、本発明の結核菌コンセンサスゲノムは検出率が高く、結核菌標準株である結核菌H37Rv株よりも優れた標準ゲノム情報を有することが示された。

【0043】

【表3】

結核菌対応配列検出率 (%)

検体	H37Rv	コンセンサスゲノム
1	95.81	99.88
2	96.34	97.47

【0044】

(結核菌コンセンサスゲノムを用いた遺伝子変異の同定)

本発明の結核菌コンセンサスゲノムを利用することで、サンプル中の菌株の同定をするだけでなく、その遺伝子変異についても検出することができる。本発明の実施例として、配列番号1の結核菌ゲノムのコンセンサスゲノムを用いて、ある結核菌のサンプルにおいて、結核菌株の遺伝子変異について同定した。サンプルの結核菌ゲノムの全遺伝子配列解析を行い、その情報をコンピューターに入力して、本発明である結核菌コンセンサスゲノムと比較処理を行った。図4では、結核菌コンセンサスゲノムとの比較による遺伝子変異に関する情報が表示されたものを示す。

【0045】

図4は、本発明の結核菌コンセンサスゲノムとあるサンプル中の結核菌の遺伝子配列をアライメント処理して、サンプル中の結核菌ゲノムにある遺伝子変異を検出したものである。左側から、コンセンサス配列中の変異の位置(結核菌コンセンサスゲノムの特定の遺伝子(塩基)とその位置)、query側の塩基(この遺伝子と対応するサンプル中の結核菌ゲノムの遺伝子配列にある塩基(欠落している変異はドットで表示))、query側の変異の位置、隣までのミスマッチまでの距離、query側の配列名が表示されている。query側の変異の位置で、1TBと連続して表示されているのは、遺伝子変異が連続して存在していることが分かり、3574TBや5393TBと表示されているのは、遺伝子変異が分散していることが分かる。query側の配列名が、KZN4207とあるのは、サンプル中の結核菌がKZN4207であることが同定されたことを示す。

【0046】

(結核菌コンセンサスゲノムを用いた結核菌の系統分析)

本発明の結核菌コンセンサスゲノムを用いて、結核菌の系統分析をした実施例を説明する。図5は、本発明の結核菌コンセンサスゲノムを利用して得られた結核菌・類縁菌の系統図を示す。系統樹の作成方法は、各菌株がいつ頃分岐したかを推定可能とするBayesの

事後確率法を用いた。横棒と共に小さく記載されている「90.0」という数字は「900年」を意味している。これまで、菌株の系統解析は、16S rRNAの塩基配列やハウスキーピング遺伝子の多型、結核菌では、スポリゴタイピングやVNTRにより実施され、近年では全ゲノムデータを利用したSNP concatenationによる方法で実施されてきた。しかし、これらは全ゲノムの一部の情報しか使用しておらず、そのため、ここの分岐部内で見ただけの場合には利にかなった結果が得られている場合でも、系統樹全体のトポロジーが正しいかどうかの判断が困難であった。それに対し、全ゲノムワイドのアライメントが可能になったことで、全ゲノム情報を真に反映した「究極の系統樹」が作製可能となったものである。

【0047】

次に、本発明の結核菌コンセンサスゲノムを用いた結核菌の系統分析の精度を説明する。図6aは、本発明の結核菌コンセンサスゲノムを用いて結核菌の系統分析をした実施例、図6bは、結核菌標準株である結核菌H37Rv株を用いて結核菌の系統分析をした実施例、図6cは、全結核菌株（本発明の結核菌コンセンサスゲノムの構築に用いた13種類の結核菌とその類縁菌）のゲノムを用いて結核菌の系統分析をした実施例である。図6aと図6cは、ほぼ同様の系統分類が得られているが、図6bでは、結核菌RGTB423株がいずれのグループにも属さず異なる系統図となった。結核菌13株については「MTB」と表示して1つのグループとすることができる（図6a、図6c）。すなわち、本発明の結核菌コンセンサスゲノムを用いた結核菌の系統分析は、結核菌標準株である結核菌H37Rv株を用いたものよりも、その精度において優位であることが示された。なお、系統分析に生じた違いは、結核菌RGTB423株が結核菌H37Rv株との比較において検出されるSNP情報に基づくものである。すなわち、本発明の結核菌コンセンサスゲノムは標準化の程度が高いために、対象とする結核菌RGTB423株に固有のSNPが検出される確率が高くなるが、結核菌H37Rv株と比較した場合には、この結核菌H37Rv株自体の固有のSNPも検出されてしまうためである。なお、図6のなかの「300」の数字は「300年」を意味している。また、図5と図6は、いずれも結核菌の系統分析を表示したのであるが、図5は無根系統樹、図6は有根系統樹とした。

【0048】

（ロングシーケンス多型解析と結核菌コンセンサスゲノムを用いた結核菌系統の解析）

次に、19種類の結核菌及びその類縁菌に対して結核菌系統の解析をするために、ロングシーケンス多型解析方法と本発明の結核菌コンセンサスゲノムを用いたバーチャル判定方法を行った。表4では、ロングシーケンス多型解析方法による系統判定の結果を菌株毎に示す。ロングシーケンス多型解析方法は、spoligotypingやVNTRとは異なる系統分類法であり、H37Rvという代表的な結核菌株とそれ以外の系統の結核菌との比較で、特定の系統にだけ欠落している配列を見いだすものである（Proc Natl Acad Sci U S A. 2006 Feb 21;103(8):2869-73. Epub 2006 Feb

13.）。この方法は、人類進化と人類との結核菌共生の関係の解析などに応用されている（Nat Genet.

2013 Oct;45(10):1176-82. doi: 10.1038/ng.2744. Epub 2013 Sep 1.）。他方で、本発明の結核菌コンセンサスゲノムを用いたバーチャル判定方法では、系統分類の方法は、PCRによる特定領域の増幅の有無に基づいた、spoligotyping法、VNTR法、Long sequence polymorphism法の3法において、PCRによる増幅の標的となっている配列の有無をバーチャルに判定することで実施した。この2つの方法での結核菌系統の解析結果では、19種類の結核菌及びその類縁菌に対して18種類については同一の結果を得たが、F11菌株においては、ロングシーケンス多型解析方法では「系統5」となり、結核菌コンセンサスゲノムを用いたバーチャル判定方法では、「系統4」であった。この違いは、ロングシーケンス多型解析方法では、「系統5」でないことを示す領域の一部が欠損していたため、「系統5」と判定された。この実施例により、従来のタイピング法の判定結果が本来の菌株の性質一致しない例外的な菌株が存在することを示すと共に、これらをバーチャルな手法で実際の実験無しに証明でき、本手法の有用性を示すことができた。

【0049】

【表 4】

ロングシーケンス多型解析による結核菌系統の解析結果

菌種	菌株	系統判定結果
<i>M. africanum</i>	GM041182	6
<i>M. bovis</i>	BCG_Mexico	4
	BCG_Moreau_RDJ	4
	BCG_Pasteur	4
	BCG_Tokyo172	4
<i>M. canettii</i>	CIPT_140010059	2
<i>M. tuberculosis</i>	CCDC5079	2
	CCDC5180	2
	CDC1551	4
	CTRI-2	4
	erdman	4
	F11	5
	H37Ra	4
	H37Rv	4
	KZN1435	4
	KZN4207	4
	KZN605	4
	RGTB327	4
	RGTB423	4

【 0 0 5 0 】

(VNTR 及びスポリゴタイピングを用いた結核菌コンセンサスゲノムによる結核菌株の同定)

次に、サンプル中の結核菌についてVNTR (variable number of tandem repeat、反復配列多型)を調べ、このデータを本発明のコンセンサスゲノムにマッピングすることで、結核菌の菌株の同定ができるかを調べた(図7)。VNTRはゲノム中に数百~数千箇所の存在が確認されている数塩基~数十塩基を単位とする繰り返し配列である。反復配列多型分析とは、結核菌ゲノム上に存在するミニサテライトDNA中の繰り返し配列のコピー数を調べることによって、結核菌をタイピングする方法である。この繰り返しの数に多型が見られ、また多型性に富むことから、これまで有用な多型マーカーとして連鎖解析等に用いられてきた。図7は、任意に選んだ結核菌について、VNTRを調べて、このデータを本発明のコンセンサスゲノムにマッピングした結果を表示したものである。コンピューター処理によって、サンプル中の結核菌のVNTRのパターンに近似するコンセンサスゲノムに含まれる結核菌の菌株のVNTRのパターンを表示させて、この2つを比較できるように並び替えることができる。各表の上段には各VNTRの名称が表示され、その下にそのVNTRの数が表示されている。表の中段は、本発明のコンセンサスゲノムにおいて同定された菌株の名称とVNTRのデータが表示され、下段にはサンプル中の結核菌のVNTRのデータが表示されている。この2つのデータが合致しない点は、斜線の囲みでマークされている。本発明のコンセンサスゲノムを利用したVNTRの比較処理により、任意の結核菌の菌株の同定をすることができる。

【 0 0 5 1 】

スポリゴタイピング法は、結核菌群のゲノムDR領域に複数存在するdirect repeats(DR)配列に挟まれたスペーサ配列を可視化し、菌株間でその有無を比較することによって、結核菌の遺伝子型を解析する手法である。DR領域は、結核菌群のみに特異的に検出される36bpからなる複数の反復配列(Directrepeat:DR)が存在し、さらに37~41bpからなるそれぞれ異なった塩基配列で構成されたスペーサー配列を介して10~50コピーほど連なっている。DR領域に含まれるDRの数は菌株間で異なり、その間にあるスペーサ配列も菌株間で異なることから、43個のスペーサ配列の有無を菌株間で比較することにより、菌株の遺伝子型を判別する。図8は、任意に選んだ結核菌について、スポリゴタイピングを調べて、このデータを本発明のコンセンサスゲノムにマッピングした結果を表示したものである。コンピューター処理によって、サンプル中の結核菌のスポリゴタイピングのパターンに近似するコンセンサスゲノムに含まれる結核菌の菌株のスポリゴタイピングのパターンを表示させて、この2つを比較できるように並び替えることができる。

【0052】

図8上段には各データの名称が表示され、その下に本発明のコンセンサスゲノムにおけるスポリゴタイピングのパターンが表示され、その下段にはサンプル中の結核菌のスポリゴタイピングのデータが表示されている。この2つのデータが合致しない点は、囲み線でマークされている。本発明のコンセンサスゲノムを利用したスポリゴタイピングの比較処理により、任意の結核菌の菌株の同定をすることができる。Distanceは、値の違う(黒か白か)の数を値の総数で割ったものである。SpolDB4 STのSTとはSequence typeの略で、ある特定パターンのスポリゴタイピングパターンを持つものに番号を付けるものである。パターンの黒・白の表示は、43個の特定遺伝子領域の有無の判定を示し、その判定は43個の特定遺伝子領域の有無をPCRとハイブリダイゼーションで行う。図8のスポリゴタイピングパターンでは、例えば1番目の領域が「有り」なら1番目は黒、2番目の領域が「無し」なら2番目は白、と表示している。

【0053】

サンプル中の結核菌のゲノムについて、VNTRパターン又はスポリゴタイピングパターンを調べて、本発明の結核菌コンセンサスゲノムにマッピングすることで結核菌の菌株を同定できることを示すことができた。すなわち、サンプル中の結核菌のゲノムの全遺伝子配列の情報を得ることなく、従来方法である結核菌のVNTRパターン又はスポリゴタイピングパターンの情報があれば、本発明の結核菌コンセンサスゲノムを利用することで菌株の同定をすることができる。

【0054】

したがって、本発明の変形例として、コンピューターによるサンプル中の結核菌の菌株を同定する方法であって、コンピューターが入力されたサンプル中の結核菌のVNTRパターン又はスポリゴタイピングパターンを記憶する工程と、コンピューターが記憶している結核菌のコンセンサスゲノム情報に含まれる特定の菌株のVNTRパターン又はスポリゴタイピングパターンと、サンプル中の結核菌のVNTRパターン又はスポリゴタイピングパターンについてアライメント処理を行う工程と、サンプル中の結核菌のゲノム情報における特定のVNTRパターン又はスポリゴタイピングパターンが、結核菌のコンセンサスゲノム情報に含まれる特定のVNTRパターン又はスポリゴタイピングパターンと一致することで、特定の菌株であることを同定する工程と、サンプル中の結核菌の菌株の同定結果を表示する工程とを具備する結核菌の菌株を同定する方法としても良い。

【0055】

(結核菌コンセンサスゲノムを利用するコンピューターシステム)

結核菌コンセンサスゲノムを利用するコンピューターシステムについて説明する。このサンプル中の結核菌の菌株及び変異を同定するコンピューターシステムは、コンピューターが入力されたサンプル中の結核菌のゲノム情報を記憶する結核菌サンプル情報記憶部と、コンピューターが結核菌コンセンサスゲノム情報を記憶するコンセンサスゲノム情報記憶部と、サンプル中の結核菌のゲノム情報に含まれる遺伝子配列と結核菌コンセンサスゲノム情報に含まれる遺伝子配列についてアライメント処理するアライメント処理部と、サ

サンプル中の結核菌の特定の遺伝子配列が、結核菌のコンセンサスゲノム情報に含まれる特定の遺伝子配列と一致することで、特定の菌株であることを同定する菌株同定処理部と、前記菌株同定処理部で同定された特定菌株に対応する結核菌コンセンサスゲノム情報に含まれる特定菌株の基準遺伝子配列を同定する基準遺伝子配列同定処理部と、前記基準遺伝子配列とサンプル中の結核菌の遺伝子配列とを比較処理して、前記基準遺伝子配列と異なる遺伝子情報を遺伝子変異情報として同定する変異遺伝子情報検出部と、サンプル中の結核菌の菌株の同定結果、及び遺伝子変異情報を表示する情報出力部と、を具備することを特徴とする。また、遺伝子変異の箇所等の任意の遺伝子配列の部位における、当該遺伝子の機能・名称等の情報も表示させることができる。また、コンピューターシステムは、解析ソフトウェアとして提供しても良い。なお、サンプル中の結核菌等の菌株の系統解析も

10

【0056】

(結核菌コンセンサスゲノムを用いた北京型、モダン型又はアンセストラル型のバーチャル判定)

結核菌のうち特にアジアで流行している北京型株は高病原性であるとされており、さらに北京型のうちモダン型は薬剤耐性化しやすく(Nat Genet. 2013 Jul;45(7):784-90)、その識別は重要である。従来、北京型の判定は主にスポリゴタイピング法またはPCR法、モダン型とアンセストラル型への分類はPCR法により実施されてきた(例えばGenome Res. 2005 Oct;15(10):1357-64.)。本実施例では、本発明の結核菌コンセンサスゲノムを用いたバーチャル判定方法を行った。表5は、その結果を示すものである。従来利用されている北京型の判定であるスポリゴタイピング法では、CDC5079とCCDC5180の2株は北京型と判定されるものであるが、本発明の結核菌コンセンサスゲノムを用いたバーチャル判定方法では、CCDC5180株のみが、北京型且つモダン型と判定された。ちなみに、PCRベースの手法においても、同様の結果を得ることができる。これは、本発明の結核菌コンセンサスゲノムに「IS6110という挿入配列の特定部位」が保存されており、この配列を指標として判定することができたため、正確に判定できたものである。IS6110は、結核菌ゲノム中に複数個存在しており、その存在位置や個数を元に結核菌の亜型分類が可能となっている。IS6110の挿入配列は、遺伝子配列2に示す。以上の結果は、複数のタイピング法の判定結果が一致しない例外的な菌株が存在することを示すと共に、これらをバーチャルな手法で実際の実験無しに証明でき、本手法の有用性を示すことができた。

20

30

【0057】

【表 5】

結核菌流行株の北京型、モダン型又はアンセストラル型のバーチャル判定結果

菌種	菌株	北京型	アンセストラル ／モダン
<i>M. africanum</i>	GM041182	No	
<i>M. bovis</i>	BCG_Mexico	No	
	BCG_Moreau_RDJ	No	
	BCG_Pasteur	No	
	BCG_Tokyo172	No	
<i>M. canettii</i>	CIPT_140010059	No	
<i>M. tuberculosis</i>	CCDC5079	No	
	CCDC5180	Yes	Modern
	CDC1551	No	
	CTRI-2	No	
	erdman	No	
	F11	No	
	H37Ra	No	
	H37Rv	No	
	KZN1435	No	
	KZN4207	No	
	KZN605	No	
	RGTB327	No	
	RGTB423	No	

【0058】

(結核菌コンセンサスゲノムを利用する薬剤応答変異遺伝子の同定)

薬剤耐性検出法は、結核菌の場合、薬剤耐性は遺伝子の点突然変異により発生するので、薬剤耐性を誘発する遺伝子の変異を検出することで薬剤耐性の有無を判定する。薬剤耐性を誘発する遺伝子の具体例は、*rpoB*, *katG*, *mabA-inhA*, *embB*, *pncA*, *gryA*, *rpsL*, *rrs* の遺伝子等があるので、これらの変異を解析する。例えば、*katG*では、232番目のアミノ酸がPからSに変化する点突然変異（この残基はH37Rvという標準株ではゲノム中の2155418番目の塩基（同残基がCからTに変異））の有無で判定を行う。本発明のソフトウェア内では、コンセンサスゲノムを用いて、それぞれの変異に寄与する残基のannotationを行った

【0059】

例えば、コンピューターに予め結核菌の薬剤応答変異遺伝子情報を記憶させておき、あるサンプル中の結核菌の遺伝子配列を本発明のコンセンサスゲノムと比較処理し、サンプル中の結核菌の遺伝子変異情報を検出し、これが特定の薬剤応答変異遺伝子情報に合致する場合にはこれを同定しても良い。すなわち遺伝子変異情報の結果の表示において、遺伝子変異情報と薬剤応答変異遺伝子情報の両方の結果を表示する方法が挙げられる。また、この薬剤応答変異遺伝子情報には、その薬剤名、薬剤に関する情報、薬剤に対する感受性又は耐性の性質、薬剤応答変異遺伝子の名称、遺伝子配列関連情報などを含まれることが挙げられる。他方で、本発明の結核菌のコンセンサスゲノム情報は、13種類の結核菌・

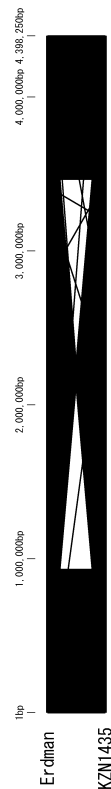
類縁菌に基づいて構築されているが、これに薬剤応答変異遺伝子を有する配列を含んだ結核菌のゲノム情報を加えたコンセンサスゲノム情報を構築しても良い。この場合には、サンプル中の結核菌の遺伝子配列情報が、薬剤応答変異結核菌の遺伝子配列情報と直接に比較処理することができることとなる。

【0060】

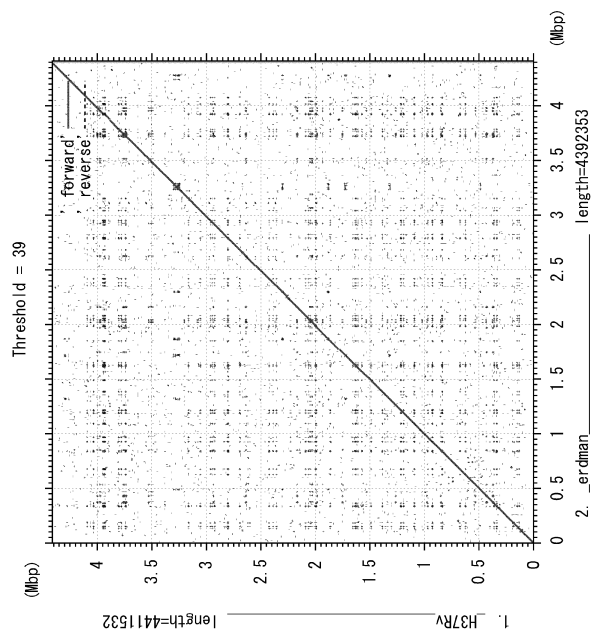
図10は、結核菌コンセンサスゲノムを利用する薬剤応答変異遺伝子情報の同定を示したものである。左の縦列より、「種の名称」「菌株の名称」が表示される。EMB（エタンブトール）、FQ（フルオロキノロン）、INH（イソニコチン酸ヒドラジッド）、PZA（ピラジナミド）、RIF（リファンピシン）、STR（硫酸ストレプトマイシン）はそれぞれ薬剤の名称であり、その下にその薬剤の代謝に関連する遺伝子の名称が表示されている。薬剤応答変異遺伝子情報の1つとして、菌株毎にそれぞれの薬剤に対する感受性又は耐性の性質が、「R」（耐性）及び「S」（感受性）として示されている。この結果により、結核菌コンセンサスゲノムを利用する薬剤応答変異遺伝子に関する情報が得られたことが示される。なお、菌株毎にそれぞれの薬剤に対する感受性又は耐性の性質が異なるのは、薬剤の代謝に関連する遺伝子に変異があるためであり、本発明ではその遺伝子変異情報も表示させることができる。

10

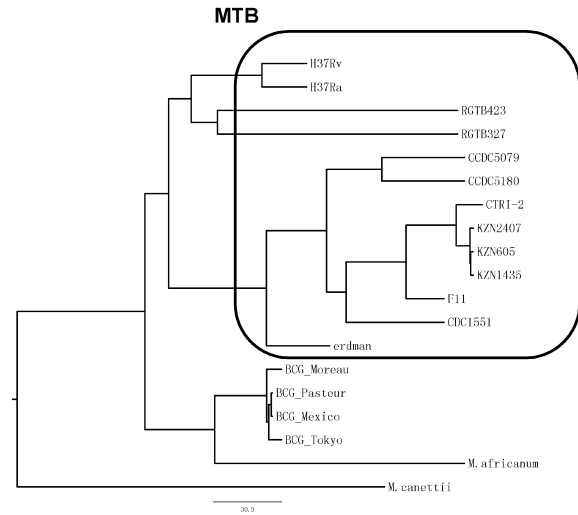
【図1】



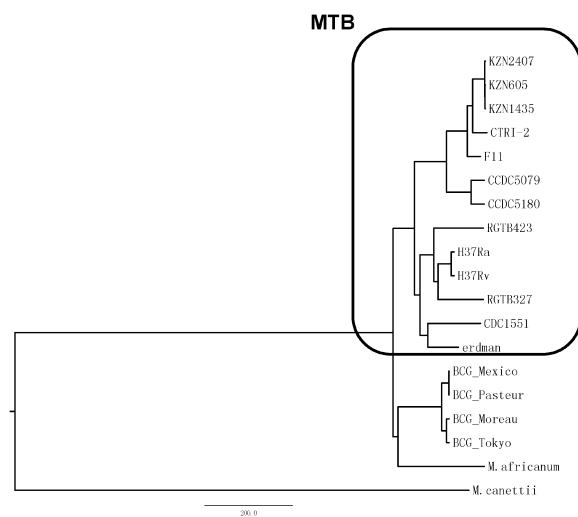
【図2a】



【図 6 a】



【図 6 c】



【配列表】

0006327473000001.app

フロントページの続き

(72)発明者 切替 照雄

東京都新宿区戸山1-21-1 国立国際医療研究センター内

(72)発明者 奥村 香世

北海道帯広市稲田町西2線11番地 国立大学法人帯広畜産大学内

審査官 宮久保 博幸

(56)参考文献 特開2010-204753 (JP, A)

特開2011-217703 (JP, A)

特開2003-135099 (JP, A)

特開2003-330934 (JP, A)

特開2012-239430 (JP, A)

VISHNOI, Anchal, et al., Comparative analysis of bacterial genomes: identification of divergent regions in mycobacterial strains using an anchor-based approach, Nucleic Acids Research, 2007年, Vol.35, No.11, p.3654-3667

KURABACHEW, Mekonnen, et al., Sequence analysis in the 23S rDNA region of Mycobacterium tuberculosis and related species, Journal of microbiological Methods, 2003年, Vol.54, p.373-380

(58)調査した分野(Int.Cl., DB名)

G06F 19/22

C12N 15/09

C12Q 1/68