

(19)日本国特許庁 (J P)

(12) 特 許 公 報 (B 2)

(11)特許番号

特許第3431135号
(P3431135)

(45)発行日 平成15年7月28日(2003.7.28)

(24)登録日 平成15年5月23日(2003.5.23)

(51)Int.Cl. ⁷	識別記号	F I	
G 0 6 F 17/30	1 7 0	G 0 6 F 17/30	1 7 0 F
	2 1 0		2 1 0 D
C 1 2 N 15/09	Z N A	C 1 2 N 15/00	Z N A A

請求項の数15(全 19 頁)

(21)出願番号 特願2000-215134(P2000-215134)

(22)出願日 平成12年7月14日(2000.7.14)

(65)公開番号 特開2002-32377(P2002-32377A)

(43)公開日 平成14年1月31日(2002.1.31)

審査請求日 平成12年7月26日(2000.7.26)

(73)特許権者 501203344
独立行政法人農業技術研究機構
茨城県つくば市観音台3-1-1

(73)特許権者 500333981
渡邊 克二
熊本県菊池郡西合志町大字須屋2421 農
林水産省九州農業試験場内

(72)発明者 渡邊 克二
熊本県菊池郡西合志町大字須屋2421 農
林水産省九州農業試験場内

(72)発明者 奥田 充
熊本県菊池郡西合志町大字須屋2421 農
林水産省九州農業試験場内

(74)代理人 100064908
弁理士 志賀 正武 (外6名)

審査官 高瀬 勤

最終頁に続く

(54)【発明の名称】 遺伝子の類縁性検索方法および遺伝子の類縁性検索システム

(57)【特許請求の範囲】

【請求項1】 プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列を表すプライマデータを基に、当該既知遺伝子から当該プライマで増幅される増幅後塩基配列データを生成する増幅配列認識部と、

制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の前記増幅後塩基配列データとが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で

切断したときの切断長理論値とが関連付けられて保持された切断長理論値データとして出力する理論値計算部と、

前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較部とを備えることを特徴とする塩基配列類縁性検索システム。

【請求項2】 前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示部を備えることを特徴とする請求項1に記載の塩基配列類縁性検索システム。

【請求項3】 前記表示部は、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする請求項2に記載の塩基配列類縁性検索システム。

【請求項4】 前記増幅配列認識部は、前記増幅前塩基配列データを読み込み、その塩基配列と前記プライマ配列とが mismatches の許容限界として指定された数値範囲内でマッチする箇所を探し出し、当該箇所を切断して残った塩基配列を前記増幅後塩基配列データとして生成することを特徴とする請求項1に記載の塩基配列類縁性検索システム。

【請求項5】 前記増幅配列認識部は、上流プライマおよび下流プライマそれぞれについて mismatches の許容限界として指定された数値範囲内でマッチする箇所を探し出すことを特徴とする請求項4に記載の塩基配列類縁性検索システム。

【請求項6】 プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列を表すプライマデータを基に、当該既知遺伝子から当該プライマで増幅される増幅後塩基配列データを生成する増幅配列認識過程と、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の前記増幅後塩基配列データとが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データを出力する理論値計算過程と、前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較過程とを有することを特徴とする塩基配列類縁性検索方法。

【請求項7】 前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示過程を有することを特徴とする請求項6に記載の塩基配列類縁性検索方法。

【請求項8】 前記表示過程では、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする請求項7に記載の塩基配列類縁性検索方法。

【請求項9】 前記増幅配列認識過程においては、前記増幅前塩基配列データを読み込み、その塩基配列と前記

プライマ配列とが mismatches の許容限界として指定された数値範囲内でマッチする箇所を探し出し、当該箇所を切断して残った塩基配列を前記増幅後塩基配列データとして生成することを特徴とする請求項6に記載の塩基配列類縁性検索方法。

【請求項10】 前記増幅配列認識過程においては、上流プライマおよび下流プライマそれぞれについて mismatches の許容限界として指定された数値範囲内でマッチする箇所を探し出すことを特徴とする請求項9に記載の塩基配列類縁性検索方法。

【請求項11】 プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列を表すプライマデータを基に、当該既知遺伝子から当該プライマで増幅される増幅後塩基配列データを生成する増幅配列認識過程と、

制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の前記増幅後塩基配列データとが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データを出力する理論値計算過程と、前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較過程との処理をコンピュータに実行させるコンピュータプログラムを記録したコンピュータ読取り可能な記録媒体。

【請求項12】 前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示過程の処理をコンピュータに実行させるコンピュータプログラムを記録した請求項11に記載のコンピュータ読取り可能な記録媒体。

【請求項13】 前記表示過程では、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする請求項12に記載のコンピュータ読取り可能な記録媒体。

【請求項14】 前記増幅配列認識過程においては、前記増幅前塩基配列データを読み込み、その塩基配列と前記プライマ配列とが mismatches の許容限界として指定された数値範囲内でマッチする箇所を探し出し、当該箇所を切断して残った塩基配列を前記増幅後塩基配列データ

として生成することを特徴とする請求項11に記載のコンピュータ読取り可能な記録媒体。

【請求項15】 前記増幅配列認識過程においては、上流プライマおよび下流プライマそれぞれについて mismatches の許容限界として指定された数値範囲内でマッチする箇所を探し出すことを特徴とする請求項14に記載のコンピュータ読取り可能な記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】この発明は、塩基配列類縁性検索方法および塩基配列類縁性検索システムに関する。特に、制限酵素による切断片の切断片長データを基にした塩基配列類縁性検索方法および塩基配列類縁性検索システムに関する。

【0002】

【従来の技術】試料として与えられたDNA (deoxyribonucleic acid, デオキシリボ核酸) 分子内の塩基配列を同定するために、その試料を分析することによって全塩基配列を出力するシーケンサ装置が広く用いられている。

【0003】また、制限酵素を用いてDNA分子を切断し、切断片の長さを基に塩基配列を推定する方法もある。

【0004】

【発明が解決しようとする課題】しかしながら、上記シーケンサ装置は高価であり、より安価に塩基配列を特定できる手段が求められている。

【0005】また、制限酵素による切断片の長さを基に塩基配列を推定する方法は、様々な制限酵素によって得られる長さデータを用いて、既知の塩基配列とのパターンマッチングを行う手間が膨大であり、効率的なデータ処理が課題となっている。

【0006】本発明は、上記事情に鑑みてなされたものであり、制限酵素による切断片長の実測データを基に、効率的に既知の塩基配列との類縁性を解析するための塩基配列類縁性検索方法および塩基配列類縁性検索システムを提供することを目的とする。

【0007】

【課題を解決するための手段】上記の課題を解決するために、本発明による塩基配列類縁性検索システムは、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データとして出力する理論値計算部と、前記切断長理

論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較部とを備えることを特徴とする。

【0008】また、本発明による塩基配列類縁性検索システムは、前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示部を備えることを特徴とする。

【0009】また、本発明による塩基配列類縁性検索システムは、前記表示部が、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする。

【0010】また、本発明による塩基配列類縁性検索システムは、前記比較部が、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とする。

【0011】また、本発明による塩基配列類縁性検索システムは、プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データを生成する増幅配列認識部を備え、前記理論値計算部は、この増幅後塩基配列データを基に前記切断長理論値を計算することを特徴とする。なお、ここで「増幅」とは、所定の上流プライマあるいは下流プライマまたはこれら両者を用いて、塩基配列の、始端から当該上流プライマ固有の塩基配列パターンの箇所までと当該下流プライマ固有の塩基配列パターンの箇所から終端までを切断し、残りの塩基配列の部分のみを抽出することを言う。

【0012】また、本発明による塩基配列類縁性検索方法は、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データを出力する理論値計算過程と、前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較過程とを有することを特徴とする。

【0013】また、本発明による塩基配列類縁性検索方法は、前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示過程を有することを特徴とする。

【0014】また、本発明による塩基配列類縁性検索方法は、前記表示過程では、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とする。

【0015】また、本発明による塩基配列類縁性検索方法は、前記比較過程では、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とする。

【0016】また、本発明による塩基配列類縁性検索方法は、プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データを生成する増幅配列認識過程を有し、前記理論値計算過程では、この増幅後塩基配列データを基に前記切断長理論値を計算することを特徴とする。

【0017】また、本発明は、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データと、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている塩基配列データとを基に、当該既知遺伝子を当該制限酵素で切断したときの切断長理論値を計算して、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持された切断長理論値データを出力する理論値計算過程と、前記切断長理論値データと、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている切断長実測値データとを比較することにより、当該既知遺伝子と当該試料との類似度を算出して解析結果データとして出力する比較過程との処理をコンピュータに実行させるコンピュータプログラムを記録したコンピュータ読取り可能な記録媒体を要旨とする。

【0018】また、本発明によるコンピュータ読取り可能な記録媒体は、前記解析結果データを基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示過程の処理をコンピュータに実行させるコンピュータプログラムを記録したことを特徴とする。

【0019】また、本発明によるコンピュータ読取り可能な記録媒体は、前記表示過程では、樹形図により、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて表示することを特徴とす

るものである。

【0020】また、本発明によるコンピュータ読取り可能な記録媒体は、前記比較過程では、前記切断長理論値データと前記切断長実測値データとを基に、非加重結合法を用いることにより前記類似度を算出することを特徴とするものである。

【0021】また、本発明によるコンピュータ読取り可能な記録媒体は、プライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データを生成する増幅配列認識過程の処理をコンピュータに実行させるコンピュータプログラムが記録されており、前記理論値計算過程では、この増幅後塩基配列データを基に前記切断長理論値を計算することを特徴とするものである。

【0022】

【発明の実施の形態】以下、図面を参照しこの発明の一実施形態について説明する。

【0023】図1は、この発明の一実施形態による塩基配列解析システム（遺伝子の類縁性検索システム）の構成を示すブロック図である。図1において、符号21は、制限酵素の種別と当該制限酵素が切断する切断箇所の塩基配列パターンと当該塩基配列パターン内における切断位置とが関連付けられて保持されている制限酵素データである。また、22bは、既知遺伝子の種別と当該既知遺伝子の塩基配列とが関連付けられて保持されている増幅後塩基配列データ（塩基配列データ）である。なお、この増幅後塩基配列データ22bは、一般に公開されている遺伝子データを手作業等の何らかの方法によって編集することにより得られる。なお、増幅前の塩基配列データを基に自動的に増幅後塩基配列データ22bを生成する方法は、後の第二実施形態において説明する。

【0024】11は、制限酵素データ21と増幅後塩基配列データ22bとを基に既知遺伝子を制限酵素で切断したときの切断長理論値を計算して、切断長理論値データ23を出力する理論値計算部である。理論値計算部11によって出力される切断長理論値データ23では、既知遺伝子の種別と制限酵素の種別と当該既知遺伝子を当該制限酵素で切断したときの切断長理論値とが関連付けられて保持されている。また、13は、この切断長理論値データ23と、実測値入力部12から入力された切断長実測値データ25とを比較することにより既知遺伝子と試料との類似度を算出して解析結果データ27として出力する比較部である。なお、この切断長実測値データ25では、制限酵素の種別と当該制限酵素を用いて試料を切断して測定した結果得られた切断長実測値とが関連付けられて保持されている。

【0025】14はデータの比較の際の許容誤差を設定する許容誤差設定部であり、許容誤差設定部によって設

定された値は許容誤差データ29として保持されている。比較部13は、この許容誤差データ29を参照して、切断長理論値データ23と切断長実測値データ25とを比較する際には、許容誤差範囲内の相違は一致しているものとして扱う。なお、比較部13による具体的な比較方法および許容誤差の適用方法については後に詳述する。

【0026】そして、15は、比較部13によって出力された解析結果データ27を基に、相互に類似度の高い試料同士または相互に類似度の高い既知遺伝子と試料とを関連付けて図形的に表示する表示部である。表示部15による具体的な表示方法については、後で説明する。

【0027】次に、切断長理論値の計算の基となる増幅後塩基配列データ22bの詳細について説明する。図2は増幅後塩基配列データ22b(塩基配列データ)の前半部分を示す図であり、配列番号1は同データの後半部分を示す図である(図3参照)。これらの図に示すように増幅後塩基配列データ22bはテキスト形式のデータであり、図2に示す前半部分には、下線を付したDNA番号(ACCESSION)と細菌名(ORGANISM)と遺伝子名(KEYWORDS)とが少なくとも含まれている。また、図3の後半部分には、塩基配列(ORIGIN)が含まれている。この塩基配列のデータ内の文字「a」はアデニン(adenine)を、「g」はグアニン(guanine)を、「c」はシトシン(cytosine)を、「t」はチミン(thymine)をそれぞれ表している。

【0028】次に制限酵素データ21の詳細について説明する。図4は、制限酵素データ21のデータ構造およびデータ例を示す表図である。図4に示すように、制限酵素データ21は、表形式のデータであり、制限酵素番号と制限酵素名と切断箇所と切断位置の列を持っている。制限酵素番号は、制限酵素の識別のために本システムによって与えられた番号である。また、切断箇所は制限酵素が切断する箇所の塩基配列パターンを表し、切断位置はこの塩基配列パターン内における切断位置の変位を表す。例えば、図4に示す制限酵素番号12のSmaIは、DNA分子の塩基配列中の「・・・ccccggg・・・」にマッチする箇所の変位3の位置に作用して、「・・・ccc」と「ggg・・・」に切断する。

【0029】図5は、理論値計算部11に対して切断長理論値計算を指示する利用者インタフェース例を示す画面図である。この画面内には、存在している塩基配列ファイル(増幅後塩基配列データ22b)と制限酵素データ21内の制限酵素名とがそれぞれリスト表示されている。利用者がこれらのリストの中からそれぞれ選択して画面下部の「切断長計算」ボタンを押すことにより、理論値計算部11は選択されたデータを基に切断片長の理論値を計算する。

【0030】理論値計算部11による計算の手順は次のとおりである。すなわち、選択された制限酵素の切断箇

所と切断位置のデータを制限酵素データ21から読み取り、この切断箇所の塩基配列パターンとマッチする部分を選択された増幅後塩基配列データ22bの塩基配列(ORIGIN)内で検索する。切断箇所が見つければ、始端からあるいは前回切断位置から今回切断位置までの塩基配列数を切断長理論値データ23に出力し、さらに検索を繰り返す。終端まで到達すれば、始端からあるいは前回切断位置から終端までの塩基配列数を切断長理論値データ23に出力する。

【0031】図6は、切断長理論値データ23のデータ構造およびデータ例を示す表図である。図6に示すように、切断長理論値データ23は、DNA番号、数値区分、細菌名、遺伝子名、制限酵素番号切断長の列を有する。また切断長理論値データ23の各行は、DNA毎(DNA番号により識別)かつ制限酵素毎(制限酵素番号により識別)かつ切断長毎に存在する。ただし、切断片を識別するデータを列として加えて、同一DNAの同一制限酵素による同一切断長を有する切断片を区別するようにしても良い。

【0032】これらの列のうち、DNA番号と細菌名と遺伝子名の値は、理論値計算部16によって増幅後塩基配列データ22bから読み取られ出力されたものである。また、制限酵素番号の値は、同じく制限酵素データ21から読み取られ出力されたものである。また、切断長の値は理論値計算部11によって前記手順で算出されたものである。なお、数値区分は理論値か実測値かの区分を示すデータであり「1」は理論値であることを表している。

【0033】次に、切断長実測値データ25の詳細について説明する。切断長を実測するため、塩基配列の特定の対象となる試料は、予めプライマによって増幅されてから、制限酵素の適用によって切断される。このようにして作られた切断片の長さの測定は従来技術によって可能であり、例えば電気泳動によって得られる画像を数値化することにより行う。より具体的には、例えば、アガロースゲル電気泳動により切断片長分布を画像として得るようにする。

【0034】上記のような方法で実測された切断長は、実測値入力部12から入力される。図7は、実測値入力部12の利用者インタフェース例を示す画面図である。図7に示す画面では、利用者は、試料を識別するためのDNA番号を入力し、制限酵素データ21を基に表示される制限酵素のリストから1つを選択し、実測結果の切断長を入力することができる。また、複数の試料をグルーピングして付与したグループ名を入力することもできる。

【0035】図8は、切断長実測値データ25のデータ構造およびデータ例を示す表図である。切断長実測値データは、図8に示すように表形式のデータであり、DNA番号、数値区分、グループ名、制限酵素番号、切断長

の列を有する。また切断長実測値データ25の各行は、試料毎(DNA番号により識別)かつ制限酵素毎(制限酵素番号により識別)かつ切断長毎に存在する。なお、数値区分は理論値か実測値かの区分を示すデータであり「2」は実測値であることを表している。

【0036】次に、比較部13が理論値と実測値との比較を行い両者の類縁性を分析する方法について説明する。本実施形態においては、比較部13は、非加重結合法(UPGMA)を用いることによって、解析結果データ27を生成する。

【0037】図9は、比較部13による塩基配列のクラスタリングの基となる切断片長のバンド分布を表す参考図である。図9において、AとBとCは、それぞれDN

$$S(A,B) = \frac{\text{AとBに共通するバンドの総数}}{\text{Aのバンド数+Bのバンド数}}$$

【0040】例えば、図9に示すAとBでは、塩基配列Aには1, 2, 4, 5, 8の5本のバンドがあり、塩基配列Bには1, 3, 4, 5, 7, 8の6本のバンドがある。従ってAとBに共通するバンドは、AおよびBそれぞれに1, 4, 5, 8の4本であり、両方をたすと共通するバンドは合計8本である。従って、

【数2】

$$S(A,B) = \frac{4 \times 2}{6 + 5}$$

であり、類似度は8/11(約0.727)と算出される。

【0041】ただし、実測値データは誤差を含むため、共通のバンドであるかどうかの判定にあたっては、比較部13は許容誤差データ29を用いることとする。つまり、設定された許容誤差範囲内であれば共通のバンドとして扱う。なお、設定される許容誤差が小さすぎると、本来マッチすべき切断長実測値と切断長理論値とがうまくマッチしないという問題が起こる。また逆に、設定される許容誤差が大きすぎると、切断長実測値が本来マッチすべきでない切断長理論値とマッチしてしまい、マッチングの精度が悪くなるという問題が起こる。従ってこれらの問題が起こらないように適切な許容誤差を設定する必要があり、例えば、10%程度の許容誤差を設定するようにする。

【0042】次に、比較部13は、算出した類似度を基に塩基配列のクラスタ分析を行う。図10は、比較部13が塩基配列間の類似度を基に平均距離法を用いてクラスタ分析を行う過程の例を示す表図であり、この図に示す例ではA, B, C, Dの4つの塩基配列を分析対象としている。図10(a)は、算出されたA, B, C, D相互間の類似度を示している。図10(a)の表の中で最も高い類似度を持っているのはAとC(類似度0.8)であるため、これらをまとめて(A+C)のクラス

A番号によって識別される塩基配列の理論値または実測値の切断長分布である。このバンド分布は、電気泳動によって得られる画像パターンと同様のものであり、縦方向が切断長の次元である。また図9では、実測値データだけでなく理論値データについても、得られるはずの仮想的な電気泳動パターンを示している。

【0038】比較部13は、これらの実測値データおよび理論値データを読み込み、2つのDNA番号で与えられるデータ間の類似度を算出する。AとBの間の類似度S(A, B)の定義は、次式で与えられる。

【0039】

【数1】

タとして次のステップに進む。

【0043】図10(b)は、図10(a)のAおよびCを単一のクラスタ(A+C)で置き換えたときの類似度を示している。ここで、塩基配列Bとクラスタ(A+C)との類似度S(B, A+C)は、S(B, A)とS(B, C)との平均により0.6であり、塩基配列Dとクラスタ(A+C)との類似度S(D, A+C)は、S(D, A)とS(D, C)との平均により0.2となっている。そして、この表の中で最も高い類似度を持っているのは(A+C)とBであるため、これらをまとめて((A+C)+B)のクラスタとして次のステップに進む。

【0044】図10(c)は、図10(b)の(A+C)およびBを単一のクラスタ((A+C)+B)で置き換えたときの類似度を示している。ここで、クラスタ((A+C)+B)と塩基配列Dとの類似度S(D, (A+C)+B)は、S(D, A+C)とS(D, B)との平均により0.3となっている。

【0045】図11は、上に説明した非加重結合法による解析結果データを、木構造を用いて図形として表した参考図である。図11において、木のノード部分に付記された数値は、そのノードに従属する終端ノード(塩基配列を表す)あるいは非終端ノード(クラスタを表す)間の類似度を表している。例えば、塩基配列Bとクラスタ(A+C)との類似度は0.6である。

【0046】表示部15は、解析結果データ27を読み込み、与えられた試料と類似度の高い既知遺伝子とを関連付けて図形的に表示する。図12は、表示部15によって樹形図として表示されている解析結果の例を示す画面図である。図12の例では、複数の試料DNA(AP1, AP2, AP3)と複数の既知遺伝子とのクラスタ分析の結果が示されている。

【0047】次に、本発明の第二実施形態について説明する。この第二実施形態は、増幅前の塩基配列を表すデ

ータ内において、プライマによって増幅される塩基配列部分を自動的に認識して抽出する機能を有することを特徴とする。図13は、この第二実施形態による塩基配列解析システムの構成を示すブロック図である。図13において、符号22aはプライマによって増幅される前の既知遺伝子の塩基配列が保持されている増幅前塩基配列データであり、31は増幅前塩基配列データ22aを読み込み、当該プライマの塩基配列パターンを表すプライマデータを基に、当該既知遺伝子を当該プライマで増幅した後の増幅後塩基配列データ22bを生成する増幅配列認識部である。

【0048】図14は増幅配列認識部31への入力データとなる増幅前塩基配列データ22aの前半部分を示し、また、配列番号2は同データ22aの後半部分を示している(図15参照)。図示するように、増幅前塩基配列データ22aは増幅後塩基配列データ22bと同様のテキスト形式のデータである。図3および図15は、同一の遺伝子(ACCESSION="M59070", KEYWORDS="16S ribosomal RNA.")のものであるが、増幅後の塩基配列(ORIGIN)は、増幅前の塩基配列(ORIGIN)の23番目の文字から1031番目の文字までの部分列であり、その長さは1009となっている。

【0049】図16は、増幅配列認識部31による認識のためのプライマ配列の入力画面を示す画面図である。この画面では、上流プライマおよび下流プライマそれぞれについて、利用者がその配列と名前とミスマッチの許容限界を入力できるようになっている。図16の例では、上流プライマとして「41f」という名前で配列番号3の「gctcagattgaactcggcg」という配列が入力されており、下流プライマとして「1066r」という名前で配列番号4の「acatttccacaacacgagctg」という配列が入力されている。そして、入力されたこれらの塩基配列パターンはプライマデータとして増幅配列認識部31によって利用される。

【0050】増幅配列認識部31は、増幅前塩基配列データ22aを読み込み、その塩基配列(ORIGIN)を、始端から順に走査する。そして、上流プライマおよび下流プライマそれぞれについて、ミスマッチの許容限界として指定された数値範囲内でマッチする箇所を探し出す。そしてその箇所よりも上流および下流をそれぞれ切断し、残った塩基配列を増幅後塩基配列データ22bの塩基配列(ORIGIN)として出力する。

【0051】上述の遺伝子の類縁性検索システムはコンピュータシステムを用いて実現されている。そして、上述した理論値計算部、実測値入力部、比較部、許容誤差設定部、表示部の各々の処理の過程は、コンピュータプログラムの形式でコンピュータ読み取り可能な記録媒体に記憶されており、このコンピュータプログラムをコンピュータが読み出して実行することによって、上記処理

が行われる。ここでコンピュータ読み取り可能な記録媒体とは、フロッピー(登録商標)ディスク、光磁気ディスク、CD-ROM、DVD-ROM、磁気ハードディスク、半導体メモリ等をいう。

【0052】なお、この遺伝子の類縁性検索システムを実現するためのコンピュータシステムとしては、パーソナルコンピュータやワークステーションなどの汎用のコンピュータを用いることが可能である。

【0053】また、遺伝子の類縁性検索システムにおいて扱うデータの形式は、上記第一および第二実施形態において記載したものに限定されず、他の形式あるいは表現のデータを用いても良い。また、上記実施形態で説明した表形式のデータの一部は非正規形であるが、正規化しても良い。例えば、図6に示す切断長理論値データにはDNA番号と細菌名と遺伝子名の列が含まれているが、この3者の関係を別表に保持するようにしても良い。

【0054】次に、本発明による遺伝子の類縁性検索システムを用いて、細菌の分類を行った事例について説明する。

【0055】生物の分類において最も元となる門(division)は植物、動物、糸糸菌などの真核細胞を有し有性生殖を行う真核生物と、単細胞である真正細菌(細菌)、古細菌の3つの門に分類されている。初期の細菌分類体系では、細菌の細胞形態や生理学的性質、そして生化学的活性等の基準を基に分類体系が出来上がってきた。複雑で煩雑な操作が要求される菌の同定を容易に実施できるように、従来技術によるシステムにおいては、同定しようとする菌の複数の生理試験結果を自動的に読み取り、予め登録しデータベース化していた既知菌株の生理試験結果と比較して類縁性を検索し菌の種名を推定していた。このような生理学的試験に基づく同定システムは、医学・臨床分野や、化粧品、食品衛生、品質管理、環境衛生等の幅広い産業分野で利用されてきている。

【0056】しかしながら、近年は、リボゾームRNAを分子時計とする系統分類体系が構築され、細菌の16S rDNAの塩基配列の相違に基づく分類体系に従い分類体系が再構築されてきている。この理由は有性生殖を行う真核生物と異なり細胞分裂で増殖する真正細菌、古細菌では明確な種の概念が設定し難く、リボゾームRNAを分子時計とする系統分類体系を基準とすることが定まったためである。旧来の生理試験結果に基づく同定法は、必ずしも現在の分類体系を正確に反映させることが困難であり、リボゾームRNAを分子時計とする系統分類体系を正確に反映した本例のような簡易分類・同定手法が必要とされる。

【0057】なお、細菌の名前は属(genus)と種(species)の二名式命名法で表記されるが、高順位の範疇として科(family)、目(class)、綱(division)、群

(group)と真正細菌門(division)の下に順次まとめられている。

【0058】本例では、次のような実験を行った。試料AP1~AP9として、様々な土壌から分離したアルキルフェノール分解菌(九州共立大学、名城大学)を用いる。また、試料MA1~MA4として、アルカリ耐性菌(山口大学)を用いる。そして、定法に従って染色体DNAを各菌株から抽出し、PCR反応を行い、増幅した16S rDNAを制限酵素で切断後各切断片の長さをデントグラムで読み取り、実測値切断長として入力した。なお、試料MA1と試料MA11とは、同じ菌株の同じDNAを用いてこれらの操作を別途行ったものである。

【0059】図17は、これらの試料から得られた実測値同士の類縁性を、本発明による遺伝子の類縁性検索システムでの解析で検定した結果を示す樹形図である。

【0060】また、図18~図27は、それぞれ、異なった試料の菌株から得られた実測値と遺伝子配列データを基に得られた理論値との類縁性の検定を行った結果を示す樹形図である。なお、図18は試料AP1について、図19は試料AP3について、図20は試料AP6について、図21は試料AP2について、図22は試料AP5について、図23は試料MA1について、図24は試料MA11について、図25は試料MA2について、図26は試料MA3について、図27は試料MA4について、それぞれ上記検定を行った結果を示す。

【0061】上記検定を行ったとき、切断長理論値データ(23)としては、357属、1233種、1503種類のDNA配列データから作成した理論値データが登録されていた。この理論値データは、すべて公開されているDNA配列データを基に本発明による遺伝子の類縁性検索システムによって算出されたものであり、細菌の種類の分類・同定の基礎とするのに十分なこれほど多種のデータを安価かつ短時間で作成することができたのは、本発明を用いたことによる大きな効果である。これに対して、例えば、従来技術による生理学的試験に基づく同定キット(市販品)の場合には、数十年にもおよぶ長い年月をかけてデータを整備することにより、ようやく1210種、200属に満たない細菌を同定できるようになっているに過ぎない。また、他のある同定キット(市販品)の場合にも、同様に約700菌種を同定できるようになっているに過ぎない。これらの従来技術の方法に基づくシステムと比べて、本発明によるシステムは、今後新たに追加されるDNA配列データも含めて、公開データを基に極めて安価かつ短時間でデータを増やすことができるという利点を持っている。

【0062】前記検定結果によると、試料MA1および試料MA11は、*Aeromonas hydrophila* (Proteobacteria subdivision, *Aeromonas* group) に最も近く、同じProteobacteria subdivisionに属するが少し離れ

たEnterobacteriaceae(腸内細菌科)に属するSerratia属、Shewanella属、またはPlesiomonas属が近いという結果が出ている。

【0063】また、試料MA2は、*Sporolactobacillus*属、または*Bacillus*属(いずれもLowGC gram positive bacteria, *Bacillus*/*Clostridium*, Bacillaceae (*Bacillus*科))に等しく近いという結果が得られている。

【0064】また一方で、試料MA3は、*Bacillus cereus*と*Bacillus thuringiensis* (LowGC gram positive bacteria, *Bacillus*/*Clostridium*, Bacillaceae)に最も近いという結果が得られている。なお、*Bacillus cereus*と*Bacillus thuringiensis*は近縁であり同一種とする意見が多い。

【0065】また、試料MA4は、*Oeskovia*属もしくは*Cellulomonas*属(共にActinobacteria, Actinobacteridae, Actinomycetales, Micrococcineae, Cellulomonadaceae)に最も近く、Actinomycetes属(Actinobacteria, Actinobacteridae, Actinomycetales, Actinomycetaceae)、*Streptomyces*属(Actinobacteria, Actinobacteridae, Actinomycetales, Streptomycetaceae)に次に近い。

【0066】試料AP1とAP3は、*Pseudomonas putida*, *Pseudomonas fulva*, *Pseudomonas straminea*, *Pseudomonas alcaligenes*, *Flavimonas oryzihabitans* (いずれもProteobacteria subdivision, Pseudomonadaceae)に近い。

【0067】また、試料AP6は、*Xanthomonas*属細菌(Proteobacteria subdivision, Xanthomonadales, *Xanthomonas* group)に近く、試料AP2とAP5は、*Phyllobacterium*属、*Rhizobium*属、*Agrobacterium*属(Proteobacteria subdivision, Rhizobiales, *Rhizobium* group)と推定された。

【0068】図28は、分離した脱窒菌134株について本方法で種名を推定し代表的な菌株の16S rDNAの部分塩基配列を決定し、公開用塩基配列データベースで相同性検索した結果最も近い16S rDNAの塩基配列構造を有する菌の種名を検索した結果を示している。

【0069】この図28では、分類されたグループ(IからXIII)、そのグループに含まれる菌数、推定される種名(RFLP)、16S rDNAの塩基配列を実際に決定し、最も近い塩基配列を有する菌の名前および相違度を示している。この場合、類似度が高い塩基配列部分の塩基配列のうち、いくつ一致したかを百分率表示で示している。複数の数値が示されているのは、そのグループに属する別の細菌の塩基配列を決定し相同性検索を別に行った結果である。

【0070】

【発明の効果】以上説明したように、この発明によると、既知の遺伝子の塩基配列と制限酵素の切断パターン

とを基に予め切断長理論値データを作成しておき、この切断長理論値データと入力される切断長実測値データとを非加重結合法などを用いて分析するため、既知遺伝子と試料との類似度を高速かつ効率的に算出することが可能となる。

【0071】また、この発明によると、多種の既知遺伝子および試料との類似度を基に、類似度の高いもの同士を関連付けて、樹形図などを用いて図形的に表示するため、解析結果を利用者に理解しやすい形で提供することが可能となる。

【0072】また、この発明によると、増幅配列認識部が増幅前塩基配列データを基に、増幅部分を自動的に認

識して増幅後塩基配列データを生成するため、多くの手間をかけることなく増幅後塩基配列データを準備することができ、塩基配列の解析をより一層効率化することが可能となる。

【0073】また、この発明によると、パーソナルコンピュータやワークステーションなどの汎用のコンピュータを用いて遺伝子の類縁性検索システムを実現することにより、専用のシーケンサ等に比べて極めて安価なシステムを提供することが可能となる。

【0074】

【配列表】

< 1 1 0 > Director General of Kyushu National Agricultural Experiment Station; Katsuji Watanabe

< 1 2 0 > Methods and systems for analyzing genetic relationship

< 1 6 0 > 4

< 2 1 0 > 1

< 2 1 1 > 1 0 0 9

< 2 1 2 > DNA

< 2 1 3 > Rhodospirillum salexigens

< 2 2 0 >

< 2 2 1 > rRNA

< 2 2 2 > (50), (51), (63), (463), (468), (574), (687), (806), (853), (854), (888), (889), (966)

< 2 2 3 > unknown

< 4 0 0 >

```

1 gctcagaacg aacgctggcg gcaggcctaa cacatgcaag tcgagcgcan nccttcgggg
61 gtnagcggcg gacgggtgag taacgcgtgg gaacctgctc agggctctgg gataactgct
121 ggaaacggca gctaataccg gatacggcgt attgggaaag aaattcggcc ttggatgggg
181 ccgcgttga ttagctagat ggtggggtaa cggcctacca tggcgacgat ccatagctgg
241 tttgagagga tgatcagcca cactgggact gagacacggc ccagactcct acgggaggca
301 gcagtgggga atcttagaca atgggggcaa ccctgatcta gccatgccgc gtgagtgatg
361 aaggccttag ggttghtaaag ctctttcagc aggaagata atgactgtac ctgcagaaga
421 agctccggct aactccgtgc cagcagccgc ggtaatcgg agngggcnag cgttgttccg
481 aattactggg cgtaaagcgc gcgtaggcgg atcggtcagt tgggggtgaa agcccggggc
541 tcaacctcgg aactgccctc aaaactaccg atcnagagtt cgggagaggt aagcgggaatt
601 cccagtgtag aggtgaaatt cgtagatatt ggaagaaca ccagtggcga agcggcctta
661 ctggaccgat actgacgctg aggtgcnaaa gcgtggggag caaacaggat tagataccct
721 ggtagtccac gccgtaaacg atgggtgcta gatgtcgggg ctcttagagt ttcggtatcg
781 cagctaaccg attaagcacc ccgcnngggg agtacggccg caagggttaa actcaaagga
841 atgacgggg gcnngcacia gcggtggagc atgtggttta attcgaanna acgcgagaa
901 ccttaccagc tcttgacatc ccgggacgac ttccagagat ggatttttc acttcgggtga
961 cccgnggaca ggtgctgcat ggctgtcgtc agctcgtgtc gtgagatgt

```

< 2 1 0 > 2

< 2 1 1 > 1 4 9 0

< 2 1 2 > DNA

< 2 1 3 > Rhodospirillum salexigens

< 2 2 0 >

< 2 2 1 > rRNA

< 2 2 2 > (1), (72), (73), (85), (485), (490), (596), (709), (828), (875), (876), (

910), (911), (988), (1045), (1332), (1333), (1408), (1437)-(1442), (1451)-(1475)

< 2 2 3 > u n k n o w n

< 4 0 0 >

1 ncaacatgag agtttgatcc tggctcagaa cgaacgctgg cggcaggcct aacacatgca
 61 agtcgagcgc annccctcgg gggtnagcgg cggacgggtg agtaacgcgt gggaaacctgc
 121 tcagggtctt gggataactg ctggaaacgg cagctaatac cggatacacc gtattgggaa
 181 agaaattcgg ccttggatgg gcccgcgttg gattagctag atggtgggtg aacgcctac
 241 catggcgacg atccatagct ggttgagag gatgatcagc cactactgga ctgagacacg
 301 gccagactc ctacgggagg cagcagtggg gaatcttaga caatgggggc aacctgatc
 361 tagccatgcc gcgtgagtga tgaaggcctt agggttgtaa agctcttca gcaggaaga
 421 taatgactgt acctgcagaa gaagctccgg ctaactccgt gccagcagcc gcggtaatc
 481 ggagngggcn agcgttgttc ggaattactg ggcgtaaac gcgcgtaggc ggatcggtca
 541 gttgggggtg aaagcccggg gctcaacctc ggaactgcc tcaaaactac cgatcnagag
 601 ttcgggagag gtaagcggaa ttcccagtgt agaggtgaaa ttcgtagata ttgggaagaa
 661 caccagtggc gaaggcggct tactggaccg atactgacgc tgaggtgcna aagcgtgggg
 721 agcaaacagg attagatacc ctggtagtcc acgccgtaaa cgatgggtgc tagatgtcgg
 781 ggctcttaga gtttcggtat cgcagctaac gcattaagca ccccgccngg ggagtacggc
 841 cgcaaggta aaactcaaag gaattgacgg gggcngcac aagcgggtga gcattgggtt
 901 taattcgaan naacgcgcag aacctacca gctctgaca tccgggagc acttccagag
 961 atggattttt tcaactcggg gaccggnga cagggtgctgc atgggtgtcg tcagctcgtg
 1021 tcgtgagatg ttgggtaag tccncaacg agcgcaacc tcgcccttag ttgccagcat
 1081 ttggttgggg actctaaggg aactgccggg gataagccgg aggaagggtg ggatgacgtc
 1141 aagtcctcat ggcccttatg ggcgggcta cacacgtgct acaatggcgg tgacagaggg
 1201 cagcagcct gcgagggtga gcgaatctt aaaagccgtc tcagttcggg ttgttctctg
 1261 caactcgaga gcatgaagg ggaatcgcta gtaatcgcg atcagcatgc cgcggtgaat
 1321 acgttcccgg gnntgtaca caccgccgt cacaccatgg gagggtgtt gacccgaaga
 1381 cggtagccta acccgaagg gggcagncg gccacggtca ggtcagcagc tggggtnnnn
 1441 nngtaacaag nnnnnnnnn nnnnnnnnn nnnngatca cctccttct

< 2 1 0 > 3

< 2 1 1 > 1 9

< 2 1 2 > DNA

< 2 1 3 > A r t i f i c i a l S e q u e n c e

< 2 2 0 >

< 2 2 3 > A r t i f i c i a l S e q u e n c e

< 4 0 0 >

1 gctcagattg aactcggcg

< 2 1 0 > 4

< 2 1 1 > 1 6

< 2 1 2 > DNA

< 2 1 3 > A r t i f i c i a l S e q u e n c e

< 2 2 0 >

< 2 2 3 > A r t i f i c i a l S e q u e n c e

< 4 0 0 >

1 acattcaca acacgagctg

【図面の簡単な説明】

【図1】 この発明の第一実施形態による塩基配列解析システムの構成を示すブロック図である。

【図2】 同実施形態による増幅後塩基配列データ(22b)の前半部分を示す図である。

【図3】 同実施形態による増幅後塩基配列データ(2

2b)の後半部分を示す図である。

【図4】 同実施形態による制限酵素データ(21)のデータ構造およびデータ例を示す表図である。

【図5】 同実施形態による理論値計算部(11)に対して切断長理論値計算を指示する利用者インターフェース例を示す画面図である。

【図6】 同実施形態による切断長理論値データ(23)のデータ構造およびデータ例を示す表図である。

【図7】 同実施形態による実測値入力部(12)の利用者インタフェース例を示す画面図である。

【図8】 同実施形態による切断長実測値データ(25)のデータ構造およびデータ例を示す表図である。

【図9】 同実施形態による塩基配列のクラスタリングの基となる切断片のバンド分布を表す参考図である。

【図10】 同実施形態により塩基配列間の類似度を基に平均距離法を用いてクラスタ分析を行う過程を示す表図である。

【図11】 同実施形態によるクラスタ分析の結果を図形化して示した参考図である。

【図12】 同実施形態による表示部(15)が解析結果を樹形図として表示する例を示す画面図である。

【図13】 この発明の第二実施形態による塩基配列解析システムの構成を示すブロック図である。

【図14】 同実施形態による塩基配列解析システムの入力データとなる増幅前塩基配列データ(22a)の前半部分を示す図である。

【図15】 同実施形態による塩基配列解析システムの入力データとなる増幅前塩基配列データ(22a)の後半部分を示す図である。

【図16】 同実施形態による増幅配列認識部(31)による認識のためのプライマ配列の入力画面を示す画面図である。

【図17】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図18】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図19】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図20】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図21】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形

図である。

【図22】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図23】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図24】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図25】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図26】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図27】 本発明による塩基配列解析システムを用いて、細菌の種の分類・同定を行った例の結果を示す樹形図である。

【図28】 分離した脱窒菌について本方法で種名を推定し代表的な菌株の16S rDNAの部分塩基配列を決定し、公開用塩基配列データベースで相同性検索した結果最も近い16S rDNAの塩基配列構造を有する菌の種名を検索した結果を示している表図である。

【符号の説明】

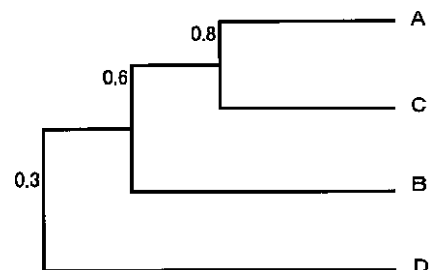
- 1 塩基配列解析システム
- 11 理論値計算部
- 12 実測値入力部
- 13 比較部
- 14 許容誤差設定部
- 15 表示部
- 21 制限酵素データ
- 22a 増幅前塩基配列データ
- 22b 増幅後塩基配列データ
- 23 切断長理論値データ
- 25 切断長実測値データ
- 27 解析結果データ
- 29 許容誤差データ
- 31 増幅配列認識部

【図8】

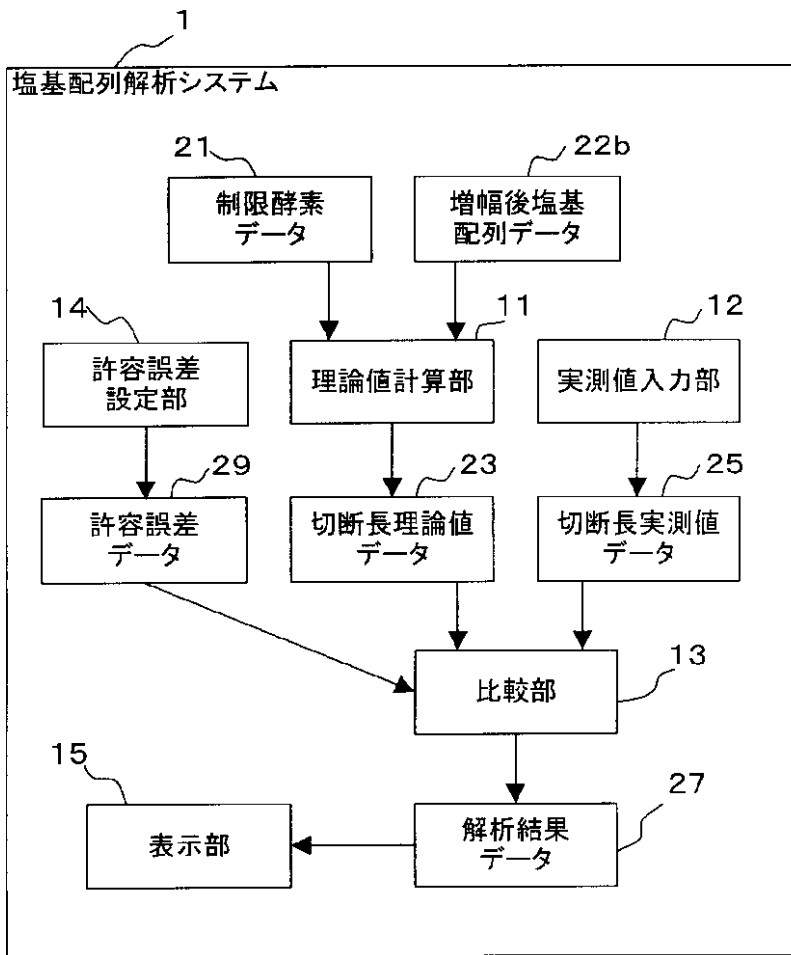
切断長実測値データ

DNA番号	数値区分	グループ名	細菌名	遺伝子名	制限酵素番号	切断長
AP1	2	AP			1	100
AP1	2	AP			1	200
AP1	2	AP			1	300

【図11】



【図1】

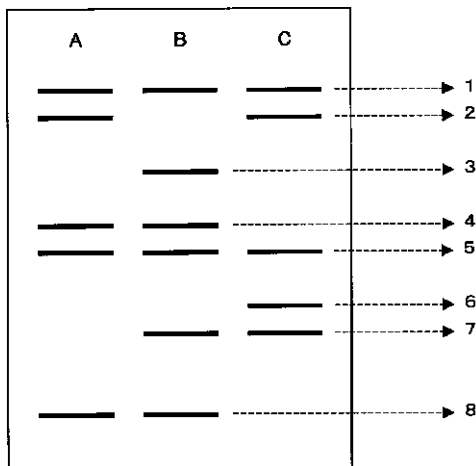


【図4】

制限酵素データ

制限酵素番号	制限酵素名	切断箇所	切断位置
1	AluI	agct	2
2	HaeIII	ggcc	2
3	RsaI	gtac	2
4	SrfI	ccngg	2
5	HhaI	gcgc	2
6	BamHI	ggatcc	1
7	EcoRI	gaatt	1
8	HindIII	aagctt	1
9	PstI	ctgacg	5
10	PvuII	cagctg	3
11	SalI	gagctc	5
12	SmaI	ccggg	3
13	XbaI	tctaga	1

【図9】



【図2】

LOCUS RSP16SRZ 1490 bp rRNA BCT 04-DEC-1995 ORIGIN
 DEFINITION Rhodospirillum salexigens 16S ribosomal RNA.
 ACCESSION M59070
 NID g175871
 VERSION M59070.1 GI:175871
 KEYWORDS 16S ribosomal RNA.
 SOURCE Rhodospirillum salexigens rRNA.
 ORGANISM Rhodothalassium salexigens
 Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 Rhodothalassium.
 REFERENCE 1 (bases 1 to 1490)
 AUTHORS Woese, C. R.
 TITLE A phylogenetic analysis of the some purple bacteria
 JOURNAL Unpublished (1991)
 FEATURES Location/Qualifiers
 source 1..1490
 /organism="Rhodothalassium salexigens"
 /db_xref="taxon:1086"
 /tissue_lib="DSM 2132"
 rRNA 1..1490
 /gene="16S rRNA"
 /product="16S ribosomal RNA"
 gene 1..1490
 /gene="16S rRNA"
 BASE COUNT 342 a 343 c 472 g 284 t 49 others

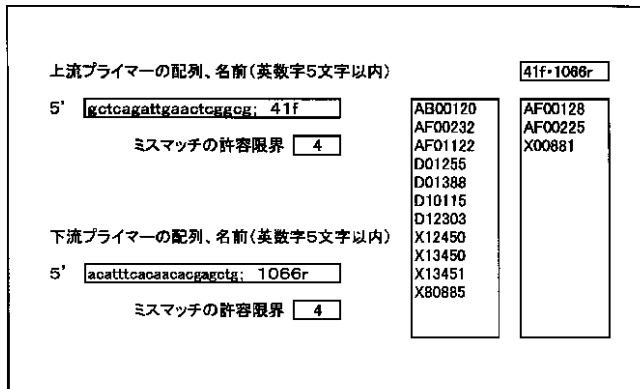
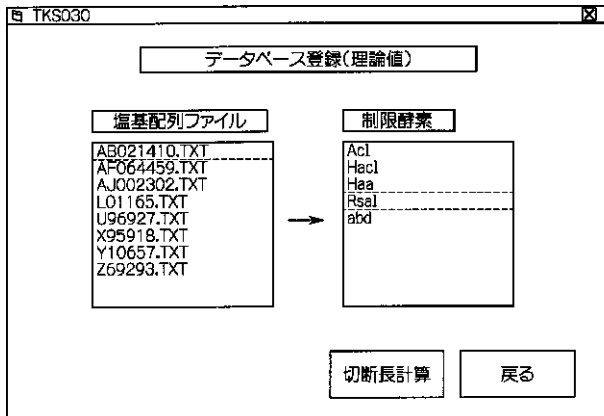
【図3】

```

1 gctcagaacg aacgctggcg gcaggcctaa cacatgcaag tcgagcgcan nccttcgggg
61 gtnagcggcg gacgggtgag taacgcgtgg gaacctgctc agggctctgg gataactgct
121 ggaacggcca gctaataccg gatacgcgtt atgggaaag aaattcggcc ttggatgggc
181 ccgcgttggg ttagctagat ggiggggtaa ccgcctacca tggcgacgat ccatacgtgg
241 tttagagaga tgatcagcca cactgggact gagacacggc ccagactcct accgaaagca
301 gcagtgggga atotttagaca atgggggcaa ccctgatcta gccatgccgc gtgagtgatg
361 aagccottag ggttgaagag ctccttcagc agggaagata atgactgtac ctgcagaaga
421 agctccgact aactcgcgtc cagcagccgc ggttaaacgg agrgggcnag cgttgttcgg
481 aattactggg cgtaaagcgc gcgtaggcgg atcgctcagt tgggggtgaa agccccgggc
541 tcaacctcgg aactgcctc aaaactaccg atcnagagtt cgggagaggt aagcgaatt
601 cccagigtg aggtgaatt cgtagatatt ggaagaaca ccagtgccga agccgctta
661 ctggacgat actgaagcgt aggtgchnaa gcgtggggag caaacaggat tagataccct
721 ggtagtccac gccgtaaagc atgggtgcta gatgicgggg cctcttaggt ttccgtatcg
781 cagctaagc attaagcacc ccgcongggg agtacggccg caagggtaaa actcaaaaga
841 attgacgggg gcnngcaca gcggggggag atgtggttta attcgaanna acgcccagaa
901 ccttaccagc ttgtgacatc ccgggacgac ttccagagat ggatttttc acttcggtga
961 cccgngaca ggtgctgcat ggcctgctgc agctcgtctc gtgagatgt
  
```

【図5】

【図16】



【図6】

切断長理論値データ

DNA番号	数値区分	グループ名	細菌名	遺伝子名	制限酵素番号	切断長
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	2	204
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	5	509
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	5	542
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	3	194
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	3	457
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	3	400
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	2	156
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	2	90
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	2	317
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	2	169
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	1	550
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	1	180
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	1	79
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	1	67
AB000278	1		Vibrio illopiscarius	16S rRNA; 16S ribosomal RNA	2	105
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	158
AB038030	1		Vibrio splendidus	16S ribosomal RNA	5	509
AB038030	1		Vibrio splendidus	16S ribosomal RNA	5	565
AB038030	1		Vibrio splendidus	16S ribosomal RNA	4	1074
AB038030	1		Vibrio splendidus	16S ribosomal RNA	3	194
AB038030	1		Vibrio splendidus	16S ribosomal RNA	3	234
AB038030	1		Vibrio splendidus	16S ribosomal RNA	3	223
AB038030	1		Vibrio splendidus	16S ribosomal RNA	3	423
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	90
AB038030	1		Vibrio splendidus	16S ribosomal RNA	1	175
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	204
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	317
AB038030	1		Vibrio splendidus	16S ribosomal RNA	2	105
AB038030	1		Vibrio splendidus	16S ribosomal RNA	1	236
AB038030	1		Vibrio splendidus	16S ribosomal RNA	1	572

【図10】

(a)

	A	B	C	D
A	—	—	—	—
B	0.7	—	—	—
C	0.8	0.5	—	—
D	0.1	0.4	0.3	—

(b)

	A+C	B	D
A+C	—	—	—
B	0.6 (*1)	—	—
D	0.2 (*2)	0.4	—

*1 $S(B, A+C) = \frac{S(B, A) + S(B, C)}{2}$

*2 $S(D, A+C) = \frac{S(D, A) + S(D, C)}{2}$

(c)

	(A+C)+B	D
(A+C)+B	—	—
D	0.3 (*3)	—

*3 $S(D, (A+C)+B) = \frac{S(D, A+C) + S(D, B)}{2}$

【図7】

TKS040

データベース登録(実測値)

分子量ファイル
Deni01(Ha).txt
DNA.xls

グループ名
AP

DNA番号
AP1

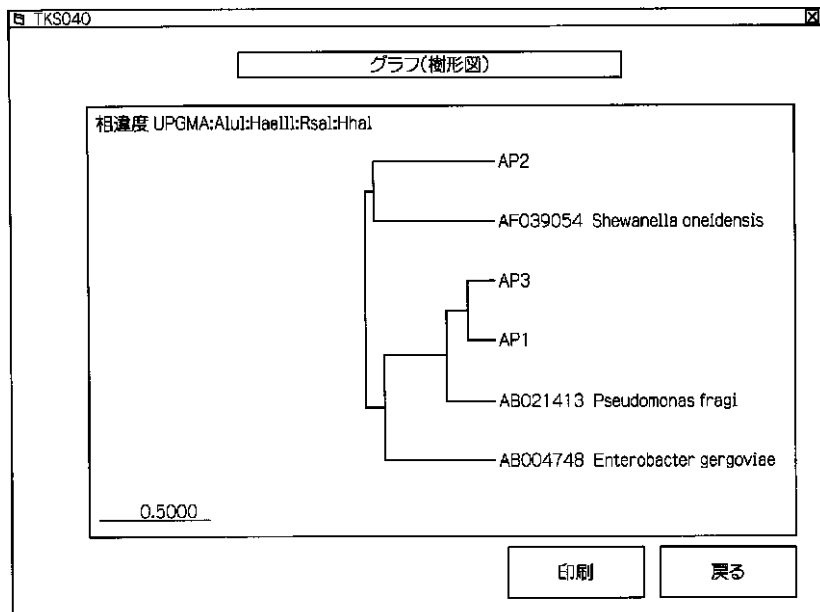
制限酵素
AluI

切断長

100	200	300		

登録 戻る

【図12】



【図18】

相違度 UPGMA:AluI:HaeIII:RsaI:HhaI



【図14】

【図15】

LOCUS RSP16SRRZ 1490 bp rRNA BCT 04-DEC-1995 ORIGIN
 DEFINITION Rhodospirillum salexigens 16S ribosomal RNA.
 ACCESSION M59070
 NID g175871
 VERSION M59070.1 GI:175871
 KEYWORDS 16S ribosomal RNA.
 SOURCE Rhodospirillum salexigens rRNA.
 ORGANISM Rhodothalassium salexigens
 Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 Rhodothalassium.
 REFERENCE 1 (bases 1 to 1490)
 AUTHORS Woese, C. R.
 TITLE A phylogenetic analysis of the some purple bacteria
 JOURNAL Unpublished (1991)
 FEATURES Location/Qualifiers
 source 1..1490
 /organism="Rhodothalassium salexigens"
 /db_xref="taxon:1086"
 /tissue_lib="DSM 2132"
 rRNA 1..1490
 /gene="16S rRNA"
 /product="16S ribosomal RNA"
 gene 1..1490
 /gene="16S rRNA"
 BASE COUNT 342 a 343 c 472 g 284 t 49 others

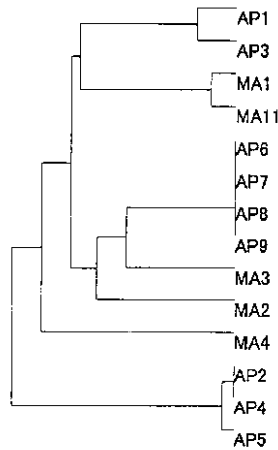
```

1 noaacatgag agtttgatcc tggctcagaa cgaacgctgg cggcaggcct aacacatgca
61 agtcgagcgc annocctcgg sggtnagcgg cggacgggtg agtaacgctg gggaaacctg
121 tcagggtctt gggataactg ctggaaacgg cagctaatac cggatacggc gtattgggaa
181 agaaattcgg ccttggatgg gcccgcttgg gattagctag atggtagggg aacggcctac
241 catggcgagc atccatagct ggtttgagag gatgatcagc cacactggga ctgagacagc
301 gcccgagact ctacggggag cagcagtgag gaactitaga caatgggggc aacctgac
361 tagccatgcc gcgtgagtga tgaaggcctt agggttgtaa agctctttca gcagggaaga
421 taalgactgt acctgcagaa gaagctccgg ctaactcogt gccagcagcc gcgtaatac
481 ggagnggcn agcgttgttc ggaattactg ggcgtaaac gcgctaggc ggaatcggtca
541 gtgggggtg aaagccggg gctcaacctc ggaactgccc tcaaaactac cgatcnagag
601 ttggggagag gtaagcggaa ttcccagtgat agaggtagaa ttctagata ttgggaagaa
661 caccagtgcc gaaagcggct tactggaccg atactgaccg tgaggtagca aagcgtgggg
721 agcaaacagg attagatacc ctggtagtcc acgcgctaaa cgatgggtgc tagatgtcgg
781 ggctctttag gtttcgggat cgcagctaac gcattaagca ccccgccngg ggaatagcgc
841 cgcaaggtta aaactcaag gaattgacgg gggcnggcac aagcgttggc gcatgtggtt
901 taattcgaan naacgocag aacctiacca gctcttgaca tcccgggacg acttccagag
961 atggattttt tcacttcggt gaccogngca caggctctgc atgctgtcgc tcagctcgtg
1021 tctgtagatg ttgggitaag tccncaacg agcgaaccc tcgoccttag ttgccagcat
1081 ttggttgggg acttaaggg aactgccgtt gataagccgg aggaaggtgg ggaatgacgc
1141 aagtcctcat ggcccttatg ggctgggcta cacacgtgct acaatggcgg tgacagaggg
1201 cagcagcctt gcgagggtag cgaatctct aaaagcgcgc tcagttcggc ttgttctctg
1261 caactcgaga gcatgaaggt ggaatcgcta gtaatcgccg atcagcatgc cgcggtgaat
1321 acgttcccgg gnntgtaca caccgccgtt cacaccatgg gagttggttt gaccogaaga
1381 cggtagccta accogaagg ggggcagnoc gccacggta ggtcagcgac tggggtnnnn
1441 nngtaacaag nnnnnnnnnn nnnnnnnnnn nnnngatac cctctctttt
  
```

//

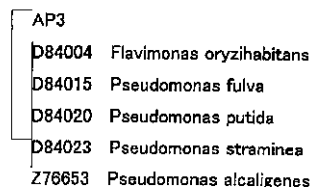
【図17】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI

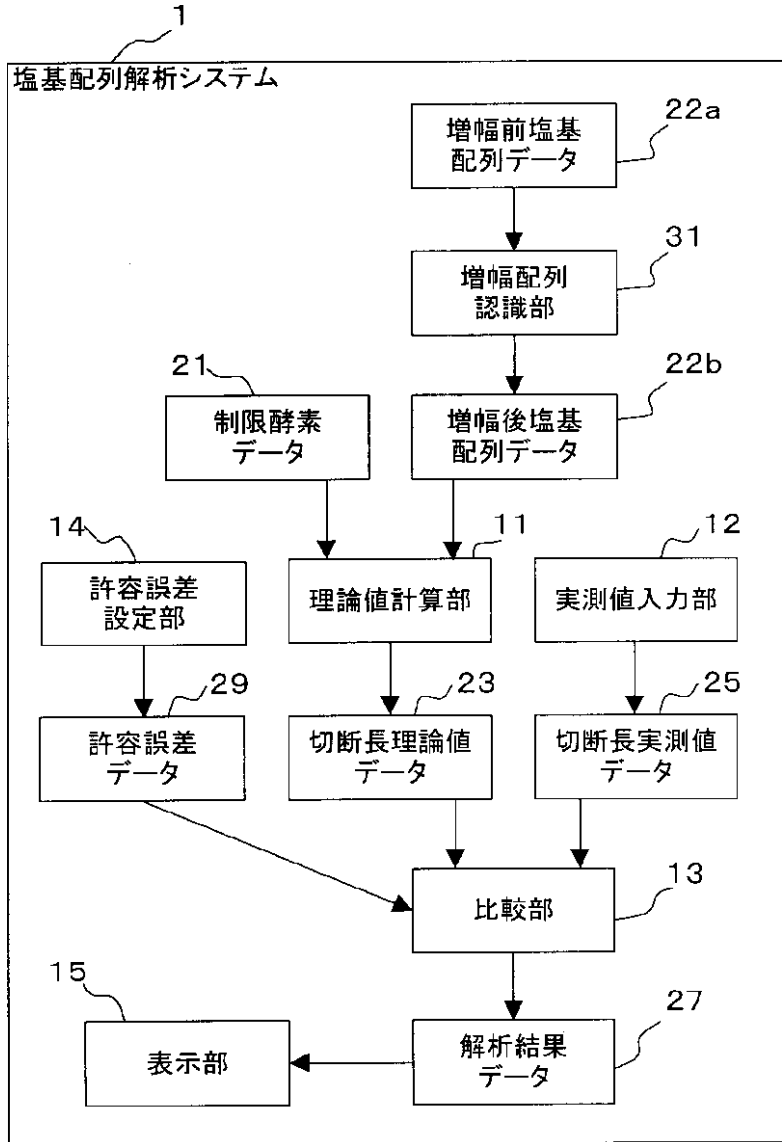


【図19】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



【図13】



【図20】

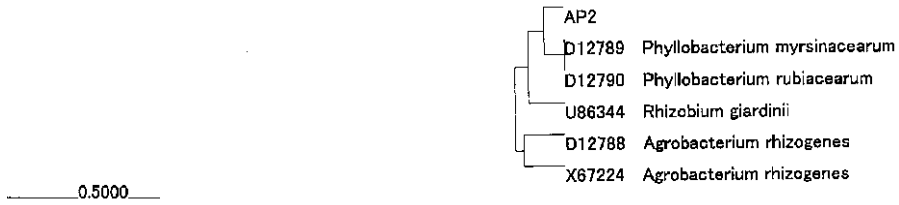
相違度 UPGMA: AluI: HaeIII: RsaI: HhaI

0.5000

- AP6
- X95919 *Xanthomonas axonopodis*
- X95921 *Xanthomonas oryzae*
- Y10761 *Xanthomonas vesicatoria*
- Y10763 *Xanthomonas theicola*
- Y10765 *Xanthomonas codiae*

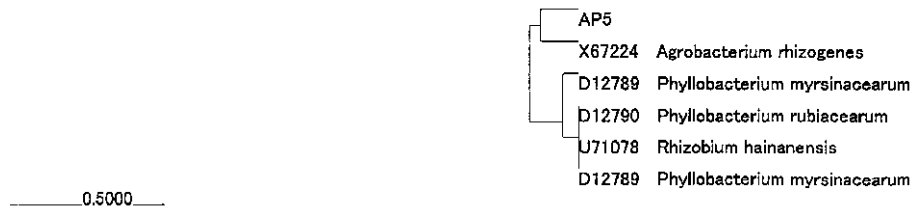
【図21】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



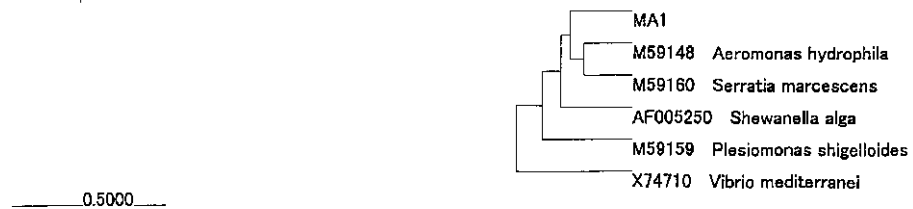
【図22】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



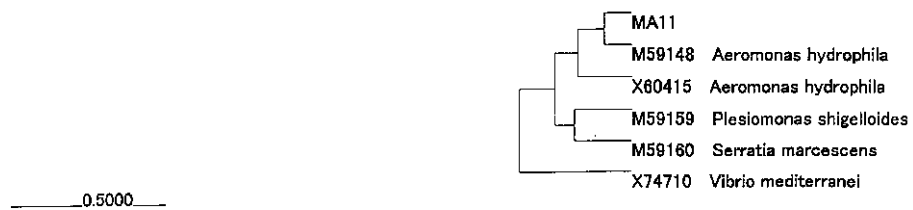
【図23】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



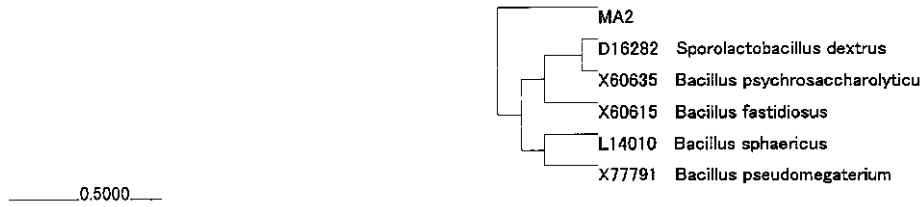
【図24】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



【図25】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



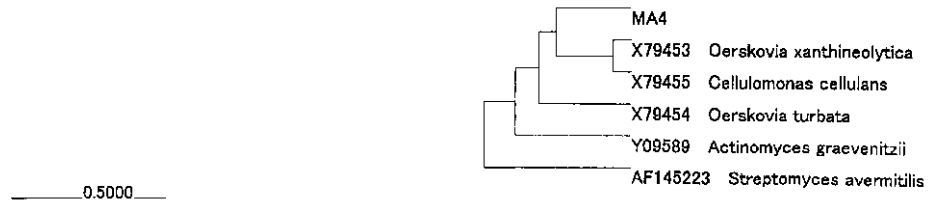
【図26】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



【図27】

相違度 UPGMA: AluI: HaeIII: RsaI: HhaI



【図28】

分離脱窒菌の16S rDNAのRFLP (制限酵素 *Hae*III, *Hha*I, *Alu*I, *Rsa*I, *Srf*I) による分類と塩基配列の相同性検索による同定結果の比較

グループ	菌数	RFLP	塩基配列 (%ホモロジー)
I	20	<i>Enterobacteriaceae</i>	<i>Klebsiella</i> 属 (100%)
II	9	<i>Burkholderia</i> 属	<i>B. vietnamiensis</i> (96.9%, 92.6%, 93.0%, 94.8%, 93.4%), <i>Burkholderia</i> 属 (100%)
III+IV	12	<i>Ralstonia</i> 属	<i>Ralstonia</i> 属 (92.0%, 94.5%, 94.5%), <i>R. paucula</i> (95.1%, 93.8%), <i>R. eutropha</i> (95.6%, 100%, 96.7%)
V	6	<i>Comamonas acidovorans</i>	<i>C. acidovorans</i> (98.2%, 100%)
VI+VIII	32	<i>Pseudomonas</i> 属	<i>P. putida</i> (97.7%, 99.0%, 99.2%), <i>P. fluorescens</i> (95.8%, 99.5%), <i>P. rhodesiae</i> (98.4%)
VII	20	<i>P. putida</i>	<i>P. putida</i> (100%)
IX	8	<i>P. rhodesiae</i>	<i>P. rhodesiae</i> (98.5%, 99.5%)
X	5	<i>P. stutzeri</i>	<i>P. stutzeri</i> (98.0%, 94.6%, 92.0%)
XI	3	<i>Acinetobacter haemolyticus</i>	<i>A. haemolyticus</i> (96.1%)
XII	18	<i>Pseudomonas</i> 属	<i>Pseudomonas</i> 属 (99.5%)
XIII	1	<i>Acivorax delafieldii</i>	<i>Acivorax delafieldii</i> (94.7%)

フロントページの続き

(56)参考文献 特開 平10 - 262699 (J P , A)
渡辺克二、奥田充、古賀伸久、制限酵素切断長多型 (R F L P) を利用した土壌細菌群集の解析手法、農業環境研究成果情報、日本、農業環境技術研究所、2000年 7月10日、第16集、p . 39 - 40
渡辺日出海、国沢隆、コンピュータ解析による大腸菌遺伝子のゲノム上の位置と転写方向の決定、情報学シンポジウム講演論文集、日本、1990年 1月17日、1990、p . 89 - 98

(58)調査した分野(Int.Cl.⁷, D B 名)
G06F 17/30
C12N 15/09
J I C S T ファイル (J O I S)