(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号

特許第5288355号
(P5288355)

(45)発行日　平成25年9月11日(2013.9.11)　(24)登録日　平成25年6月14日(2013.6.14)

(51)Int.Cl.		F I		
G06F 19/22	(2011.01)	G06F	19/22	
G06F 17/30	(2006.01)	G06F	17/30	170F
C12M 1/00	(2006.01)	C12M	1/00	ZNAA
C12Q 1/68	(2006.01)	C12Q	1/68	Z
C12N 15/09	(2006.01)	C12N	15/00	A

請求項の数 9　(全 21 頁)

(21)出願番号　特願2009-539130 (P2009-539130)
(86)(22)出願日　平成20年10月31日 (2008.10.31)
(86)国際出願番号　PCT/JP2008/069897
(87)国際公開番号　WO2009/057757
(87)国際公開日　平成21年5月7日 (2009.5.7)
審査請求日　平成23年5月17日 (2011.5.17)
(31)優先権主張番号　特願2007-283480 (P2007-283480)
(32)優先日　平成19年10月31日 (2007.10.31)
(33)優先権主張国　日本国(JP)

（出願人による申告）平成１９年度農林水産技術会議事務局「アグリ・ゲノム研究の総合的な推進（有用遺伝子活用のためのイネゲノム研究・ゲノム育種による効率的品種育成技術の開発）」委託事業、産業技術力強化法第１９条の適用を受けるもの

(73)特許権者　501167644
独立行政法人農業生物資源研究所
茨城県つくば市観音台２丁目１－２
(74)代理人　100089118
弁理士　酒井　宏明
(72)発明者　宮尾　安▲藝▼雄
茨城県つくば市観音台２丁目１－２　独立行政法人農業生物資源研究所内

審査官　宮久保　博幸

最終頁に続く

(54)【発明の名称】塩基配列決定プログラム、塩基配列決定装置および塩基配列決定方法

(57)【特許請求の範囲】
【請求項１】
制御手段と記憶手段とを備えた情報処理装置に実行させる塩基配列決定プログラムであって、
前記記憶手段は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶し、
前記制御手段に、
対象とする前記親個体の前記塩基配列である対象塩基配列に基づいて、前記記憶手段に記憶した前記データベースから、当該対象塩基配列に繋ぎ合せる前記塩基配列を検索する検索ステップと、
前記検索ステップで検索した前記塩基配列に基づいて、前記親個体間における前記遺伝的多型を検出する検出ステップと、
前記検索ステップで検索した前記塩基配列に同一の前記親個体から検索したものが複数存在するか否かを確認する確認ステップと、
前記検出ステップで前記遺伝的多型が検出され且つ前記確認ステップで複数存在すると確認されなかった場合、前記検索ステップで検索した前記後世代個体の前記塩基配列に基づいて、前記遺伝的多型が検出された部位での前記後世代個体の遺伝子型を調査する調査ステップと、
前記調査ステップで調査した前記後世代個体の前記遺伝子型と既に調査済みの前記後世代個体の前記遺伝子型とが一致するか否かを判定する判定ステップと、

前記判定ステップで一致すると判定された場合、前記検索ステップで検索した前記親個体の前記塩基配列に基づいて前記対象塩基配列を伸長する、および、前記確認ステップで複数存在すると確認された場合、前記検索ステップで検索した同一の前記親個体の前記塩基配列ごとに前記対象塩基配列を別々に伸長する伸長ステップと、
を実行させること
を特徴とする塩基配列決定プログラム。
【請求項２】
前記データベースは、前記個体の前記塩基配列と前記個体を一意に識別するための個体識別情報とを相互に関連付けてなるリレーショナルデータベースであり、
前記検索ステップは、データベース言語であるＳＱＬで前記塩基配列を検索すること
を特徴とする請求項１に記載の塩基配列決定プログラム。
【請求項３】
前記記憶手段は、前記データベースに記憶した前記塩基配列に対して作成された、前方一致検索が可能なインデックスをさらに記憶し、
前記検索ステップは、前記インデックスを参照して前記塩基配列を前方一致検索すること
を特徴とする請求項２に記載の塩基配列決定プログラム。
【請求項４】
制御手段と記憶手段とを備えた塩基配列決定装置であって、
前記記憶手段は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶し、
前記制御手段は、
対象とする前記親個体の前記塩基配列である対象塩基配列に基づいて、前記記憶手段に記憶した前記データベースから、当該対象塩基配列に繋ぎ合せる前記塩基配列を検索する検索手段と、
前記検索手段で検索した前記塩基配列に基づいて、前記親個体間における前記遺伝的多型を検出する検出手段と、
前記検索手段で検索した前記塩基配列に同一の前記親個体から検索したものが複数存在するか否かを確認する確認手段と、
前記検出手段で前記遺伝的多型が検出され且つ前記確認手段で複数存在すると確認されなかった場合、前記検索手段で検索した前記後世代個体の前記塩基配列に基づいて、前記遺伝的多型が検出された部位での前記後世代個体の遺伝子型を調査する調査手段と、
前記調査手段で調査した前記後世代個体の前記遺伝子型と既に調査済みの前記後世代個体の前記遺伝子型とが一致するか否かを判定する判定手段と、
前記判定手段で一致すると判定された場合、前記検索手段で検索した前記親個体の前記塩基配列に基づいて前記対象塩基配列を伸長する、および、前記確認手段で複数存在すると確認された場合、前記検索手段で検索した同一の前記親個体の前記塩基配列ごとに前記対象塩基配列を別々に伸長する伸長手段と、
を備えたこと
を特徴とする塩基配列決定装置。
【請求項５】
前記データベースは、前記個体の前記塩基配列と前記個体を一意に識別するための個体識別情報とを相互に関連付けてなるリレーショナルデータベースであり、
前記検索手段は、データベース言語であるＳＱＬで前記塩基配列を検索すること
を特徴とする請求項４に記載の塩基配列決定装置。
【請求項６】
前記記憶手段は、前記データベースに記憶した前記塩基配列に対して作成された、前方一致検索が可能なインデックスをさらに記憶し、
前記検索手段は、前記インデックスを参照して前記塩基配列を前方一致検索すること
を特徴とする請求項５に記載の塩基配列決定装置。

【請求項７】
制御手段と記憶手段とを備えた情報処理装置で実行する塩基配列決定方法であって、
前記記憶手段は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶し、
前記制御手段で、
対象とする前記親個体の前記塩基配列である対象塩基配列に基づいて、前記記憶手段に記憶した前記データベースから、当該対象塩基配列に繋ぎ合せる前記塩基配列を検索する検索ステップと、
前記検索ステップで検索した前記塩基配列に基づいて、前記親個体間における前記遺伝的多型を検出する検出ステップと、
前記検索ステップで検索した前記塩基配列に同一の前記親個体から検索したものが複数存在するか否かを確認する確認ステップと、
前記検出ステップで前記遺伝的多型が検出され且つ前記確認ステップで複数存在すると確認されなかった場合、前記検索ステップで検索した前記後世代個体の前記塩基配列に基づいて、前記遺伝的多型が検出された部位での前記後世代個体の遺伝子型を調査する調査ステップと、
前記調査ステップで調査した前記後世代個体の前記遺伝子型と既に調査済みの前記後世代個体の前記遺伝子型とが一致するか否かを判定する判定ステップと、
前記判定ステップで一致すると判定された場合、前記検索ステップで検索した前記親個体の前記塩基配列に基づいて前記対象塩基配列を伸長する、および、前記確認ステップで複数存在すると確認された場合、前記検索ステップで検索した同一の前記親個体の前記塩基配列ごとに前記対象塩基配列を別々に伸長する伸長ステップと、
を実行すること
を特徴とする塩基配列決定方法。
【請求項８】
前記データベースは、前記個体の前記塩基配列と前記個体を一意に識別するための個体識別情報とを相互に関連付けてなるリレーショナルデータベースであり、
前記検索ステップは、データベース言語であるＳＱＬで前記塩基配列を検索すること
を特徴とする請求項７に記載の塩基配列決定方法。
【請求項９】
前記記憶手段は、前記データベースに記憶した前記塩基配列に対して作成された、前方一致検索が可能なインデックスをさらに記憶し、
前記検索ステップは、前記インデックスを参照して前記塩基配列を前方一致検索すること
を特徴とする請求項８に記載の塩基配列決定方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、ゲノム全体の塩基配列を決定する塩基配列決定プログラム、塩基配列決定装置および塩基配列決定方法に関するものである。
【背景技術】
【０００２】
近年、ヒトを含む多くの生物の全ゲノム塩基配列が決定されている。これらの塩基配列の決定には、塩基配列決定装置（シーケンサー）が用いられている。現在、塩基配列決定装置は、サンプルを解析して、１つのサンプルあたり５００～１５００ｂｐ程度の塩基配列を得ることができる。全ゲノム塩基配列は、非常に多くのサンプルを解析して個々の塩基配列データの同じ部分（相同部分）をオーバーラップさせて繋ぎ合わせることにより、完成する。塩基配列決定装置から得られた５００～１５００ｂｐ程度の塩基配列に基づくオーバーラップさせる部分の検出とその繋ぎ合わせには、Ｐｈｒｅｄ／Ｐｈｒａｐ（非特許文献１）、Ｃａｐ３（非特許文献２）、Ａｒａｃｈｎｅ（非特許文献３）等のプログラ

ムを用いるのが一般的である。また、ＢＬＡＴという、ＢＬＡＳＴのようなアライメントツールが存在する（非特許文献４）。ＢＬＡＴでは、ゲノムをオーバーラップしないＫ－ｍｅｒに分解し、ＲＡＭ上にインデックスを置くことで、処理を高速化している。

【０００３】

Ｐｈｒｅｄ／Ｐｈｒａｐは、事実上標準となっているプログラムであり、塩基配列の重なり合いをＳｍｉｔｈ－Ｗａｔｅｒｍａｎアルゴリズムで計算し、それぞれの塩基の品質データを勘案しながら連結した塩基配列を出力する。Ｃａｐ３は、個々の塩基配列の末端領域に存在する不確かな部分を排除しながら塩基配列を連結することにより、より確度の高い塩基配列を出力する。しかし、Ｐｈｒｅｄ／ＰｈｒａｐおよびＣａｐ３は、まったく同一の繰り返し配列が存在するとそれを見分けることが出来ない。一方、Ａｒａｃｈｎｅは、各サンプルの塩基配列を両端から解析し、その解析情報を加えて塩基配列を連結する。そのため、Ａｒａｃｈｎｅは、繰り返し配列が存在していても、比較的正確に塩基配列を連結することができる。

【０００４】

【非特許文献１】Ｅｗｉｎｇ Ｂ, Ｇｒｅｅｎ Ｐ.，「Ｂａｓｅ－ｃａｌｌｉｎｇ ｏｆ ａｕｔｏｍａｔｅｄ ｓｅｑｕｅｎｃｅｒ ｔｒａｃｅｓ ｕｓｉｎｇ ｐｈｒｅｄ. ＩＩ. Ｅｒｒｏｒ ｐｒｏｂａｂｉｌｉｔｉｅｓ.」, Ｇｅｎｏｍｅ Ｒｅｓ., ８（３）, １８６－１９４, １９９８.

【非特許文献２】Ｈｕａｎｇ Ｘ, Ｍａｄａｎ Ａ.，「ＣＡＰ３： Ａ ＤＮＡ ｓｅｑｕｅｎｃｅ ａｓｓｅｍｂｌｙ ｐｒｏｇｒａｍ.」, Ｇｅｎｏｍｅ Ｒｅｓ., ９（９）, ８６８－８７７, １９９９.

【非特許文献３】Ｂａｔｚｏｇｌｏｕ Ｓ ｅｔ ａｌ.，「ＡＲＡＣＨＮＥ： ａ ｗｈｏｌｅ－ｇｅｎｏｍｅ ｓｈｏｔｇｕｎ ａｓｓｅｍｂｌｅｒ.」, Ｇｅｎｏｍｅ Ｒｅｓ., １２（１）, １７７－１８９, ２００２.

【非特許文献４】Ｗ. Ｊａｍｅｓ Ｋｅｎｔ，「ＢＬＡＴ － Ｔｈｅ ＢＬＡＳＴ－Ｌｉｋｅ Ａｌｉｇｎｍｅｎｔ Ｔｏｏｌ」, Ｇｅｎｏｍｅ Ｒｅｓ., １２, ６５６－６６４, ２００２

【発明の開示】

【発明が解決しようとする課題】

【０００５】

しかしながら、これらのプログラムは、ゲノムを分割した１００ｋｂｐ程度のＢＡＣ（バクテリア人工染色体）クローンの解析（換言すると、１００ｋｂｐ程度のクローンを細分化して得られた塩基配列の繋ぎ合わせ）には適しているが、これらのプログラムでは、一度にゲノム全体の塩基配列を決定しようとすると、ゲノム上に散在する繰り返し配列のため、各塩基配列を正しく連結することができない、という問題点があった。すなわち、これらのプログラムでは、ゲノム全体を一度に細分化して得られた塩基配列を繋ぎ合わせてゲノム全体の塩基配列を再構築することは、ゲノム上に繰り返し配列が存在することから困難である、という問題点があった。

【０００６】

本発明は上記問題点に鑑みてなされたもので、数十塩基程度の短い大量の塩基配列から既存の塩基配列を参照することなく全ゲノム塩基配列を構築することができる塩基配列決定プログラム、塩基配列決定装置および塩基配列決定方法を提供することを目的とする。

【課題を解決するための手段】

【０００７】

上記目的を達成するために、本発明にかかる塩基配列決定プログラムは、制御手段と記憶手段とを備えた情報処理装置に実行させる塩基配列決定プログラムであって、前記記憶手段は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶し、前記制御手段に、前記記憶手段に記憶した前記データベースを参照しながら、複数の前記塩基配列を、前記遺伝的多型を指標にして繋ぎ合わせる塩基配列決定方法を実行させることを特徴とする。

【０００８】
また、本発明にかかる塩基配列決定プログラムは、制御手段と記憶手段とを備えた情報処理装置に実行させる塩基配列決定プログラムであって、前記記憶手段は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶し、前記制御手段に、対象とする前記親個体の前記塩基配列である対象塩基配列に基づいて、前記記憶手段に記憶した前記データベースから、当該対象塩基配列に繋ぎ合せる前記塩基配列を検索する検索ステップと、前記検索ステップで検索した前記塩基配列に基づいて、前記親個体間における前記遺伝的多型を検出する検出ステップと、前記検索ステップで検索した前記塩基配列に同一の前記親個体から検索したものが複数存在するか否かを確認する確認ステップと、前記検出ステップで前記遺伝的多型が検出され且つ前記確認ステップで複数存在すると確認されなかった場合、前記検索ステップで検索した前記後世代個体の前記塩基配列に基づいて、前記遺伝的多型が検出された部位での前記後世代個体の遺伝子型を調査する調査ステップと、前記調査ステップで調査した前記後世代個体の前記遺伝子型と既に調査済みの前記後世代個体の前記遺伝子型とが一致するか否かを判定する判定ステップと、前記判定ステップで一致すると判定された場合、前記検索ステップで検索した前記親個体の前記塩基配列に基づいて前記対象塩基配列を伸長する、および、前記確認ステップで複数存在すると確認された場合、前記検索ステップで検索した同一の前記親個体の前記塩基配列ごとに前記対象塩基配列を別々に伸長する伸長ステップと、を実行させることを特徴とする。
【０００９】
また、本発明にかかる塩基配列決定プログラムは、前記に記載の塩基配列決定プログラムにおいて、前記データベースは、前記個体の前記塩基配列と前記個体を一意に識別するための個体識別情報とを相互に関連付けてなるリレーショナルデータベースであり、前記検索ステップは、データベース言語であるＳＱＬで前記塩基配列を検索することを特徴とする。
【００１０】
また、本発明にかかる塩基配列決定プログラムは、前記に記載の塩基配列決定プログラムにおいて、前記記憶手段は、前記データベースに記憶した前記塩基配列に対して作成された、前方一致検索が可能なインデックスをさらに記憶し、前記検索ステップは、前記インデックスを参照して前記塩基配列を前方一致検索することを特徴とする。
【００１１】
また、本発明は塩基配列決定装置に関するものであり、本発明にかかる塩基配列決定装置は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶した記憶手段と、前記記憶手段に記憶した前記データベースを参照しながら、複数の前記塩基配列を、前記遺伝的多型を指標にして繋ぎ合わせる制御手段とを備えたことを特徴とする。
【００１２】
また、本発明にかかる塩基配列決定装置は、制御手段と記憶手段とを備えた塩基配列決定装置であって、前記記憶手段は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶し、前記制御手段は、対象とする前記親個体の前記塩基配列である対象塩基配列に基づいて、前記記憶手段に記憶した前記データベースから、当該対象塩基配列に繋ぎ合せる前記塩基配列を検索する検索手段と、前記検索手段で検索した前記塩基配列に基づいて、前記親個体間における前記遺伝的多型を検出する検出手段と、前記検索手段で検索した前記塩基配列に同一の前記親個体から検索したものが複数存在するか否かを確認する確認手段と、前記検出手段で前記遺伝的多型が検出され且つ前記確認手段で複数存在すると確認されなかった場合、前記検索手段で検索した前記後世代個体の前記塩基配列に基づいて、前記遺伝的多型が検出された部位での前記後世代個体の遺伝子型を調査する調査手段と、前記調査手段で調査した前記後世代個体の前記遺伝子型と既に調査済みの前記後世代個体の前記遺伝子型とが一致するか否かを判定する判定手段と、前記判定手段で一致すると判定された場合、前

記検索手段で検索した前記親個体の前記塩基配列に基づいて前記対象塩基配列を伸長する、および、前記確認手段で複数存在すると確認された場合、前記検索手段で検索した同一の前記親個体の前記塩基配列ごとに前記対象塩基配列を別々に伸長する伸長手段と、を備えたことを特徴とする。

【００１３】

また、本発明にかかる塩基配列決定装置は、前記に記載の塩基配列決定装置において、前記データベースは、前記個体の前記塩基配列と前記個体を一意に識別するための個体識別情報とを相互に関連付けてなるリレーショナルデータベースであり、前記検索手段は、データベース言語であるＳＱＬで前記塩基配列を検索することを特徴とする。

【００１４】

また、本発明にかかる塩基配列決定装置は、前記に記載の塩基配列決定装置において、前記記憶手段は、前記データベースに記憶した前記塩基配列に対して作成された、前方一致検索が可能なインデックスをさらに記憶し、前記検索手段は、前記インデックスを参照して前記塩基配列を前方一致検索することを特徴とする。

【００１５】

また、本発明は塩基配列決定方法に関するものであり、本発明にかかる塩基配列決定方法は、制御手段と記憶手段とを備えた情報処理装置で実行する塩基配列決定方法であって、前記記憶手段は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶し、前記制御手段で、前記記憶手段に記憶した前記データベースを参照しながら、複数の前記塩基配列を、前記遺伝的多型を指標にして繋ぎ合わせることを特徴とする。

【００１６】

また、本発明にかかる塩基配列決定方法は、制御手段と記憶手段とを備えた情報処理装置で実行する塩基配列決定方法であって、前記記憶手段は、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを記憶し、前記制御手段で、対象とする前記親個体の前記塩基配列である対象塩基配列に基づいて、前記記憶手段に記憶した前記データベースから、当該対象塩基配列に繋ぎ合せる前記塩基配列を検索する検索ステップと、前記検索ステップで検索した前記塩基配列に基づいて、前記親個体間における前記遺伝的多型を検出する検出ステップと、前記検索ステップで検索した前記塩基配列に同一の前記親個体から検索したものが複数存在するか否かを確認する確認ステップと、前記検出ステップで前記遺伝的多型が検出され且つ前記確認ステップで複数存在すると確認されなかった場合、前記検索ステップで検索した前記後世代個体の前記塩基配列に基づいて、前記遺伝的多型が検出された部位での前記後世代個体の遺伝子型を調査する調査ステップと、前記調査ステップで調査した前記後世代個体の前記遺伝子型と既に調査済みの前記後世代個体の前記遺伝子型とが一致するか否かを判定する判定ステップと、前記判定ステップで一致すると判定された場合、前記検索ステップで検索した前記親個体の前記塩基配列に基づいて前記対象塩基配列を伸長する、および、前記確認ステップで複数存在すると確認された場合、前記検索ステップで検索した同一の前記親個体の前記塩基配列ごとに前記対象塩基配列を別々に伸長する伸長ステップと、を実行することを特徴とする。

【００１７】

また、本発明にかかる塩基配列決定方法は、前記に記載の塩基配列決定方法において、前記データベースは、前記個体の前記塩基配列と前記個体を一意に識別するための個体識別情報とを相互に関連付けてなるリレーショナルデータベースであり、前記検索ステップは、データベース言語であるＳＱＬで前記塩基配列を検索することを特徴とする。

【００１８】

また、本発明にかかる塩基配列決定方法は、前記に記載の塩基配列決定方法において、前記記憶手段は、前記データベースに記憶した前記塩基配列に対して作成された、前方一致検索が可能なインデックスをさらに記憶し、前記検索ステップは、前記インデックスを参照して前記塩基配列を前方一致検索することを特徴とする。

【発明の効果】
【００１９】
本発明によれば、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを参照しながら、複数の塩基配列を、遺伝的多型を指標にして繋ぎ合わせる。具体的には、本発明によれば、（１）遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースから、対象とする親個体の塩基配列である対象塩基配列に基づいて、当該対象塩基配列に繋ぎ合せる塩基配列を検索し、（２）検索した塩基配列に基づいて、親個体間における遺伝的多型を検出し、（３）検索した塩基配列に同一の親個体から検索したものが複数存在するか否かを確認し、（４）遺伝的多型が検出され且つ複数存在すると確認されなかった場合、検索した後世代個体の塩基配列に基づいて、遺伝的多型が検出された部位での後世代個体の遺伝子型を調査し、（５）調査した後世代個体の遺伝子型と既に調査済みの後世代個体の遺伝子型とが一致するか否かを判定し、（６）一致すると判定された場合、検索した親個体の塩基配列に基づいて対象塩基配列を伸長する、および、複数存在すると確認された場合、検索した同一の親個体の塩基配列ごとに対象塩基配列を別々に伸長する。
【００２０】
これにより、数十塩基程度の短い大量の塩基配列から既存の塩基配列を参照することなく全ゲノム塩基配列を構築することができるという効果を奏する。換言すると、ゲノム全体を一度に細分化して得られた数十塩基程度の短い大量の塩基配列を、ゲノム上に散在する繰り返し配列も含め正しく連結することができ、その結果、一度にゲノム全体の塩基配列を決定することができるという効果を奏する。すなわち、ゲノム全体を一度に細分化して得られた塩基配列を繋ぎ合わせてゲノム全体の塩基配列を再構築することができるという効果を奏する。また、ゲノム全体の塩基配列だけでなく、各後世代個体の遺伝的多型の情報も同時に得ることができるという効果を奏する。
【図面の簡単な説明】
【００２１】
【図１】図１は、本発明の基本原理を示す原理構成図である。
【図２】図２は、本発明の基本原理を示す原理構成図である。
【図３】図３は、本発明の基本原理を示す原理構成図である。
【図４】図４は、塩基配列決定装置１００の構成を示すブロック図である。
【図５】図５は、塩基配列データベース１０６ａの構成要素であるデータテーブル１０６ａ１に格納される情報の一例を示す図である。
【図６】図６は、塩基配列決定装置１００の制御部１０２で行うデータベース構築処理の一例を示すフローチャートである。
【図７】図７は、塩基配列決定装置１００の制御部１０２で行う塩基配列決定処理の一例を示すフローチャートである。
【図８】図８は、塩基配列決定処理にて作成される情報の出力例を示す図である。
【図９】図９は、塩基配列決定処理にて作成される情報の出力例を示す図である。
【図１０】図１０は、塩基配列決定装置１００の制御部１０２で行う出力結果分析処理の一例を示すフローチャートである。
【符号の説明】
【００２２】
１００　塩基配列決定装置
　１０２　制御部
　　１０２ａ　検索部
　　１０２ｂ　伸長部
　　１０２ｃ　検出部
　　１０２ｄ　確認部
　　１０２ｅ　調査部

１０２ｆ　判定部
１０４　通信インターフェース部
１０６　記憶部
１０６ａ　塩基配列データベース
１０６ａ１　データテーブル
１０６ａ２　インデックスファイル
１０６ｂ　対象塩基配列ファイル
１０６ｃ　遺伝子型ファイル
１０６ｄ　正解塩基配列ファイル
１０８　入出力インターフェース部
１１０　入力装置
１１２　出力装置
３００　ネットワーク

【発明を実施するための最良の形態】

【００２３】

以下に、本発明にかかる塩基配列決定プログラム、塩基配列決定装置および塩基配列決定方法の実施の形態を図面に基づいて詳細に説明する。なお、本実施の形態により本発明が限定されるものではない。

【００２４】

［１．本発明の概要］

本発明の概要について図１から図３を参照して説明する。図１から図３は、本発明の基本原理を示す原理構成図である。

【００２５】

本発明では、遺伝的多型（以下では、単に「多型」と記す場合がある。）を持つ２種類の親個体由来のゲノムＤＮＡと、その後世代の複数の個体由来のゲノムＤＮＡを予め取得する。具体的には、親個体１および親個体２の塩基配列と、その孫世代（Ｆ２）個体３～６の塩基配列を取得し、取得した塩基配列を、図１のＭＡ１に示すテーブル（塩基配列と個体番号とを相互に関連付けてなるテーブル：テーブル名「ｓｅｑ」，カラム名「ｓｅｑ」，個体番号「ｎｏ」）にまとめる。なお、本説明では、便宜的に、「個々の塩基配列の長さを３０塩基とし、１塩基のオーバーラップで全ての塩基配列データが個体ごとに得られている」と仮定する。

【００２６】

本発明は、基本的に、下記の〔操作１〕～〔操作３〕を繰り返し実行する。これにより、図１のＭＡ２に示すように、アトランダムに選んだ親個体１の塩基配列「ｇｃａｃｇｔｃｇａｇｇａａｔｇｃｇｃｇａｇｃｃｇａｃａａｃｇ」を３′側に１塩基ずつ伸長させる。

〔操作１〕親個体１の塩基配列が設定された対象塩基配列（ディフォルトとして、図１のＭＡ１に示すテーブル内の親個体１の塩基配列からアトランダムに選んだ３０塩基長の塩基配列「ｇｃａｃｇｔｃｇａｇｇａａｔｇｃｇｃｇａｇｃｃｇａｃａａｃｇ」を設定）の３′側に続く次の塩基を、図１のＭＡ１に示すテーブルから、データベース言語であるＳＱＬ文「ＳＥＬＥＣＴ　ｓｅｑ，　ｎｏ　ＦＲＯＭ　ｓｅｑ　ＷＨＥＲＥ　ｓｅｑ　ＬＩＫＥ‘ｃａｃｇｔｃｇａｇｇａａｔｇｃｇｃｇａｇｃｃｇａｃａａｃｇ％’；」で前方一致検索する。なお、親個体２のゲノム塩基配列を決める場合は、対象塩基配列に親個体２の塩基配列を用いて親個体２由来の塩基配列を検索する。

〔操作２〕検索結果として、親個体１の塩基配列「ｃａｃｇｔｃｇａｇｇａａｔｇｃｇｃｇａｇｃｃｇａｃａａｃｇｃ　１」が返ってきた場合、対象塩基配列の３′側に続く次の塩基が「ｃ」であると決定する。

〔操作３〕決定した塩基「ｃ」を対象塩基配列「ｇｃａｃｇｔｃｇａｇｇａａｔｇｃｇｃｇａｇｃｃｇａｃａａｃｇ」の３′側に繋ぎ合わせることで、対象塩基配列を伸長する。

【００２７】

しかし、上記操作を実行している際に、〔操作１〕での検索の結果、例えば親個体１の塩基配列「ｇｔｃｃｇｃｇｃｔｃｇｇｇｃｔｃｃｔｔｃａｃｃｔｇｃｔｃｇａ　１」と親個体１の塩基配列「ｇｔｃｃｇｃｇｃｔｃｇｇｇｃｔｃｃｔｔｃａｃｃｔｇｃｔｃｇｇ１」が得られた場合、対象塩基配列の３′側に続く次の塩基が「ａ」または「ｇ」のどちらであるかは、検索したこれら塩基配列だけでは判らない。具体的には、２９塩基長のクエリー塩基配列「ｇｔｃｃｇｃｇｃｔｃｇｇｇｃｔｃｃｔｔｃａｃｃｔｇｃｔｃｇ」と前方一致する３０塩基長の塩基配列が同一個体（親個体１）中のゲノムから複数検索された場合、すなわち同一個体（親個体１）のゲノム上において互いに異なる位置に存在する複数個の塩基配列（複数種の塩基配列）が同時に検出された場合、これだけでは、当該検索された塩基配列のどれを選択すればよいかは判らない。仮に、検索した塩基配列ごとに上記操作で伸長を続けていくと、図１のＭＡ３に示すように、全く異なった塩基配列（「ａｔｇｃｃｇａｃｇ」と「ｇｃｇｇｃｇｃｃｇ」）で伸長されていく。

【００２８】

そこで、本発明は、対象塩基配列の３′側に続く次の塩基が「ａ」または「ｇ」のどちらであるかを判定するために、図３に示すように、「ａ」を繋ぎ合わせた対象塩基配列と「ｇ」を繋ぎ合わせた対象塩基配列とに分岐して上記操作を実行することでそれぞれの伸長を続けながら、その分岐前後での多型の分離、すなわち変異の連鎖関係を調べ、これに基づいてそれぞれの繋ぎ合わせの正当性を検証する。換言すると、本発明は、「ａ」を繋ぎ合わせた対象塩基配列と「ｇ」を繋ぎ合わせた対象塩基配列とに分岐して上記操作を実行することでそれぞれの伸長を続けながら、各対象塩基配列について分岐後最初に多型を検出した際、その最初に検出した多型（分岐後の多型）と元の多型（分岐前の多型）との違いを調べ、これに基づいてそれぞれの繋ぎ合わせの正当性を検証する。分岐後、次の分岐までに多型が見つからない場合は、多型が見つかるまで分岐を繰り返し、すべての分岐の組み合わせの中から変異の連鎖関係が一致する分岐を選択する。

【００２９】

例えば、分岐前に検出された多型部位を含む親個体１および２の塩基配列ならびに当該多型部位に対応する孫個体３～６の塩基配列が、図２のＭＡ４に示すものであったとする。この場合、親個体１の３′末端での２塩基は「ｇｇ」（これを親１型（Ａ型）とする）であり、親個体２の３′末端での２塩基は「ａａ」（これを親２型（Ｂ型）とする）である。そして、孫個体３の３′末端での２塩基は「ｇａ」であり、孫個体４の３′末端での２塩基は「ｇｇ」であり、孫個体５の３′末端での２塩基は「ａａ」であり、孫個体６の３′末端での２塩基は「ｇａ」である。つまり、孫個体３の遺伝子型はヘテロ型（Ｈ型）、孫個体４の遺伝子型はＡ型、孫個体５の遺伝子型はＢ型、孫個体６の遺伝子型はＨ型である。以上より、多型が検出された部位での遺伝子型は「ＡＢＨＡＢＡ」と表すことができる。

【００３０】

次に、分岐後にそれぞれの対象塩基配列を伸長させて最初に多型を検出した結果、その多型部位を含む親個体１および２の塩基配列ならびに当該多型部位に対応する孫個体３～６の塩基配列が、それぞれ図２のＭＡ５およびＭＡ６に示すものであったとする。図２のＭＡ５およびＭＡ６に示す各塩基配列についても、分岐前と同様に、多型部位での各個体の遺伝子型を調べる。すると、図２のＭＡ５に示す塩基配列に対しては、遺伝子型は「ＡＢＨＡＢＨ」と表すことができ、図２のＭＡ６に示す塩基配列に対しては、遺伝子型は「ＡＢＡＨＨＢ」と表すことができる。なお、図２のＭＡ５に示す塩基配列は、図１のＭＡ３の左側に示すように「・・・ｔｇｃｔｃｇ」の後に「ａｔｇｃｃｇａｃｇ」と続き、分岐後最初の多型が検出されるまで伸長した結果である。また、図２のＭＡ６に示す塩基配列は、図１のＭＡ３の右側に示すように「・・・ｔｇｃｔｃｇ」の後に「ｇｃｇｇｃｇｃｃｇ」と続き、分岐後最初の多型が検出されるまで伸長した結果である。

【００３１】

そして、本発明は、分岐前の遺伝子型「ＡＢＨＡＢＨ」と分岐後の遺伝子型とを比較し、それが一致する対象塩基配列を、伸長させるべき正しいものとして判定する。つまり、

本発明は、分岐前の多型部位での遺伝子型とその次の分岐後の多型部位での遺伝子型とを比較する。本説明では、図１のＭＡ３の左側に示すように「・・・ｔｇｃｔｃｇ」の後に「ａｔｇｃｃｇａｃｇ」と続けて伸長するのが正解であると判定する。

【００３２】

以上説明したように、本発明は、多型を持つ２種類の親系統の塩基配列とその後世代（例えば孫世代（Ｆ２））個体の塩基配列をデータベース化し、多型を指標にして短い塩基配列を繋ぎ合わせる。本発明は、遺伝的変異を持つ２つの親系統とその分離後世代の複数の個体の塩基配列をそれぞれ解析し、塩基配列の繋ぎ合わせの指標として変異部分の分離を参照し、その分離の正当性を確認しながら塩基配列の断片を繋ぎ合わせる。本発明は、後世代の塩基配列も同時に解析して変異の連鎖関係を検証しながら、塩基配列をアセンブルする。

【００３３】

これにより、ゲノム上に複数存在する同一の塩基配列を見分けることが可能となり、その結果、精度の高いアセンブルが実現できる。数十塩基程度の短い大量の塩基配列から既存の塩基配列を参照することなく全ゲノム塩基配列を構築することができる。換言すると、ゲノム全体を一度に細分化して得られた数十塩基程度の短い大量の塩基配列を、ゲノム上に散在する繰り返し配列も含め正しく連結することができ、その結果、一度にゲノム全体の塩基配列を決定することができる。すなわち、ゲノム全体を一度に細分化して得られた塩基配列を繋ぎ合わせてゲノム全体の塩基配列を再構築することができる。また、ゲノム全体の塩基配列だけでなく、各後世代個体の遺伝的多型の情報も同時に得ることができる。

【００３４】

現在、一度に数千万サンプルを解析して１つのサンプルあたり３０塩基程度の塩基配列を得ることが可能なシーケンサーが開発されてきているが、一般に、短い塩基配列を繋ぎ合わせると、ゲノム上に多数存在する繰り返し配列部分で正しい配列の判別が困難となる。これまでは、一回の解析で決定できる断片長をより長くすることで、繋ぎ合わせの間違いを回避する努力がなされてきた。

【００３５】

しかし、本発明によれば、１つの断片あたりの長さは重要ではなく、このようなシーケンサーで得られた３０塩基程度（例えば３０～７０塩基）の短い断片長でも確実にオーバーラップするだけのデータ数が確保できれば、相同な部分が複数の染色体に散在していても遺伝的変異を指標に正しい配列を見分けることができる。ゆえに、本発明によれば、繰り返し配列部分も正しく繋ぎ合わせることができ、その結果、比較的短い大量の塩基配列を繋ぎ合わせてゲノム全体の塩基配列を再構築することができる。

【００３６】

また、本発明によれば、３０塩基程度の短い塩基配列をリレーショナルデータベースに多量に格納し、データベースに格納した塩基配列に対して、例えばＢ－ｔｒｅｅ（Ｂ木）型インデックスなどのような前方一致検索が可能なインデックスを作成し、ＳＱＬ文で前方一致検索を繰り返しながら塩基配列を伸長する。これにより、次に繋げる塩基配列の候補を高速に検索することができ、その結果、ゲノムＤＮＡ全体のアセンブルを正確且つ高速に実現することができる。つまり、本発明によれば、正確且つ高速に塩基配列を繋ぎ合わせることができる。

【００３７】

また、本発明によれば、事前に、変異をスキャンし、スキャンした変異を含む全ての３０塩基に対して遺伝子型を決定してから、上述した操作で塩基配列を伸長してもよい。これにより、さらに効率よく塩基配列のアセンブルを実現することができる。

【００３８】

また、本説明では、一塩基の精度で全ゲノムをカバーするデータが充分量得られていることを仮定したが、本発明によれば、それぞれの個体についてゲノムを数倍カバーする程度のデータでも、対象塩基配列の５´末端から一塩基ずつずらしながら１７塩基長程度の

部分塩基配列を得て、得た部分塩基配列をクエリー配列としてそれにヒットする塩基配列のアライメントを作成しながら対象塩基配列を伸長することが可能である。なお、親個体１由来の１７～２９塩基長のクエリー配列で３０塩基長の塩基配列データを前方一致検索し、両親個体および孫個体由来データで一致した塩基配列でアライメントを作成した場合において、それぞれの個体間で異なる（ミスマッチする）塩基が検出された場合は、検出された多型塩基の個体間での分離を、上述した判定の指標に用いてもよい。

【００３９】

また、本発明によれば、交配可能なあらゆる生物のゲノム塩基配列をこれまでに無い効率で決定でき、同時にマップされた全ての多型情報も得ることができる。本発明によれば、クローン化できない部分も精度よく繋ぎ合せることができる。なお、解析集団の詳細な表現型の分離データが得られれば、本発明により、ＱＴＬ（Ｑｕａｎｔｉｔａｔｉｖｅ Ｔｒａｉｔ Ｌｏｃｕｓ）の原因遺伝子を特定することも原理的には可能である。本発明により決定された各種生物のゲノム塩基配列の産業上の利用価値は、計り知れない。

【００４０】

［２．システム構成］

つぎに、本実施の形態にかかる塩基配列決定装置１００の構成について、図４および図５を参照して説明する。図４は、塩基配列決定装置１００の構成を示すブロック図であり、該構成のうち本発明に関係する部分のみを概念的に示している。

【００４１】

塩基配列決定装置１００は、当該塩基配列決定装置を統括的に制御するＣＰＵ（Ｃｅｎｔｒａｌ Ｐｒｏｃｅｓｓｉｎｇ Ｕｎｉｔ）等の制御部１０２と、ルータ等の通信装置および専用線等の有線または無線の通信回線を介して当該塩基配列決定装置をネットワーク３００に通信可能に接続する通信インターフェース部１０４と、各種のデータベースやテーブルやファイルなどを格納する記憶部１０６と、入力装置１１０や出力装置１１２に接続する入出力インターフェース部１０８と、で構成されており、これら各部は任意の通信路を介して通信可能に接続されている。

【００４２】

記憶部１０６は、ストレージ手段であり、例えば、ＲＡＭ（Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）やＲＯＭ（Ｒｅａｄ Ｏｎｌｙ Ｍｅｍｏｒｙ）等のメモリ装置や、ＨＤ（Ｈａｒｄ Ｄｉｓｋ）のような固定ディスク装置や、フレキシブルディスクや、光ディスク等を用いることができる。そして、記憶部１０６は、図示の如く、塩基配列データベース１０６ａと、対象塩基配列ファイル１０６ｂと、遺伝子型ファイル１０６ｃと、正解塩基配列ファイル１０６ｄと、を格納する。

【００４３】

塩基配列データベース１０６ａは、遺伝的多型を持つ複数の親個体および複数のその後世代個体（例えば孫世代）のそれぞれに由来する複数の塩基配列、およびこれらの塩基配列に対して作成されたインデックス（例えばＢ－ｔｒｅｅ（Ｂ木）型インデックスなど）を格納する。塩基配列データベース１０６ａは、データテーブル１０６ａ１およびインデックスファイル１０６ａ２で構成されている。ここで、塩基配列データベース１０６ａの構成要素であるデータテーブル１０６ａ１に格納される情報について図５を参照して説明する。図５は、データテーブル１０６ａ１に格納される情報の一例を示す図である。図５に示すように、データテーブル１０６ａ１は、個体の塩基配列と、個体を一意に識別するための個体番号と、を相互に関連付けて格納するリレーショナルデータベースである。

【００４４】

図４に戻り、塩基配列データベース１０６ａの構成要素であるインデックスファイル１０６ａ２は、データテーブル１０６ａ１に記憶した塩基配列に対して作成された、前方一致検索が可能なインデックス（例えばＢ－ｔｒｅｅ（Ｂ木）型インデックスなど）を格納する。対象塩基配列ファイル１０６ｂは、伸長する対象となる親個体の塩基配列である対象塩基配列を格納する。なお、ディフォルトの対象塩基配列は、ゲノム塩基配列を決定したい、いずれかの親個体の塩基配列からアトランダムに設定する。遺伝子型ファイル１０

６ｃは、後述する調査部１０２ｅで調査した各個体の遺伝子型を格納する。正解塩基配列ファイル１０６ｄは、対象塩基配列の伸長を続けて最終的に正しいものとして確定した塩基配列である正解塩基配列を格納する。

【００４５】

通信インターフェース部１０４は塩基配列決定装置１００とネットワーク３００（またはルータ等の通信装置）との間における通信を媒介する。すなわち、通信インターフェース部１０４は他の端末と通信回線を介してデータを通信する機能を有する。

【００４６】

入出力インターフェース部１０８は入力装置１１０や出力装置１１２に接続する。ここで、出力装置１１２には、モニタ（家庭用テレビを含む）の他、スピーカやプリンタを用いることができる（なお、以下で、出力装置１１２をモニタとして記載する場合がある。）。また、入力装置１１０には、キーボードやマウスやマイクの他、マウスと協働してポインティングデバイス機能を実現するモニタを用いることができる。

【００４７】

制御部１０２は、ＯＳ（Ｏｐｅｒａｔｉｎｇ　Ｓｙｓｔｅｍ）等の制御プログラム、各種の処理手順等を規定したプログラムおよび所要データを格納するための内部メモリを有し、これらのプログラムに基づいて種々の処理を実行するための情報処理を行う。そして、制御部１０２は、図示の如く、大別して、検索部１０２ａと、伸長部１０２ｂと、検出部１０２ｃと、確認部１０２ｄと、調査部１０２ｅと、判定部１０２ｆと、を備えている。

【００４８】

検索部１０２ａは、対象塩基配列に基づいて、記憶部１０６に記憶した塩基配列データベース１０６ａのデータテーブル１０６ａ１から、当該対象塩基配列に繋ぎ合せる塩基配列を検索する。検索部１０２ａは、記憶部１０６に記憶した塩基配列データベース１０６ａのインデックスファイル１０６ａ２を参照して、データベース言語であるＳＱＬで、対象塩基配列に繋ぎ合せる塩基配列を前方一致検索してもよい。

【００４９】

伸長部１０２ｂは、検索部１０２ａで検索した親個体の塩基配列に基づいて対象塩基配列を伸長する。伸長部１０２ｂは、後述する判定部１０２ｆで一致すると判定された場合、検索部１０２ａで検索した親個体の塩基配列に基づいて対象塩基配列を伸長する。伸長部１０２ｂは、後述する確認部１０２ｄで複数種存在すると確認された場合、検索部１０２ａで検索した同一の親個体の塩基配列ごとに対象塩基配列を別々に伸長する。

【００５０】

検出部１０２ｃは、検索部１０２ａで検索した塩基配列に基づいて、親個体間における遺伝的多型を検出する。確認部１０２ｄは、検索部１０２ａで検索した塩基配列に同一の親個体から検索したものが複数存在するか否かを確認する。調査部１０２ｅは、検出部１０２ｃで遺伝的多型が検出され且つ確認部１０２ｄで複数種存在すると確認されなかった場合、検索部１０２ａで検索した各個体（親個体や後世代個体）の塩基配列に基づいて、遺伝的多型が検索された部位での各個体（親個体や後世代個体）の遺伝子型を調査する。判定部１０２ｆは、調査部１０２ｅで調査した各個体（親個体および／または後世代個体）の遺伝子型と、既に調査部１０２ｅで調査済みの各個体（親個体および／または後世代個体）の遺伝子型とが一致するか否かを判定する。

【００５１】

ネットワーク３００は、塩基配列決定装置１００と外部システムとを相互に接続する機能を有し、例えば、インターネットや、イントラネットや、ＬＡＮ（有線／無線の双方を含む）や、ＶＡＮや、パソコン通信網や、公衆電話網（アナログ／デジタルの双方を含む）や、専用回線網（アナログ／デジタルの双方を含む）や、ＣＡＴＶ網や、ＩＭＴ２０００方式、ＧＳＭ方式またはＰＤＣ／ＰＤＣ－Ｐ方式等の携帯回線交換網／携帯パケット交換網や、無線呼出網や、Ｂｌｕｅｔｏｏｔｈ（登録商標）等の局所無線網や、ＰＨＳ網や、ＣＳ、ＢＳまたはＩＳＤＢ等の衛星通信網等のうちいずれかを含んでもよい。これによ

り、塩基配列決定装置１００は、有線・無線を問わず任意のネットワークを介して、各種データを送受信することができる。

【００５２】

［３．システムの処理］

つぎに、塩基配列決定装置１００の制御部１０２で行う各種処理を、図６から図１０を参照して説明する。

【００５３】

［３－１．データベース構築処理］

まず、塩基配列決定装置１００の制御部１０２で行うデータベース構築処理を、図６を参照して説明する。図６は、塩基配列決定装置１００の制御部１０２で行うデータベース構築処理の一例を示すフローチャートである。

【００５４】

まず、制御部１０２は、遺伝的に異なる２つの親個体（親１，親２）および複数のその孫個体（孫３，孫４，孫５，孫６，・・・）をシーケンサーで解析して得た大量（数千万～数億）の３０塩基長の塩基配列を取得する（ステップＳＡ－１）。

【００５５】

つぎに、制御部１０２は、ステップＳＡ－１で取得した塩基配列と当該塩基配列の解析元の個体を識別するための個体番号とのペアからなるテーブルを作成し、塩基配列データベース１０６ａのデータテーブル１０６ａ１に格納する（ステップＳＡ－２）。

【００５６】

つぎに、制御部１０２は、ステップＳＡ－１で取得した塩基配列に対してＢ－ｔｒｅｅ型インデックスを作成し、作成したＢ－ｔｒｅｅ型インデックスを塩基配列データベース１０６ａのインデックスファイル１０６ａ２に格納する（ステップＳＡ－３）。

【００５７】

［３－２．塩基配列決定処理］

つぎに、塩基配列決定装置１００の制御部１０２で行う塩基配列決定処理を、図７～図９を参照して説明する。図７は、塩基配列決定装置１００の制御部１０２で行う塩基配列決定処理の一例を示すフローチャートである。

【００５８】

まず、制御部１０２は、塩基配列データベース１０６ａのデータテーブル１０６ａ１に格納されている親個体１の塩基配列の中から３０塩基長の対象塩基配列をアトランダムに選択し、選択した対象塩基配列のコピーを正解塩基配列として正解塩基配列ファイル１０６ｄに格納する（ステップＳＢ－１）。

【００５９】

つぎに、制御部１０２は、検索部１０２ａで、ステップＳＢ－１で決定した対象塩基配列の５′側の一塩基を除く２９塩基（対象塩基配列の３′側から２９塩基）をクエリーとして設定し、設定したクエリーと前方一致するレコードを、ＳＱＬ文で、塩基配列データベース１０６ａのデータテーブル１０６ａ１から、インデックスファイル１０６ａ２に格納されているＢ－ｔｒｅｅ型インデックスを参照して抽出（検索）する（ステップＳＢ－２）。

【００６０】

つぎに、制御部１０２は、ステップＳＢ－２でレコードが抽出されなかった場合（ステップＳＢ－３：Ｎｏ）には、遺伝子型ファイル１０６ｃに格納されている遺伝子型と正解塩基配列ファイル１０６ｄに格納されている正解塩基配列とを出力装置１１２に出力して（ステップＳＢ－４）、本処理を終了する。

【００６１】

つぎに、制御部１０２は、ステップＳＢ－２でレコードが抽出された場合（ステップＳＢ－３：Ｙｅｓ）には、検出部１０２ｃで、ステップＳＢ－２で抽出した親個体１，２のレコードの５′側から３０番目の塩基（３′側から一番目の塩基）を調べる（ステップＳＢ－５）。

【００６２】
つぎに、制御部１０２は、ステップＳＢ－５で調べた親個体１，２の塩基が単一であると検出された、すなわち多型でないと検出された場合（ステップＳＢ－６：Ｙｅｓ）には、伸長部１０２ｂで、調べた親個体１の塩基を対象塩基配列の３´側および正解塩基配列ファイル１０６ｄに格納されている正解塩基配列の３´側に加え（ステップＳＢ－７）、ステップＳＢ－２の処理に戻る。
【００６３】
つぎに、制御部１０２は、ステップＳＢ－５で調べた親個体１，２の塩基が単一でなく（ステップＳＢ－６：Ｎｏ）、さらに確認部１０２ｄでそれが親個体１内で単一でないと確認された、すなわち３０塩基目が互いに異なる、親個体１の複数のレコード（親個体１の複数種のレコード）が検索された場合（ステップＳＢ－８：Ｎｏ）には、伸長部１０２ｂで、調べたその塩基の数だけ対象塩基配列のコピーを作成し、作成した各々の対象塩基配列の３´側に当該調べたそれぞれの塩基を加える（繋ぎ合わせる）ことで、複数の対象塩基配列を更新し（ステップＳＢ－９）、ステップＳＢ－２の処理に戻る。
【００６４】
つぎに、制御部１０２は、ステップＳＢ－５で調べた親個体１，２の塩基が単一でなく（ステップＳＢ－６：Ｎｏ）さらに確認部１０２ｄでそれが親個体１内で単一であると確認された場合（ステップＳＢ－８：Ｙｅｓ）、すなわち多型であると検出された場合には、調査部１０２ｅで、ステップＳＢ－２で検索された両親個体（親１，親２）および複数の孫個体（孫３，孫４，孫５，孫６，・・・）の塩基配列の５´側から３０番目（３´末端）の塩基を調べ、調べた各塩基に基づいて各孫個体の遺伝子型を調べ、調べた遺伝子型を遺伝子型ファイル１０６ｃに格納する（ステップＳＢ－１０）。
【００６５】
つぎに、制御部１０２は、判定部１０２ｆにより、ステップＳＢ－１０で調べた遺伝子型と以前に遺伝子型ファイル１０６ｃに格納した遺伝子型とが一致すると判定された場合（ステップＳＢ－１１：Ｙｅｓ）には、ステップＳＢ－７の処理に戻る。
【００６６】
つぎに、制御部１０２は、判定部１０２ｆにより、ステップＳＢ－１０で調べた遺伝子型と以前に遺伝子型ファイル１０６ｃに格納した遺伝子型とが一致すると判定されなかった場合（ステップＳＢ－１１：Ｎｏ）には、遺伝子型ファイル１０６ｃに格納されている遺伝子型と正解塩基配列ファイル１０６ｄに格納されている分岐する前までの正解塩基配列とを出力装置１１２に出力して（ステップＳＢ－１２）、本処理を終了する。
【００６７】
以上説明したように、本処理では、伸長により現れる複数の分岐ポイントの中から正解塩基配列を得る。すなわち、本処理では、分岐ポイントで、抽出された各塩基を３´末端に持つ対象塩基配列を、多型情報と共に個別のファイルに保存する。そして、各ファイルの対象塩基配列の３´側２９塩基をクエリーとして個々の対象塩基配列を伸長する。そして、その先に出現する多型が元の多型と一致しない場合には、そのファイルを削除し、新たに次のファイルで同様の処理を行う。一方、その先に出現する多型が元の多型と一致した場合には、当該一致したそのファイルのみを残して他のファイルを削除し、残したファイルからさらなる伸長を行う。
【００６８】
また、本処理では、図８のＭＢ１に示すように、親個体１の塩基配列（Ｓｔａｒｔ「ｇｃａｃｇｔｃｇａｇｇａａｔｇｃｇｃｇａｇｃｃｇａｃａａｃｇ」）の５´側から３０塩基目が多型であった、つまりこの３０塩基目が親個体間で異なった（例えば、親個体１（Ａ）由来の３０塩基目は「ｇ（＝１）」で親個体２（Ｂ）由来の３０塩基目は「ａ（＝２）」であった）場合、孫個体の多型および遺伝子型を調べる。そして、調べた結果、孫個体１由来の３０塩基目が「ａ」および「ｇ」で遺伝子型がヘテロ型（Ｈ型）で、孫個体２由来の３０塩基目が「ａ」および「ａ」で遺伝子型が親２型（Ｂ型）で、孫個体３由来の３０塩基目が「ａ」および「ａ」で遺伝子型が親２型（Ｂ型）で、孫個体４由来の３０

塩基目が「ａ」および「ｇ」で遺伝子型がヘテロ型（Ｈ型）であった場合、この位置での親個体および孫個体の遺伝子型は「ＨＨＢＢＨ」と表される（この遺伝子型における先頭の「Ｈ」は親個体の遺伝子型を表し、これに続く「ＨＢＢＨ」は孫個体親個体１から４の遺伝子型を表す。本説明では、親個体１，２の多型を一緒に判断しているため、親個体の遺伝子型は、多型部位では常に「Ｈ」となる。）。なお、図８の例では、２倍体のイネゲノムデータを模しているため、番号「１」が親個体１由来のデータ、番号「２」が親個体２由来のデータ、番号「３および４」が孫個体１由来のデータ、番号「５および６」が孫個体２由来のデータ、番号「７および８」が孫個体３由来のデータ、そして、番号「９および１０」が孫個体４由来のデータを表す。また、各親個体内では多型は分離し得ないので、図８の例では、親個体１に番号「１および２」を割り当てず、親個体１には番号「１」のみ、親個体２には番号「２」のみを割り当てている。また、実際の孫個体のデータとしては、この場合、３種の組み合わせ（「ａ」のみの場合、「ｇ」のみの場合、「ａ」と「ｇ」がほぼ同じ回数出現する場合）が得られる。そして、「ａ」のみの場合を親個体１（Ａ型）、「ｇ」のみの場合を親個体２（Ｂ型）、「ａ」と「ｇ」がほぼ同じ回数現れた場合をヘテロ型（Ｈ型）とする。本処理では、理解し易くするため、孫個体に関して、姉妹染色体それぞれから由来するデータに関して別々（例えば３と４）にデータベースに入力している。

【００６９】

そして、本処理では、親個体１から２種類の異なる塩基配列が検索された時のその異なる３０塩基目を分岐ポイントとして、図８のＭＢ２に示すＢｒａｎｃｈ配列「ｇｔｃｃｇｃｇｃｔｃｇｇｇｃｔｃｃｔｔｃａｃｃｔｇｃｔｃｇａ」および図８のＭＢ３に示すＢｒａｎｃｈ配列「ｇｔｃｃｇｃｇｃｔｃｇｇｇｃｔｃｃｔｔｃａｃｃｔｇｃｔｃｇｇ」を、親個体１，２間において新たな多型を検出するまで個別に伸長させ続ける。そして、本処理では、親個体１，２間において新たな多型を検出した位置（図８のＭＢ２に示すＦｏｌｌｏｗｉｎｇ　ｍｕｔａｔｉｏｎ配列「ｔｔｃｇｇｇｇｔｇｇａｃａｃｇｇｇｃｇａｃａｔｇａａｃｇａｇ」の５′側から３０塩基目「ｇ」、および図８のＭＢ３に示すＦｏｌｌｏｗｉｎｇ　ｍｕｔａｔｉｏｎ配列「ｃｇｇｃｇｔｔｃｇｔｇａｔｇｇｔｇｔａｃｇｃａｇａｇｇａｇｇ」の５′側から３０塩基目「ｇ」）で、再び、孫個体の多型および遺伝子型を調べる。そして、調べた結果、例えば、図８のＭＢ２に示すＦｏｌｌｏｗｉｎｇ　ｍｕｔａｔｉｏｎ配列に対しては親個体および孫個体の遺伝子型が「ＨＨＢＢＨ」と表され、図８のＭＢ３に示すＦｏｌｌｏｗｉｎｇ　ｍｕｔａｔｉｏｎ配列に対しては親個体および孫個体の遺伝子型が「ＨＢＢＨＡ」と表された場合、本処理では、対象塩基配列を、先に調べた遺伝子型「ＨＨＢＢＨ」と一致する図８のＭＢ２に示すＦｏｌｌｏｗｉｎｇ　ｍｕｔａｔｉｏｎ配列のように伸長するのが正しいと判定する。そして、本処理では、最終的に、例えば図８のＭＢ４に示す正解塩基配列を出力する。

【００７０】

また、本処理では、図９のＭＣ１に示すように、図９のＭＣ１の最上段に示す親個体１由来の塩基配列の３′側２９塩基をクエリーとし、それを一塩基ずつ３′側へずらしながら、それに繋がる塩基配列を繰り返し検索する。そして、本処理では、親個体１から異なる塩基配列が複数種検索された場合、図９のＭＣ２に示す塩基配列の３′末端および図９のＭＣ３に示す塩基配列の３′末端のように検索された塩基別に、先に調査した遺伝子型「ＨＨＢＢＨ」と共にそれぞれの塩基配列を個別の正解塩基配列ファイルに出力し、個々の正解塩基配列ファイルに対してさらに伸長を続ける。そして、個々の正解塩基配列ファイルごとに、次に検出された多型部位の遺伝子型が「ＨＨＢＢＨ」と一致するかを調べ、一致しない場合は当該ファイルを削除し、一致した場合は当該ファイルを残し、残したそのファイルに対してさらに伸長を進める。

【００７１】

［３－３．出力結果分析処理］

つぎに、塩基配列決定装置１００の制御部１０２で行う出力結果分析処理を、図１０を参照して説明する。図１０は、塩基配列決定装置１００の制御部１０２で行う出力結果分

析処理の一例を示すフローチャートである。

【００７２】

まず、制御部１０２は、上述した塩基配列決定処理で出力された正解塩基配列および遺伝子型情報に基づいて、当該正解塩基配列を遺伝子型で複数のグループに分類する（ステップＳＣ－１）。

【００７３】

つぎに、制御部１０２は、ステップＳＣ－１で分類したグループごとに、正解塩基配列を整列（ソート）する（ステップＳＣ－２）。

【００７４】

つぎに、制御部１０２は、ステップＳＣ－１で分類した複数のグループにおいて遺伝子地図上で隣り合うクループ同士を、遺伝子型に基づいて連結する（ステップＳＣ－３）。

【００７５】

［４．本実施の形態のまとめ、及び他の実施の形態］

以上説明したように、塩基配列決定装置１００によれば、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースを参照しながら、複数の塩基配列を、遺伝的多型を指標にして繋ぎ合わせる。具体的には、本発明によれば、遺伝的多型を持つ複数の親個体および複数のその後世代個体のそれぞれに由来する複数の塩基配列を格納したデータベースデータベースから、対象とする親個体の塩基配列である対象塩基配列に基づいて、当該対象塩基配列に繋ぎ合せる塩基配列を検索し、検索した塩基配列に基づいて親個体間における遺伝的多型を検出し、検索した塩基配列に同一の親個体から検索したものが複数種存在する（検索した塩基配列に同一の親個体から検索された互いに異なるものが複数個存在する）か否かを確認し、遺伝的多型が検出され且つ複数種存在すると確認されなかった場合、遺伝的多型が検出された部位での後世代個体の遺伝子型をデータベースに基づいて調査し、調査した後世代個体の遺伝子型と既に調査済みの後世代個体の遺伝子型とが一致するか否かを判定し、一致すると判定された場合、検索した親個体の塩基配列に基づいて対象塩基配列を伸長する。また、複数種存在すると確認された場合、検索した同一の親個体の塩基配列ごとに対象塩基配列を別々に伸長する。

【００７６】

これにより、数十塩基程度の短い大量の塩基配列から既存の塩基配列を参照することなく全ゲノム塩基配列を構築することができる。換言すると、ゲノム全体を一度に細分化して得られた数十塩基程度の短い大量の塩基配列を、ゲノム上に散在する繰り返し配列も含め正しく連結することができ、その結果、一度にゲノム全体の塩基配列を決定することができる。すなわち、ゲノム全体を一度に細分化して得られた塩基配列を繋ぎ合わせてゲノム全体の塩基配列を再構築することができる。また、ゲノム全体の塩基配列だけでなく、各後世代個体の遺伝的多型の情報も同時に得ることができる。

【００７７】

また、塩基配列決定装置１００によれば、データベースは、個体の塩基配列と個体番号とを相互に関連付けてなるリレーショナルデータベースであり、ＳＱＬで塩基配列を検索する。これにより、次に繋ぎ合せる塩基配列を効率よく検索することができる。

【００７８】

また、塩基配列決定装置１００によれば、データベースに記憶した塩基配列に対して作成されたＢ－ｔｒｅｅ型インデックスをさらに記憶し、当該Ｂ－ｔｒｅｅ型インデックスを参照して塩基配列を前方一致検索する。これにより、次に繋ぎ合せる塩基配列を高速に検索することができる。

【００７９】

また、本発明は、上述した実施の形態以外にも、特許請求の範囲の書類に記載した技術的思想の範囲内において種々の異なる実施の形態にて実施されてよいものである。例えば、塩基配列決定装置１００は、当該塩基配列決定装置とは別筐体で構成されるクライアント端末からの要求に応じて処理を行い、その処理結果を当該クライアント端末に返却する

ように構成してもよい。また、本実施の形態において説明した各処理のうち、自動的に行なわれるものとして説明した処理の全部または一部を手動的に行うこともでき、あるいは、手動的に行なわれるものとして説明した処理の全部または一部を公知の方法で自動的に行うこともできる。この他、上記文書中や図面中で示した処理手順、制御手順、具体的名称、各種の登録データや検索条件等のパラメータを含む情報、画面例、データベース構成については、特記する場合を除いて任意に変更することができる。

【００８０】

また、塩基配列決定装置１００に関して、図示の各構成要素は機能概念的なものであり、必ずしも物理的に図示の如く構成されていることを要しない。例えば、塩基配列決定装置１００の各部または各装置が備える処理機能、特に制御部１０２にて行なわれる各処理機能については、その全部または任意の一部を、ＣＰＵおよび当該ＣＰＵにて解釈実行されるプログラムにて実現することができ、あるいは、ワイヤードロジックによるハードウェアとして実現することも可能である。なお、本発明にかかるプログラムは、後述する記録媒体に記録されており、必要に応じて塩基配列決定装置１００に機械的に読み取られる。すなわち、ＲＯＭまたはＨＤなどの記憶部１０６などには、ＯＳと協働してＣＰＵに命令を与え、各種処理を行うためのコンピュータプログラムが記録されている。このコンピュータプログラムは、ＲＡＭ等にロードされることによって実行され、ＣＰＵと協働して制御部１０２を構成する。また、このコンピュータプログラムは、塩基配列決定装置１００に対して任意のネットワーク３００を介して接続されたアプリケーションプログラムサーバに記録されてもよく、必要に応じてその全部または一部をダウンロードすることも可能である。

【００８１】

また、本発明にかかるプログラムを、コンピュータ読み取り可能な記録媒体に格納することもできる。ここで、この「記録媒体」とは、フレキシブルディスク、光磁気ディスク、ＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ、ＣＤ－ＲＯＭ、ＭＯ、ＤＶＤ等の任意の「可搬用の物理媒体」や、各種コンピュータシステムに内蔵されるＲＯＭ、ＲＡＭ、ＨＤ等の任意の「固定用の物理媒体」、あるいは、ＬＡＮ、ＷＡＮ、インターネットに代表されるネットワークを介してプログラムを送信する場合の通信回線や搬送波のように、短期にプログラムを保持する「通信媒体」を含むものとする。また、「プログラム」とは、任意の言語や記述方法にて記述されたデータ処理方法であり、ソースコードやバイナリコード等の形式を問わない。なお、「プログラム」は必ずしも単一的に構成されるものに限られず、複数のモジュールやライブラリとして分散構成されるものや、ＯＳに代表される別個のプログラムと協働してその機能を達成するものをも含む。なお、実施の形態に示した各装置において記録媒体を読み取るための具体的な構成、読み取り手順、あるいは、読み取り後のインストール手順等については、周知の構成や手順を用いることができる。

【００８２】

また、塩基配列決定装置１００は、既知のパーソナルコンピュータ、ワークステーション等の情報処理端末等の情報処理装置にプリンタやモニタやイメージスキャナ等の周辺装置を接続し、該情報処理装置に本発明にかかる塩基配列決定方法を実現させるソフトウェア（プログラム、データ等を含む）を実装することにより実現してもよい。

【００８３】

さらに、塩基配列決定装置１００の分散・統合の具体的形態は図示のものに限られず、その全部または一部を、各種の負荷等に応じた任意の単位で、機能的または物理的に分散・統合して構成することができる。例えば、各データベースを独立したデータベース装置として独立に構成してもよく、また、処理の一部をＣＧＩ（Ｃｏｍｍｏｎ　Ｇａｔｅｗａｙ　Ｉｎｔｅｒｆａｃｅ）を用いて実現してもよい。

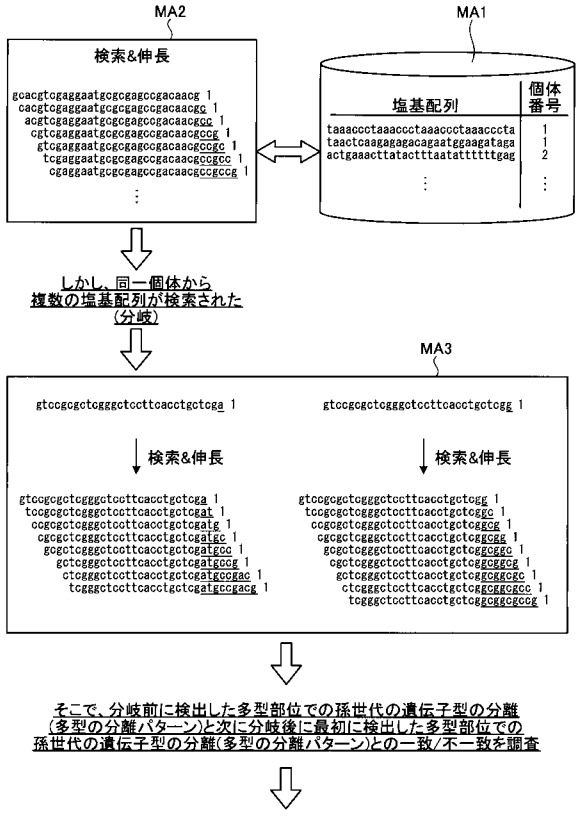
【産業上の利用可能性】

【００８４】

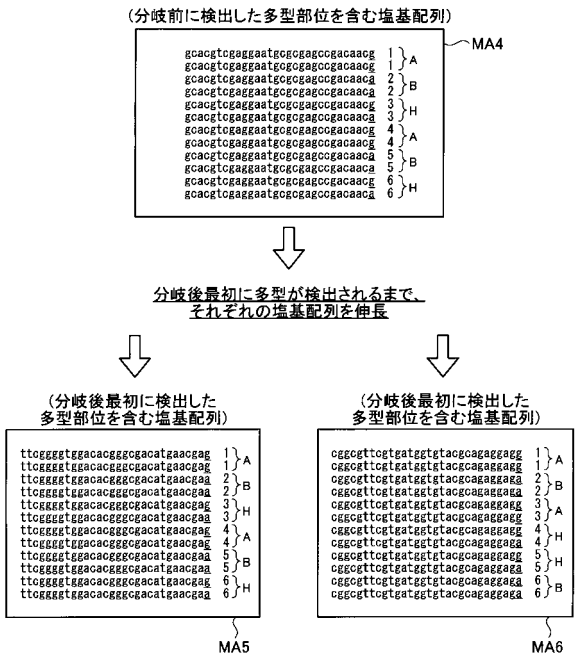
以上のように、本発明にかかる塩基配列決定プログラム、塩基配列決定装置および塩基配列決定方法は、医療や製薬や創薬や生物学研究などの様々な分野において極めて有用で

ある。

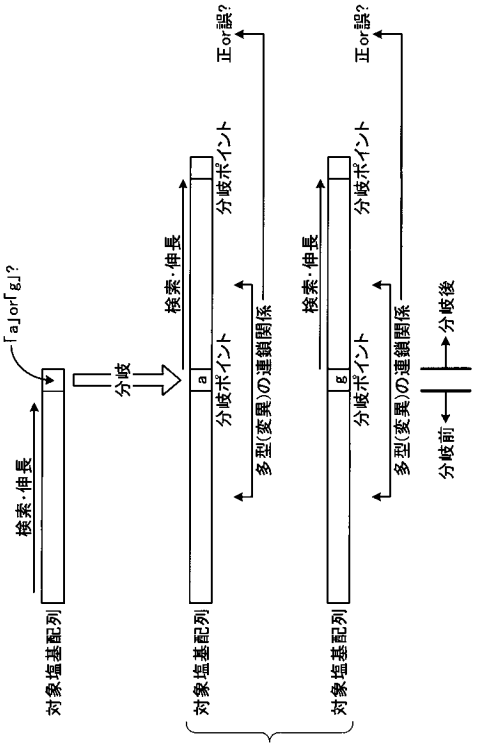
【図１】


MA2
検索&伸長
MA1
塩基配列
個体番号
しかし、同一個体から
複数の塩基配列が検索された
（分岐）
MA3
検索&伸長
検索&伸長
そこで、分岐前に検出した多型部位での孫世代の遺伝子型の分離
（多型の分離パターン）と次に分岐後に最初に検出した多型部位での
孫世代の遺伝子型の分離（多型の分離パターン）との一致/不一致を調査

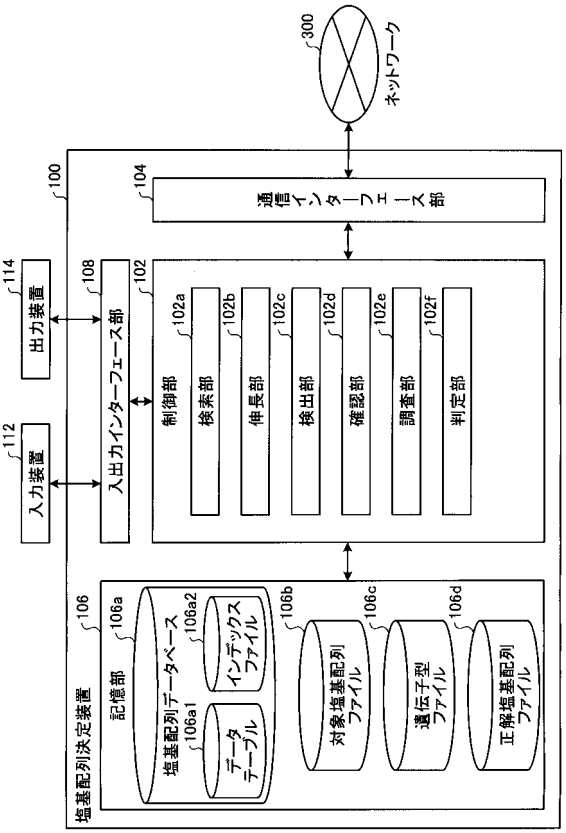
【図２】


（分岐前に検出した多型部位を含む塩基配列）
MA4
分岐後最初に多型が検出されるまで、
それぞれの塩基配列を伸長
（分岐後最初に検出した
多型部位を含む塩基配列）
（分岐後最初に検出した
多型部位を含む塩基配列）
MA5
MA6

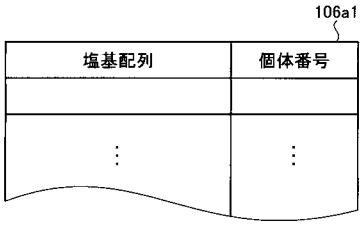
【図３】


「a」or「g」?
検索・伸長
対象塩基配列
分岐
検索・伸長
分岐ポイント
a
多型(変異)の連鎖関係
正or誤?
対象塩基配列
検索・伸長
分岐ポイント
g
多型(変異)の連鎖関係
正or誤?
分岐後
分岐前
対象塩基配列

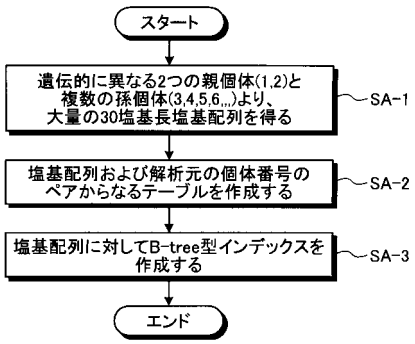
【図４】


300
ネットワーク
100
104
通信インターフェース部
114
出力装置
108
102
112
入力装置
入出力インターフェース部
制御部
102a
検索部
102b
伸長部
102c
検出部
102d
確認部
102e
調査部
102f
判定部
塩基配列決定装置
106
記憶部
106a
塩基配列データベース
106a1
データテーブル
106a2
インデックスファイル
106b
対象塩基配列ファイル
106c
遺伝子型ファイル
106d
正解塩基配列ファイル

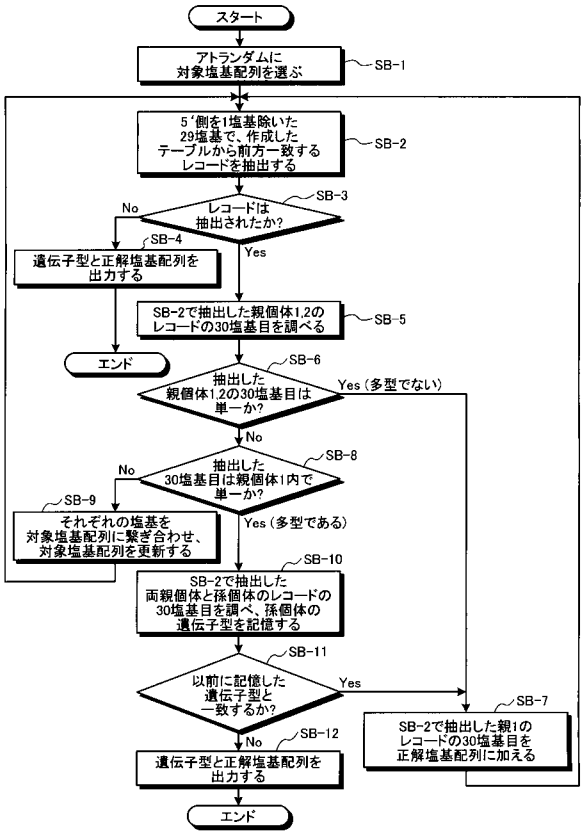
【図５】


106a1
塩基配列
個体番号

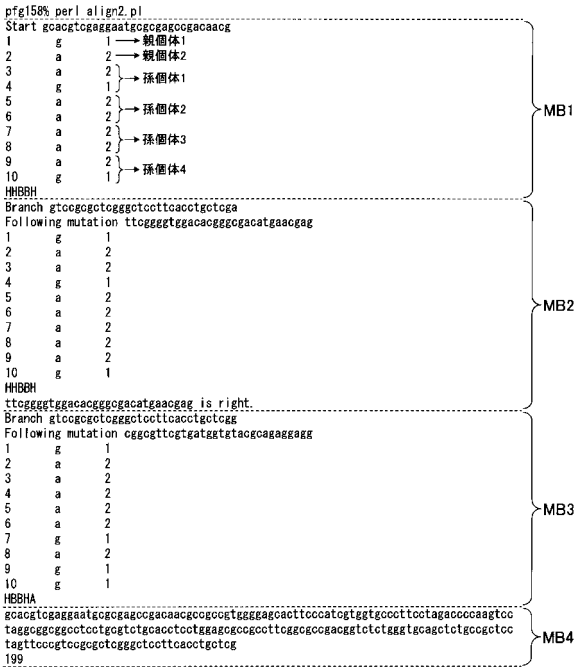
【図６】


スタート
遺伝的に異なる2つの親個体(1,2)と
複数の孫個体(3,4,5,6,,,)より、
大量の30塩基長塩基配列を得る
SA-1
塩基配列および解析元の個体番号の
ペアからなるテーブルを作成する
SA-2
塩基配列に対してB-tree型インデックスを
作成する
SA-3
エンド

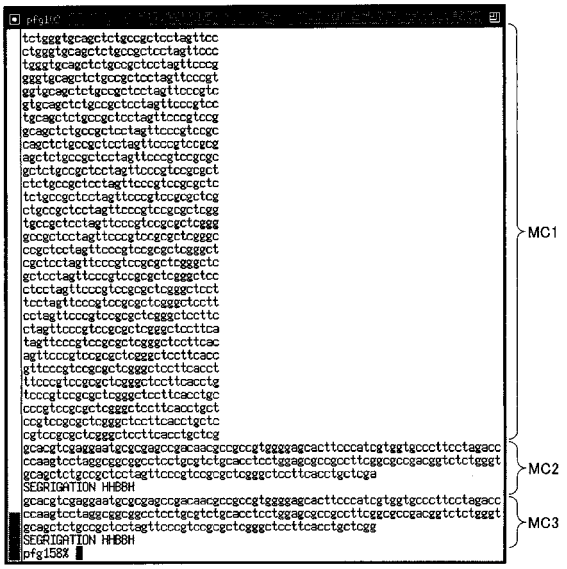
【図７】


スタート
アトランダムに
対象塩基配列を選ぶ
SB-1
5'側を1塩基除いた
29塩基で、作成した
テーブルから前方一致する
レコードを抽出する
SB-2
SB-3
レコードは
抽出されたか?
No
Yes
SB-4
遺伝子型と正解塩基配列を
出力する
SB-2で抽出した親個体1,2の
レコードの30塩基目を調べる
SB-5
エンド
SB-6
抽出した
親個体1,2の30塩基目は
単一か?
Yes (多型でない)
No
SB-8
抽出した
30塩基目は親個体1内で
単一か?
No
SB-9
それぞれの塩基を
対象塩基配列に繋ぎ合わせ、
対象塩基配列を更新する
Yes (多型である)
SB-10
SB-2で抽出した
両親個体と孫個体のレコードの
30塩基目を調べ、孫個体の
遺伝子型を記憶する
SB-11
以前に記憶した
遺伝子型と
一致するか?
Yes
No
SB-7
SB-2で抽出した親1の
レコードの30塩基目を
正解塩基配列に加える
SB-12
遺伝子型と正解塩基配列を
出力する
エンド

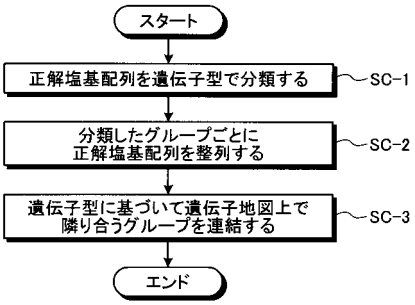
【図８】


pfg158% perl align2.pl
Start gcacgtcgaggaatgcgcgagccgacaacgccgccgtggggagcacttcccatcgtggtgcccttcctagaccccaagtcc
1 g 1 → 親個体1
2 a 2 → 親個体2
3 a 2 → 孫個体1
4 g 1
5 a 2 → 孫個体2
6 a 2
7 a 2 → 孫個体3
8 a 2
9 a 2 → 孫個体4
10 g 1
HHBBH
MB1
Branch gtcgcggctcgggctccttcacctgctcga
Following mutation ttcggggtggacacgggcgacatgaacgag
1 g 1
2 a 2
3 a 2
4 g 1
5 a 2
6 a 2
7 a 2
8 a 2
9 a 2
10 g 1
HHBBH
MB2
ttcggggtggacacgggcgacatgaacgag is right.
Branch gtcgcggctcgggctccttcacctgctcgg
Following mutation cggcgttcgtgatggtgtacgcagaggagg
1 g 1
2 a 2
3 a 2
4 a 2
5 a 2
6 a 2
7 g 1
8 a 2
9 g 1
10 g 1
HBBHA
MB3
gcacgtcgaggaatgcgcgagccgacaacgccgccgtggggagcacttcccatcgtggtgcccttcctagaccccaagtcc
taggcggcggcctcctgcgtctgcacctcctggagcgccgccttcggcgccgacggtctctgggtgcagctctgccgctcc
tagttcccgtccgcgctcgggctccttcacctgctcg
199
MB4

【図９】


pfg158%
tctgggtgcagctctgccgctcctagttcc
ctgggtgcagctctgccgctcctagttccc
tgggtgcagctctgccgctcctagttcccg
gggtgcagctctgccgctcctagttcccgt
ggtgcagctctgccgctcctagttcccgtc
gtgcagctctgccgctcctagttcccgtcc
tgcagctctgccgctcctagttcccgtccg
gcagctctgccgctcctagttcccgtccgc
cagctctgccgctcctagttcccgtccgcg
gctctgccgctcctagttcccgtccgcgc
agctctgccgctcctagttcccgtccgcgc
gctctgccgctcctagttcccgtccgcgct
ctctgccgctcctagttcccgtccgcgctc
tctgccgctcctagttcccgtccgcgctcg
ctgccgctcctagttcccgtccgcgctcgg
tgccgctcctagttcccgtccgcgctcggg
gccgctcctagttcccgtccgcgctcgggc
ccgctcctagttcccgtccgcgctcgggct
cgctcctagttcccgtccgcgctcgggctc
gctcctagttcccgtccgcgctcgggctcc
ctcctagttcccgtccgcgctcgggctcct
tcctagttcccgtccgcgctcgggctcctt
cctagttcccgtccgcgctcgggctccttc
ctagttcccgtccgcgctcgggctccttca
tagttcccgtccgcgctcgggctccttcac
agttcccgtccgcgctcgggctccttcacc
gttcccgtccgcgctcgggctccttcacct
ttcccgtccgcgctcgggctccttcacctg
tcccgtccgcgctcgggctccttcacctgc
cccgtccgcgctcgggctccttcacctgct
ccgtccgcgctcgggctccttcacctgctc
cgtccgcgctcgggctccttcacctgctcg
MC1
gcacgtcgaggaatgcgcgagccgacaacgccgccgtggggagcacttcccatcgtggtgcccttcctagacc
ccaagtcctaggcggcggcctcctgcgtctgcacctcctggagcgccgccttcggcgccgacggtctctgggt
gcagctctgccgctcctagttcccgtccgcgctcgggctccttcacctgctcga
SEGRIGATION HHBBH
MC2
gcacgtcgaggaatgcgcgagccgacaacgccgccgtggggagcacttcccatcgtggtgcccttcctagacc
ccaagtcctaggcggcggcctcctgcgtctgcacctcctggagcgccgccttcggcgccgacggtctctgggt
gcagctctgccgctcctagttcccgtccgcgctcgggctccttcacctgctcgg
SEGRIGATION HHBBH
MC3
pfg158%

【図１０】


スタート
正解塩基配列を遺伝子型で分類する
SC-1
分類したグループごとに
正解塩基配列を整列する
SC-2
遺伝子型に基づいて遺伝子地図上で
隣り合うグループを連結する
SC-3
エンド

フロントページの続き

(56)参考文献 特開２００３－５３０６３１（ＪＰ，Ａ）
特開２００３－１５７２６７（ＪＰ，Ａ）
特開平１０－０４０２５７（ＪＰ，Ａ）
Sundquist, A., Whole-Genome Sequencing and Assembly with High-Throughput, Short-Read Technologies, PLOS ONE, ２００７年　5月３０日, Vol.2, No.5, p.e484
Fasulo, D., Efficiently detecting polymorphisms during the fragment assembly process, Bioinformatics, ２００２年, Vol.18, Suppl.1, p.S294-302
Vinson, J.P., Assembly of polymorphic genomes: algorithms and application to Ciona savignyi, Genome Research, ２００５年　8月, Vol.15, No.8, p.1127-1135
Kim, J.H., Accuracy assessment of diploid consensus sequences, IEEE/ACM transactions on computational biology and bioinformatics, ２００７年　3月, Vol.4, No.1, p.88-97

(58)調査した分野(Int.Cl., ＤＢ名)
Ｇ０６Ｆ　１９／１０
Ｃ１２Ｍ　１／００
Ｃ１２Ｑ　１／６８
Ｇ０６Ｆ　１７／３０
Ｃ１２Ｎ　１５／０９
ＪＳＴＰｌｕｓ／ＪＭＥＤＰｌｕｓ／ＪＳＴ７５８０（ＪＤｒｅａｍＩＩＩ）
ＰｕｂＭｅｄ