

(19) 日本国特許庁(JP)

再公表特許(A1)

(11) 国際公開番号

W02009/057757

発行日 平成23年3月10日 (2011.3.10)

(43) 国際公開日 平成21年5月7日 (2009.5.7)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 19/22 (2011.01)	G06F 19/00 622	4B024
G06F 17/30 (2006.01)	G06F 17/30 170F	4B029
C12M 1/00 (2006.01)	C12M 1/00 ZNAA	4B063
C12Q 1/68 (2006.01)	C12Q 1/68 Z	5B075
C12N 15/09 (2006.01)	C12N 15/00 A	

審査請求 未請求 予備審査請求 未請求 (全 26 頁)

出願番号 特願2009-539130 (P2009-539130)  
 (21) 国際出願番号 PCT/JP2008/069897  
 (22) 国際出願日 平成20年10月31日 (2008.10.31)  
 (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)

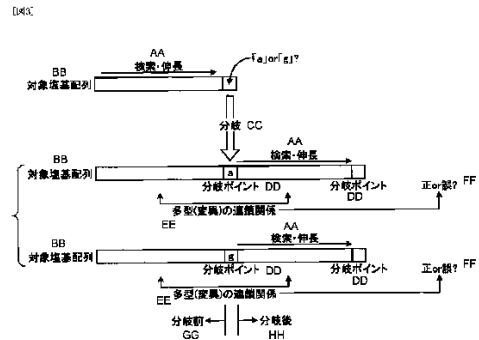
(71) 出願人 501167644  
 独立行政法人農業生物資源研究所  
 茨城県つくば市観音台2丁目1-2  
 (74) 代理人 100089118  
 弁理士 酒井 宏明  
 (72) 発明者 官尾 安▲藝▼雄  
 茨城県つくば市観音台2丁目1-2 独立  
 行政法人農業生物資源研究所内  
 Fターム(参考) 4B024 AA19 AA20 CA01 HA11  
 4B029 AA07 FA15  
 4B063 QA13 QQ42 QR32 QS39  
 5B075 ND03 UU19

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) 【要約】

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課題とする。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。



AA SEARCH/EXTENSION  
 BB TARGET BASE SEQUENCE  
 CC BRANCHING  
 DD BRANCHING POINT  
 EE CHAIN REACTION AMONG POLYMORPHISMS (MUTATIONS)  
 FF CORRECT OR ERROR?  
 GG BEFORE BRANCHING  
 HH AFTER BRANCHING

**【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、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

を特徴とする塩基配列決定プログラム。

10

**【請求項 2】**

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わ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、

20

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、

30

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 3】**

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 2 に記載の塩基配列決定プログラム。

40

**【請求項 4】**

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 3 に記載の塩基配列決定プログラム。

**【請求項 5】**

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

50

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  
を特徴とする塩基配列決定装置。

【請求項 6】

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わ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 7】

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 6 に記載の塩基配列決定装置。

【請求項 8】

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 7 に記載の塩基配列決定装置。

【請求項 9】

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

を特徴とする塩基配列決定方法。

【請求項 10】

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わせる前記塩基配列を検索する

10

20

30

40

50

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 11】

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 10 に記載の塩基配列決定方法。

#### 【請求項 12】

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 11 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

10

20

30

40

50

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.

10

【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

20

【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、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。

30

【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。

40

【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わ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

50

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。

10

**【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。

20

**【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わ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。

30

40

**【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

50

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。

10

【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わ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。

20

30

【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。

40

【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

50

て、当該対象塩基配列に繋ぎ合わ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

10

20

30

40

50



- 1 0 6 c 遺伝子型ファイル
- 1 0 6 d 正解塩基配列ファイル
- 1 0 8 入出力インターフェース部
- 1 1 0 入力装置
- 1 1 2 出力装置

3 0 0 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。

10

【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。

20

【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「gcacgtcggaggaaatgcgcgagccgacaaacg」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「gcacgtcggaggaaatgcgcgagccgacaaacg」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 'cacgtcggaggaaatgcgcgagccgacaaacg%」;」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。

30

〔操作2〕検索結果として、親個体1の塩基配列「cacgtcggaggaaatgcgcgagccgacaaacg 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」と決定する。

40

〔操作3〕決定した塩基「c」を対象塩基配列「gcacgtcggaggaaatgcgcgagccgacaaacg」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「gtccgcgctcgggctcccttcacctgctcga 1」と親個体1の塩基配列「gtccgcgctcgggctcccttcacctgctcgg 1」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「gtccgcgctcgggctcccttcacctgctcgg」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

50

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 M A 3 に示すように、全く異なった塩基配列（「 a t g c c g a c g 」と「 g c g g c g c c g 」）で伸長されていく。

#### 【 0 0 2 8 】

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。

#### 【 0 0 2 9 】

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 M A 4 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「 g g 」（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「 a a 」（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「 g a 」であり、孫個体 4 の 3' 末端での 2 塩基は「 g g 」であり、孫個体 5 の 3' 末端での 2 塩基は「 a a 」であり、孫個体 6 の 3' 末端での 2 塩基は「 g a 」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「 A B H A B A 」と表すことができる。

#### 【 0 0 3 0 】

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 M A 5 および M A 6 に示すものであったとする。図 2 の M A 5 および M A 6 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 M A 5 に示す塩基配列に対しては、遺伝子型は「 A B H A B H 」と表すことができ、図 2 の M A 6 に示す塩基配列に対しては、遺伝子型は「 A B A H H B 」と表すことができる。なお、図 2 の M A 5 に示す塩基配列は、図 1 の M A 3 の左側に示すように「 . . . t g c t c g 」の後に「 a t g c c g a c g 」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 M A 6 に示す塩基配列は、図 1 の M A 3 の右側に示すように「 . . . t g c t c g 」の後に「 g c g g c g c c g 」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。

#### 【 0 0 3 1 】

そして、本発明は、分岐前の遺伝子型「 A B H A B H 」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 M A 3 の左側に示すように「 . . . t g c t c g 」の後に「 a t g c c g a c g 」と続けて伸長するのが正解であると判定する。

#### 【 0 0 3 2 】

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（ F 2 ））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

10

20

30

40

50

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。

10

#### 【0034】

現在、一度に数百万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。

20

#### 【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。

30

#### 【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。

40

#### 【0039】

50

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わせることができる。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。

10

【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】

20

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。

30

【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106c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。

40

【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。

50

## 【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。

10

## 【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わ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わ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。

20

## 【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。

30

## 【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。

40

## 【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】

50

### [ 3 - 1 . データベース構築処理 ]

まず、塩基配列決定装置 1 0 0 の制御部 1 0 2 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 1 0 0 の制御部 1 0 2 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。

#### 【 0 0 5 4 】

まず、制御部 1 0 2 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 3 0 塩基長の塩基配列を取得する（ステップ S A - 1 ）。

#### 【 0 0 5 5 】

つぎに、制御部 1 0 2 は、ステップ S A - 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 1 0 6 a のデータテーブル 1 0 6 a 1 に格納する（ステップ S A - 2 ）。

#### 【 0 0 5 6 】

つぎに、制御部 1 0 2 は、ステップ S A - 1 で取得した塩基配列に対して B - t r e e 型インデックスを作成し、作成した B - t r e e 型インデックスを塩基配列データベース 1 0 6 a のインデックスファイル 1 0 6 a 2 に格納する（ステップ S A - 3 ）。

#### 【 0 0 5 7 】

### [ 3 - 2 . 塩基配列決定処理 ]

つぎに、塩基配列決定装置 1 0 0 の制御部 1 0 2 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 1 0 0 の制御部 1 0 2 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。

#### 【 0 0 5 8 】

まず、制御部 1 0 2 は、塩基配列データベース 1 0 6 a のデータテーブル 1 0 6 a 1 に格納されている親個体 1 の塩基配列の中から 3 0 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 1 0 6 d に格納する（ステップ S B - 1 ）。

#### 【 0 0 5 9 】

つぎに、制御部 1 0 2 は、検索部 1 0 2 a で、ステップ S B - 1 で決定した対象塩基配列の 5 ´ 側の一塩基を除く 2 9 塩基（対象塩基配列の 3 ´ 側から 2 9 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 1 0 6 a のデータテーブル 1 0 6 a 1 から、インデックスファイル 1 0 6 a 2 に格納されている B - t r e e 型インデックスを参照して抽出（検索）する（ステップ S B - 2 ）。

#### 【 0 0 6 0 】

つぎに、制御部 1 0 2 は、ステップ S B - 2 でレコードが抽出されなかった場合（ステップ S B - 3 : N o ）には、遺伝子型ファイル 1 0 6 c に格納されている遺伝子型と正解塩基配列ファイル 1 0 6 d に格納されている正解塩基配列とを出力装置 1 1 2 に出力して（ステップ S B - 4 ）、本処理を終了する。

#### 【 0 0 6 1 】

つぎに、制御部 1 0 2 は、ステップ S B - 2 でレコードが抽出された場合（ステップ S B - 3 : Y e s ）には、検出部 1 0 2 c で、ステップ S B - 2 で抽出した親個体 1 , 2 のレコードの 5 ´ 側から 3 0 番目の塩基（3 ´ 側から一番目の塩基）を調べる（ステップ S B - 5 ）。

#### 【 0 0 6 2 】

つぎに、制御部 1 0 2 は、ステップ S B - 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 S B - 6 : Y e s ）には、伸長部 1 0 2 b 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 1 0 6 d に格納されている正解塩基配列の 3 ´ 側に加え（ステップ S B - 7 ）、ステップ S B - 2 の処理に戻る。

#### 【 0 0 6 3 】

10

20

30

40

50

つぎに、制御部 102 は、ステップ S B - 5 で調べた親個体 1, 2 の塩基が単一でなく (ステップ S B - 6 : N o)、さらに確認部 102 d 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 S B - 8 : N o) には、伸長部 102 b 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 S B - 9)、ステップ S B - 2 の処理に戻る。

【0064】

つぎに、制御部 102 は、ステップ S B - 5 で調べた親個体 1, 2 の塩基が単一でなく (ステップ S B - 6 : N o) さらに確認部 102 d 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 S B - 8 : Y e s)、すなわち多型であると検出された場合には、調査部 102 e で、ステップ S B - 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 106 c に格納する (ステップ S B - 10)。

10

【0065】

つぎに、制御部 102 は、判定部 102 f により、ステップ S B - 10 で調べた遺伝子型と以前に遺伝子型ファイル 106 c 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 S B - 11 : Y e s) には、ステップ S B - 7 の処理に戻る。

【0066】

つぎに、制御部 102 は、判定部 102 f により、ステップ S B - 10 で調べた遺伝子型と以前に遺伝子型ファイル 106 c 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 S B - 11 : N o) には、遺伝子型ファイル 106 c に格納されている遺伝子型と正解塩基配列ファイル 106 d 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 S B - 12)、本処理を終了する。

20

【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。

30

【0068】

また、本処理では、図 8 の M B 1 に示すように、親個体 1 の塩基配列 (S t a r t 「g c a c g t c g a g g a a t g c g c g a g c c g a c a a c g」) 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「H H B B H」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「H B B H」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

40

50

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。

10

## 【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 MB 2 に示す Branch 配列「gtccgcgctcgggctccttcacctgctcga」および図 8 の MB 3 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 MB 2 に示す Following mutation 配列「ttcggggtggacacgggacacatgaacgag」の 5' 側から 30 塩基目「g」、および図 8 の MB 3 に示す Following mutation 配列「cggcgctcgtgatgggtgtacggagaggagg」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 MB 2 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 MB 3 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 MB 2 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 MB 4 に示す正解塩基配列を出力する。

20

## 【0070】

また、本処理では、図 9 の MC 1 に示すように、図 9 の MC 1 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 MC 2 に示す塩基配列の 3' 末端および図 9 の MC 3 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。

30

## 【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。

40

## 【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

50



を整理（ソート）する（ステップ S C - 2）。

【 0 0 7 4 】

つぎに、制御部 1 0 2 は、ステップ S C - 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グループ同士を、遺伝子型に基づいて連結する（ステップ S C - 3）。

【 0 0 7 5 】

[ 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 1 0 0 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わ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。

10

20

【 0 0 7 6 】

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。

【 0 0 7 7 】

また、塩基配列決定装置 1 0 0 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、S Q L で塩基配列を検索する。これにより、次に繋ぎ合わせる塩基配列を効率よく検索することができる。

30

【 0 0 7 8 】

また、塩基配列決定装置 1 0 0 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 - t r e e 型インデックスをさらに記憶し、当該 B - t r e e 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、次に繋ぎ合わせる塩基配列を高速に検索することができる。

【 0 0 7 9 】

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 1 0 0 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。

40

【 0 0 8 0 】

50

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。

10

#### 【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。

20

#### 【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。

30

#### 【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。

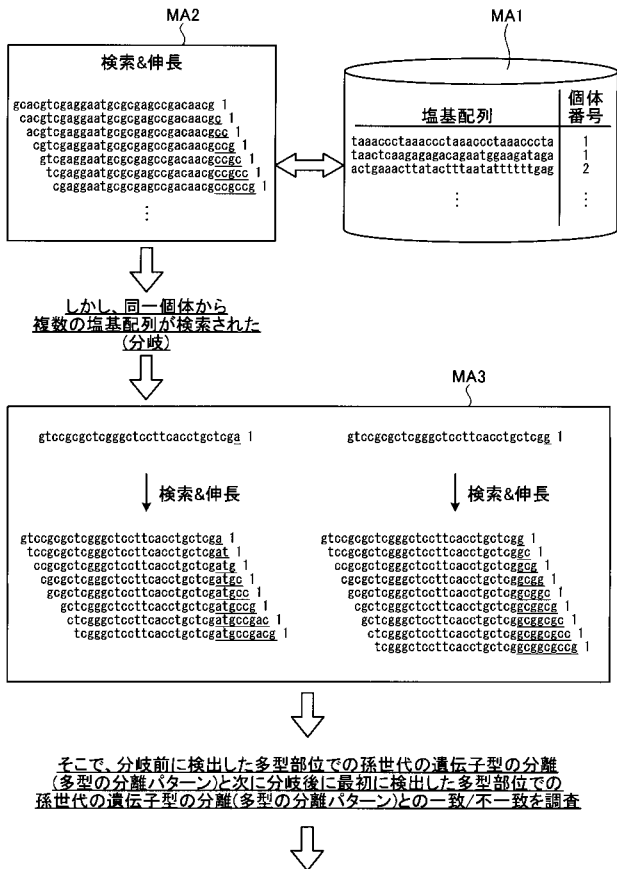
40

#### 【産業上の利用可能性】

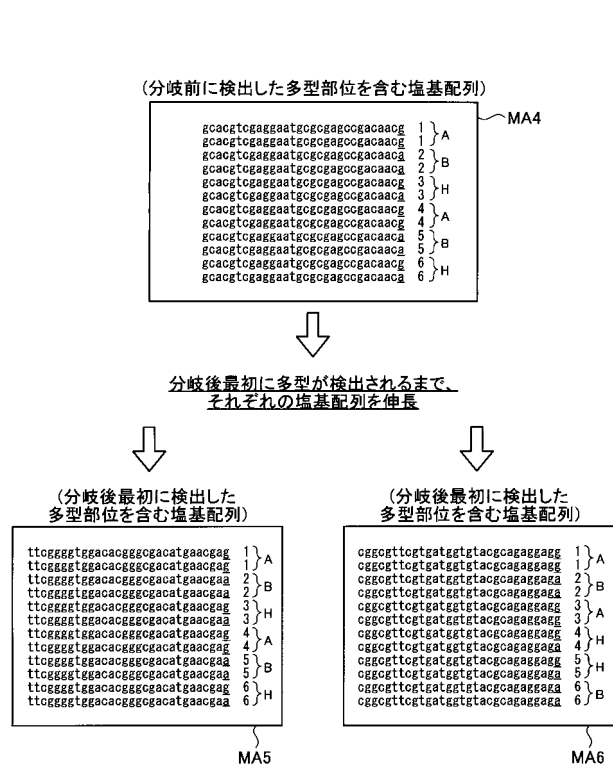
#### 【0084】

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。

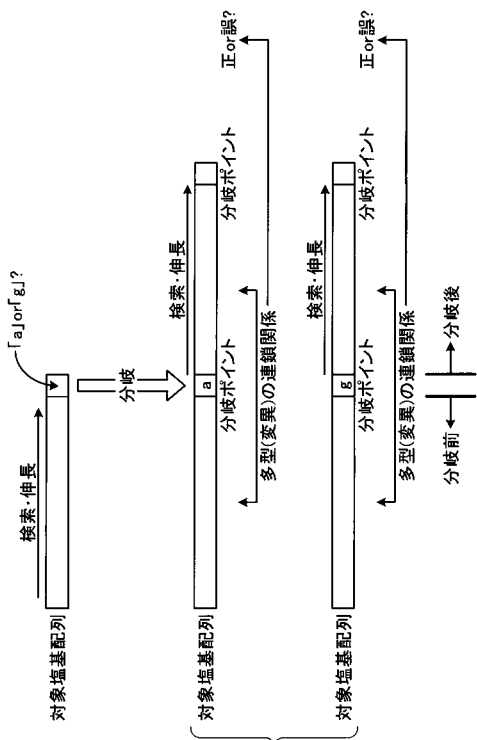
【図1】



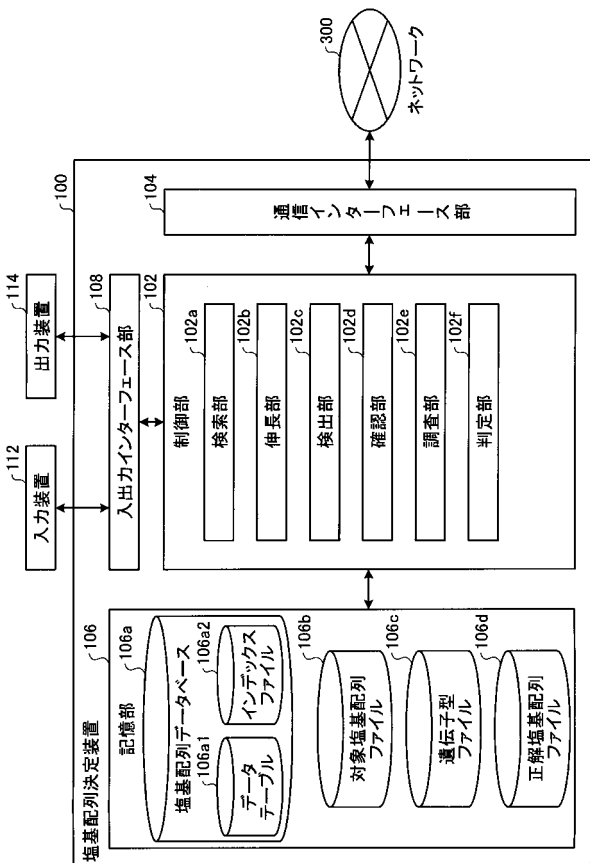
【図2】



【図3】



【図4】

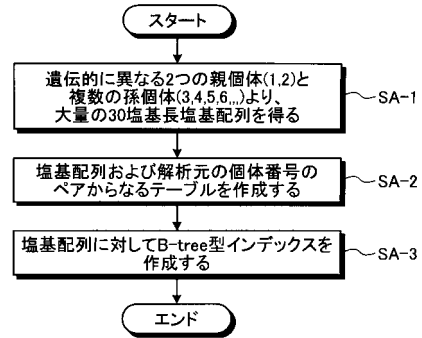


【 図 5 】

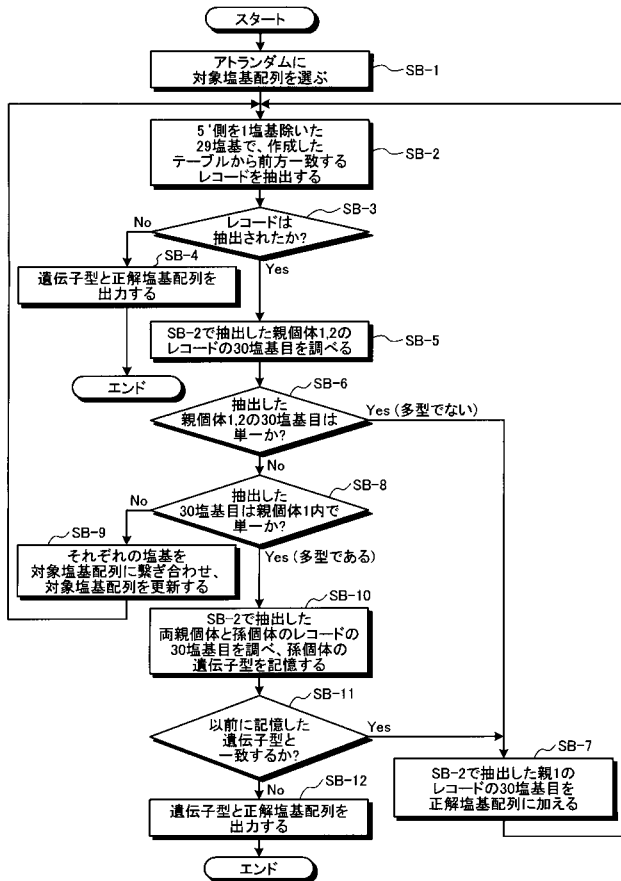
塩基配列	個体番号
⋮	⋮

106a1

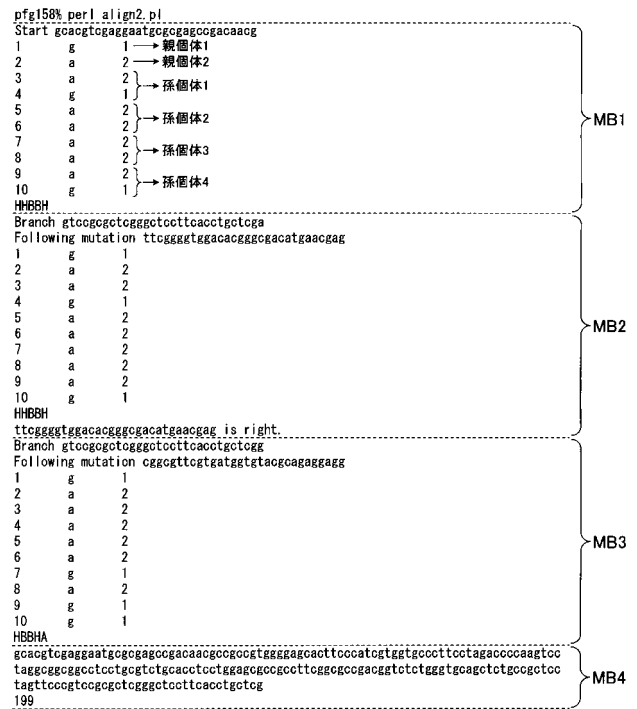
【 図 6 】



【 図 7 】



【 図 8 】



【 図 9 】

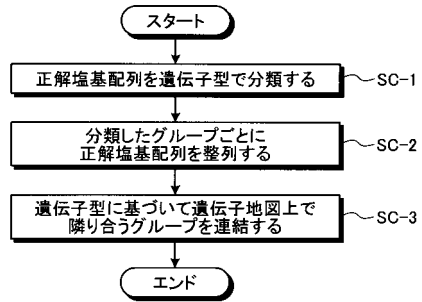
```

tctgggtgcagctctgocgctcttagttcc
ctgggtgcagctctgocgctcttagttcc
ttgggtgcagctctgocgctcttagttccg
gggtgcagctctgocgctcttagttccgt
gtgcagctctgocgctcttagttccgtc
gtgcagctctgocgctcttagttccgtcc
tgagctctgocgctcttagttccgtccg
gcagctctgocgctcttagttccgtccgc
cagctctgocgctcttagttccgtccgcg
agctctgocgctcttagttccgtccgcg
gctctgocgctcttagttccgtccgcgct
ctctgocgctcttagttccgtccgcgctc
tctgocgctcttagttccgtccgcgctcg
ctgocgctcttagttccgtccgcgctcgg
tgocgctcttagttccgtccgcgctcggg
gcgctcttagttccgtccgcgctcgggc
cgcctcttagttccgtccgcgctcgggct
cgcctcttagttccgtccgcgctcgggctc
ctcttagttccgtccgcgctcgggctcct
tcttagttccgtccgcgctcgggctcctt
ccttagttccgtccgcgctcgggctccttc
ctagttccgtccgcgctcgggctccttca
tagttccgtccgcgctcgggctccttca
agtccgtccgcgctcgggctccttcaac
gttccgtccgcgctcgggctccttcaact
ttccgtccgcgctcgggctccttcaactg
ttccgtccgcgctcgggctccttcaactgc
ccgtccgcgctcgggctccttcaactgct
cgtccgcgctcgggctccttcaactgctc
cgtccgcgctcgggctccttcaactgctcg
gcagctcaggaatgcagcgcgacacgcgcgctggggagcacttccatcgtgtcccttcttagacc
ccagctcctaggcgcgctcctctgcctctgacactcctggagcgcgctcggcgcgacgctctgggt
gcagctctgocgctcttagttccgtccgcgctcgggctccttcaactgctoga
SEGR10ATI0N HH88H
gcagctcaggaatgcagcgcgacacgcgcgctggggagcacttccatcgtgtcccttcttagacc
ccagctcctaggcgcgctcctctgcctctgacactcctggagcgcgctcggcgcgacgctctgggt
gcagctctgocgctcttagttccgtccgcgctcgggctccttcaactgctcgg
SEGR10ATI0N HH88H
pfg158x

```

MC1  
MC2  
MC3

【 図 1 0 】



## 【 国際調査報告 】

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/069897

A. CLASSIFICATION OF SUBJECT MATTER G06F19/00(2006.01)i, C12M1/00(2006.01)i, C12Q1/68(2006.01)i, G06F17/30(2006.01)i, C12N15/09(2006.01)n According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06F19/00, C12M1/00, C12Q1/68, G06F17/30, C12N15/09 Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched Jitsuyo Shinan Koho 1922-1996 Jitsuyo Shinan Toroku Koho 1996-2008 Kokai Jitsuyo Shinan Koho 1971-2008 Toroku Jitsuyo Shinan Koho 1994-2008 Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) PubMed, JSTPlus (JDreamII), JMEDPlus (JDreamII)		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SUNDQUIST, A., Whole-genome sequencing and assembly with high-throughput, short-read technologies, PLoS ONE, 2007.05.30, Vol.2, issue 5, p.e484	1-12
A	FASULO, D., Efficiently detecting polymorphisms during the fragment assembly process, Bioinformatics, 2002, Vol.18, Suppl.1, p.S294-302	1-12
A	VINSON, J.P., Assembly of polymorphic genomes: algorithms and application to Ciona savignyi, Genome research, 2005.08, Vol.15, No.8, p.1127-35	1-12
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.		
* Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier application or patent but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art "&" document member of the same patent family		
Date of the actual completion of the international search 28 November, 2008 (28.11.08)		Date of mailing of the international search report 09 December, 2008 (09.12.08)
Name and mailing address of the ISA/ Japanese Patent Office		Authorized officer
Facsimile No.		Telephone No.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2008/069897

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	KIM, J.H., Accuracy assessment of diploid consensus sequences, IEEE/ACM transactions on computational biology and bioinformatics, 2007.03, Vol.4, No.1, p.88-97	1-12
A	JP 2003-530631 A (PE Corp.(NY)), 14 October, 2003 (14.10.03), Full text; all drawings & US 6714874 B1 & EP 1285390 A & WO 2001/063543 A2 & AU 3655501 A & CA 2400890 A	1-12
A	JP 2003-157267 A (Hitachi, Ltd.), 30 May, 2003 (30.05.03), Full text; all drawings & US 2003/0104408 A1	1-12
A	JP 10-040257 A (Hitachi, Ltd.), 13 February, 1998 (13.02.98), Full text; all drawings (Family: none)	1-12

国際調査報告		国際出願番号 PCT/JP2008/069897									
A. 発明の属する分野の分類 (国際特許分類 (IPC)) Int.Cl. G06F19/00(2006.01)i, C12M1/00(2006.01)i, C12Q1/68(2006.01)i, G06F17/30(2006.01)i, C12N15/09(2006.01)n											
B. 調査を行った分野 調査を行った最小限資料 (国際特許分類 (IPC)) Int.Cl. G06F19/00, C12M1/00, C12Q1/68, G06F17/30, C12N15/09											
最小限資料以外の資料で調査を行った分野に含まれるもの <table border="0"> <tr> <td>日本国実用新案公報</td> <td>1922-1996年</td> </tr> <tr> <td>日本国公開実用新案公報</td> <td>1971-2008年</td> </tr> <tr> <td>日本国実用新案登録公報</td> <td>1996-2008年</td> </tr> <tr> <td>日本国登録実用新案公報</td> <td>1994-2008年</td> </tr> </table>				日本国実用新案公報	1922-1996年	日本国公開実用新案公報	1971-2008年	日本国実用新案登録公報	1996-2008年	日本国登録実用新案公報	1994-2008年
日本国実用新案公報	1922-1996年										
日本国公開実用新案公報	1971-2008年										
日本国実用新案登録公報	1996-2008年										
日本国登録実用新案公報	1994-2008年										
国際調査で使用了電子データベース (データベースの名称、調査に使用了用語) PubMed, JSTPlus(JDreamII), JMEDPlus(JDreamII)											
C. 関連すると認められる文献											
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号									
A	SUNDQUIST, A., Whole-genome sequencing and assembly with high-throughput, short-read technologies, PLoS ONE, 2007.05.30, Vol.2, issue 5, p.e484	1-12									
A	FASULO, D., Efficiently detecting polymorphisms during the fragment assembly process, Bioinformatics, 2002, Vol.18, Suppl.1, p.S294-302	1-12									
<input checked="" type="checkbox"/> C欄の続きにも文献が列挙されている。 <input type="checkbox"/> パテントファミリーに関する別紙を参照。											
* 引用文献のカテゴリー 「A」特に関連のある文献ではなく、一般的技術水準を示すもの 「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの 「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す) 「O」口頭による開示、使用、展示等に言及する文献 「P」国際出願日前で、かつ優先権の主張の基礎となる出願		の日の後に公表された文献 「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの 「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの 「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの 「&」同一パテントファミリー文献									
国際調査を完了した日 28.11.2008		国際調査報告の発送日 09.12.2008									
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号100-8915 東京都千代田区霞が関三丁目4番3号		特許庁審査官 (権限のある職員) 宮久保 博幸	5L 3136								
		電話番号 03-3581-1101	内線 3562								



国際調査報告		国際出願番号 PCT/JP2008/069897
C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求の範囲の番号
A	VINSON, J.P., Assembly of polymorphic genomes: algorithms and application to <i>Ciona savignyi</i> , Genome research, 2005.08, Vol.15, No.8, p.1127-35	1-12
A	KIM, J.H., Accuracy assessment of diploid consensus sequences, IEEE/ACM transactions on computational biology and bioinformatics, 2007.03, Vol.4, No.1, p.88-97	1-12
A	JP 2003-530631 A (ピーイー コーポレーション (エヌワイ)) 2003.10.14, 全文, 全図 & US 6714874 B1 & EP 1285390 A & WO 2001/063543 A2 & AU 3655501 A & CA 2400890 A	1-12
A	JP 2003-157267 A (株式会社日立製作所) 2003.05.30, 全文, 全図 & US 2003/0104408 A1	1-12
A	JP 10-040257 A (株式会社日立製作所) 1998.02.13, 全文, 全図 (フ ァミリーなし)	1-12

## フロントページの続き

(81)指定国 AP(BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL, NO, PL, PT, RO, SE, SI, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(特許庁注：以下のものは登録商標)

1 . G S M

(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

(注)この公表は、国際事務局(WIPO)により国際公開された公報を基に作成したものである。なおこの公表に係る日本語特許出願(日本語実用新案登録出願)の国際公開の効果は、特許法第184条の10第1項(実用新案法第48条の13第2項)により生ずるものであり、本掲載とは関係ありません。