

【公報種別】特許法第17条の2の規定による補正の掲載

【部門区分】第6部門第3区分

【発行日】令和1年9月5日(2019.9.5)

【公表番号】特表2018-535484(P2018-535484A)

【公表日】平成30年11月29日(2018.11.29)

【年通号数】公開・登録公報2018-046

【出願番号】特願2018-520481(P2018-520481)

【国際特許分類】

G 16 B 30/00 (2019.01)

G 06 F 16/00 (2019.01)

【F I】

G 06 F 19/22

G 06 F 17/30 170 F

G 06 F 17/30 414 A

【手続補正書】

【提出日】令和1年7月25日(2019.7.25)

【手続補正1】

【補正対象書類名】特許請求の範囲

【補正対象項目名】全文

【補正方法】変更

【補正の内容】

【特許請求の範囲】

【請求項1】

検索配列を基準データに適合させるための方法であって、

演算装置により、

a) 基準データをメモリ内に保存することと、

b) 前記基準データに基づいて階層的索引表を作成することであって、前記作成することが、前記階層的索引表内の複数のレベルで複数のエントリを作成することを含み、

前記複数のレベルは第1のレベルを含み、第1のレベルのエントリは、第1の長さを有する前記基準データの複数の副配列のそれぞれのために作成され、

それぞれのレベルnのエントリに対して、nは少なくとも1つがゼロでない正の整数であり、前記作成は、閾値より大きい前記それぞれのレベルnのエントリの適合基準に応答して、前記階層的索引表内の前記それぞれのレベルnのエントリに対するn+1レベルのエントリを作成することを含み、

前記階層的索引表内にn+1レベルのエントリを作成することは、前記n+1レベルに対応する長さを有する前記基準データの複数の副配列のそれぞれに対するn+1レベルのエントリを作成することを含み、

c) 検索配列を特定する入力を受信することと、

d) 前記基準データへの前記検索配列の1つまたは複数の適合について前記階層的索引表を検索することとを含み、

前記階層的索引表の前記検索は、前記階層的索引表のその後のレベルで前記検索配列の小区分の適合を反復的に検索することを含む、

方法。

【請求項2】

前記第1の長さを有する前記基準データの前記複数の副配列、および前記n+1レベルに対応する長さを有する前記基準データの前記複数の副配列は、それぞれの長さの副配列の網羅的なセットをそれぞれ含む、請求項1に記載の方法。

【請求項3】

前記階層的索引表の任意のそれぞれのレベルでそれぞれのエントリを前記作成することが、

前記基準データ内の前記それぞれのエントリに対応する前記それぞれの長さの前記副配列の適合について検索することと、

前記階層的索引表のそれぞれのレベル内の前記それぞれのエントリに情報を保存することであって、前記情報が、前記基準データ内の前記基準データ内のそれぞれの長さの前記副配列の適合の数を特定し、前記情報が、前記基準データ内の前記適合の各々の場所をさらに特定し、保存することと、
により実行される、請求項2に記載の方法。

【請求項4】

$n + 1$ レベルに対して、前記 $n + 1$ レベルの前記それぞれの副配列の適合について前記検索することが、前記 n レベルにおける前記対応するエントリに関連付けられた場所で実行される、請求項3に記載の方法。

【請求項5】

各エントリに関連付けられた前記適合の数を示す前記データが、第1のデータ構造内に保存され、各エントリに関連付けられた前記適合の各々の前記データ場所が、第2のデータ構造内に保存され、前記第1及び第2のデータ構造が、前記階層的索引表内に各々含まれる、請求項3に記載の方法。

【請求項6】

各それぞれの n レベルのエントリに対して、前記それぞれの n レベルのエントリに対応する $n + 1$ レベルのエントリを参照するメモリ内にポインタを保存することをさらに含む、請求項1に記載の方法。

【請求項7】

前記基準データが、基準ゲノムを含み、前記基準データを検索することが、短い読み出し(SR)を前記基準ゲノムと整列させることを含む、請求項1に記載の方法。

【請求項8】

短い読み出しを基準ゲノムと整列させるためのプログラム命令を含むコンピュータ読み取り可能な記憶媒体であって、前記プログラム命令が、

a) 前記基準ゲノムをメモリ内に保存することと、

b) 前記基準ゲノムに基づいて階層的検索表を作成することであって、前記作成することが、前記階層的索引表において複数のレベルのそれぞれで複数のエントリを作成することを含み、前記複数のレベルは第1のレベルを含み、第1のレベルのエントリは、第1の長さを有する前記基準データの複数の副配列のそれぞれのために作成され、各それぞれのエントリが、前記それぞれのエントリに関連付けられた塩基対の配列の前記基準ゲノム内の場所に関連する情報を含み、各それぞれのレベル n のエントリについて、少なくとも1つのゼロでない正の整数 n に対して、前記作成することは、閾値より大きい前記それぞれのレベル n のエントリの適合基準に応答して、前記階層的索引表内のそれぞれのレベル n のエントリに対する前記 $n + 1$ レベルのエントリを作成することを含む、作成することと、

前記階層的索引表内に $n + 1$ レベルのエントリを作成することは、前記 $n + 1$ レベルに対応する長さを有する前記基準データの複数の副配列のそれぞれのために $n + 1$ レベルのエントリを作成することを含み、

c) 短い読み出しを特定する入力を受信することと、

d) 前記基準ゲノムへの前記短い読み出しの1つまたは複数の適合に対して前記階層的索引表を検索することであって、前記階層的索引表の前記検索は、前記階層的索引表のその後のレベルで前記短い読み出しの小区分の適合を反復的に検索することを含む、検索することと、

を行うように実行可能である、コンピュータ読み取り可能な記憶媒体。

【請求項9】

前記第1の長さを有する前記基準データの前記複数の副配列、および前記 $n + 1$ レベル

に対応する長さを有する前記基準データの前記複数の副配列は、それぞれの長さの副配列の網羅的なセットをそれぞれ含む、請求項8に記載の記憶媒体。

【請求項10】

前記階層的索引表の任意のそれぞれのレベル内のそれぞれのエントリを前記作成することが、

前記基準データ内の前記それぞれのエントリに対応する前記それぞれの長さの前記それぞれの副配列の適合について検索することと、

前記階層的索引表のそれぞれのレベルの前記それぞれのエントリ内の情報を保存することであって、前記情報が、前記基準ゲノム内のそれぞれの長さの前記それぞれの副配列の適合の数を特定し、前記情報が、前記基準ゲノムでの前記適合の各々の前記場所をさらに特定する、保存することと、により実行される、請求項8に記載の記憶媒体。

【請求項11】

$n + 1$ レベルに対して、前記 $n + 1$ レベルの前記それぞれの副配列の適合について前記検索することが、前記 n レベル内の前記対応するエントリに関連付けられた場所で実行される、請求項10に記載の方法。

【請求項12】

各エントリに関連付けられた前記適合の数を示す前記データが、第1のデータ構造内に保存され、各エントリに関連付けられた前記適合の各々の前記データ場所が、第2のデータ構造内に保存され、前記第1及び第2のデータ構造が、前記階層的索引表内に各々含まれる、請求項10に記載の方法。

【請求項13】

それぞれの n レベルのエントリに対して、前記それぞれの n レベルのエントリに対応する $n + 1$ レベルのエントリを参照するメモリ内にポインタを保存することをさらに含む、請求項8に記載の方法。

【請求項14】

検索配列を基準データに適合させるための方法であって、

演算装置により、

a) 基準データをメモリ内に保存することと、

b) 前記基準データに基づいて階層的索引表を作成することであって、

第1の長さを有する前記基準データのそれぞれの副配列の各網羅的なセットに対して、

i) 前記基準データ内の前記第1の長さの前記それぞれの副配列の適合について検索すること、ならびに

前記階層的索引表の第1のレベルにおける前記それぞれのエントリ内の情報を保存することであって、前記情報が、前記基準データ内の前記第1の長さの前記それぞれの副配列の適合の数を特定し、前記情報が、前記基準データ内での前記適合の各々の前記場所をさらに特定する、保存すること、により、前記階層的索引表の第1のレベル内のそれぞれのエントリを作成すること、

ii) 前記それぞれの第1のエントリ内の前記適合の数を第1の閾値と比較すること、ならびに

iii) 前記第1の閾値より大きい適合の数を有する前記第1のレベル内の各それぞれのエントリに対して、第2の長さを有する副配列の第2のセットに対する前記階層的索引表の第2のレベル内のそれぞれのエントリを前記作成することを実行することであって、前記第1のレベル内の各それぞれのエントリが、前記第2のレベル内のその対応するエントリに関連付けられ、副配列の前記第2のセットが、前記第2の長さを有する前記基準データの副配列の各網羅的なセットを含み、前記第2の長さを有する前記それぞれの副配列の適合について前記検索することが、前記第1のレベル内の前記対応するエントリに関連付けられた場所で実行される、実行すること、を含む、作成することと、

c) 検索配列を特定する入力を受信することと、

d) 前記基準データへの前記検索配列の小区分の1つまたは複数の適合について前記階

層的索引表を検索することであって、前記階層的索引表の前記検索は、前記階層的索引表のその後のレベルで前記検索配列の小区分の適合を反復的に検索することを含む、検索することと、を実行することを含む、検索配列を基準データに適合させるための方法。

【請求項 15】

前記基準データ及び各副配列が、数字的にコード化される、請求項14に記載の方法。

【請求項 16】

前記階層的索引表の第2のレベル内のそれぞれのエントリを前記比較すること及び前記作成することが、第2の反復を含み、前記方法が、

前記階層的索引表の1つ以上のそれぞれの追加の閾値、長さ、副配列のセット、及びレベルを使用して、1つ以上の追加の反復を実行することをさらに含む、請求項14に記載の方法。

【請求項 17】

前記基準データが、基準ゲノムを含み、前記基準データを検索することが、短い読み出し(SR)を前記基準ゲノムと整列させることを含む、請求項14に記載の方法。

【請求項 18】

各エントリに関連付けられた前記場所の数を示す前記データが、第1のデータ構造内に保存され、各エントリに関連付けられた前記基準データ内の前記場所を示す前記データが、第2のデータ構造内に保存され、前記第1及び第2のデータ構造が、前記階層的索引表内に各々含まれる、請求項14に記載の方法。

【請求項 19】

前記第2のレベル内の前記対応するエントリの前記第1のレベル内の前記それぞれのエントリとの前記関連付けが、前記第1のレベル内の前記それぞれのエントリ内に含まれる情報を含み、前記情報を、

前記第2のレベル内の前記対応するエントリを指摘するリンク情報と、前記第2の長さとを含む、請求項14に記載の方法。

【請求項 20】

前記第2のその後の反復の各々が、

前記それぞれの閾値より大きくない適合の数を有すると発見された各それぞれの副配列に対して、

前記それぞれの副配列に関連付けられた前記エントリ内のストップ命令を保存することであって、前記ストップ命令が、前記ストップ命令に関連付けられた前記副配列に基づいてさらなるエントリの前記作成を防止するために、前記方法により使用可能である、保存すること、をさらに含む、請求項16に記載の方法。

【請求項 21】

前記検索配列に関連付けられた前記索引表内の第1のエントリが、前記索引表の端末エントリではないとの判定に基づいて、前記検索配列の前記小区部の前記長さを増加することと、

前記第1のエントリが、前記索引表の端末エントリではないとの判定に基づいて、増加された長さの前記検索配列の前記小区分に適合する前記索引表内の第2のエントリを調べることと、をさらに含む、請求項14に記載の方法。

【請求項 22】

前記基準データを検索する間に発見された各適合を評価することであって、

各適合に対して、前記検索配列と前記基準データとの間の塩基誤差の特定に基づいて、検索配列と前記基準データとの間の不適合を判定すること、を含む、評価すること、をさらに含む、請求項14に記載の方法。

【請求項 23】

前記評価することが、

前記検索配列内の少なくとも1つのインデルを判定することであって、

前記検索配列のLo端及びHi端不適合の数に基づいて、前記検索配列内のアンカ位置を判定することと、

前記アンカ位置を使用して、前記少なくとも 1 つのインデルの長さ及び種類を判定することと、

前記少なくとも 1 つのインデルの開始位置を判定することであって、

前記適合場所で前記検索配列と前記基準データとの間の第 1 の移動誤差合計を演算すること、

前記適合場所からのオフセットで前記検索配列と前記基準データとの間の第 2 の移動誤差合計を演算することであって、前記オフセットが、前記少なくとも 1 つのインデルの前記種類及び長さに基づいている、演算すること、ならびに

前記第 1 及び第 2 の移動誤差合計の最小値に基づいて、前記少なくとも 1 つのインデルの開始場所を判定すること、を含む、判定することと、をさらに含む、請求項 2 2 に記載の方法。