

【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】平成28年11月10日 (2016.11.10)

【公開番号】特開2015-197899(P2015-197899A)

【公開日】平成27年11月9日 (2015.11.9)

【年通号数】公開・登録公報2015-069

【出願番号】特願2014-77278(P2014-77278)

【国際特許分類】

G 0 6 F 19/22 (2011.01)

【F I】

G 0 6 F 19/22

【手続補正書】

【提出日】平成28年9月27日 (2016.9.27)

【手続補正 1】

【補正対象書類名】明細書

【補正対象項目名】0 0 2 9

【補正方法】変更

【補正の内容】

【0 0 2 9】

図 3 の S 1 0 4 において、リード辞書作成部 2 1 は、BW 文字列 3 1 1 を Wavelet Tree 形式に変換することにより、検索を効率的に行うためのリード配列辞書 1 4 を作成する。

図 5 (a) は、Wavelet Tree 形式に変換するために参照される二分木 3 2 0 を示す。この二分木 3 2 0 では、文字列に使用される全文字 (\$、&、A、C、G、T、N) 3 2 1 が根となる。

二分木 3 2 0 は、文字列に使用される全文字 (\$、&、A、C、G、T、N) 3 2 1 を、再帰的に二分類し、分類の末端で高々 2 種類の文字しか含まないように分類する方法を示した二分木である。

二分木 3 2 0 の根では、全文字 3 2 1 を、A と、T (W) 3 2 4 と、それ以外 (S) 3 2 5 と、に分類する。S に分類された文字 (\$、&、C、G、N) 3 2 5 も、同様に M と K に二分類する。以下、これを再帰的に繰り返し、分類の末端で高々 2 種類の文字しか含まないようにする。但し、ペアを表す符号 & (3 0 2) と終端記号 \$ (3 0 3) は、分類の末端 3 3 1 に一緒に現れるように分類する。

【手続補正 2】

【補正対象書類名】明細書

【補正対象項目名】0 0 3 2

【補正方法】変更

【補正の内容】

【0 0 3 2】

図 3 の S 1 0 5 において、リード辞書作成部 2 1 は、Wavelet Tree 3 4 0 から作成されたリード配列辞書 1 4 を記憶手段に出力する。ここで、リード辞書作成部 2 1 は、Wavelet Tree 3 4 0 に加えて、その Wavelet Tree 3 4 0 の rank 関数や select 関数を計算するための、わずかな (BW 文字列 3 1 1 に対して 3.5 % 程度のデータ量の) 補助データを、リード配列辞書 1 4 に付加してもよい。

rank(p,c) とは、配列要素 0 ~ p のうちの文字「c」の出現回数を返す関数である。

select(i,c) とは、(i+1) 番目の文字「c」が出現する配列位置を返す関数である。

補助データは、例えば、参考文献「Kouichi Kimura, Yutaka Suzuki, Sumio Sugano, and Asako Koike. Journal of Computational Biology. November 2009, 16(11): 1601-1613.」に記載の「hierarchical binary string」である。この補助データは、BW 文字列

3 1 1 から、任意に与えられた塩基配列と一致するリード配列断片を全て求める、などの検索を効率的に行うためのデータである。