

【特許請求の範囲】

【請求項 1】

塩基配列情報に含まれる核酸構成塩基の偏在性を表す方法であって：
 塩基配列情報を、所定の分節化開始塩基から、所定の分節化塩基数で重複なく順次分割して分節群を作成する手順；
 前記各分節において、前記塩基配列情報を、所定の換算表に従って核酸構成塩基のみからなる情報に換算し、各核酸構成塩基の出現数を計数する手順；
 前記各分節において、前記核酸構成塩基の出現数を使用して、偏在性表示塩基を作成する手順；及び、
 前記各分節において作成された偏在性表示塩基を前記塩基配列情報の並びに従って配列化して、偏在性表示配列を作成する手順；
 を含む、前記方法。

10

【請求項 2】

前記偏在性表示塩基ごとに所定の色、パターン又はテクスチャを割り当て、偏在性表示視覚化情報を作成する手順をさらに含む、請求項1記載の方法。

【請求項 3】

前記偏在性表示視覚化情報が、当該塩基配列情報の特性プロファイルに基づいて、直鎖状又は環状で出力される、請求項2記載の方法。

【請求項 4】

前記所定の換算表が、下記の換算表である、請求項1～3のいずれか1項記載の方法：

20

【表 1】

塩基	換算される核酸構成塩基
A	A
C	C
G	G
T	T
U	T
M	1/2 A、1/2 C
R	1/2 A、1/2 G
W	1/2 A、1/2 T
S	1/2 C、1/2 G
Y	1/2 C、1/2 T
K	1/2 G、1/2 T
V	1/3 A、1/3 C、1/3 G
H	1/3 A、1/3 C、1/3 T
D	1/3 A、1/3 G、1/3 T
B	1/3 C、1/3 G、1/3 T
N	1/4 A、1/4 C、1/4 G、1/4 T

30

40

。

【請求項 5】

前記偏在性表示塩基の作成手順が、下記工程を含む、請求項1～4のいずれか1項記載の

50

方法：

当該分節において、前記核酸構成塩基出現数を使用して、各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する工程

【数 1】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し； p はGC含量(%)を示し；及び、 L は当該分節内の塩基の総数；を示す)；

10

当該分節において、前記各核酸構成塩基の χ^2 値を、下記式に従って算出する工程

【数 2】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2$$

(式中、 B は任意の1つの核酸構成塩基を示し、 BR は前記 B 以外の全ての核酸構成塩基群を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及び L は当該分節内の塩基の総数を示す)；及び、

20

(i) 前記 χ_B^2 群の全てが、条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たさない場合には、当該分節に、偏在性表示塩基として「N」を割り当て、
(ii) 前記 χ_B^2 群の少なくとも1つの χ_B^2 (群)が条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の全てが条件(B)：「当該出現数が、その期待数よりも多い」を満たさない場合、前記条件(A)を満たしかつ前記条件(B)を満たさない核酸構成塩基(群)以外の全てからなる核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当て、及び、

30

(iii) 前記 χ_B^2 群のうち少なくとも1つの χ_B^2 (群)が条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の少なくとも1つが条件(B)：「当該出現数が、その期待数よりも多い」を満たす場合、前記条件(A)を満たしかつ前記条件(B)を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる、前記工程。

【請求項 6】

前記偏在性表示塩基の作成手順が、下記工程を含む、請求項1～4のいずれか1項記載の方法：

当該分節において、前記核酸構成塩基出現数を使用して、各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する工程

40

【数 3】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し； p はGC含量(%)を示し；及び、 L は当該分節内の塩基の総数；を示す)；
当該分節において、前記各核酸構成塩基の χ^2 値を、下記式に従って算出する工程

【数 4】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2$$

(式中、 B は任意の1つの核酸構成塩基を示し、 BR は前記 B 以外の全ての核酸構成塩基群を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及び L は当該分節内の塩基の総数を示す) 10

；及び、
(i) 前記 χ_B^2 群の全てが、条件(A):「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たさない場合には、当該分節に偏在性表示塩基として「N」を割り当て、

(ii) 前記 χ_B^2 群の少なくとも1つの χ_B^2 (群)が条件(A):「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の全てが条件(B):「当該出現数が、その期待数よりも多い」を満たさない場合には、当該分節に偏在性表示塩基として「N」を割り当て、及び、

(iii) 前記 χ_B^2 群のうち少なくとも1つの χ_B^2 (群)が条件(A):「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の少なくとも1つが条件(B):「当該出現数が、その期待数よりも多い」を満たす場合、前記条件(A)を満たしかつ前記条件(B)を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる、前記工程。 20

【請求項 7】

前記偏在性表示塩基の作成手順が、下記工程を含む、請求項1～4のいずれか1項記載の方法：

当該分節において、核酸構成塩基の出現数をそれぞれ計数して、第1、第2、第3及び第4頻出塩基を決定する工程；

前記頻出塩基の出現数を使用して、所定の有意水準において有意に出現頻度が高い塩基(群)を判定するための検定を実行する工程；及び、

前記検定により有意に出現頻度が高いと判定された塩基(群)が存在する場合には該当する塩基(群)に対応する偏在性表示塩基を所定の対応表に従って当該分節に割り当て、全ての核酸構成塩基の出現数に有意差はないと判断された場合には、当該分節に偏在性表示塩基として「N」を割り当てる工程。 30

【請求項 8】

前記検定が、統計学的検定に基づく検定である、請求項7記載の方法。

【請求項 9】

前記検定が、下記工程を含む仮説検定である、請求項7記載の方法：

第1仮説を「当該分節において、第1頻出塩基の出現数はその期待値に等しい」とする第1仮説検定を実行する工程；

前記第1仮説検定で前記第1仮説が棄却される場合には前記第1頻出塩基が当該分節において有意に出現頻度が高い塩基であると判定し、前記第1仮説検定で前記第1仮説が棄却されない場合には第2仮説を「当該分節において、第1及び第2頻出塩基の出現数の合計はその期待値に等しい」とする第2仮説検定を実行する工程； 40

前記第2仮説検定で前記第2仮説が棄却される場合には前記第1及び第2頻出塩基からなる塩基(群)が当該分節において有意に出現頻度が高い塩基(群)であると判定し、前記第2仮説検定で前記第2仮説が棄却されない場合には第3仮説を「当該分節において、第1、第2及び第3頻出塩基の出現数の合計はその期待値に等しい」とする第3仮説検定を実行する工程；及び、

前記第3仮説検定で前記第3仮説が棄却される場合には前記第1、第2及び第3頻出塩基からなる塩基(群)が当該分節において有意に出現頻度が高い塩基(群)であると判定し、前 50

記第3仮説検定で前記第3仮説が棄却されない場合には全ての核酸構成塩基の出現数に有意差はないと判定する工程。

【請求項10】

前記仮説検定が、正規分布表、t分布表、F分布表、又はカイ二乗分布表からなる群から選択される分布表を使用する、請求項9記載の方法。

【請求項11】

前記対応表が、下記の対応表である、請求項5～10のいずれか1項記載の方法：

【表2】

核酸構成塩基（群）	偏在性表示塩基	
A	A	10
C	C	
G	G	
T	T	
A及びC	M	20
A及びG	R	
A及びT	W	
C及びG	S	
C及びT	Y	
G及びT	K	
A及びC及びG	V	
A及びC及びT	H	30
A及びG及びT	D	
C及びG及びT	B	

。

【請求項12】

請求項1記載の方法に従って同一の分節化塩基数を使用して作成された複数の偏在性表示配列間の類似性を比較解析する手順を含む、塩基配列情報の比較解析方法。

【請求項13】

前記複数の偏在性表示配列が、所定のデータベースに格納されている、請求項12記載の方法。

【請求項14】

請求項2記載の方法に従って同一の分節化塩基数を使用して作成された複数の偏在性表示視覚化情報間の類似性を、所定のクラスターと関連付けて表示する手順を含む、塩基配列情報の比較解析方法。

【請求項15】

前記複数の偏在性表示視覚化情報が、所定のデータベースに格納されている、請求項14記載の方法。

【請求項16】

塩基配列情報に含まれる核酸構成塩基の偏在性を表す方法をコンピュータに実行させるためのプログラムであって、前記方法が：

入力された塩基配列情報を、所定の分節化開始塩基から、所定の分節化塩基数で重複なく順次分割して分節群を作成する手順；

前記各分節において、前記塩基配列情報を、所定の換算表に従って核酸構成塩基のみから

なる情報に換算し、各核酸構成塩基の出現数を計数する手順；

前記各分節において、前記核酸構成塩基の出現数を使用して、偏在性表示塩基を作成する手順；

前記各分節において作成された偏在性表示塩基を前記塩基配列情報の並びに従って配列化して、偏在性表示配列を作成する手順；及び、

前記偏在性表示配列を出力する手順；

を含む、前記プログラム。

【請求項 17】

前記偏在性表示塩基ごとに所定の色、パターン又はテクスチャを割り当て、偏在性表示視覚化情報を作成する手順をさらに含む、請求項16記載の方法。

10

【請求項 18】

前記偏在性表示視覚化情報が、当該塩基配列情報の特性プロファイルに基づいて、直鎖状又は環状で出力される、請求項17記載のプログラム。

【請求項 19】

前記所定の換算表が、下記の換算表である、請求項16～18のいずれか1項記載のプログラム；

【表 3】

塩基	換算される核酸構成塩基
A	A
C	C
G	G
T	T
U	T
M	1/2 A、1/2 C
R	1/2 A、1/2 G
W	1/2 A、1/2 T
S	1/2 C、1/2 G
Y	1/2 C、1/2 T
K	1/2 G、1/2 T
V	1/3 A、1/3 C、1/3 G
H	1/3 A、1/3 C、1/3 T
D	1/3 A、1/3 G、1/3 T
B	1/3 C、1/3 G、1/3 T
N	1/4 A、1/4 C、1/4 G、1/4 T

20

30

40

。

【請求項 20】

前記偏在性表示塩基の作成手順が、下記工程を含む、請求項16～19のいずれか1項記載のプログラム；

当該分節において、前記核酸構成塩基出現数を使用して、各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する工程

【数 5】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し； p はGC含量(%)を示し；及び、 L は当該分節内の塩基の総数；を示す)；前記各核酸構成塩基の χ^2 値を、下記式に従って算出する工程

【数 6】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2$$

10

(式中、 B は任意の1つの核酸構成塩基を示し、 BR は前記 B 以外の全ての核酸構成塩基群を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及び L は当該分節内の塩基の総数を示す)；及び、

(i) 前記 χ_B^2 群の全てが、条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たさない場合には、当該分節に、偏在性表示塩基として「N」を割り当て、
(ii) 前記 χ_B^2 群の少なくとも1つの χ_B^2 (群)が条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の全てが条件(B)：「当該出現数が、その期待数よりも多い」を満たさない場合、前記条件(A)を満たしかつ前記条件(B)を満たさない核酸構成塩基(群)以外の全てからなる核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当て、及び、

20

(iii) 前記 χ_B^2 群のうち少なくとも1つの χ_B^2 (群)が条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の少なくとも1つが条件(B)：「当該出現数が、その期待数よりも多い」を満たす場合、前記条件(A)を満たしかつ前記条件(B)を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる、前記工程。

30

【請求項 21】

前記偏在性表示塩基の作成手順が、下記工程を含む、請求項16～19のいずれか1項記載のプログラム：

当該分節において、前記核酸構成塩基出現数を使用して、各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する工程

【数 7】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

40

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し； p はGC含量(%)を示し；及び、 L は当該分節内の塩基の総数；を示す)；前記各核酸構成塩基の χ^2 値を、下記式に従って算出する工程

【数 8】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2$$

50

(式中、 B は任意の1つの核酸構成塩基を示し、 B_R は前記 B 以外の全ての核酸構成塩基を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{B_R} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{B_R} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及び L は当該分節内の塩基の総数を示す)；及び、

(i) 前記 χ^2_B 群の全てが、条件(A):「当該 χ^2_B 値が、所定の有意水準のカイ二乗値より大きい」を満たさない場合には、当該分節に偏在性表示塩基として「N」を割り当て、

(ii) 前記 χ^2_B 群の少なくとも1つの χ^2_B (群)が条件(A):「当該 χ^2_B 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の全てが条件(B):「当該出現数が、その期待数よりも多い」を満たさない場合には、当該分節に偏在性表示塩基として「N」を割り当て、及び、

(iii) 前記 χ^2_B 群のうち少なくとも1つの χ^2_B (群)が条件(A):「当該 χ^2_B 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の少なくとも1つが条件(B):「当該出現数が、その期待数よりも多い」を満たす場合、前記条件(A)を満たしかつ前記条件(B)を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる、前記工程。

【請求項22】

前記偏在性表示塩基の作成手順が、下記工程を含む、請求項16～19のいずれか1項記載のプログラム；

当該分節において、核酸構成塩基の出現数をそれぞれ計数して、第1、第2、第3及び第4頻出塩基を決定する工程；

前記頻出塩基の出現数を使用して、所定の有意水準において有意に出現頻度が高い核酸構成塩基(群)を判定するための検定を実行する工程；及び、

前記検定により有意に出現頻度が高いと判定された核酸構成塩基(群)が存在する場合には該当する核酸構成塩基(群)に対応する偏在性表示塩基を所定の対応表に従って当該分節に割り当て、全ての核酸構成塩基の出現数に有意差はないと判断された場合には、当該分節に偏在性表示塩基として「N」を割り当てる工程。

【請求項23】

前記検定が、統計学的検定に基づく検定である、請求項22記載のプログラム。

【請求項24】

前記検定が、下記工程を含む仮説検定である、請求項22記載のプログラム；

第1仮説を「当該分節において、第1頻出塩基の出現数はその期待値に等しい」とする第1仮説検定を実行する工程；

前記第1仮説検定で前記第1仮説が棄却される場合には前記第1頻出塩基が当該分節において有意に出現頻度が高い塩基であると判定し、前記第1仮説検定で前記第1仮説が棄却されない場合には第2仮説を「当該分節において、第1及び第2頻出塩基の出現数の合計はその期待値に等しい」とする第2仮説検定を実行する工程；

前記第2仮説検定で前記第2仮説が棄却される場合には前記第1及び第2頻出塩基からなる塩基(群)が当該分節において有意に出現頻度が高い塩基(群)であると判定し、前記第2仮説検定で前記第2仮説が棄却されない場合には第3仮説を「当該分節において、第1、第2及び第3頻出塩基の出現数の合計はその期待値に等しい」とする第3仮説検定を実行する工程；及び、

前記第3仮説検定で前記第3仮説が棄却される場合には前記第1、第2及び第3頻出塩基からなる塩基(群)が当該分節において有意に出現頻度が高い塩基(群)であると判定し、前記第3仮説検定で前記第3仮説が棄却されない場合には全ての核酸構成塩基の出現数に有意差はないと判定する工程。

【請求項25】

前記仮説検定が、正規分布表、t分布表、F分布表、又はカイ二乗分布表からなる群から選択される分布表を使用する、請求項24記載のプログラム。

【請求項 26】

前記対応表が、下記の対応表である、請求項16～25のいずれか1項記載のプログラム：

【表 4】

核酸構成塩基（群）	偏在性表示塩基
A	A
C	C
G	G
T	T
A及びC	M
A及びG	R
A及びT	W
C及びG	S
C及びT	Y
G及びT	K
A及びC及びG	V
A及びC及びT	H
A及びG及びT	D
C及びG及びT	B

10

20

。

【請求項 27】

請求項16記載の方法に従って同一の分節化塩基数を使用して作成された複数の偏在性表示配列間の類似性を比較解析する手順を含む、塩基配列情報の比較解析用プログラム。

【請求項 28】

前記複数の偏在性表示配列が、所定のデータベースに格納されている、請求項27記載のプログラム。

30

【請求項 29】

請求項17記載の方法に従って同一の分節化塩基数を使用して作成された複数の偏在性表示視覚化情報間の類似性を、所定のクラスターと関連付けて表示する手順を含む、塩基配列情報の比較解析用プログラム。

【請求項 30】

前記偏在性視覚化情報が、所定のデータベースに格納されている、請求項29記載のプログラム。

【請求項 31】

請求項16～30のいずれか1項記載のプログラムを記録したコンピュータ読み取り可能な記録媒体。

40

【請求項 32】

塩基配列情報に含まれる核酸構成塩基の偏在性を表すための装置であって：

塩基配列情報の入力手段；

前記塩基配列情報を、所定の分節化開始塩基から、所定の分節化塩基数で重複なく順次分割して分節群を作成する手段；

前記各分節において、前記塩基配列情報を、所定の換算表に従って核酸構成塩基のみからなる情報に換算し、各核酸構成塩基の出現数を計数する手段；

前記各分節において、前記核酸構成塩基の出現数を使用して、偏在性表示塩基を作成する

50

手段；

前記各分節において作成された偏在性表示塩基を前記塩基配列情報の並びに従って配列化して、偏在性表示配列を作成する手段；及び、

前記偏在性表示配列の出力手段；

を含む、前記装置。

【請求項 33】

前記偏在性表示塩基ごとに所定の色、パターン又はテクスチャを割り当て、偏在性表示視覚化情報を作成する手段をさらに含む、請求項32記載の装置。

【請求項 34】

前記偏在性表示視覚化情報が、当該塩基配列情報の特性プロファイルに基づいて、直鎖状又は環状で出力される、請求項33記載の装置。

10

【請求項 35】

前記所定の換算表が、下記の換算表である、請求項32～34のいずれか1項記載の装置：

【表 5】

塩基	換算される核酸構成塩基
A	A
C	C
G	G
T	T
U	T
M	1/2 A、1/2 C
R	1/2 A、1/2 G
W	1/2 A、1/2 T
S	1/2 C、1/2 G
Y	1/2 C、1/2 T
K	1/2 G、1/2 T
V	1/3 A、1/3 C、1/3 G
H	1/3 A、1/3 C、1/3 T
D	1/3 A、1/3 G、1/3 T
B	1/3 C、1/3 G、1/3 T
N	1/4 A、1/4 C、1/4 G、1/4 T

20

30

40

。

【請求項 36】

前記偏在性表示塩基の作成手段が、下記手段を含む、請求項32～35のいずれか1項記載の装置：

当該分節において、前記核酸構成塩基出現数を使用して、各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する手段

【数 9】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し； p はGC含量(%)を示し；及び、 L は当該分節内の塩基の総数；を示す)；前記各核酸構成塩基の χ^2 値を、下記式に従って算出する手段

【数 10】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2$$

10

(式中、 B は任意の1つの核酸構成塩基を示し、 BR は前記 B 以外の全ての核酸構成塩基群を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及び L は当該分節内の塩基の総数を示す)；及び、

(i) 前記 χ_B^2 群の全てが、条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たさない場合には、当該分節に、偏在性表示塩基として「N」を割り当て、
(ii) 前記 χ_B^2 群の少なくとも1つの χ_B^2 (群)が条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の全てが条件(B)：「当該出現数が、その期待数よりも多い」を満たさない場合、前記条件(A)を満たしかつ前記条件(B)を満たさない核酸構成塩基(群)以外の全てからなる核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当て、及び、

20

(iii) 前記 χ_B^2 群のうち少なくとも1つの χ_B^2 (群)が条件(A)：「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の少なくとも1つが条件(B)：「当該出現数が、その期待数よりも多い」を満たす場合、前記条件(A)を満たしかつ前記条件(B)を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる、前記手段。

30

【請求項 37】

前記偏在性表示塩基の作成手段が、下記手段を含む、請求項32～35のいずれか1項記載の装置：

当該分節において、前記核酸構成塩基出現数を使用して、各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する手段

【数 11】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

40

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し； p はGC含量(%)を示し；及び、 L は当該分節内の塩基の総数；を示す)；前記各核酸構成塩基の χ^2 値を、下記式に従って算出する手段

【数 12】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2$$

50

(式中、 B は任意の1つの核酸構成塩基を示し、 B_R は前記 B 以外の全ての核酸構成塩基を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{B_R} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{B_R} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及び L は当該分節内の塩基の総数を示す)；及び、

(i) 前記 χ^2_B 群の全てが、条件(A):「当該 χ^2_B 値が、所定の有意水準のカイ二乗値より大きい」を満たさない場合には、当該分節に偏在性表示塩基として「N」を割り当て、

(ii) 前記 χ^2_B 群の少なくとも1つの χ^2_B (群)が条件(A):「当該 χ^2_B 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の全てが条件(B):「当該出現数が、その期待数よりも多い」を満たさない場合には、当該分節に偏在性表示塩基として「N」を割り当て、及び、

(iii) 前記 χ^2_B 群のうち少なくとも1つの χ^2_B (群)が条件(A):「当該 χ^2_B 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の少なくとも1つが条件(B):「当該出現数が、その期待数よりも多い」を満たす場合、前記条件(A)を満たしかつ前記条件(B)を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる、前記手段。

【請求項38】

前記P値の入力手段をさらに含む、請求項32～37のいずれか1項記載の装置。

【請求項39】

前記偏在性表示塩基の作成手段が、下記手段を含む、請求項32～35のいずれか1項記載の装置：

当該分節において、核酸構成塩基の出現数をそれぞれ計数して、第1、第2、第3及び第4頻出塩基を決定する手段；

前記頻出塩基の出現数を使用して、所定の有意水準において有意に出現頻度が高い核酸構成塩基(群)を判定するための検定を実行する手段；及び、

前記検定により有意に出現頻度が高いと判定された核酸構成塩基(群)が存在する場合には該当する核酸構成塩基(群)に対応する偏在性表示塩基を所定の対応表に従って当該分節に割り当て、全ての核酸構成塩基の出現数に有意差はないと判断された場合には、当該分節に偏在性表示塩基として「N」を割り当てる手段。

【請求項40】

前記検定が、統計学的検定に基づく検定である、請求項39記載の装置。

【請求項41】

前記検定が、下記手段を含む仮説検定である、請求項39記載の装置：

第1仮説を「当該分節において、第1頻出塩基の出現数はその期待値に等しい」とする第1仮説検定を実行する手段；

前記第1仮説検定で前記第1仮説が棄却される場合には前記第1頻出塩基が当該分節において有意に出現頻度が高い塩基であると判定し、前記第1仮説検定で前記第1仮説が棄却されない場合には第2仮説を「当該分節において、第1及び第2頻出塩基の出現数の合計はその期待値に等しい」とする第2仮説検定を実行する手段；

前記第2仮説検定で前記第2仮説が棄却される場合には前記第1及び第2頻出塩基からなる塩基(群)が当該分節において有意に出現頻度が高い塩基(群)であると判定し、前記第2仮説検定で前記第2仮説が棄却されない場合には第3仮説を「当該分節において、第1、第2及び第3頻出塩基の出現数の合計はその期待値に等しい」とする第3仮説検定を実行する手段；及び、

前記第3仮説検定で前記第3仮説が棄却される場合には前記第1、第2及び第3頻出塩基からなる塩基(群)が当該分節において有意に出現頻度が高い塩基(群)であると判定し、前記第3仮説検定で前記第3仮説が棄却されない場合には全ての核酸構成塩基の出現数に有意差はないと判定する手段。

【請求項42】

前記仮説検定が、正規分布表、t分布表、F分布表、又はカイ二乗分布表からなる群から選択される分布表を使用する、請求項41記載の装置。

【請求項 4 3】

前記対応表が、下記の対応表である、請求項32～42のいずれか1項記載の装置：

【表 6】

核酸構成塩基（群）	偏在性表示塩基
A	A
C	C
G	G
T	T
A及びC	M
A及びG	R
A及びT	W
C及びG	S
C及びT	Y
G及びT	K
A及びC及びG	V
A及びC及びT	H
A及びG及びT	D
C及びG及びT	B

。

【請求項 4 4】

(i) 塩基配列情報を取得又は出力するための、及び(ii) 本発明の偏在性表示塩基、偏在性表示配列若しくは偏在性表示視覚化情報を取得又は出力するための、インターネット通信手段をさらに含む、請求項32～43のいずれか1項記載の装置。

【請求項 4 5】

前記分節化開始塩基の入力手段をさらに含む、請求項32～44のいずれか1項記載の装置

。

【請求項 4 6】

前記分節化塩基数の入力手段をさらに含む、請求項32～45のいずれか1項記載の装置。

【請求項 4 7】

前記有意水準の入力手段をさらに含む、請求項36～46のいずれか1項記載の装置。

【請求項 4 8】

請求項32記載の手段を使用して同一の分節化塩基数で作成された複数の偏在性表示配列間の類似性を比較解析する手段を含む、塩基配列情報の比較解析用装置。

【請求項 4 9】

前記複数の偏在性表示配列が、所定のデータベースに格納されている、請求項48記載の装置。

【請求項 5 0】

請求項33記載の手段を使用し、同一の分節化塩基数を使用して作成された複数の偏在性表示視覚化情報間の類似性を、所定のクラスターと関連付けて表示する手段を含む、塩基配列情報の比較解析用装置。

【請求項 5 1】

10

20

30

40

50

前記偏在性視覚化情報が、所定のデータベースに格納されている、請求項50記載の装置

。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、塩基配列情報に含まれる核酸構成塩基の偏在性を表す偏在性表示配列及び偏在性表示視覚化情報を作成する方法に関する。また、偏在性表示配列を使用する比較解析及び偏在性表示視覚化情報を使用するクラスタリング解析にも関する。さらに、前記方法をコンピュータに実行させるためのプログラム、前記プログラムを記録したコンピュータ読み取り可能な記録媒体、並びに該方法を実質的に実行可能な装置にも関する。

10

【背景技術】

【0002】

ゲノムが保持している遺伝情報は、当該ゲノム核酸に含まれる塩基A（アデニン）、G（グアニン）、C（シトシン）、及びT（チミン）又はU（ウラシル）（DNAの場合にはT、RNAの場合にはU）の並びを示す塩基配列として表される。ゲノム核酸塩基配列は生物種によって異なるが、近縁の種ほど互いの配列が似ており、かつ遠縁の種ほど互いの配列が異なることは当業者に周知である。ゆえに、未知の生物由来の塩基配列情報に基づいてその分類をする場合には、当該生物の塩基配列を既知の他の生物の塩基配列と比較解析することが一般的に実施されている。

20

【0003】

代表的な生物学的分類法には、生物の形態的差異、生理生化学的性状の差異などの他に、塩基配列を利用する以下の方法が含まれる：（a）GC含量を使用する分類法、（b）DNA-DNAハイブリダイゼーションを使用する分類法、及び（c）rRNAなど特定の遺伝子をコードするDNAの相同性を使用する分類法（非特許文献1を参照されたい）。（a）GC含量を使用する分類法に関して、GC含量は、当該塩基配列に含まれるグアニン及びシトシンの合計のモル百分率として決定され、かつ細菌分類の標準的記載の一部とみなされている。一般的に、よく定義された（well-defined）種の群におけるGC含量のばらつきは3%以内であり、よく定義された属の群におけるGC含量のばらつきは10%以内であることが知られている。（b）DNA-DNAハイブリダイゼーションは、2種の全ゲノム間の配列類似性の間接的パラメータを得るための方法である。当該方法によると、ゲノム間での塩基対ミスマッチの1%につき、当該ゲノム間の熱安定性が1~2.2%減少することが確立されている。そのため、当該方法により得られた融解温度は、系統解析用データとして使用されている（例えば、非特許文献2を参照されたい）。（c）rRNAをコードするDNAの相同性を使用する分類法は、主に16S又は23S rRNAをコードするDNA配列同士を比較することにより、複数の生物種間の近縁関係を解析する方法である。

30

【0004】

しかしながら、上記方法には、それぞれ以下のような問題点がある：（a）GC含量は細菌分類における標準的記載の一部ではあるが、当該細菌が含まれている種又は属に含まれる細菌群のGC含量と比較することによって結果的に見出される特徴であって、GC含量自体は当該細菌を特徴づけるための指標ではないこと；（b）DNA-DNAハイブリダイゼーションは、in vitroで実施される方法であるため条件検討が難しく、かつ再現性があまりよくないこと（例えば、非特許文献1を参照されたい）；及び、（c）rRNAが系統分類学的解析における最良の標的であることは当業者に周知であるが（例えば、非特許文献1を参照されたい）、rRNA遺伝子に導入された変異はゲノム全体の変異の割合を必ずしも反映していないこと（例えば、非特許文献3を参照されたい）。

40

【0005】

一方、生物学的分類法においてゲノム核酸塩基配列の全体同士の比較解析結果を使用することが理想的な分類法を提供し得ることは、当業者に理解されている。しかしながら：（i）ゲノムを構成する塩基配列数はしばしば長大であること（例えば、ヒトのゲノムは約30億塩基対から構成され、原核生物である大腸菌のゲノムでさえも約500万塩基対から

50

構成されている（例えばNCBIのホームページ<<http://www.ncbi.nlm.nih.gov/>>を参照されたい））；及び、（ii）in silicoでの塩基配列情報の比較解析において、当業者が一般的に使用するFASTAやBLASTなどの比較解析ソフトウェアは、ダイナミックプログラミングアルゴリズムに基づいているために、入力ゲノム配列長の倍数に対して指数関数的に計算が複雑化すること（例えば、非特許文献4を参照されたい）；から、ゲノム核酸塩基配列の全体をin silicoでの比較解析に使用することは、実質的に困難であった。

【非特許文献1】P. VANDAMME, B. POT, M. GILLIS, P. DE VOS, K. KERSTERS, 及びJ. S WINGSの論文、MICROBIOLOGICAL REVIEWS, Vol. 60, No. 2, 1996, 407-438頁

【非特許文献2】ゲノム 第3版—新しい生命情報システムへのアプローチ（第3版） T.A . ブラウン(著)、村松正實（監訳）、木南 凌（監訳）、メディカルサイエンスインターナショナル、2007年

【非特許文献3】Oshima K及びNishida H.の論文、J Mol Evol. 2007 Sep;65(3):249-58

【非特許文献4】ゲノム配列から機能解析へ 第2版、マウント デービッド W. (著)、岡崎 康司（監訳）、坊農 秀雅（監訳）、メディカル・サイエンス・インターナショナル、2005年

【発明の開示】

【発明が解決しようとする課題】

【0006】

先に記載した公知技術は全て、一塩基レベルでの塩基配列情報に基づく方法を開示している。

一方で、ゲノムには、塩基配列は異なるものの同一の機能を果たす配列が存在することは周知であり、遺伝子コード領域における例としては、プロモータ、イニシエーター、TAボックス及びエンハンサーなどの転写調節エレメント；並びに、1種類のアミノ酸を複数のコドンが指定する縮重コドン；などの存在が挙げられる。また、ゲノム内の遺伝子非コード領域に導入された変異は、個体レベルでの表現型の変化には何ら寄与しないと考えられている。それゆえ、既存の解析技術を使用して、未知のゲノム塩基配列全体を比較解析に供することは、機能又は表現型に関連するか否かにかかわらず同等に解析されるため、解析に要する時間に比して効果が低かった。

【0007】

しかしながら、先に記載したように、近縁の種ほど互いの配列が似ており、かつ遠縁の種ほど互いの配列が異なることは当業者に周知であることから、上記に記載したような同機能異配列又は同表現型異配列は、近縁であるほど互いに大部分の配列が一致し、かつ少数の配列のみが変異した配列となることもまた周知であるといえる。それゆえ、任意の塩基配列情報の一定の領域内に存在する核酸構成塩基の出現数のプロファイルは、前記同機能異配列又は同表現型異配列に関する問題を解決するために有用な情報となり得る。さらに、複数の塩基配列情報から前記プロファイルを作成してそれらを比較解析することにより、異なる配列であるにもかかわらず同じ機能又は同じ表現型に寄与する領域をスクリーニングすることが可能になる。

【0008】

そこで本発明は、前記一定領域内に存在する核酸構成塩基の出現数のプロファイルから、一塩基レベルでは異なる配列であっても実質的に同じ機能又は同じ表現型を有する塩基配列情報を簡便にスクリーニングすることができる方法及び手段を提供することを課題とする。

また、任意の塩基配列情報を生物学的に同定又は分類することを目的として、当該塩基配列情報の特徴を表すための方法及び手段を提供することを課題とする。

【課題を解決するための手段】

【0009】

上記課題を解決するための手段として、本発明は、ゲノム核酸塩基配列情報などの長大な塩基配列情報に含まれる、核酸構成塩基の偏在性を表す偏在性表示配列及び偏在性表示視覚化情報を作成する方法、該方法をコンピュータに実行させるためのプログラム、前記

プログラムを記録したコンピュータ読み取り可能な記録媒体、並びに該方法を実質的に実行可能な装置を提供する。また本発明は、前記偏在性表示配列及び偏在性表示視覚化情報の比較解析方法、該方法をコンピュータに実行させるためのプログラム、前記プログラムを記録したコンピュータ読み取り可能な記録媒体、並びに該方法を実質的に実行可能な装置も提供する。

【0010】

詳細に述べると、本発明は、塩基配列情報に含まれる核酸構成塩基の偏在性を表す偏在性表示配列を作成する方法であって：塩基配列情報を、所定の分節化開始塩基から、所定の分節化塩基数で重複なく順次分割して分節群を作成する手順；前記各分節において、前記塩基配列情報を、所定の換算表に従って核酸構成塩基のみからなる情報に換算し、各核酸構成塩基の出現数を計数する手順；前記各分節において、前記核酸構成塩基の出現数を使用して、偏在性表示塩基を作成する手順；及び、前記各分節において作成された偏在性表示塩基を前記塩基配列情報の並びに従って配列化して、偏在性表示配列を作成する手順；を含む、前記方法を提供する。さらに、本発明は、前記偏在性表示塩基ごとに所定の色、パターン又はテクスチャを割り当て、偏在性表示視覚化情報を作成する手順を含む方法も提供する。

10

【発明の効果】

【0011】

本発明の偏在性表示塩基は、当該偏在性表示塩基に対応する分節を作成した元の塩基配列領域内に含まれる核酸構成塩基、又は混合塩基から所定の割合で換算された核酸構成塩基の存在比率に基づいて決定される。特に、当該偏在性表示塩基に対応する分節を作成した元の塩基配列領域内において、本発明の基準に基づき有意に存在比率の高い核酸構成塩基（群）を示すことができる。それゆえ、偏在性表示塩基は、当該分節内の核酸構成塩基の存在比率に基づく特徴を反映する記号であり得る。また、偏在性表示配列は、当該ゲノムの全体にわたって、核酸構成塩基の存在比率が特徴的に偏っている（特に、当該核酸構成塩基の存在比率が有意に多い）領域を概略的に把握するのに有用であり得る。すなわち、本発明は、当該塩基配列に含まれる生物学的に有用な情報を保持したまま、前記塩基配列にかかる情報量を低減することができる。

20

本発明の偏在性表示塩基及び偏在性表示配列は、当業者に周知の塩基記号及び混合塩基記号の記載様式に対応することができる。それゆえ、本発明の偏在性表示塩基及び偏在性表示配列を使用する解析には、既存の塩基配列解析ソフトウェアを使用することができる。

30

【0012】

本発明により、任意の塩基配列情報内の一定の領域内に存在する核酸構成塩基の出現数のプロファイルから、前記塩基配列情報に含まれる核酸構成塩基の偏在性を表す偏在性表示配列及び偏在性表示視覚化情報を作成することができる。当該偏在性表示配列は、元の塩基配列情報に含まれる核酸構成塩基の偏在性を示すために、核酸構成塩基に含まれる特徴領域をきわめて容易に知ることができる。例えば、元の塩基配列情報におけるGCリッチ領域（一般的には、遺伝子が多く存在する領域として知られる）を知りたい場合には偏在性表示配列中の偏在性表示塩基Sを探せばよく、元の塩基配列情報におけるATリッチ領域（一般的には、非遺伝子領域又はセントロメア領域などとして知られる）を知りたい場合には偏在性表示配列中の偏在性表示塩基Wを探せばよい。これら偏在性表示塩基に対応する元の塩基配列情報が、前記所望の特徴を有する配列領域である。

40

【0013】

また、ある程度の長さを有する遺伝子領域又は非遺伝子領域、若しくはトランスポゾンなどの可動性因子領域、あるいはその他の任意の領域に含まれる塩基配列から得られる偏在性表示配列が特有な偏在性表示配列を示す場合、当該特有の偏在性表示配列を解析対象から得られた偏在性表示配列において検索することにより、一塩基レベルでの比較解析よりもきわめて容易に前記領域を検索することが可能となる。また、本発明の偏在性表示視覚化情報は、前記特有の偏在性表示配列のパターンを視覚的に認識するための手段を提供

50

するので、特別な解析ソフトなどを使用せずにきわめて容易に前記領域を検索することが可能である。また、解析対象の偏在性表示配列を、分節化開始塩基及び分節化塩基数などのパラメータを変更した一連の参照偏在性表示配列（群）との比較解析に供することにより、前記領域の検索精度を高めることができる。

【 0 0 1 4 】

また、偏在性表示配列は、既存の混合塩基の表記方法に対応しているために、当該配列を既存の比較解析ソフトを使用する比較解析方法又はアルゴリズム若しくは比較解析プログラムに供することができる。これにより、偏在性表示配列レベルでの類似性をin silicoで比較解析することができ、それにより好ましくは塩基配列情報の詳細な比較解析を実施する前の一次スクリーニングなどの用途に使用することができる。当該方法は特に、一塩基レベルでは異なる配列であっても実質的に単一の同機能又は同表現型、若しくは関連遺伝子群が連続的に配置された領域を有する塩基配列情報をスクリーニングすることに適する。

10

【 0 0 1 5 】

また、本発明により作成された偏在性表示視覚化情報を使用することにより、元の塩基配列情報に含まれる核酸構成塩基の偏在性をパターン表示することが可能となる。これにより、本発明に供した塩基配列情報全体にわたる配列的特徴をきわめて容易に把握することができる。また複数の偏在性表示視覚化情報を、好ましくは既存の生物学的分類ごとに並べて表示することにより、塩基配列情報全体にわたる類似性を比較することができる。これにより、好ましくは塩基配列情報の詳細な比較解析を実施する前の一次スクリーニングなどの用途に使用することができる。当該方法は特に、一塩基レベルでは異なる配列であっても実質的に同機能又は同表現型を有する塩基配列情報をスクリーニングすることに適する。

20

【 0 0 1 6 】

また、偏在性表示塩基からなる配列情報は、元の塩基配列情報を分節化塩基数で割った値にほぼ等しいデータ量であることから、当該塩基配列情報の保存及び送受信にかかるデータ量を減少させることができる。特に本発明の際立った効果は、当該核酸に含まれる塩基配列の機能性に関与する変異のみを対象とし、機能性に関与しない特に一塩基レベルでの変異を「ノイズ」として除去することができることである。それゆえ、計算手順を複雑化させるが、表現型及び核酸配列の機能に関与しない特に一塩基レベルでの変異を「無視」し、比較解析などの解析に供することができる。それゆえ、本発明の複数の偏在性表示配列を使用する比較解析は、元の塩基配列情報を使用した場合の比較解析に要する計算量に比べて大幅に減少させることができる。

30

これにより、既存の技術では実質的に不可能であった、ゲノムなどの長大な塩基配列情報同士を比較解析することが可能となる。

【 0 0 1 7 】

それゆえ、本発明は、以下は限定の意図のない例示であるが、例えば：未同定種における種の同定；種の近縁関係の判定；有用形質又は有害形質を有するゲノム領域の推測；及び、長大なゲノムから解析対象領域を絞り込む際のスクリーニング；などに、直接的又は間接的に好ましく適用することができる。

40

【 発明を実施するための最良の形態 】

【 0 0 1 8 】

以下に記載する本発明の実施態様は、本発明の特許請求の範囲に関する理解を深めるために記載しているものであり、本発明の特許請求の範囲を限定することを意図するものではない。本明細書に記載の特許請求の範囲を逸脱しない範囲において、本発明の主題を達成し得る様々な態様、修飾、及び変更が可能であることは、当業者に理解されるであろう。

【 0 0 1 9 】

（定義）

本明細書で使用する用語「核酸塩基配列」、「塩基配列」及び「核酸配列」は、互換的

50

かつ同義的に使用することができる。

本明細書で使用する用語「塩基」は、核酸構成塩基及び混合塩基を包含する、全ての塩基をいうことができる。本明細書で使用する用語「塩基」は、文脈により、物質としての塩基、及び情報としての塩基記号の両方をいうことができる。本明細書で使用する用語「塩基」は、文脈により、単数又は複数であり得、核酸構成塩基若しくは混合塩基の個々の塩基あるいはそれらの複数を包含する塩基群をいうこともできる。

【0020】

塩基配列情報は任意の媒体で提供されてよく、限定の意図なく例示すると、紙などに記載又は印刷された情報、及びコンピュータで使用可能な情報などであってよい。コンピュータで使用する塩基配列情報は、以下は限定の意図のない例示であるが、入力手段によって入力された塩基配列情報、所定のデータベースから得られた塩基配列情報、内部記録媒体（例えば、ハードディスクドライブ（HDD）など）から得られた塩基配列情報、外部記録媒体（例えば、CD-ROM/RAM、DVD-ROM/RAM、外付けHDD又はUSBメモリなど）から得られた塩基配列情報、及びインターネット通信手段（例えば、通信インターフェイス）を介して得られた塩基配列情報であってよい。

10

【0021】

塩基配列情報は、下記表に記載のサイトを含む、インターネット上のサイトから入手可能な情報であってよい（下記表は、ゲノム配列から機能解析へ 第2版、マウント デービッド W.（著）、岡崎 康司（監訳）、坊農 秀雅（監訳）、メディカル・サイエンス・インターナショナル、2005年、465頁から引用した）。

20

【表 1】

データベース	ウェブアドレス
A. thaliana information resource TAIR	http://www.arabidopsis.org/
Caenorhabditis elegans (worm) database	http://www.wormbase.org/
Cold Spring Harbor Laboratory Genome Projects	http://www.cshl.org/
DOE Human Genome	http://www.doegenomes.org/
Drosophila genome project Berkeley	http://www.fruitfly.org/
Drosophila: Flybase, a genomic database	http://flybase.bio.indiana.edu/
E. coli genome project	http://www.genome.wisc.edu/
ENSEMBL e! genome browser (Sanger Inst., EMBL-EBI)	http://www.ensembl.org/
Genome list at NIH	http://molbio.info.nih.gov/molbio/db.html
Genome database in support of human genome project	http://www.gdb.org
Gramene—cereal grasses	http://www.gramene.org/
Human Genome Browser (Golden Path, UCSC)	http://genome.ucsc.edu/goldenPath/hgTracks.html
Mouse (Mus musculus) genome informatics	http://www.informatics.jax.org/
National Center for Biotechnology Information (NCBI)	http://www.ncbi.nlm.nih.gov/
NIH Human Genome Res. Initiative NHGRI	http://www.genome.gov/
Parasite genome databases	http://www.ebi.ac.uk/parasites/parasite-genome.html
PIR Georgetown University genomes	http://www-nbrf.georgetown.edu/pir/genome.html
Prokaryotic genomes P. Bork laboratory	http://www.bork.embl-heidelberg.de/Genome/
Rice (Oryza sativa) genome project	http://rgp.dna.affrc.go.jp/
S. cerevisiae (budding yeast) database (SGD)	http://genome-www.stanford.edu/Saccharomyces/
Stanford University genome resources	http://genome-www.stanford.edu/
The Institute of Genomics Research (TIGR)	http://www.tigr.org/
The Sanger Institute	http://sanger.ac.uk/
U.S. Dept. of Energy Joint Genome Initiative	http://www.jgi.doe.gov/
Washington University Genome Center	http://genome.wustl.edu/
Whitehead and Broad Institutes for Biomedical Research	http://www.broad.mit.edu/

10

20

30

40

塩基配列情報は、XML形式、HTML形式、GenBank形式、EMBL形式、SwissProt形式、FASTA形式、PIR形式、GCG形式、GDE形式、AceDB形式、GFF形式、及び任意のテキスト形式、並びにその他の形式を含むがこれらに限定されない、任意の形式で提供されてよい。

【 0 0 2 2 】

本明細書で使用する用語「核酸構成塩基」とは、A、G、C、及びTのそれぞれの塩基、又は該塩基群をいう。塩基配列情報に塩基Uが含まれる場合には実質的にTとみなし、かつTとして取り扱うことができる。

本明細書で使用する用語「混合塩基」とは、核酸構成塩基の群を表すことができる塩基情報又は塩基記号をいう。具体的には、混合塩基は、当業者に周知の、下記表に記載の対

50

応関係を有する。

【表 2】

混合塩基	核酸構成塩基
M	A又はC
R	A又はG
W	A又はT
S	C又はG
Y	C又はT
K	G又はT
V	A又はC又はG
H	A又はC又はT
D	A又はG又はT
B	C又はG又はT
N	A又はC又はG又はT

10

【0023】

20

本明細書で使用する用語「特性プロファイル」とは、当該塩基配列情報を特徴づけるための名称、注釈、説明、又はパラメータを含むが、これらに限定されない情報セットの全て又はその一部を含む情報という。それゆえ、「特性プロファイル」は、塩基配列情報に含まれる塩基配列以外の情報、すなわち：当該塩基配列を構成する塩基数（L値）及びGC含量（P値）；並びに、当該塩基配列に付随する様々な情報、例えば、当該塩基配列を定義するための名前及びID、当該核酸の起源生物、当該塩基配列の形状又は形態（例えば、直鎖状、環状、一本鎖、二本鎖など）、当該核酸を構成する塩基（例えば、DNA、RNA、人工塩基など）、当該起源生物の生物学的分類、遺伝子座、並びに当該核酸に含まれる遺伝子名；などを含むがこれらに限定されない情報のセットの全て又はその一部を含む情報という。特性プロファイルは、XML形式、HTML形式、GenBank形式、EMBL形式、SwissProt形式、FASTA形式、PIR形式、GCG形式、GDE形式、AceDB形式、GFF形式、及び任意のテキスト形式、並びにその他の形式を含むがこれらに限定されない、任意の形式で提供されてよい。

30

【0024】

用語「分節化開始塩基」とは、前記「分節化」を実行する際に基準となる第1番目の塩基をいい、数値で表現される。分節化開始塩基は、分節化される塩基配列情報よりも小さな、任意の正の整数であってよい。

本明細書で使用する用語「分節化塩基数」とは、塩基配列情報を分節化の分割間隔を規定する数値をいう。分節化塩基数は、分節化される塩基配列情報よりも小さな、任意の正の整数であってよい。

【0025】

40

本明細書で使用する用語「分節化」とは、対象の塩基配列情報を、所定の分節化開始塩基から、所定の分節化塩基数で重複なく順次分割して分節群を作成することをいう。それゆえ、用語「分節」とは、前記分節化により作成される塩基配列情報の各々の断片をいう。分節化開始塩基の前に塩基が存在する場合、すなわち分節化開始塩基が「1」でない場合、分節化開始塩基より前の塩基は分節として扱ってもよく、又は分節として扱わなくてもよい。塩基配列情報が環状の核酸由来の配列である場合、当該塩基配列情報の最後の塩基から第1番目以下の塩基にかけて連続的に分節化塩基数として計数して分節化に使用してもよい。

【0026】

本明細書で使用する用語「換算表」とは、混合塩基を含む塩基の出現数を、核酸構成塩

50

基の出現数へと換算するために使用する換算表をいう。

本明細書で使用する用語「出現数」とは、対象となる分節内に存在する核酸構成塩基の出現数をいう。本発明によると、分節内に混合塩基が存在する場合には、当該混合塩基は所定の換算表に従って、核酸構成塩基の出現数として換算される。それゆえ、出現数は、0、又は正の整数、若しくは正の実数であり得る。

本明細書で使用する用語「核酸構成塩基の出現数のプロファイル」とは、任意の塩基配列情報の一定の領域内、特に本発明の分節内に存在する、核酸構成塩基 A、G、C 及び T のそれぞれの出現数のセットからなる情報をいう。それゆえ、核酸構成塩基の出現数のプロファイルは、分節化によって生じた各分節ごとに特徴づけられる情報であり得る。

【0027】

用語「第1頻出塩基」、「第2頻出塩基」、「第3頻出塩基」及び「第4頻出塩基」は、本発明の偏在性表示配列を作成する過程で使用される用語であり：第1頻出塩基は、対象となる分節における出現数が最も多い核酸構成塩基を意味し；第2頻出塩基は、対象となる分節における出現数が2番目に多い核酸構成塩基を意味し；第3頻出塩基は、対象となる分節における出現数が3番目に多い核酸構成塩基を意味し；及び第4頻出塩基は、対象となる分節における出現数が4番目に多い、すなわち出現数が最も少ない核酸構成塩基；を意味する。用語「頻出塩基」は、ある分節において、複数の核酸構成塩基間での出現数が互いに等しい場合、該複数の核酸構成塩基からなる群を意味することができる。用語「頻出塩基」が複数の核酸構成塩基からなる群をいう場合、本発明の検定には、前記複数の核酸構成塩基群の出現数の合計が使用され得る。それゆえ、頻出塩基は、ある分節において、第3頻出塩基まで又は第2頻出塩基までの場合があり得、全ての頻出塩基が等しい場合には第1頻出塩基のみの場合すらあり得る。頻出塩基が複数の核酸構成塩基をいう場合、頻出塩基は、混合塩基で表され得る。

【0028】

本明細書で使用する用語「偏在性」とは、当該塩基配列情報に含まれる塩基の偏りをいう。好ましい実施態様において、用語「偏在性」は、本発明に従って作成された分節内において、有意に存在数の多い核酸構成塩基（群）をいう。本発明によると、塩基配列情報に含まれる全ての塩基はいったん核酸構成塩基に変換されるが、最終的な出力情報は、核酸構成塩基のみならず、混合塩基の形態でも出力され得る。それゆえ、本明細書で使用する用語「偏在性」は、当該分節に含まれる単一の核酸構成塩基の偏在性を表すことができ、複数の核酸構成塩基の偏在性も表すことができる。

【0029】

本明細書で使用する用語「偏在性表示塩基」とは、本発明に従って得られる各々の分節において、当該各分節に含まれる核酸構成塩基のうち、有意に存在数の多い核酸構成塩基（群）を塩基記号の形態で表現したものをいう。

本発明の検定に使用される有意水準は、当該検定に使用され得る任意の数値であり得る。

本発明の検定に使用される自由度は、当該検定に使用され得る任意の数値であり得る。

本発明に使用される検定は、任意に、片側検定又は両側検定であり得る。

【0030】

本発明に使用する用語「対応表」とは、本発明の偏在性表示塩基の作成において、当該分節において有意に存在が多いと判断された核酸構成塩基（群）に対応する記号を割り当てるために使用する表をいう。典型的実施態様において、本発明の偏在性表示塩基は、当業者に周知の塩基記号の形態であり得、A、C、G、T、M、R、W、S、Y、K、V、H、D、B、又はNの形態であり得る。

本明細書で使用する用語「偏在性表示配列」とは、偏在性表示塩基を、元の塩基配列情報の並びに従って並べた配列情報をいう。それゆえ、典型的実施態様において、本発明の偏在性表示配列は、A、C、G、T、M、R、W、S、Y、K、V、H、D、B、又はNからなる偏在性表示塩基群から構成され得る。

【0031】

本明細書で使用する用語「偏在性表示視覚化情報」とは、本発明に従って得られた偏在性表示塩基ごとにそれぞれ対応する、所定の色、パターン又はテクスチャを割り当てて作成された情報をいう。偏在性表示塩基に割り当てられる色は任意の色であってよいが、各偏在性表示塩基によってそれぞれ異なることを条件とする。偏在性表示塩基に割り当てられるパターンは、任意のパターンであってよいが、各偏在性表示塩基によってそれぞれ異なることを条件とする。偏在性表示塩基に割り当てられるテクスチャは、任意のテクスチャであってよいが、各偏在性表示塩基によってそれぞれ異なることを条件とする。

【0032】

偏在性表示視覚化情報は、入力された塩基配列情報の特性プロファイルに含まれる情報に基づき、直線状又は環状などを含む任意の形態で出力されてよい。好ましい実施態様において、偏在性表示視覚化情報は、当該偏在性表示視覚化情報を作成するのに使用した塩基配列情報の特性プロファイルに含まれる塩基配列の形態情報と一致した形態で出力される。例えば、偏在性表示視覚化情報は、当該偏在性表示視覚化情報の作成に使用された塩基配列情報を模式的に表す長方形（例えば、直鎖状ゲノムの場合）又は環（例えば、環状ゲノム又はプラスミドの場合）を、その偏在性表示配列に含まれる塩基数で等分し、当該各等分された面積部分にそれぞれの偏在性表示塩基に対応する所定の色、パターン又はテクスチャを順次割り当てることによって作成することができる。

【0033】

本明細書で使用する用語「クラスター」とは、本発明の偏在性表示視覚化情報を分類するための単位をいう。クラスターは、既定のクラスター（群）が設定されていてもよく、又は入力手段を使用して任意に設定してもよい。クラスターには、任意の分類カテゴリーを設定することができる。クラスターは、以下は限定の意図のない例示であるが：特定の界、門、綱、目、科、属、及び種、並びに各々の中間的又は細分化された既存の生物学的分類；特定の病原性、及び発酵能などを含むが、これらに限定されない表現型；海水、淡水、高温、低温、高塩濃度、低塩濃度などを含むが、これらに限定されない生育環境；細菌、酵母、ウイルス、真核生物、原核生物、単細胞生物、多細胞生物、及び哺乳動物などを含むがこれらに限定されない任意の分類カテゴリー；を含むが、これらに限定されないカテゴリーであってよい。クラスターは、偏在性表示視覚化情報の作成に使用した元の塩基配列情報の起源情報に基づいて設定してよい。

【0034】

本明細書で使用する用語「クラスターと関連付けて出力（表示）する」又は「クラスターリング」とは、偏在性表示視覚化情報とクラスターとを関連づけて表示若しくは出力することをいう。前記クラスターは単数又は複数であってよい。前記クラスターが複数である場合には、該クラスター同士は包含関係を示すように表示又は出力されてよい。「偏在性表示視覚化情報とクラスターとを関連づける」とは、各々の偏在性表示視覚化情報を、元の塩基配列情報の特性プロファイルに含まれる情報に基づいて適切なクラスターへと包含させることをいう。例えば、元の塩基配列情報がバチルス（*Bacillus*）属由来であった場合、当該塩基配列情報から作成された偏在性表示視覚化情報は：生物学的分類クラスターに関しては、バチルス属のクラスター、バチルス科のクラスター、及びバチルス目のクラスターなどに包含させることができ；又は、細菌クラスター、原核生物クラスター、又は単細胞生物クラスター；などに包含させることができる。「クラスター同士の包含関係を関連づける」とは、より上位概念のクラスターの中に、その下位概念にあたるクラスターを包含させることをいう。例えば、生物学的分類をクラスターとした場合には、「種」レベルのクラスターは「属」レベルのクラスターに包含され、「属」レベルのクラスターは「科」レベルのクラスターに包含され、及び「科」レベルのクラスターは「目」レベルのクラスターに包含され得る。

【0035】

本明細書に記載する用語「データベース」とは、塩基配列情報、塩基配列情報に付随する特性プロファイル、本発明の偏在性表示配列及び偏在性表示視覚化情報、本発明の偏在性表示配列及び偏在性表示視覚化情報を作成する際の各種パラメータ並びにそれらの中間

10

20

30

40

50

データを含む、本発明のあらゆる段階で得られる全てのデータを格納、記録又は保存、及び保持することができるデータベースをいう。データベースは、格納、記録又は保存、及び保持されるデータの種類に従ってそれぞれ独立に設定されてよい。本発明において設定され得るデータベースを限定の意図なく例示すると、塩基配列情報データベース、偏在性表示配列データベース（分節化塩基数、分節化開始塩基などのパラメータを変更した一連の偏在性表示配列を含む）、偏在性表示視覚化情報データベース（分節化塩基数、分節化開始塩基などのパラメータを変更した一連の偏在性表示視覚化情報を含む）などが挙げられる。データベースは、メモリ又は内部記録媒体などのコンピュータ内部の媒体内に作成されてよく、外部記録媒体又はインターネット上などのコンピュータを使用してアクセスできる媒体内に作成されてよく、若しくは紙などの物理的媒体に記載又は記録された形態であってもよい。

10

【 0 0 3 6 】

（本発明の方法）

一実施態様において、本発明は、塩基配列情報に含まれる核酸構成塩基の偏在性を表す方法であって：塩基配列情報を、所定の分節化開始塩基から、所定の分節化塩基数で重複なく順次分割して分節群を作成する手順；前記各分節において、前記塩基配列情報を、所定の換算表に従って核酸構成塩基のみからなる情報に換算し、各核酸構成塩基の出現数を計数する手順；前記各分節において、前記核酸構成塩基の出現数を使用して、偏在性表示塩基を作成する手順；及び、前記各分節において作成された偏在性表示塩基を前記塩基配列情報の並びに従って配列化して、偏在性表示配列を作成する手順；を含む、前記方法を提供する。別の実施態様において、本発明は、前記偏在性表示塩基ごとに所定の色、パターン又はテクスチャを割り当て、偏在性表示視覚化情報を作成する手順をさらに含む方法を提供する。

20

【 0 0 3 7 】

本発明の方法の一般的実施態様において、所定の換算表は、下記の換算表である。

【表 3】

塩基	換算される核酸構成塩基	
A	A	
C	C	
G	G	
T	T	
U	T	10
M	1/2 A、1/2 C	
R	1/2 A、1/2 G	
W	1/2 A、1/2 T	
S	1/2 C、1/2 G	
Y	1/2 C、1/2 T	
K	1/2 G、1/2 T	
V	1/3 A、1/3 C、1/3 G	20
H	1/3 A、1/3 C、1/3 T	
D	1/3 A、1/3 G、1/3 T	
B	1/3 C、1/3 G、1/3 T	
N	1/4 A、1/4 C、1/4 G、1/4 T	

【0038】

本発明の方法の別の実施態様において、偏在性表示塩基の作成手順は、下記工程を含む
 ；
 当該分節において、前記核酸構成塩基出現数を使用して、各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する工程

【数 1】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し；PはGC含量(%)を示し；及び、Lは当該分節内の塩基の総数；を示す)；
 前記各核酸構成塩基の χ^2 値を、下記式に従って算出する工程

【数 2】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2$$

(式中、 $_B$ は任意の1つの核酸構成塩基を示し、 $_{BR}$ は前記 $_B$ 以外の全ての核酸構成塩基を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及びLは当該分節内の塩基の総数を示す)；

及び、

(i) 前記 χ_B^2 群の全てが、条件(A): 「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たさない場合には、当該分節に、偏在性表示塩基として「N」を割り当て、
 (ii) 前記 χ_B^2 群の少なくとも1つの χ_B^2 (群) が条件(A): 「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の全てが条件(B): 「当該出現数が、その期待数よりも多い」を満たさない場合、前記条件(A)を満たしかつ前記条件(B)を満たさない核酸構成塩基(群)以外の全てからなる核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当て、及び、

(iii) 前記 χ_B^2 群のうち少なくとも1つの χ_B^2 (群) が条件(A): 「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の少なくとも1つが条件(B): 「当該出現数が、その期待数よりも多い」を満たす場合、前記条件(A)を満たしかつ前記条件(B)を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる、工程。 10

【0039】

本発明の方法の別の実施態様において、偏在性表示塩基の作成手順は、下記工程を含む。

当該分節において、前記核酸構成塩基出現数を使用して、各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する工程

【数3】

20

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し；PはGC含量(%)を示し；及び、Lは当該分節内の塩基の総数；を示す)；前記各核酸構成塩基の χ^2 値を、下記式に従って算出する工程

【数4】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2 \quad 30$$

(式中、 B は任意の1つの核酸構成塩基を示し、 BR は前記 B 以外の全ての核酸構成塩基を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及びLは当該分節内の塩基の総数を示す)；及び、

(i) 前記 χ_B^2 群の全てが、条件(A): 「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たさない場合には、当該分節に偏在性表示塩基として「N」を割り当て、
 (ii) 前記 χ_B^2 群の少なくとも1つの χ_B^2 (群) が条件(A): 「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の全てが条件(B): 「当該出現数が、その期待数よりも多い」を満たさない場合には、当該分節に偏在性表示塩基として「N」を割り当て、及び、 40

(iii) 前記 χ_B^2 群のうち少なくとも1つの χ_B^2 (群) が条件(A): 「当該 χ_B^2 値が、所定の有意水準のカイ二乗値より大きい」を満たし、かつ、前記条件(A)を満たした核酸構成塩基(群)の少なくとも1つが条件(B): 「当該出現数が、その期待数よりも多い」を満たす場合、前記条件(A)を満たしかつ前記条件(B)を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる、工程。

【0040】

50

本発明の方法の別の実施態様において、偏在性表示塩基の作成手順は、下記工程を含む。

当該分節において、核酸構成塩基の出現数をそれぞれ計数して、第1、第2、第3及び第4頻出塩基を決定する工程；

前記頻出塩基の出現数を使用して、所定の有意水準において有意に出現頻度が高い核酸構成塩基（群）を判定するための検定を実行する工程；及び、

前記検定により有意に出現頻度が高いと判定された核酸構成塩基（群）が存在する場合には該当する核酸構成塩基（群）に対応する偏在性表示塩基を所定の対応表に従って当該分節に割り当て、全ての核酸構成塩基の出現数に有意差はないと判断された場合には、当該分節に偏在性表示塩基として「N」を割り当てる工程。

10

【0041】

本発明の方法の別の実施態様において、検定は、統計学的検定に基づく検定である。

本発明の方法の別の実施態様において、検定は、下記工程を含む仮説検定である。

第1仮説を「当該分節において、第1頻出塩基の出現数はその期待値に等しい」とする第1仮説検定を実行する工程；

前記第1仮説検定で前記第1仮説が棄却される場合には前記第1頻出塩基が当該分節において有意に出現頻度が高い塩基であると判定し、前記第1仮説検定で前記第1仮説が棄却されない場合には第2仮説を「当該分節において、第1及び第2頻出塩基の出現数の合計はその期待値に等しい」とする第2仮説検定を実行する工程；

前記第2仮説検定で前記第2仮説が棄却される場合には前記第1及び第2頻出塩基からなる塩基（群）が当該分節において有意に出現頻度が高い塩基（群）であると判定し、前記第2仮説検定で前記第2仮説が棄却されない場合には第3仮説を「当該分節において、第1、第2及び第3頻出塩基の出現数の合計はその期待値に等しい」とする第3仮説検定を実行する工程；及び、

20

前記第3仮説検定で前記第3仮説が棄却される場合には前記第1、第2及び第3頻出塩基からなる塩基（群）が当該分節において有意に出現頻度が高い塩基（群）であると判定し、前記第3仮説検定で前記第3仮説が棄却されない場合には全ての核酸構成塩基の出現数に有意差はないと判定する工程。

【0042】

本発明の方法の別の実施態様において、仮説検定は、正規分布表、t分布表、F分布表、又はカイ二乗分布表からなる群から選択される分布表を使用する。

30

本発明の方法の一般的実施態様において、対応表は、下記に記載の対応表である。

【表 4】

核酸構成塩基 (群)	偏在性表示塩基	
A	A	
C	C	
G	G	
T	T	
A及びC	M	10
A及びG	R	
A及びT	W	
C及びG	S	
C及びT	Y	
G及びT	K	
A及びC及びG	V	
A及びC及びT	H	20
A及びG及びT	D	
C及びG及びT	B	

【0043】

一実施態様において、本発明は、本明細書に記載の方法に従い、同一の分節化塩基数を使用して作成された複数の偏在性表示配列間の類似性を比較解析する手順を含む、塩基配列情報の比較解析方法を提供する。本発明による塩基配列情報の比較解析方法の別の実施態様において、複数の偏在性表示配列は、所定のデータベースに格納されている。

一実施態様において、本発明は、本明細書に記載の方法に従い、同一の分節化塩基数を使用して作成された複数の偏在性表示視覚化情報間の類似性を、所定のクラスターと関連付けて表示する手順を含む、塩基配列情報の比較解析方法を提供する。本発明の塩基配列情報の比較解析方法の別の実施態様において、複数の偏在性表示視覚化情報は、所定のデータベースに格納されている。

【0044】

一実施態様において、本発明は、本発明の塩基配列情報に含まれる核酸構成塩基の偏在性を表す方法をコンピュータに実行させるプログラムを提供する。別の実施態様において、本発明は、本発明による本発明の塩基配列情報に含まれる核酸構成塩基の偏在性を表す装置を提供する。別の実施態様において、本発明は、P値の入力手段をさらに含む。別の実施態様において、本発明は、所定の有意水準のカイ二乗値の入力手段をさらに含む。別の実施態様において、本発明は、(i) 塩基配列情報を取得又は出力するための、及び(i) 本発明の偏在性表示塩基、偏在性表示配列若しくは偏在性表示視覚化情報を取得又は出力するための、インターネット通信手段をさらに含む。別の実施態様において、本発明は、P値の入力手段をさらに含む。別の実施態様において、本発明は、分節化開始塩基の入力手段をさらに含む。別の実施態様において、本発明は、分節化塩基数の入力手段をさらに含む。別の実施態様において、本発明は、有意水準の入力手段をさらに含む。

【0045】

(本発明のプログラム、コンピュータ読み取り可能な記録媒体、及び装置)

本発明は、本発明の方法をコンピュータに実行させるプログラムを提供する。本発明のプログラムは、コンピュータ読み取り可能な記録媒体に記録させて提供することもできる。また、本発明は、本発明の方法を実質的に実行可能な装置も提供する。

【 0 0 4 6 】

(本 発 明 の 構 成)

図1は、本発明の装置の構成図である。図中、メモリ(101)、CPU(102)、入力装置(103)、内部記録媒体(104)、通信インターフェイス(105)、及び出力装置(107)はバス(100)を介して互いに接続され、任意に双方向通信(情報の入出力)が可能である。インターネット(106)は、本発明の装置の通信インターフェイス(105)を介して任意に接続されている。外部記録媒体(108)は、本発明の装置のバス(100)を介して任意に接続されている。入力装置(103)は、キーボード、マウス、又は音声認識入力装置などを含むが、これらに限定されない入力装置であってよい。内部記録媒体(104)は、HDDを含むがこれに限定されない、当該コンピュータ内に搭載されている記録媒体であってよい。通信インターフェイス(105)は、インターネットに接続するための双方向の通信を可能にするデバイスであってよく、モデム、ネットワークインターフェース(例えばイーサネット(登録商標)カード)、通信ポート、PCMCIAスロット及びカードなどを含むがこれらに限定されない通信インターフェイスであり得る。外部記録媒体(108)は、CD-ROM/RAM、DVD-ROM/RAM、外付けHDD又はUSBメモリを含むがこれらに限定されない、当該コンピュータに外付け可能な記録媒体であってよい。

10

【 0 0 4 7 】

メモリ(101)は、本発明の方法をコンピュータに実行させるプログラム、該プログラムを実行するために必要な情報(例えば、塩基配列情報、特性プロファイル、分節化塩基数、分節化開始塩基、所定の分布表、所定の対応表、所定の換算表; 偏在性表示視覚化情報の作成に使用する所定の色、パターン及びテキスト情報; 入力情報から出力情報までの中間データ群、出力データ、及び各種パラメータなど)、及び該プログラムの手順を実行することにより得られた各種中間情報及び出力情報(例えば、各分節における核酸構成塩基の出現数及び期待数、各分節における各核酸構成塩基の χ^2 値、各分節における第1~第4頻出塩基、各分節に割り当てられた偏在性表示塩基、偏在性表示配列、及び偏在性表示視覚化情報など)などを含む情報を記憶(格納)並びに保持できる。本発明の方法をコンピュータに実行させるプログラムは、入力装置(103)、内部記録媒体(104)、通信インターフェイス(105)、又は外部記録媒体(108)を介してメモリ(101)に入力できる。本発明のプログラムは、該プログラムを記録した記録媒体を介して別のコンピュータにインストールして使用することもできる。また、インターネットを介して、本発明のプログラムをダウンロード及びインストールして使用することもできる。

20

30

【 0 0 4 8 】

CPU(102)では、メモリ(101)に入力された情報に基づき、本発明の方法に従う計数及び演算手順が実行され、その結果得られた情報は再びメモリ(101)へと格納される。本発明のプログラムの手順を実行することにより得られた各種中間情報及び出力情報は、バス(100)を介して、メモリ(101)から出力装置(107)、内部記録媒体(104)、又は外部記録媒体(108)などに出力される。ある実施態様において、各種中間情報及び出力情報は、通信インターフェイス(105)を介して出力される。出力装置(107)は、モニタ、ディスプレイ、プロジェクタ又はプリンタなどを含むが、これらに限定されない出力装置であってよい。内部記録媒体(104)は、HDDなどを含むがこれに限定されない、コンピュータ内部に搭載されている記録媒体であってよい。外部記録媒体(108)は、CD-ROM/RAM、DVD-ROM/RAM、外付けHDD又はUSBメモリなどを含むがこれらに限定されない、コンピュータの外部から接続して使用される記録媒体であってよい。通信インターフェイス(105)は、モデム、ネットワークインターフェース(例えばイーサネット(登録商標)カード)、通信ポート、PCMCIAスロット及びカードなどを含むがこれらに限定されない通信インターフェイスであってよい。

40

【 0 0 4 9 】

(本 発 明 の 方 法 の 実 施 態 様)

本発明による、塩基配列情報に含まれる核酸構成塩基の偏在性を表す方法の実施態様を、図2~4にフローチャート形式で例示的に記載している。以下順次、図2、図3及び図4を

50

参照しながら説明する。以下に記載の全てのステップは、本発明のプログラムに基づき、CPU (102) で実行され得る。

はじめに、塩基配列情報に含まれる核酸構成塩基の偏在性を表す第1の方法を、図2を参照しながら説明する。ステップ202において、解析対象の塩基配列情報を入力する。塩基配列情報は、所定のデータベース、又はメモリ (101)、内部記録媒体 (104) 若しくは外部記録媒体 (108) などに格納することができる。塩基配列情報は、以下は限定の意図のない例示であるが、手動 (すなわち図1に103として示したような入力手段を介してコンピュータに配列を入力すること) で、インターネット (通信インターフェイス (105) を使用) を介して、メモリ (101) から、内部記録媒体 (104) から、又は外部記録媒体 (108) から、入力された情報であってよい。塩基配列は、以下は限定の意図のない例示であるが、任意の生物、ウイルス、プラスミド、人工染色体 (BAC及びYACなどを含む)、及びそれらの組換え体由来の塩基配列の全体及び部分の塩基配列、合成配列由来の塩基配列の全体及び部分の塩基配列、並びに他の全ての塩基配列の全体及び部分の塩基配列であってよい。当該塩基配列情報の特性プロファイルには、例えば塩基配列、該配列を構成する塩基数 (L値)、及びGC含量 (P値) などが含まれていてもよい。さらに、特性プロファイルには、当該塩基配列に付随する様々な情報、例えば、当該塩基配列を定義するための名前及びID、当該核酸の起源生物、当該起源生物の生物学的分類、遺伝子座、並びに当該核酸に含まれる遺伝子名を含むがこれらに限定されない情報も含まれていてよい。

10

【0050】

ステップ203において、ステップ202にて入力された塩基配列情報を分節化する。分節化は、当該塩基配列情報を、所定の分節化開始塩基から、所定の分節化塩基数で重複なく順次分割して分節群を作成することにより実行される。

20

一実施態様において、分節化開始塩基は、規定値であってよい。別の実施態様において、分節化塩基は、当該塩基配列情報に含まれる塩基数を超えない範囲で入力された任意の数値であってよい。一実施態様において、分節化開始塩基は、当該塩基配列情報の第1番目の塩基であってよい。別の実施態様において、分節化開始塩基は、分節化塩基数よりも小さな任意の塩基番号の塩基であってよい。他の実施態様において、分節化開始塩基は、所定の間隔で、第1の分節化開始塩基から下流 (3'方向) 又は必要に応じて上流 (5'方向) に順次変更することができる。前記間隔は、元の塩基配列情報、又は他の比較対象の塩基配列情報などとの関連から条件検討を実施し、最適値又はそれに準じる適切な値を設定することができる。所定の間隔で順次分節化開始塩基を変更させた一連の分節化は、コンピュータに実行させることができる。分節化開始塩基は、塩基配列情報の機能、又は染色体上の位置に応じて決定されてよい。本発明の特定の実施態様において、分節化開始塩基は、以下は限定の意図のない例示であるが：塩基配列情報中に含まれる複製開始点、転写開始点、又は翻訳開始点に対応する数値；若しくは、染色体の第1番目の塩基に対応する数値；などであってよい。当該分節化開始塩基を順次変更した一連の分節化データは、所定のデータベース、又はメモリ (101)、内部記録媒体 (104) 若しくは外部記録媒体 (108) などに一時的又は持続的に記録されてよい。

30

【0051】

一実施態様において、分節化塩基数は、規定値であってよい。別の実施態様において、分節化塩基数は、当該塩基配列情報に含まれる塩基数を超えない範囲で任意に入力された数値であってよい。他の実施態様において、分節化塩基数を順次変更した一連の分節化が可能であり、前記一連の分節化は、所定の比率又は倍率に基づいていてよい。分節化塩基数を順次変更した一連の分節化は、コンピュータに実行させることができる。分節化塩基数は、元の塩基配列情報、又は他の比較対象の塩基配列情報などとの関連から条件検討を実施し、最適値又はそれに準じる適切な値を設定することができる。当該分節化塩基数を順次変更した一連の分節化データは、所定のデータベース、又はメモリ (101)、内部記録媒体 (104) 若しくは外部記録媒体 (108) などに一時的又は持続的に記録されてよい。

40

【0052】

さらに、分節化開始塩基及び分節化塩基数の両方を、それぞれ順次変更した一連の分節

50

化も可能である。例えば、分節化開始塩基を所定の値に固定しておき分節化塩基数を所定の比率又は倍率で順次変更する一連の分節化が可能であり、分節化塩基数を所定の値に固定しておき分節化開始塩基を所定の間隔で順次変更した一連の値での分節化も可能である。その結果、分節内塩基数が一定でかつ分節化開始塩基が順次異なる一連の分節群、分節化開始塩基が一定でかつ分節内塩基数が異なる一連の分節群、及びそれらの組み合わせとして、分節化開始塩基が順次異なりかつ分節内塩基数が異なる一連の分節群を作成することができる。これらの分節化データは、所定のデータベース、又はメモリ（101）、内部記録媒体（104）若しくは外部記録媒体（108）などに一時的又は持続的に記録されてよい。

ステップ203で得られる全てのデータは、所定のデータベース、又はメモリ（101）、内部記録媒体（104）若しくは外部記録媒体（108）などに格納することができる。

【0053】

ステップ204において、前記ステップ203の分節化により得られた各分節において、各核酸構成塩基の出現数を計数する。典型的実施態様において、当該分節内に混合塩基が含まれている場合、下記表に従って核酸構成塩基へと換算される。

【表5】

塩基	換算される核酸構成塩基
A	A
C	C
G	G
T	T
U	T
M	1/2 A、1/2 C
R	1/2 A、1/2 G
W	1/2 A、1/2 T
S	1/2 C、1/2 G
Y	1/2 C、1/2 T
K	1/2 G、1/2 T
V	1/3 A、1/3 C、1/3 G
H	1/3 A、1/3 C、1/3 T
D	1/3 A、1/3 G、1/3 T
B	1/3 C、1/3 G、1/3 T
N	1/4 A、1/4 C、1/4 G、1/4 T

すなわち、当該分節に含まれる全ての塩基の出現数は、核酸構成塩基 A、G、C 及び T の出現数へと変換される。各核酸構成塩基の出現数は、0、又は正の整数、若しくは正の実数であり得る。

【0054】

ステップ205では、各分節において、核酸構成塩基 A、G、C 及び T の χ^2 値をそれぞれ算出する。本発明の実施態様において、前記各核酸構成塩基の χ^2 値は、下記式に従って算出される。

【数 5】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し； p はGC含量(%)を示し；及び、 L は当該分節内の塩基の総数；を示す)；

当該分節において、前記各核酸構成塩基の χ^2 値を、下記式に従って算出するステップ

【数 6】

$$\chi_B^2 = \frac{(O_B - E_B)^2}{E_B} + \frac{(O_{BR} - E_{BR})^2}{E_{BR}} = \frac{(O_B - E_B)^2}{E_B} + \frac{1}{L - E_B} [(L - O_B) - (L - E_B)]^2$$

(式中、 B は任意の1つの核酸構成塩基を示し、 BR は前記 B 以外の全ての核酸構成塩基群を示し、 O_B は当該分節における前記核酸構成塩基の出現数を示し、 O_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の出現数の合計を示し、 E_B は当該分節における前記核酸構成塩基の期待数を示し、 E_{BR} は当該分節における前記核酸構成塩基以外の全ての核酸構成塩基群の期待数の合計を示し、及び L は当該分節内の塩基の総数を示す)

。

上記各核酸構成塩基の χ^2 値の算出ステップの一実施態様において、 P 値は、既定値で50であってよく、又は既定値で50以外の任意の数値であってよい。別の実施態様において、 P 値は、当該塩基配列情報に付随する特性プロファイルに含まれる P 値の値に基づいて設定されてもよく、又は任意に設定されてよい。ステップ205で得られる、各分節において算出される各核酸構成塩基の χ^2 値、及びその算出過程に含まれる値、並びに各種パラメータ値を含む全てのデータは、所定のデータベース、又はメモリ(101)、内部記録媒体(104)若しくは外部記録媒体(108)などに格納することができる。

【0055】

ステップ206~210では、各分節において、所定の有意水準を基準として有意に存在比率の高い核酸構成塩基(群)を決定するための検定を実施する。以下に、ステップ206~210を個別に説明する。

ステップ206は、当該各分節において、条件(A)：「当該核酸構成塩基の χ^2 値が、所定の有意水準のカイ二乗値より大きい」を満たす核酸構成塩基が1つ以上存在するかを判断するステップである。ステップ206において、全ての核酸構成塩基の χ^2 値が、前記所定の有意水準の χ^2 値以下であった場合、すなわち全ての核酸構成塩基が前記条件(A)を満たさない場合、当該分節に含まれる各核酸構成塩基の存在比率には有意な差はないと判断してステップ207へと移行し、当該分節に偏在性表示塩基として「N」が割り当てられる(ステップ207)。一方、前記ステップ206で、当該分節に含まれる核酸構成塩基のうち少なくとも1つの χ^2 値が、所定の有意水準のカイ二乗値より大きい場合、すなわち前記条件(A)を満たす核酸構成塩基が1つ以上存在する場合には、ステップ208へと移行する。

【0056】

ステップ208は、当該各分節において、条件(B)：「当該核酸構成塩基の出現数が、その期待数よりも多い」を満たす核酸構成塩基が1つ以上存在するかを判断するステップである。ステップ208において、前記ステップ206で χ^2 値が所定の有意水準よりも大きいと判断された核酸構成塩基(群)の全ての出現数がその期待値以下であった場合、すなわち前記条件(A)を満たす核酸構成塩基(群)が存在したが、その全てが前記条件(B)を満たさない場合には、ステップ209に移行する。

【0057】

ステップ209では、前記条件(A)を満たすが前記条件(B)を満たさない核酸構成塩基(群)以外の全てからなる核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性

10

20

30

40

50

表示塩基として割り当てる。例えば、前記条件(A)を満たすが前記条件(B)を満たさない核酸構成塩基がAであった場合、当該分節にはG、C及びTを意味する偏在性表示塩基「B」が割り当てられ；前記条件(A)を満たすが前記条件(B)を満たさない核酸構成塩基群がA及びGであった場合、当該分節にはC及びTを意味する偏在性表示塩基「Y」が割り当てられる。一方、前記ステップ208において、前記ステップ206で χ^2 値が所定の有意水準よりも大きいと判断された全ての核酸構成塩基（群）の出現数がその期待値よりも大きい核酸構成塩基（群）が1つ以上存在する場合、すなわち前記条件(A)及び(B)の両方を満たす核酸構成塩基（群）が1つ以上存在する場合、ステップ210に移行する。

【0058】

ステップ210では、前記条件(A)及び(B)の両方を満たす核酸構成塩基（群）を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる。例えば、前記条件(A)及び(B)の両方を満たす核酸構成塩基がAであった場合、当該分節には偏在性表示塩基「A」が割り当てられ；前記条件(A)及び(B)の両方を満たす核酸構成塩基群がA及びGであった場合、当該分節にはA及びGを意味する偏在性表示塩基「R」が割り当てられる。

【0059】

すなわち、本発明による塩基配列情報に含まれる核酸構成塩基の偏在性を表す第1の実施態様は、当該各分節において、(i)全ての核酸構成塩基群が条件(A)を満たさない場合には、偏在性表示塩基として「N」を割り当て；(ii)条件(A)及び(B)の両方を満たした核酸構成塩基（群）を、所定の対応表に従う偏在性表示塩基の作成対象とし；(iii)条件(A)は満たすが、条件(B)を満たさない核酸構成塩基（群）以外の核酸構成塩基（群）の全てを、所定の対応表に従う偏在性表示塩基の作成対象とする；実施態様である。

ステップ206～210で得られる全ての偏在性表示塩基及びその中間データ並びに各種パラメータを含む全てのデータは、所定のデータベース、又はメモリ（101）、内部記録媒体（104）若しくは外部記録媒体（108）などに格納することができる。

【0060】

ステップ211は、前記ステップ206～210において、各分節に割り当てられた偏在性表示塩基群を、当該塩基配列情報の塩基配列の順序（すなわち、作成された分節の順序）に従って整列させるステップである。ステップ211によって得られた偏在性表示塩基の並びからなる配列を偏在性表示配列という。ステップ211で得られる偏在性表示配列は、所定のデータベース、又はメモリ（101）、内部記録媒体（104）若しくは外部記録媒体（108）などに格納することができる。

【0061】

ステップ212は、分節化塩基数を変更して偏在性表示塩基及び偏在性表示配列を作成するかを判断するステップである。分節化塩基数を変更する場合には、ステップ203へと戻る。分節化塩基数を変更する場合の手順については、先のステップ203の記載において実質的に説明してあるとおりである。ゆえに、分節化塩基数を順次変更しながらステップ203～212を繰り返すことにより、入力された塩基配列情報について、同一の分節化開始塩基を有し、かつ異なる分節化塩基数を有する一連の偏在性表示配列が得られる。ステップ212、又はステップ203～212を繰り返すことにより得られた一連の偏在性表示配列は、所定のデータベース、又はメモリ（101）、内部記録媒体（104）若しくは外部記録媒体（108）などに格納することができる。ステップ212において、分節化塩基数を変更しない場合には、ステップ213へと移行する。前記分節化塩基数を変更せずにステップ213へと移行する場合には、(i)設定された1つの分節化塩基数の処理が終了した場合、及び(ii)設定された一連の分節化塩基数の全てについての処理が終了した場合、の両方が含まれ得る。

【0062】

ステップ213は、分節化開始塩基を変更して偏在性表示塩基及び偏在性表示配列を作成するかを判断するステップである。分節化開始塩基を変更する場合には、ステップ203へと戻る。分節化開始塩基を変更する場合の手順については、先のステップ203の記載において実質的に説明してあるとおりである。ゆえに、分節化開始塩基を順次変更しながらステップ203～212を繰り返すことにより、入力された塩基配列情報について、異なる分節化

開始塩基を有し、かつ同一の分節化塩基数を有する一連の偏在性表示配列が得られる：ここで、ステップ212で分節化塩基数を変更していた場合には、変更された分節化塩基数及び分節化開始塩基を組み合わせた数の偏在性表示配列が作成される。ステップ213、又はステップ203～213を繰り返すことにより得られた一連の偏在性表示配列は、所定のデータベース、又はメモリ（101）、内部記録媒体（104）若しくは外部記録媒体（108）などに格納することができる。ステップ213において、分節化開始塩基を変更しない場合には、ステップ214へと移行する。前記分節化開始塩基を変更せずにステップ214へと移行する場合には、（i）設定された1つの分節化開始塩基の処理が終了した場合、及び（ii）設定された一連の分節化開始塩基の全てについての処理が終了した場合、の両方が含まれ得る。

【0063】

ステップ214は、本発明による偏在性表示視覚化情報を作成するかを判断するステップである。偏在性表示視覚化情報を作成しない場合にはステップ215へと移行し、偏在性表示配列を出力する（ステップ215）。出力は、モニタ、ディスプレイ、プロジェクタ又はプリンタなどを含むが、これらに限定されない出力装置によってなされてよく、又は所定のデータベース、内部記録媒体（104）若しくは外部記録媒体（108）などに出力されてよい：ここで、所定のデータベース、又は内部記録媒体（104）若しくは外部記録媒体（108）などに出力される場合には、実質的に偏在性表示配列が格納されることを意味し得る。偏在性表示配列は、通信インターフェイス（105）を介してインターネット上に出力することもできる。出力対象は、ステップ203～213を繰り返すことにより得られた一連の偏在性表示配列、及びその中間データ並びに各種パラメータを含む全てのデータであり得る。本発明の偏在性表示配列は、当業界で一般的に使用されている混合塩基記号も含む塩基記号の一字表記形式に対応しているため、コンピュータ上で既存の解析ソフト等を使用する解析等に使用することができ、又はインターネット上の所望のサイトでの解析等に使用することができる。本発明の偏在性表示配列を核酸構成塩基の偏在性に関する解析に使用することより、解析対象の塩基配列情報を一塩基レベルで解析するのに比べて計算量を大幅に低減することができ、それゆえ解析時間を大幅に短縮することができる。

【0064】

さらに偏在性表示配列を使用する解析は、本発明に従って分節化開始塩基及び分節化塩基数を順次変更した一連の偏在性表示配列群を使用することにより、解析精度を上げることが可能である。以下に、図6を参照しながら分節化開始塩基を変更する場合について説明する。例えば、あるゲノムXに、配列Yが挿入されて、別のゲノムZとなっている場合、前記ゲノムZは、前記配列Yの開始点を基準として当該配列Yの塩基数分、ゲノムXからずれた配列を有するゲノムとなる（図6A）。次に、ゲノムX及びゲノムZを分節する場合について説明する。ゲノムX及びゲノムZについて同じ分節化開始塩基及び分節化塩基数を使用して分節化を行うと、ゲノムXについては分節1～6、ゲノムZについては分節I～VIIIが作成される（図6B）。ここで挿入配列Yは分節化塩基数とは異なる長さを有しているため、ゲノムZには、ゲノムXの分節3～6に対応する分節は存在しなくなる。その結果、ゲノムXの分節3～6に対応する領域は、ゲノムZ内に実際には存在しているにもかかわらず、ゲノムZの偏在性表示配列としては存在しなくなる。ここで、分節化開始塩基を配列Yの塩基数分だけ下流側にずらして分節化を実行すると、ゲノムXの分節3～6に対応する領域には、分節V～分節VIIIが該当することになる（図6B）。実際には、本発明の方法には核酸構成塩基の出現数に関して検定を行っているため、この分節化開始塩基の位置については、分節化塩基数に依存する関数で表され得るある程度の「ゆらぎ」が許容されると考えられる。すなわち、分節化に使用する塩基配列情報が同じである場合には、分節化開始塩基が対応箇所に完全に一致していなくても、当該対応領域において同じ偏在性表示配列が作成される蓋然性は高いと考えられる。

【0065】

次に分節化塩基数の変更が、本発明の偏在性表示配列を使用する解析の精度に影響を及ぼす可能性があることについて説明する。分節化に最適な分節化塩基数は、解析対象の塩基配列情報に含まれる塩基数と分節化塩基数とのバランスに従って決定され得る。すなわ

10

20

30

40

50

ち、塩基配列情報の総塩基数により近くなりすぎると偏在性表示配列が短くなりすぎ、当該塩基配列情報の特異性を反映しない配列となる傾向があり得る。一方、より1に近い数を分節化塩基数として設定すると偏在性表示配列が長くなりすぎて一塩基レベルでの情報に近くなり、本発明の効果が低くなる傾向があり得る。ゆえに、分節化塩基数は、解析対象とする塩基配列情報によって、適切な値が設定されてよい。偏在性表示配列を使用する比較解析を実行する場合、比較解析に使用される複数の偏在性表示配列は、分節化開始塩基については異なっていてよいが、分節化塩基数は同じであることが好ましい。このように、本発明に使用するパラメータを適宜変更することによって、複数の偏在性表示配列及び偏在性表示視覚化情報を使用する比較解析精度の向上を図ることができる。

一方、ステップ214において、偏在性表示視覚化情報を作成する場合には、ステップ217へと移行する。 10

【0066】

ステップ217は、偏在性表示配列から偏在性表示視覚化情報を作成するステップである。偏在性表示視覚化情報の作成は、偏在性表示塩基ごとに所定の色、パターン又はテクスチャを割り当てることによって実行されてよいが、各偏在性表示塩基ごとに異なる色、パターン又はテクスチャが割り当てられることを条件とする。偏在性表示塩基ごとに割り当てられる色、パターン又はテクスチャは既定であってよく、又は都合に合わせて任意に設定されてよい。偏在性表示視覚化情報の出力形態は、当該塩基配列情報の特性プロファイルに基づいて、直鎖状又は環状を含む形態であってよい。好ましくは、前記出力形態は、当該塩基配列情報の特性プロファイルに一致する形態で出力されてよい。偏在性表示塩基とパターンとの対応関係について、例示的实施態様を図5に示す。図5には、偏在性表示塩基 A、C、G、T、M、R、W、S、Y、K、V、H、D、B、及び N と、前記各偏在性表示塩基に割り当てられるパターンとの対応関係の一例を示している。当該偏在性表示視覚化情報は、環状ゲノムを有する生物種を塩基配列情報とした場合が想定されているため、環状で示されている。 20

【0067】

ステップ218は、偏在性表示視覚化情報を出力するステップである。偏在性表示視覚化情報の出力は、モニタ、ディスプレイ、プロジェクタ又はプリンタなどを含むが、これらに限定されない出力装置によってなされてよく、所定のデータベース、又は内部記録媒体(104)若しくは外部記録媒体(108)などに出力されてよい。ここで、所定のデータベース、又は内部記録媒体(104)若しくは外部記録媒体(108)などに出力される場合には、実質的に偏在性表示視覚化情報が格納されることを意味し得る。偏在性表示視覚化情報は、通信インターフェイス(105)を介してインターネット上に出力することもできる。出力対象は、ステップ203~213を繰り返すことにより得られた一連の偏在性表示視覚化情報、及びその中間データ並びに各種パラメータを含む全てのデータであり得る。偏在性表示視覚化情報は、同様に本発明に従って作成された他の偏在性表示視覚化情報とともに、同一画面上に出力又は表示することができる。この場合、当該複数の偏在性表示視覚化情報の分節化開始塩基及び分節化塩基数の少なくとも一方には、共通の数値を使用していることが好ましい。上記のように本発明に従うと、複数の偏在性表示視覚化情報間における核酸構成塩基の偏在性の類似性を容易にかつ的確に把握することができる。偏在性表示視覚化情報の作成にかかる分節化開始塩基及び分節化塩基数を変更することにより、核酸構成塩基の偏在性の類似性に関する解析精度の向上を図ることができる。複数の偏在性表示視覚化情報は、それぞれ所望のクラスターと関連付けて出力することができる。この場合に使用されるクラスターは、予め設定されていたクラスターであってよく、又は任意に設定されたクラスターであってもよい。クラスターは、単数又は複数であってよい。クラスターが複数である場合、それらのクラスター自身同士の包含関係も示されてよい。例えば、生物学的分類において、より下位に位置づけられている「種」の集合は、より上位に位置づけられている「属」の集合に包含されていることを示すように出力又は表示されてよい。 40

【0068】

次に、本発明による塩基配列情報に含まれる核酸構成塩基の偏在性を表す第2の実施態様を、図3を参照しながら説明する。ステップ301、302、303、304、305、306及び307については、それぞれステップ201、202、203、204、205、206及び207に実質的に対応するので、先に記載したステップ201～207の説明を参照されたい。

ステップ308は、当該各分節において、条件(B)：「当該核酸構成塩基の出現数が、その期待数よりも多い」を満たす核酸構成塩基が1つ以上存在するかを判断するステップである。前記ステップ308において、前記ステップ306で χ^2 値が所定の有意水準よりも大きいと判断された核酸構成塩基(群)の全ての出現数とその期待値以下であった場合、すなわち前記条件(A)を満たす核酸構成塩基(群)が前記条件(B)を満たさない場合にはステップ309に移行し、当該分節に偏在性表示塩基として「N」を割り当てられる。

10

【0069】

一方、前記ステップ306で χ^2 値が所定の有意水準よりも大きいと判断された全ての核酸構成塩基(群)の出現数とその期待値よりも大きい核酸構成塩基(群)が1つ以上存在する場合、すなわち前記条件(A)及び(B)の両方を満たす核酸構成塩基(群)が1つ以上存在する場合、ステップ310に移行する。前記ステップ310において、前記条件(A)及び(B)の両方を満たす核酸構成塩基(群)を、所定の対応表に従って、当該分節に偏在性表示塩基として割り当てる。例えば、前記条件(A)及び(B)の両方を満たす核酸構成塩基がAであった場合、当該分節には偏在性表示塩基「A」が割り当てられ；前記条件(A)及び(B)の両方を満たす核酸構成塩基群がA及びGであった場合、当該分節にはA及びGを意味する偏在性表示塩基「R」が割り当てられる。

20

【0070】

すなわち、本発明による、塩基配列情報に含まれる核酸構成塩基の偏在性を表す第2の実施態様は、当該分節において、(i)前記条件(A)及び(B)の両方を満たす核酸構成塩基(群)を、所定の対応表に従う偏在性表示塩基の作成対象とし；及び(ii)前記条件(A)及び(B)のうち少なくとも1つも満たさない場合には、偏在性表示塩基として「N」を割り当てる；実施態様である。

ステップ306～310で得られる全ての偏在性表示塩基及びその中間データ並びに各種パラメータを含む全てのデータは、所定のデータベース、又はメモリ(101)、内部記録媒体(104)若しくは外部記録媒体(108)などに格納することができる。

以降のステップ311、312、313、314、315、316、317、318及び319は、それぞれステップ211、212、213、214、215、216、217、218及び219に実質的に対応するので、先に記載したステップ211～219の説明を参照されたい。

30

【0071】

次に、本発明による塩基配列情報に含まれる核酸構成塩基の偏在性を表す第3の実施態様を、図4-1及び4-2を参照しながら説明する。ステップ401、402及び403については、それぞれステップ201、202及び203に実質的に対応するので、先に記載したステップ201～203の説明を参照されたい。

ステップ404は、各分節において、各核酸構成塩基の出現数を、当該出現数が多い方から順に第1、第2、第3及び第4頻出塩基として計数するステップである。例えば、当該分節の分節化塩基数が1000であって、Aの出現数が400、Cの出現数が200、Gの出現数が150、及びTの出現数が250である場合、当該分節の第1頻出塩基はAであり、第2頻出塩基はTであり、第3頻出塩基はCであり、及び第4頻出塩基はGである。異なる核酸構成塩基間で出現数が等しい値であった場合、当該出現数が等しい核酸構成塩基群は、混合塩基で表すことができる。例えば、当該分節の分節化塩基数が1000であって、Aの出現数が350、Cの出現数が200、Gの出現数が200、及びTの出現数が250である場合、当該分節の第1頻出塩基はAであり、第2頻出塩基は「S」(すなわち、C及びG)であり、第3頻出塩基はTであり得る。出現数が等しい核酸構成塩基群が3つ及び4つ(すなわち、全ての核酸構成塩基の出現数が等しい)であった場合にも同様に、頻出塩基は混合塩基として処理することができる。

40

【0072】

50

ステップ405は、各分節において、当該分節における第1頻出塩基の出現数を使用して、第1仮説を「当該分節において、第1頻出塩基の出現数はその期待値に等しい」とする第1仮説検定を実行するステップである。ステップ405を含む以下の全ての仮説検定には、任意の有意水準及び任意の自由度を使用する、正規分布表、t分布表、F分布表、又はカイニ乗分布表を含むがこれらに限定されない分布表を使用する仮説検定が包含され得る。ここでは、有意水準0.05、自由度1のカイニ乗分布表を使用する場合、すなわちカイニ乗値が3.841の場合を例に挙げて説明する。はじめに、前記核酸構成塩基出現数を使用して、各分節における各核酸構成塩基の期待数 E_A 、 E_T 、 E_G 及び E_C を、下記式に従って算出する。

【数7】

$$\left\{ E_A = \frac{0.25 \times (100 - p)}{50} L, E_T = \frac{0.25 \times (100 - p)}{50} L, E_G = \frac{0.25 \times p}{50} L, E_C = \frac{0.25 \times p}{50} L \right\}$$

10

(式中、 E_A 、 E_T 、 E_G 及び E_C は、それぞれ、当該分節における核酸構成塩基A、T、G及びCの期待数を示し；PはGC含量(%)を示し；及び、Lは当該分節内の塩基の総数；を示す)。

【0073】

次に、第n頻出塩基の χ^2 値を算出するための一般式を下記に示す。

【数8】

$$\chi_{FBn}^2 = \frac{(O_{FBn} - E_{FBn})^2}{E_{FBn}} + \frac{(O_{FBnR} - E_{FBnR})^2}{E_{FBnR}} = \frac{(O_{FBn} - E_{FBn})^2}{E_{FBn}} + \frac{1}{L - E_{FBn}} [(L - O_{FBn}) - (L - E_{FBn})]^2$$

20

(式中、 $_{FBn}$ は第n頻出塩基を示し、 $_{FBnR}$ は前記 $_{FBn}$ 以外の全ての核酸構成塩基群を示し、 χ_{FBn}^2 は当該分節における前記第n頻出塩基の χ^2 値を示し、 O_{FBn} は当該分節における前記第n頻出塩基の出現数を示し、 O_{FBnR} は当該分節における前記第n頻出塩基以外の全ての核酸構成塩基群の出現数の合計を示し、及びLは当該分節内の塩基の総数を示す)。

【0074】

次にステップ406において、各分節の前記第1頻出塩基の χ^2 値である χ_{FB1}^2 が、所定の有意水準のカイニ乗値、すなわちここでは3.841以下であるかを判定する。全ての分節において、第1頻出塩基の出現数は、核酸構成塩基4種類のうち、最も出現数の多い核酸構成塩基(群)の出現数である。それゆえ、第1頻出塩基の出現数は、必然的に、その期待値に等しいか又はその期待値よりも大きい値であり得る。ゆえに、前記判定において、 χ_{FB1}^2 が3.841より大きい場合、前記第1仮説は棄却され、当該第1頻出塩基の出現数は当該分節において有意に出現数が多いと判断される。その結果として、前記第1頻出塩基に対応する偏在性表示塩基が当該分節に割り当てられる(ステップ407)。一方、前記判定において、 χ_{FB1}^2 が3.841以下である場合には、前記第1仮説は棄却されず、ステップ408(ステップ4081)へと移行する。

30

【0075】

ステップ4082は、各分節において、当該分節における第1及び第2頻出塩基の出現数の合計値を使用して、第2仮説を「当該分節において、第1及び第2頻出塩基の出現数の合計はその期待値に等しい」とする第2仮説検定を実行するステップである。ステップ4082では、前記第n頻出塩基の χ^2 値を算出するための一般式に従って χ_{FB1}^2 及び χ_{FB2}^2 を算出し、 $\chi_{FB1}^2 + \chi_{FB2}^2$ と3.841との大小関係を比較する判定を実行する。前記判定において、 $\chi_{FB1}^2 + \chi_{FB2}^2$ が3.841より大きい場合、前記第2仮説は棄却され、当該第1及び第2頻出塩基の出現数は当該分節において有意に出現数が多いと判断される。その結果として、前記第1及び第2頻出塩基に対応する偏在性表示塩基が当該分節に割り当てられる(ステップ4084)。一方、前記判定において、 $\chi_{FB1}^2 + \chi_{FB2}^2$ が3.841以下である場合には、前記第2仮説は棄却されず、ステップ4085へと移行する。

40

【0076】

ステップ4085は、各分節において、当該分節における第1、第2及び第3頻出塩基の出現

50

数の合計値を使用して、第3仮説を「当該分節において、第1、第2及び第3頻出塩基の出現数の合計はその期待値に等しい」とする第3仮説検定を実行するステップである。ステップ4085では、前記第n頻出塩基の χ^2 値を算出するための一般式に従って χ_{FB1}^2 、 χ_{FB2}^2 及び χ_{FB3}^2 を算出し、 $\chi_{FB1}^2 + \chi_{FB2}^2 + \chi_{FB3}^2$ と3.841との大小関係と比較する判定を実行する。前記判定において、 $\chi_{FB1}^2 + \chi_{FB2}^2 + \chi_{FB3}^2$ が3.841より大きい場合、前記第3仮説は棄却され、当該第1、第2及び第3頻出塩基の出現数は当該分節において有意に出現数が多いと判断される。その結果として、前記第1、第2及び第3頻出塩基に対応する偏在性表示塩基が当該分節に割り当てられる（ステップ4084）。一方、前記判定において、 $\chi_{FB1}^2 + \chi_{FB2}^2 + \chi_{FB3}^2$ が3.841以下である場合には、当該分節においては全ての核酸構成塩基の出現数に有意差はないと判定し、当該分節に偏在性表示塩基として「N」を割り当てる（ステップ4088）。

10

【0077】

前記第1、第2及び第3仮説検定を実行することにより各分節に割り当てられた偏在性表示塩基は、それぞれ次のステップ409で、入力された塩基配列情報の並びに従って偏在性表示配列として整列される。ステップ409、410、411、412、413、414、415、416及び417は、それぞれステップ211、212、213、214、215、216、217、218及び219に実質的に対応するので、先に記載したステップ211～219についての説明を参照されたい。

【0078】

（本発明において使用するパラメータ）

分節化開始塩基は、塩基配列情報の起源、長さ、及び比較解析における比較対象との関連などに応じて、最適化を図ることができる。分節化開始塩基は、既定値として予め設定された数値であってよく、又は入力手段で入力された数値であってもよい。本発明のある実施態様において、分節化開始塩基は既定値で1である。分節化開始塩基は、所定の間隔で、第1分節化開始塩基から上流又は下流に順次変更することができる。当該所定の間隔は、入力される塩基配列情報に応じて最適化を図ることができる。また、同一の塩基配列情報から所定の間隔で分節化開始塩基を変更した一連の偏在性表示配列を作成することができ、このような一連の偏在性表示配列は、プログラムのコンピュータに実行させることができる。

20

【0079】

分節化開始塩基は、塩基配列情報の機能、又は染色体上の位置に応じて決定されてよい。本発明の特定の実施態様において、分節化開始塩基は、以下は限定の意図のない例示であるが：塩基配列情報中に含まれる複製開始点、転写開始点、又は翻訳開始点に対応する数値；若しくは、染色体の第1番目の塩基に対応する数値；などであってよい。

30

分節化開始塩基は、本発明の比較解析に使用される偏在性表示配列間で同じでよい。分節化開始塩基は、本発明の比較解析に使用される偏在性表示配列間で異なっていてよい。分節化開始塩基は、本発明のクラスタリングに使用される偏在性表示視覚化情報間で同じでよい。分節化開始塩基は、本発明のクラスタリングに使用される偏在性表示視覚化情報間で異なっていてよい。

【0080】

本発明に使用する分節化塩基数は、塩基配列情報の起源、長さ、及び比較解析における比較対象との関連などに応じて、最適化を図ることができる。分節化塩基数は、既定値として予め設定された数値であってよく、又は入力手段で入力された数値であってもよい。分節化塩基数は、入力される塩基配列情報によって最適化を図ることができる。分節化塩基数は、本発明の比較解析に使用される偏在性表示配列間で同じでよい。分節化塩基数は、本発明の比較解析に使用される偏在性表示配列間で異なっていてよい。分節化塩基数は、本発明のクラスタリングに使用される偏在性表示視覚化情報間で同じでよい。分節化塩基数は、本発明のクラスタリングに使用される偏在性表示視覚化情報間で異なっていてよい。しかしながら、本発明に従って、複数の偏在性表示配列又は複数の偏在性表示視覚化情報を使用して解析を行う場合には、当該複数の偏在性表示配列又は複数の偏在性表示視覚化情報に使用される分節化塩基数は、それぞれ同じであることが好ましい。

40

50

【0081】

本発明に使用するP値は、既定値として予め設定された数値であってよく、又は入力手段で入力された数値であってもよい。P値は、当該塩基配列のGC含量と同一の数値であってよい。P値には、特性プロファイルに基づき、当該塩基配列のGC含量と同一の数値が自動的に又は手動で割り当てられてよい。P値は、既定値で50、又は他の任意の数値であってよい。P値は、本発明の比較解析に使用される複数の偏在性表示配列の間で同じでよい。P値は、本発明の比較解析に使用される複数の偏在性表示配列の間で異なっていてよい。P値は、本発明のクラスタリングに使用される複数の偏在性表示配列の間で同じでよい。P値は、本発明のクラスタリングに使用される複数の偏在性表示配列の間で異なっていてよい。

10

【0082】

本発明に使用する検定の有意水準は、塩基配列情報の起源、長さ、及び比較解析における比較対象との関連などに応じて、最適化を図ることができる。有意水準は、既定値として予め設定された数値であってよく、又は入力手段で入力された数値であってもよい。有意水準は、本発明に使用される、統計学的に使用することができる分布表において許容され得る任意の数値であってよい。本発明のある実施態様において、有意水準は既定値で0.25～0.005の範囲に含まれる任意の値であり得る。本発明の別の実施態様において、有意水準は、既定値で0.25、0.2、0.15、0.1、0.05、0.025、0.01、又は0.005であり得、好ましい実施態様において0.05であり得る。有意水準は、本発明の比較解析に使用される複数の偏在性表示配列の間で同じでよい。有意水準は、本発明の比較解析に使用される複数の偏在性表示配列の間で異なっていてよい。有意水準は、本発明のクラスタリングに使用される複数の偏在性表示配列の間で同じでよい。有意水準は、本発明のクラスタリングに使用される複数の偏在性表示配列の間で異なっていてよい。しかしながら、本発明に従って、複数の偏在性表示配列又は複数の偏在性表示視覚化情報を使用して解析を行う場合には、当該複数の偏在性表示配列又は複数の偏在性表示視覚化情報に使用される有意水準は、それぞれ同じであることが好ましい。

20

本発明の検定に使用される自由度は、当該検定に使用され得る任意の数値であり得る。自由度は、本発明に使用する検定の種類などに従って任意に設定できる。

本発明に使用される検定は、任意に、片側検定又は両側検定であってよい。

30

【0083】

偏在性表示配列は、偏在性表示配列を作成した元の塩基配列情報の特性プロファイルとともに出力されてよい。出力される特性プロファイルは、塩基配列情報に含まれる塩基配列以外の情報、すなわち、当該塩基配列を構成する塩基数(L値)、GC含量(P値)、並びに当該塩基配列に付随する様々な情報、例えば、当該塩基配列を定義するための名前及びID、当該核酸の起源生物、当該起源生物の生物学的分類、遺伝子座、並びに当該核酸に含まれる遺伝子名を含むがこれらに限定されない情報セットの全て又はその部分的セットであってよい。出力される特性プロファイルの情報項目は、予め設定されていてもよく、又は入力手段で任意に設定されてもよい。偏在性表示配列情報は、XML形式、HTML形式、Gen Bank形式、EMBL形式、SwissProt形式、FASTA形式、PIR形式、GCG形式、GDE形式、AceDB形式、GFF形式、及び任意のテキスト形式、並びにその他の形式を含むがこれらに限定されない、任意の形式で出力されてよい。

40

【0084】

偏在性表示配列の比較解析は、作成された偏在性表示配列の内容、長さ、及び比較対象との関連などに応じて、最適化を図ることができる。偏在性表示配列の比較解析には、ドットマトリクス法、ダイナミックプログラミング法、又はk-タプル法に基づく方法又はアルゴリズムなどを含むが、これらに限定されない方法又はアルゴリズムが使用可能である。好ましい実施態様において、偏在性表示配列の比較解析には、BLAST、FASTA、DNAsis(登録商標)、又はGENETYX(登録商標)などを含むが、これらに限定されないプログラム又はソフトウェアが使用可能である。

50

【0085】

偏在性表示視覚化情報の作成において、偏在性表示配列への割り当てに使用される色、パターン又はテクスチャは、偏在性表示配列に含まれる偏在性表示塩基の構成及びその比率、並びに比較対象との関連などに応じて、最適化を図ることができる。好ましい実施態様において、偏在性表示配列への割り当てに使用される色は、コンピュータで出力可能な色である。別の好ましい実施態様において、偏在性表示配列への割り当てに使用される色は、Photoshop（登録商標）などを含むがこれに限定されない任意のソフトウェアで出力可能な色である。好ましい実施態様において、偏在性表示配列への割り当てに使用されるパターンは、コンピュータで出力可能なパターンである。別の好ましい実施態様において、偏在性表示配列への割り当てに使用されるパターンは、Photoshop（登録商標）などを含むがこれに限定されない任意のソフトウェアで出力可能なパターンである。好ましい実施態様において、偏在性表示配列への割り当てに使用されるテクスチャは、コンピュータで出力可能なテクスチャである。ある実施態様において、偏在性表示配列への割り当てに使用されるテクスチャは、別の好ましいPhotoshop（登録商標）などを含むがこれに限定されない任意の任意のソフトウェアで出力可能なテクスチャである。

10

20

30

40

50

【0086】

偏在性表示視覚化情報は、偏在性表示視覚化情報を作成した元の塩基配列情報の特性プロフィールとともに出力されてよい。出力される特性プロフィールは、塩基配列情報に含まれる塩基配列以外の情報、すなわち、当該塩基配列を構成する塩基数（L値）、GC含量（P値）、並びに当該塩基配列に付随する様々な情報、例えば、当該塩基配列を定義するための名前及びID、当該核酸の起源生物、当該起源生物の生物学的分類、遺伝子座、並びに当該核酸に含まれる遺伝子名を含むがこれらに限定されない情報セットの全て又はその部分的セットであってよい。出力される特性プロフィールの情報項目は、予め設定されていてもよく、又は入力手段で任意に設定されてもよい。偏在性表示視覚化情報は、偏在性表示塩基と、該偏在性表示塩基に割り当てられた色、パターン又はテクスチャとの対応関係を示す凡例と共に出力されてもよい。

【0087】

偏在性表示視覚化情報のクラスタリングに使用するクラスターは、元の塩基配列情報の起源、長さ、及び比較解析における比較対象との関連などに応じて、最適化を図ることができる。クラスタリングに使用可能なクラスターは、限定の意図なく例示すると：特定の界、門、綱、目、科、属、及び種、並びに各々の中間的又は細分化された既存の生物学的分類；特定の病原性、及び発酵能などを含むが、これらに限定されない表現型；海水、淡水、高温、低温、高塩濃度、低塩濃度などを含むが、これらに限定されない生育環境；任意の分類カテゴリー、例えば、細菌、酵母、ウイルス、真核生物、原核生物、単細胞生物、多細胞生物、及び哺乳動物などの分類カテゴリーを含むが、これらに限定されない任意のカテゴリー；である。

本明細書に記載した本発明の方法の説明は、本発明のプログラム及び本発明の装置に使用される方法又は手段についても実質的に同様に適用することができるので、当業者は本明細書の説明により本発明のプログラム及び装置についても実質的な理解を得られるであろう。

【実施例】

【0088】

以下に記載する本発明の実施例は、本発明の特許請求の範囲に関する理解を深めるために記載しているものであり、本発明の特許請求の範囲を限定することを意図するものではない。本明細書に記載の特許請求の範囲を逸脱しない範囲において、本発明の主題を達成し得る様々な態様、修飾、及び変更が可能であることは、当業者に理解されるであろう。

【0089】

本発明の具体的実施態様を、以下の実施例で説明する。本実施例における記載は、本発明の範囲を限定する意図はなく、本発明を説明するための例示にすぎないことは理解されるべきである。なお、以下の実施例に使用したPC環境は、特に他に記載がない限り、Apple社製Mac Pro 2×2.66 GHz Dual-Core Intel Xeon、OS：MacOS（登録商標）10.5.2であり

、偏在性表示配列の作成にはPerl : 5.8.8 (Mac OS (登録商標) 10.5.2に標準搭載) を、及び偏在性表示視覚化情報の作成にはJava (登録商標) を使用した。また、以下の実施例に使用した塩基配列情報は、特に他に記載がない限り、日本時間2008年5月28日にNCBIのftpサーバー (ftp://ftp.ncbi.gov.genomes/) からダウンロードした、GeneBank形式のBacteriaゲノム配列がZip圧縮されているファイル (all.gb.k.tar.gz) に収録されていた配列である。

【 0 0 9 0 】

(実施例1) : 大腸菌K12株完全ゲノム配列からの偏在性表示配列の作成

本発明のプログラムを備える装置を使用して、大腸菌K12株完全ゲノム配列 (Escherichia coli K12 substr. MG1655, complete genome (ACCESSION : Acc# NC_000913 ; VERSION : NC_000913.2, GI:49175990)) (配列番号 : 1) から偏在性表示配列の作成を実行した。ダウンロードした塩基配列情報は、内部記憶媒体 (HDD) に保存した。当該塩基配列情報は、4,639,675塩基対から構成されており、混合塩基は存在しなかった。当該塩基配列情報を、分節化開始塩基1、分節化塩基数1,000、及びP=50で分節化し、各分節内の核酸構成塩基を計数した。以下、計数領域1-1000からなる分節を F_{1-1000} と表記し、他の分節も同様に、計数領域に対応する数値範囲をFの右下に付して表記した : ここで、「F」は分節 (fragment) を意味し、「F」の後の数字範囲は、当該分節を構成する塩基配列情報中の計数領域に対応している。前記処理により得られた、各分節内に含まれる核酸構成塩基の出現数を下記の表に示す。

【表 6】

分節	Aの出現数	Tの出現数	Gの出現数	Cの出現数
F_{1-1000}	258	235	253	254
$F_{1001-2000}$	226	242	278	254
$F_{2001-3000}$	226	255	287	232
:	:	:	:	:
$F_{4638001-4639000}$	291	286	200	223
$F_{4639001-4639675}$	181	163	180	151

【 0 0 9 1 】

当該 F_{1-1000} において、Aは258回、Tは235回、Gは253回、及びCは254回計数された、同様に、 $F_{1001-2000}$ において、Aは226回、Tは242回、Gは278回、及びCは254回計数され ; $F_{2001-3000}$ において、Aは226回、Tは255回、Gは287回、及びCは232回計数され ; $F_{3001-4000} \sim F_{4637001-4638000}$ についても同様に計数を実行し (表中では省略) ; $F_{4638001-4639000}$ において、Aは291回、Tは286回、Gは200回、及びCは223回計数された。最終末端に位置する $F_{4639001-4639675}$ は、分節内塩基数が675残基であり、Aは181回、Tは163回、Gは180回、及びCは151回計数された。

【 0 0 9 2 】

次に、各々の分節内において、各核酸構成塩基のカイ二乗値を算出した。ここでは例として、 F_{1-1000} について記載する。 F_{1-1000} における各塩基のカイ二乗値は、下記式に従って算出した。

【数 9】

$$\left\{ \begin{array}{l} \chi_{A_{1-1000}}^2 = \frac{(258-250)^2}{250} + \frac{1}{1000-250} [(1000-258)-(1000-250)]^2 = 0.341 \\ \chi_{T_{1-1000}}^2 = \frac{(235-250)^2}{250} + \frac{1}{1000-250} [(1000-235)-(1000-250)]^2 = 1.200 \\ \chi_{G_{1-1000}}^2 = \frac{(253-250)^2}{250} + \frac{1}{1000-250} [(1000-253)-(1000-250)]^2 = 0.048 \\ \chi_{C_{1-1000}}^2 = \frac{(254-250)^2}{250} + \frac{1}{1000-250} [(1000-254)-(1000-250)]^2 = 0.085 \end{array} \right\} \quad 10$$

同様に、各分節において、それぞれの核酸構成塩基のカイ二乗値を算出した結果を下記の表に示す。

【表 7】

分節	Aの χ^2 値	Tの χ^2 値	Gの χ^2 値	Cの χ^2 値
F ₁₋₁₀₀₀	0.341	1.2	0.048	0.085
F ₁₀₀₁₋₂₀₀₀	3.072	0.341	4.181	0.085
F ₂₀₀₁₋₃₀₀₀	2.981	0.661	6.565	2.325
⋮	⋮	⋮	⋮	⋮
F ₄₆₃₈₀₀₁₋₄₆₃₉₀₀₀	8.965	6.912	13.333	3.888
F ₄₆₃₉₀₀₁₋₄₆₃₉₆₇₅	1.186	0.261	1	2.489

20

【0093】

30

次に、各分節内において、有意に出現数の多かった核酸構成塩基（群）を決定するための検定を実行した。本実施例では、図2のフローチャートに記載の方法を使用した。具体的には、各分節における各核酸構成塩基について、次の2ステップの検定を実行した：当該分節において、条件(A)を満たす核酸構成塩基が1つ以上存在することを判断するステップであって、前記条件(A)が「当該核酸構成塩基の χ^2 値が、所定の有意水準のカイ二乗値より大きい」である、前記ステップ；及び条件(B)を満たす核酸構成塩基が1つ以上存在することを判断するステップであって、前記条件(B)が「当該核酸構成塩基（群）の出現数が、その期待数よりも多い」である、前記ステップ。当該各分節への偏在性表示塩基の割り当ては、下記判断に従った：全ての核酸構成塩基について条件(A)が満たされない場合、当該分節に偏在性表示塩基として「N」を割り当て；条件(A)を満たすが条件(B)を満たさない核酸構成塩基（群）が存在する場合には、当該核酸構成塩基（群）以外の全てからなる核酸構成塩基（群）を、本明細書に記載の対応表に従って、当該分節に偏在性表示塩基として割り当て；及び、条件(A)を満たしかつ条件(B)も満たす核酸構成塩基（群）が存在する場合には、当該核酸構成塩基（群）を、本明細書に記載の対応表に従って、当該分節に当偏在性表示塩基として割り当てた。加えて、以下の実施例において、有意水準は0.05及び自由度1に対応するカイ二乗値3.841を使用した。

40

【0094】

はじめに、F₁₋₁₀₀₀を例に挙げ解説する。F₁₋₁₀₀₀において：(A) 当該分節の全ての核酸構成塩基の χ^2 値は、所定の有意水準3.841以下であったため、当該分節には偏在性表示塩基として「N」を割り当てた。次に、F₁₀₀₁₋₂₀₀₀において、(A) 核酸構成塩基Gの χ^2

50

値は、所定の有意水準3.841のカイ二乗値より大きく、かつ（B）前記核酸構成塩基Gの出現数はその期待数よりも多いので、当該分節には偏在性表示塩基として「G」を割り当てた。 $F_{2001-3000}$ において、（A）核酸構成塩基Gの χ^2 値は、所定の有意水準3.841のカイ二乗値より大きく、かつ（B）前記核酸構成塩基Gの出現数はその期待数よりも多いので、当該分節には偏在性表示塩基として「G」を割り当てた。他の分節についても同様に検定を実施した（表中では結果は省略されている）。 $F_{4638001-4639000}$ において、（A）当該分節の全ての核酸構成塩基の χ^2 値は、所定の有意水準3.841のカイ二乗値より大きかったが、（B）出現数がその期待数よりも多かったのは核酸構成塩基A及びTのみであったので、当該分節には偏在性表示塩基として「W」を割り当てた。最後に、 $F_{4639001-4639675}$ において、（A）当該分節の全ての核酸構成塩基の χ^2 値は、所定の有意水準のカイ二乗値3.841以下であったので、当該分節には偏在性表示塩基として「N」を割り当てた。上記検定に従って各々の分節で得られた偏在性表示塩基を前記塩基配列情報の並びに従って並べた偏在性表示配列を図7に示す。図7から明らかなように、4,639,675塩基対から構成される *Escherichia coli* K12 substr. MG1655, complete genome (ACCESSION: Acc# NC_000913; VERSION: NC_000913.2, GI:49175990) (配列番号: 1) は、本発明により、偏在性表示塩基4640残基からなる偏在性表示配列に置換することができた。

【0095】

（実施例2）：分節化開始塩基を変更した場合における、偏在性表示配列及び偏在性表示視覚化情報パターンの変化

図8は、*Escherichia coli* K12 substr. MG1655について、分節化塩基数30,000及び $P=50$ で、分節化開始塩基を3,000づつ順次変更した偏在性表示配列を示す。図中、記号A～Jで表されている偏在性表示視覚化情報は、それぞれ分節化開始塩基が1 (A)、3,001 (B)、6,001 (C)、9,001 (D)、12,001 (E)、15,001 (F)、18,001 (G)、21,001 (H)、24,001 (I)、及び27,001 (J) の場合に対応している。これらの記号A～Jは、図9の偏在性表示視覚化情報にも対応している。図9には、前記偏在性表示配列を使用して作成された偏在性表示視覚化情報を示している。図8及び9の結果から、分節化開始塩基を変更することによって、分節化塩基数が同じであるにもかかわらず、異なるパターンの偏在性表示配列及び偏在性表示配列が得られることが示された。この結果は、本明細書の先に記載したように、本発明に使用するパラメータを適宜変更することが、複数の偏在性表示配列及び偏在性表示視覚化情報間での比較解析における精度を向上させ得ることを示していると考えられる。

【0096】

（実施例3）：複数のゲノム塩基配列情報からの偏在性表示配列及び偏在性表示視覚化情報の作成

実施例3では、本発明に従って18種類のゲノム塩基配列情報から、本発明の偏在性表示配列及び偏在性表示視覚化情報を作成した。当該実施例において、図10は、前記18種類のゲノム塩基配列情報に関する特性プロファイルの一部である、配列名（ゲノム名）、アクセス番号、GC含量、及び（既存の分類学的）分類を示し；図11は、前記18種類の塩基配列情報について、それぞれ本発明に従って、分節化開始塩基1かつ分節化塩基数30,000で作成した偏在性表示配列を示し；図12は、前記18種類の偏在性表示配列を使用して作成した偏在性表示視覚化情報（ $P=50$ ）；及び図13は、前記18種類の偏在性表示配列を使用して作成した偏在性表示視覚化情報（ $P=$ 当該生物の塩基配列情報のGC含量（%））を示している。

【0097】

実施例3では、*Escherichia coli* K12 substr. MG1655, complete genome (配列番号: 1) (図中、番号1) を基準的な解析対象とした。前記18種の塩基配列情報について：番号2及び3の塩基配列情報は、前記番号1の *Escherichia coli* K12 substr. MG1655 と同属 (*Escherichia* 属) であり；番号4～6の塩基配列情報は、前記番号1と同科別属 (*Shigella* 属) であり；番号7～9の塩基配列情報は、前記番号1と同科別属 (*Salmonella* 属) であり；番号10～12の塩基配列情報は、前記番号1とGC含量が類似している別の網に分類されている

生物由来であり；番号13～15の塩基配列情報は、前記番号1よりGC含量が高い生物由来であり；及び、番号16～18の塩基配列情報は、前記番号1よりGC含量が低い生物由来である。図10～13において、記載されている番号が同じ場合、該同じ番号の塩基配列情報、偏在性表示配列及び偏在性表示視覚化情報は、同一の塩基配列情報由来であることを示す。なお、図12及び13に示されている偏在性表示配列と偏在性表示視覚化情報との塩基ごとの対応関係は、図5に記載の凡例に従っている。

【0098】

図10は、ゲノム塩基配列情報及びその特性プロファイルの一部である、配列名（ゲノム名）、アクセス番号、GC含量、及び（既存の分類学的）分類を示している。本発明の偏在性表示配列及び偏在性表示視覚化情報の作成に使用した塩基配列は、添付の配列表に、配列番号：1～18として示している。図11は、前記18種の塩基配列情報を使用して作成した偏在性表示配列を示している。図11において、偏在性表示配列は、各番号の塩基配列情報について、各塩基の出現頻度が均一であるとして検定を行った場合（すなわち、 $P=50$ ）、及び各塩基の出現頻度をゲノムのGC含量として検定を行った場合（すなわち、 P は、それぞれ当該生物の塩基配列情報のGC含量（％）に等しい）の両方について示している。各生物に固有のGC含量が50％よりも離れているほど、 P 値の設定条件の変更が、本発明に従って作成される偏在性表示配列及び偏在性表示視覚化情報に及ぼす影響が大きい傾向があることが示された（図11、12及び13）。

【0099】

図12及び13は、図11に示した偏在性表示配列群に、図5の凡例に示すパターンをそれぞれ当てはめ、それぞれのゲノムの形状である環状で表記した図である。図12は P 値が50の場合、及び図13は P 値がそれぞれの塩基配列情報のGC含量である場合の結果を示している。以下、特に明記しない限り、図12及び13の両方を参照しながら説明する。

Escherichia coli K12 substr. MG1655（図中、番号1）と同属である（それゆえ、全体的なゲノム塩基配列情報の類似性が高い）*Escherichia coli* str. K-12 substr. DH10B（図中、番号2）及び*Escherichia coli* O157:H7 str. Sakai（図中、番号3）は、互いにきわめて類似したパターンを示した。この結果により、本発明の偏在性表示視覚化情報及び偏在性表示配列が、既存の技術では困難であったゲノム-ゲノム間の全体的解析を比較的容易に実行しかつ出力し得ること、及びその評価がきわめて容易であり得ることが示された。

【0100】

一方、*Shigella flexneri* 2a str. 2457T（図中、番号4）、*Shigella flexneri* 5 str. 8401（図中、番号5）及び*Shigella sonnei* Ss046（図中、番号6）の偏在性表示視覚化情報は、当該偏在性表示視覚化情報間においてパターンが類似していたのみならず、番号1～3の偏在性表示視覚化情報ともかなり類似したパターンを示した。すなわち、本発明によると、番号1～6の塩基配列情報は、それぞれが互いに全体的にかなり類似しているという結果が得られた。この結果については、当業界において*Escherichia*属と*Shigella*属とは同属とみなすべきであるという議論が存在すること（例えば、Johnson, J.の論文『分岐点におけるシゲラ及び大腸菌：マキアベリ的マスカレーダーか、分類学的反逆か？（*Shigella and Escherichia coli at the crossroads: Machiavellian masqueraders or taxonomic treachery?*）』J. Med. Microbiol. 2000 49: 583-585.；Wang L, Qu W, Reeves PR.の論文『4種のシゲラ・ボイディイのO抗原遺伝子座の配列解析：大腸菌とシゲラとの密接な関連性（Sequence analysis of four *Shigella boydii* O-antigen loci: implication for *Escherichia coli* and *Shigella* relationships.）』Infect Immun. 2001, 69: 6923-30.；及び、Pupo, G.M., Lan, R. 及びReeves, P.R.の論文『大腸菌のシゲラクロンの複数の独立複製開始点、及び多くのそれらの特徴の収束進化（Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics.）』Proc. Natl Acad. Sci. USA, 2000, 97: 10567-10572.；を参照されたい）を考慮すると、本発明は、既存の生物学的分類法では分類困難な生物種を区別することに重要な知見を提供し得る可能性がある。

【0101】

さらに、*Salmonella enterica* subsp. *enterica* serovar Paratyphi B str. SPB7 (図中、番号7)、*Salmonella enterica* subsp. *enterica* serovar Typhi Ty2 (図中、番号8)、及び*Salmonella typhimurium* LT2 (図中、番号8) の偏在性表示視覚化情報も、それらの偏在性表示視覚化情報パターンが互いに類似しており、かつ番号1~6の偏在性表示視覚化情報にもある程度の類似性を示した。また、番号1~3由来の偏在性表示視覚化情報に対する類似性は、番号7~9由来の偏在性表示視覚化情報よりも、番号4~6由来の偏在性表示視覚化情報の方がより類似しているようであった。これらの結果から、本発明によると、既存の生物学的分類において同じエンテロバクター科に含まれる番号1~9の生物のうち、*Escherichia*属及び*Shigella*属がより比較的近縁であり得ること、並びに*Salmonella*属は前記*Escherichia*属及び*Shigella*属とはより比較的遠縁であり得ることが示された。(Lawrence, J.G. 及び Ochman, H. の論文 (1998) 『大腸菌ゲノムの分子考古学 (Molecular archaeology of the *Escherichia coli* genome.) 』Proc. Natl Acad. Sci. USA, 95, 941394 17 ; Reid SD, Herbelin CJ, Bumbaugh AC, Selander RK, Whittam TS. の論文 『病原性大腸菌における毒性因子の平行進化 (Parallel evolution of virulence in pathogenic *Escherichia coli*.) 』Nature. 2000 406: 64-7. ; 及び、Naimuddin M, Kurazono T, Nishigaki K. の論文 『ゲノムプロファイリングによって示された共通保存的遺伝子片は、進化のトレーサーとして扱うことができる (Commonly conserved genetic fragments revealed by genome profiling can serve as tracers of evolution.) 』Nucleic Acids Res. 2002 30: e42. ; を参照されたい)。

10

20

上記結果から、本発明によってエンテロバクター科に含まれる番号19の生物のうち、*Escherichia*属及び*Shigella*属は類縁性が非常に高いこと、並びに*Salmonella*属も*Escherichia*属及び*Shigella*属と近縁であり得ることが示され、これらの結果は、当業界で示されているこれらの属の類縁性と同等の結果であった。

【0102】

番号10~12の塩基配列情報は、番号1の塩基配列情報に比較してGC含量が近いものの、網レベルで異なる生物由来の塩基配列情報である。本発明に従って作成された番号10~12の塩基配列情報由来の偏在性表示視覚化情報のパターンはいずれも、番号1の偏在性表示視覚化情報のパターンとはいずれも異なっていることが示された。この結果は既存の生物学的分類とよい相関を示しており、本発明の偏在性表示配列及び偏在性表示視覚化情報は、少なくとも網レベルでの生物学的分類に寄与し得ることが示された。

30

【0103】

番号13~18の塩基配列情報についても、番号1の塩基配列情報とは網レベルで異なり、その偏在性表示視覚化情報のパターンはいずれも番号1の偏在性表示視覚化情報のパターンとは異なっていることが示された。それゆえ、この結果も既存の生物学的分類とよい相関を示し、本発明の偏在性表示配列及び偏在性表示視覚化情報は、少なくとも網レベルでの生物学的分類に寄与し得ることがさらに示された。

【0104】

GC含量と偏在性表示視覚化情報パターンとの関連性について、GC含量が50から離れている番号13~18の塩基配列情報から偏在性表示視覚化情報を作成する際にP値を50に設定したところ、P値を当該GC含量に設定した場合に比較して、当該偏在性表示視覚化情報パターンは単調になる傾向があることが示された (図12及び13)。この結果は、GC含量に差のある塩基配列情報間での比較解析を実行する際には、P値を当該生物のGC含量に合わせた設定で実行する方が好ましい可能性があることを示唆している可能性がある。

40

上記のように、本発明の偏在性表示配列及び偏在性表示視覚化情報を使用して得られた近縁関係の比較解析結果は、既存の生物学的分類とかなりよい相関関係を示し得ることが実証された。それゆえ、本発明による塩基配列情報に含まれる (核酸構成) 塩基の偏在性を表す方法は、生物学的分類において、有用な情報を提供し得ることが示された。

【図面の簡単な説明】

【0105】

50

【図1】本発明の装置の構成図。図中、メモリ(101)、CPU(102)、入力装置(103)、内部記録媒体(104)、通信インターフェイス(105)、及び出力装置(107)はバス(100)を介して互いに接続されている。インターネット(106)は、本発明の装置の通信インターフェイス(105)を介して任意に接続されている。外部記録媒体(108)は、本発明の装置のバス(100)を介して任意に接続されている。

【図2】検定にカイ二乗値を使用する、本発明の方法のフローチャートを示す。

【図3】検定にカイ二乗値を使用する、本発明の別の方法のフローチャートを示す。

【図4】検定に仮説検定を使用する、本発明の方法のフローチャートを示す。

【図5】偏在性表示視覚化情報の作成における、偏在性表示塩基とパターンとの相関関係の具体的実施態様を示す。図5は、以下の図9、12及び13における偏在性表示視覚化情報の凡例としても参照される。

10

【図6】分節化開始塩基の変更と、分節の対応関係との関連性についての説明図である。A:ゲノムZは、分節化塩基数とは異なる長さを有する配列Yが、ゲノムXに挿入された塩基配列情報である。B:ゲノムXは、分節化開始塩基1から所定の分節化塩基数で分節化した場合、分節1~6を生じる。一方、ゲノムZを前記所定の分節化塩基数で分節化すると、分節I~VIIIを生じる。しかし、挿入配列Yの長さは分節化塩基数とは異なるため、挿入配列Y以降の分節はゲノムXの分節には対応しないものとなる。そこで、分節化開始塩基を、挿入配列Yの長さ分だけ移動させると、ゲノムZの分節v~viiiは、ゲノムXの分節3~6に対応するものとなる。

【図7】Escherichia coli K12 substr. MG1655, complete genome (ACCESSION: Acc# NC_000913; VERSION: NC_000913.2, GI:49175990) (配列番号:1)を使用して、本発明に従って、分節化開始塩基1及び分節化塩基数1,000で作成した偏在性表示配列を示す。

20

【図8】分節化開始塩基を変更した場合における、偏在性表示配列及び偏在性表示視覚化情報パターンの変化を示す。Escherichia coli K12 substr. MG1655, complete genome (ACCESSION: Acc# NC_000913; VERSION: NC_000913.2, GI:49175990) (配列番号:1)について、本発明に従い、分節化塩基数30,000及びP=50で、分節化開始塩基を1から3,000ずつ順次変更して偏在性表示配列を作成した。図中、記号A~Jで表されている偏在性表示視覚化情報は、それぞれ分節化開始塩基1(A)、3,001(B)、6,001(C)、9,001(D)、12,001(E)、15,001(F)、18,001(G)、21,001(H)、24,001(I)、及び27,001(J)に対応する。

30

【図9】図8に記載した偏在性表示配列の各偏在性表示塩基に、図5の凡例に示されているパターンを当てはめて作成された偏在性表示視覚化情報を示す。

【図10】本発明の偏在性表示配列及び偏在性表示視覚化情報の作成に使用した、18種類の塩基配列情報に関する特性プロファイルの一部である、配列名(ゲノム名)、アクセッション番号、GC含量、及び(既存の生物学的)分類を示す。

【図11】図10に記載した18種類の塩基配列情報から得られた偏在性表示配列を示す。各偏在性表示配列は、分節化開始塩基=1、分節化塩基数=30,000及びP値=50の場合(左欄)、並びに分節化開始塩基=1、分節化塩基数=30,000及びP値=当該塩基配列情報のGC含量(%)である場合(右欄)の両方について示している。

【図12】図11に示されている、分節化開始塩基=1、分節化塩基数=30,000及びP値=50で作成された偏在性表示配列の各偏在性表示塩基に、図5の凡例に示されているパターンを当てはめて作成された偏在性表示視覚化情報を示す。

40

【図13】図11に示されている、分節化開始塩基=1、分節化塩基数=30,000及びP値=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列の各偏在性表示塩基に、図5の凡例に示されているパターンを当てはめて作成された偏在性表示視覚化情報を示す。

【配列表のフリーテキスト】

【0106】

(配列番号:1)本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、Escherichia coli str. K-12 substr. MG1655, complete genome由来の塩基配列情報である。

50

(配列番号：2) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Escherichia coli* str. K-12 substr. DH10B, complete genome由来の塩基配列情報である。

(配列番号：3) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Escherichia coli* O157:H7 str. Sakai, complete genome. (配列番号：3) 由来の塩基配列情報である。

(配列番号：4) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Shigella flexneri* 2a str. 2457T, complete genome由来の塩基配列情報である。

(配列番号：5) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Shigella flexneri* 5 str. 8401, complete genome由来の塩基配列情報である。

(配列番号：6) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Shigella sonnei* Ss046, complete genome由来の塩基配列情報である。

(配列番号：7) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Salmonella enterica* subsp. *enterica* serovar Paratyphi B str. SPB7, complete genome由来の塩基配列情報である。

(配列番号：8) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Salmonella enterica* subsp. *enterica* serovar Typhi Ty2, complete genome由来の塩基配列情報である。

【0107】

(配列番号：9) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Salmonella typhimurium* LT2, complete genome由来の塩基配列情報である。

(配列番号：10) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Herpetosiphon aurantiacus* ATCC 23779, complete genome由来の塩基配列情報である。

(配列番号：11) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Nitrosomonas europaea* ATCC 19718, complete genome由来の塩基配列情報である。

(配列番号：12) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Prochlorococcus marinus* str. MIT 9313, complete genome由来の塩基配列情報である。

(配列番号：13) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966, complete genome由来の塩基配列情報である。

(配列番号：14) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Geobacter sulfurreducens* PCA, complete genome由来の塩基配列情報である。

(配列番号：15) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Mesorhizobium* sp. BNC1, complete genome由来の塩基配列情報である。

(配列番号：16) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Alkaliphilus oremlandii* OhILAs, complete genome由来の塩基配列情報である。

【0108】

(配列番号：17) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Listeria monocytogenes* str. 4b F2365, complete genome由来の塩基配列情報である。

(配列番号：18) 本発明の偏在性表示配列及び偏在性表示視覚化情報を作成するために使用した、*Staphylococcus aureus* subsp. *aureus* USA300_TCH1516, complete genome由来の塩基配列情報である。

(配列番号：19) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数1,000、及びP=50で作成された

10

20

30

40

50

偏在性表示配列である。

(配列番号:20) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:21) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基3,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:22) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基6,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

10

(配列番号:23) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基9,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:24) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基12,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

【0109】

(配列番号:25) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基15,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

20

(配列番号:26) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基18,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:27) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基21,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:28) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基24,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:29) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基27,001、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

30

(配列番号:30) *Escherichia coli* str. K-12 substr. DH10B, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:31) *Escherichia coli* O157:H7 str. Sakai, complete genome. (配列番号:3) 由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:32) *Shigella flexneri* 2a str. 2457T, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

40

【0110】

(配列番号:33) *Shigella flexneri* 5 str. 8401, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:34) *Shigella sonnei* Ss046, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:35) *Salmonella enterica* subsp. *enterica* serovar Paratyphi B str. SPB 7, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,

50

000、及びP=50で作成された偏在性表示配列である。

(配列番号:36) *Salmonella enterica* subsp. *enterica* serovar Typhi Ty2, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:37) *Salmonella typhimurium* LT2, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:38) *Herpetosiphon aurantiacus* ATCC 23779, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

10

(配列番号:39) *Nitrosomonas europaea* ATCC 19718, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:40) *Prochlorococcus marinus* str. MIT 9313, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

【0111】

(配列番号:41) *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

20

(配列番号:42) *Geobacter sulfurreducens* PCA, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:43) *Mesorhizobium* sp. BNC1, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:44) *Alkaliphilus oremlandii* OhILAs, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:45) *Listeria monocytogenes* str. 4b F2365, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

30

(配列番号:46) *Staphylococcus aureus* subsp. *aureus* USA300_TCH1516, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=50で作成された偏在性表示配列である。

(配列番号:47) *Escherichia coli* str. K-12 substr. MG1655, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:48) *Escherichia coli* str. K-12 substr. DH10B, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

40

【0112】

(配列番号:49) *Escherichia coli* O157:H7 str. Sakai, complete genome. (配列番号:3)由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:50) *Shigella flexneri* 2a str. 2457T, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:51) *Shigella flexneri* 5 str. 8401, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量

50

(%)で作成された偏在性表示配列である。

(配列番号:52) *Shigella sonnei* Ss046, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:53) *Salmonella enterica* subsp. *enterica* serovar *Paratyphi* B str. SPB 7, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:54) *Salmonella enterica* subsp. *enterica* serovar *Typhi* Ty2, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

10

(配列番号:55) *Salmonella typhimurium* LT2, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:56) *Herpetosiphon aurantiacus* ATCC 23779, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

【0113】

(配列番号:57) *Nitrosomonas europaea* ATCC 19718, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

20

(配列番号:58) *Prochlorococcus marinus* str. MIT 9313, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:59) *Aeromonas hydrophila* subsp. *hydrophila* ATCC 7966, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:60) *Geobacter sulfurreducens* PCA, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:61) *Mesorhizobium* sp. BNC1, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

30

(配列番号:62) *Alkaliphilus oremlandii* OhILAs, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:63) *Listeria monocytogenes* str. 4b F2365, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

(配列番号:64) *Staphylococcus aureus* subsp. *aureus* USA300_TCH1516, complete genome由来の塩基配列情報を使用して、分節化開始塩基1、分節化塩基数30,000、及びP=当該塩基配列情報のGC含量(%)で作成された偏在性表示配列である。

40

【 図 1 】

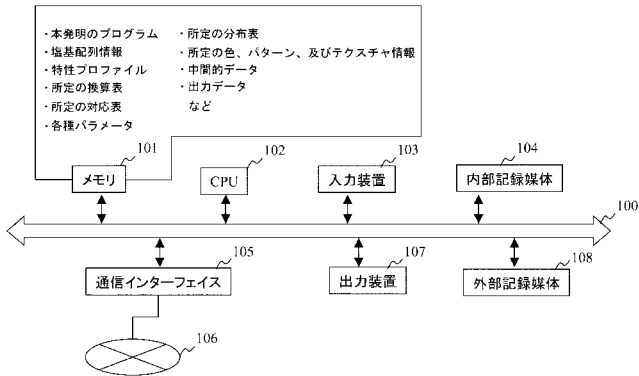


图 1

【 図 2 】

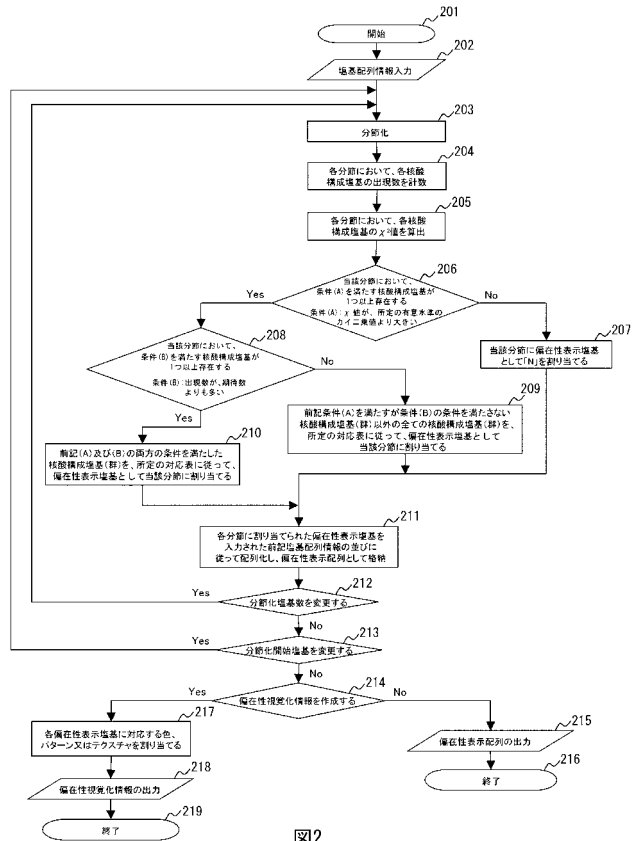


図2

【 図 3 】

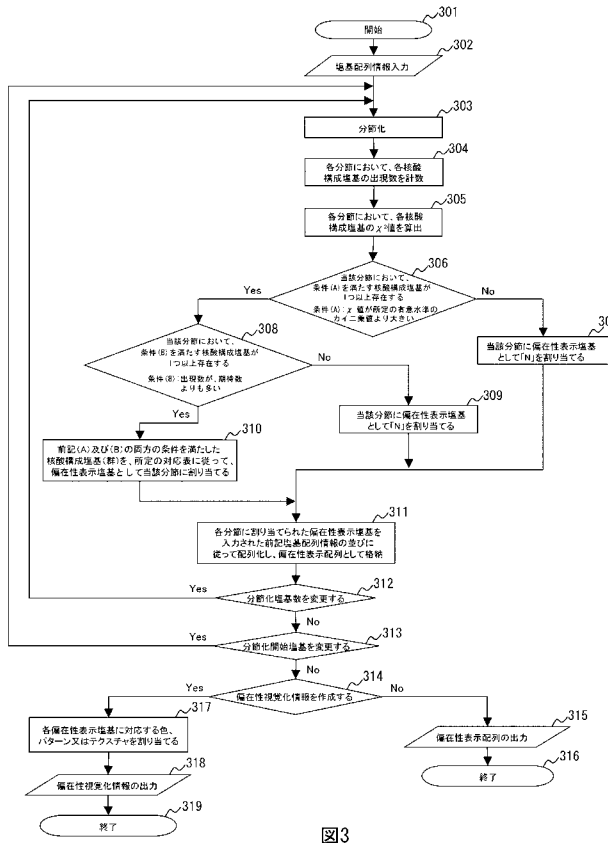


图3

【 図 4 - 1 】

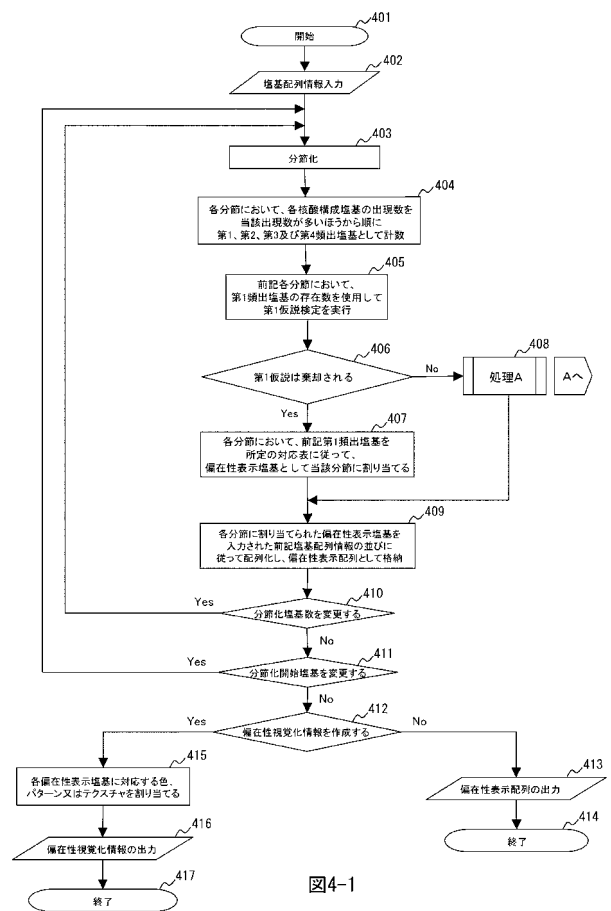


图4-1

【図 4 - 2】

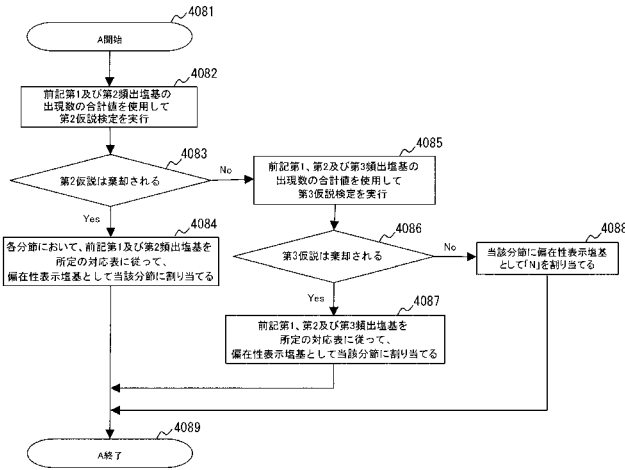
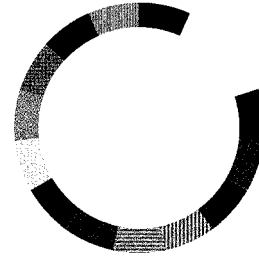


図4-2

【図 5】



■ A	T	G	■ C
■ W	■ S	■ M	■ R
■ Y	■ K	■ V	■ H
■ D	■ B	■ N	

図5

【図 6】

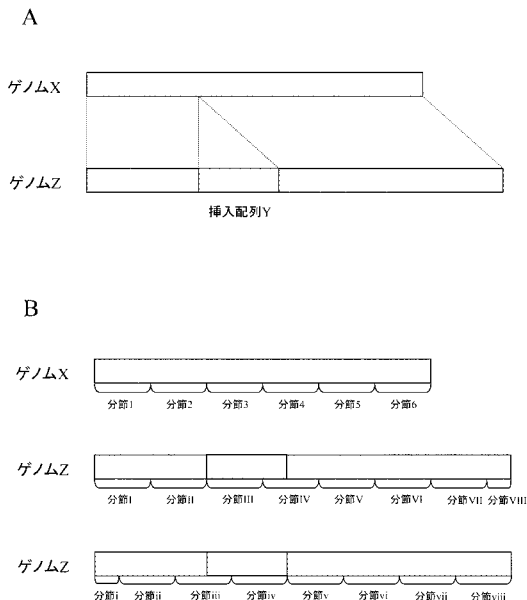


図6

【図 7】

NGGGGKNGV YTAAGTNTK NBGGGGGGT GGGGTGNC CATGGGKGN GNGGMBGAA CCBGCGCG NCCSGTKC BBNWTDGN GGGGGKGGG GGGGGKNTVG
 CHNUNNNHC VAVVGGGVC GGGGGKNC GGNNNNNH WTTATNGA NGGGGKGB GGGGKGVN YGNGGGMN NAGMBGCM NGGGGGGG NGGGMNCG
 TCGAGAGCA MNCNTMNC NCSNNMTC GGNNNNNH GTGGKCSAA AGGTAGCNA SSSSSSSCN GDNWTTSCS SGTNNBGCN NNNNNNNH NGGTSSSBG
 MNNMAGGG GCGAAGNCG CCGTCCCGG TAASCENGG GGGGGKGCN NGWNGGGMN RENDNNGN WYAGKATNG VYKNGGTG GGGGGKGGG NTGGGTCNB
 MCGNANVY NAGVGGNNG GYACBMNGC GGGGGKGNH ACCERTANG NCSNNNNH NCAGGAGNC NGCCBANG GGGSSNNCG GTGGGKATN TGANNYSGG
 CGGTGACTHT NTNNITMTR WATTTATT GAWTTNCSS CVNENYGG GTANADNCG CNGGGKCG GGGKGGGG GGGGATNG GTGAASSSSG NNGNANAGN
 NNNGCYCHN NGCBNTNTN GCAATATNG NCGGNNNN BNGGAGHNN NNGAHACNA NCCSSCMG SGGNHHNG AGGGATTA TTTNTTNGS GGGGGKGNH
 TGTNRGNA VAKCTGTCBG GANNAGGNW NNTGAVGCG GAGGCGNNT TYVGAGCCN BENGGBNG NNGNENAA TADNMGYCM CNGSVGGGV GDNWTTBNG
 CNANCAAY NTGNTKGGT CCGDAGCCB VCAAGANAG CGDGAIVEC NOTCCGNG NNNANWTTAT NNNANWTTA NNNNBTBGN NGGGGGGG NNNNNNCA
 NNCCMBAD NGNNNTVY GGGNMAAG DVAAGNNH ANSGGGHT CNBNNNNH NNGCBNGG GGNWNGMS SCNCCSTG TGVMBNHN TWBTBNG
 NTRABDAG NGONNNNNH NGVAKCTB VYAGCGVIA NSGNNNGG GGGGGBNBN TNGGNNGT CCGAGKGN ATAGVCMN GGBNANNH ATAGCGTWA
 NTNNWTTB ANBNTNTN EVAANGGA NGVNTDGN NNNNNBCH VCGGDTBT AANBNTT GGGVGGNN DTBNTNTA ATNNNNH AATTDCNNS
 NNNTTNGG NATTGNNNH CNBNNNGC NGCCNACN NSSGNNHT AGTGGGGG GGGGNGHT TACNNNNH GNNNTTGT TTTTITTD BATTNGSG
 ATNATTNT NGGAMNTN CAYCGGNG GGGGKAAA GGBATTGK NYCGGNNN TONNNHTT AANNNTTN GTTGAHAKH BANTWSDA AGAGKSSCB
 NABNAGGG CCTTCCCT NNNNNNNH WTTDTATBT NNNTTBT TCCGTCN NNGVANNL CCNNNTMT TANNNNH NABNNTTW TENHADA
 ATNANNNV AGGTBNKKA HATCCABON NNNCTNNAH TMCATCTBA BCTTGVBGN NTNCHNTN CTNNNNSS GAGACCTT TBTCTANC TWNNTCB
 MNCNCKHT TTGAGNNGV MNCCTNCA TCCNCTTN NNNANNTT TANNNTAW NNTCCGCC ATNCGGVV NNCNCCNN AATTTHANN TNNNNNT
 TATNABHN CTNNKNC DNTNNGCA VETBAATT CBNTTGBH HANTNNH CBNNNNGH NCCNCCN TNCCTNCA ACCCCCCC NCCBNHNN
 NNCNTHNC MTAACNTB TMAADTNN WNGGACNG TTBVNTCB CBTANNNN NNTNAAAA TTHNNNNH HCHTTCN ACCNNCAT ANNANTNG
 CNGVANNH BNNNNNNH NTNCCGCC BCCCCNAC BBBCGTBCC NTNAGCN CNVSSGGS GNNGTAVN ANBNTNGV CBNNTWAB NNNANCA
 NNGGZGVA CACCCGCC NNVCNGB CTGATNG NBNNNNNH NATNCHC BNNNTTN NBNNNNNH NBNNTBTS CCHNCCSG CACNCCCN
 NNGVAGNHA NNGGAGNCCC NNCNCCNC CTGVBHTC ASSNNMCH NNNNNNT KTCNCCCN AATTANCA SSCNCCCG SCNNANAM NACBTBNN
 BANNTHNC CACNCGV GGVNBNH NNNNNCHN HADABNNH CATATTNA TNNNNH NNNNNH CCCCNC ATGNNNNH AACTCCGT
 NNNNNHCC CNGBNYCHN AGCCSGAD AANSCCCC CSNCHNG CCGNNCGT CCCCCHNN KCGTKBGC NGTNNCHNG NNNNNABAC NMACCCCN
 BENCCNCC NNTCCSCB BBNHNNV GTCHNCC CNGNCCN CSMNNTT VCNENNNH GSANTCNG NNAACCHT CNGNACW NNCNCCBN
 ANNTNNH NNTTBGCA VNTCAETT TNNNNHSS CCGTNNH NNNGBTTA GTHNCCN NCMWNC AGCCNCG NCCSCNCC GGNNSNN
 NNTBNC CNGBNCHN TMAATCSC NNCABATN NNHATNCH CBNNNCCN NCTNNTBN GTTBNNNNH TNGCNCAV NCCNNNVC CNAACG
 CCGNABBN ACNTNNNNH NNNTTNG NNNNNNNH NCGANNNNH CTTNNNT TNCNCCN VAGCCGAV CBNNNNNH SGNANNNH NNNCHASH
 GACATCGV NCHGNTNN GATTAAC NNCBNNH CCCCCHN NNCNCCNC NNCNCHN NNNNNH CCBNNH ANCTNNNNH NNNNNH
 TACCNCCN SNNHSSCN VAVCVCHN NNCCTNCG GTBNNNTN CNGNNNNH TYVCHNGG TNAWNNCC CNAHAGHG CBNNNHNA NNGGCGCN
 CACNCCN NCTNNNTA NNCNTHNC CNVCCCTG GAGVNNCC NNCSSNNH NNNNNH CBNNNNNH NNNNNH CBNNNH CBNNNH
 ATVNTNB DTCVCTCN NNNNNNNH NACNNNTN CBNNNNNH NNNNTBNN BNNNTCCN CBNNNCC TNGGGGCB GNCNCCCN NNCNCCCN
 CBNNNNNH ANCCNNGV CNGTCCCG ACVCNCCG GTNCCNCC VCNNNNNH NNNCCCG CCMCTTNG CNGTGGTC KCCSSSSSS GAGACSCCN
 NNTNNNTT CCGNANNA TNNNTBN CBNNNHSH ANNCNCCN ANVCSCCG CCGNNNT CCGCTNCA NVCNNAH NCCNTNCA ANNNNNH
 NNTTNGSA CACBACCC SSSADNNH NGVANSNG CCGVNNAC ANNNNNH NNNNTBCH NNCNCCCN VCGCMAGG CNAATTAT MANACCC
 STVNNHAC NAGGAKAC CACACBCC CNGCTGTT KNNATDTH TAAACNCC NCCNCCBA ANNNNNH CNGNNHNA GAGGNNNG VGGSGGN
 CBNNTTGG GGGGGKGN NGGTTBCHN CNGGNNNNH NNGGNNH CCGTGGGG VNGGKCV GTAGAGGG NNNNNH CNGNNNT TAAACNCC
 BNGTNTBC CSBGNANC NNCNCCN CBNNNNT TNCACANT CBNCCNCC NAGNCCBG TCGGNNV NNGGNNH BBNNNKGG AGANNNTN
 VGGNNHNC CCGGNNCC GGGGNNNG ANTAATW NNGGNNH CNGTGGTA AVNCCN NCGNNNNH GNNNNH CCGTNGCN NVCTTGG
 GNNKCCCG SSSDCHNT TSSSSSSSS GCGGAGT BCBAGNNT GTTGGATT TNNNNAH TNGNNTA CNTNNCCG NTXXKCCG NNGGCGGG
 GGGGKGN TGGGNTBG GGNANNT VCNNAATN NNGGCGAT CCKZGANC TNGNTGBN NNNWATT NNCNCCCN VNNNNNNH YATWNTGS
 SSSGNNNA NNGGNNH CAAATNNH NAGANNH VGGNNKAG AACNNAG NNNNNNT BCGTGGSG GGGNNNT NNNNNNNH NNCNCHN
 CBBGNNNG CCGGNNNN (配列番号:19)

図7

【 図 8 】

記号	分岐化開始塩基	塩性変異型配列
A	1	GGGGGGGGSGTGGGSGGSHVSGWSGSGGSGGGGSGWGDGNGSGSTGKRVYVHPCCC YMCCTTCTCHHWYSCYCCCCCCCCCSCSSCCSSCCANSSTCTCCWCCCCCSCMCCMYC SCCCTCCGCTCCGCGGGGGGGGGGGGGGAGGSGTNDGG (配列番号: 22)
B	3, 001	GSGGGGGGGTGGGGSGGTTTSGCCGSGGGGGGGGGSGWGGWGGGTTGKRGKRWMMCC YCMCPTCTAAHWSHSCCCCCCCCCCSCGSCSCCANCSCSCGSCGSCSCCCCCC CSCCTCTCNCWSGCGGGGGGGGGGGGGGAGGGSNDGG (配列番号: 23)
C	6, 001	GSGGGGGGRTSGGGGGSGTTTSGCCGSGGGSGSGSGSGWGTGGTGTGGRNMAWCCC YMCCTCTCNMAACSSGCCCCCHMCHTSCSCSCSCGCCSCCCCCCCCCCSCMCCMYC CSCCCCCCSCGCGGGGGGGGGGGGGGSHNDDGTGAGSG (配列番号: 22)
D	9, 001	CGAGCGGCGGCTGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG TCCTVCTCCNAWSTCCGCCCCCHMCMSCSCSCSCGCCSCCCCCCCCCCSCMCCMYC CSCCCCCCSCGCGGGGGGGGGGGGGSGGSDTSGGCTNARG (配列番号: 23)
E	12, 001	GSGGTGGRSGSGGGGGSGHVGGBSRGGGSGSGSGRSGTGGMWNGGGTGGRRTVMAWYCT WCCGCTGCGGCTGCTCCCCCHMCMSCSCSCGCCSCCCCCCCCCCSCMCCMYC CSCSCSCSNAWCGCGGSGSGGGSGSGGSGSGWGGGNTAGSG (配列番号: 23)
F	15, 001	GSGCGGASGTGGGGGGSGGGGGGGGGGGGGGSGGGGGGMMNASTGGGANNAWCYC BHWCYCYCHMWSCSCCCCCCHMCMSCSCSCGCCSCCCCCCCCCCSCMCCMYC CSCSCSNAWCGCGGSGSGGGSGSGGGTGTGCGTGG (配列番号: 25)
G	18, 001	CGGCGGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG BHWCCCTCYCMHSCGCCCCCCCHMCMSCSCSCGCCSCCCCCCCCCCSCMCCMYC CSCSCSNAWCGCGGCGGASGGGSGSGHSGGNTNAGSA (配列番号: 26)
H	21, 001	SSGGGKGNSTGGGSGSHSGTSGVSGSGGGGGGKYDGGSGGCGTGGGTGARGATWACY HWSYSGTGCACBACCCTCCCMSCSCSCGCCSCCCCCCCCCCSCMCCMYC CSCCTWACACWCCCGGSGAGSGGGGSGASGNNAGG (配列番号: 27)
I	24, 001	SSGGGGGSGTGGGGSGGSGTSGVSGSGGGGGGGGGGGGGGSGTGGTGGTGGTGGTGGT HMCWCCYCMHSCGCCCCCCCHMCMSCSCSCGCCSCCCCCCCCCCSCMCCMYC CSCCTWACACWCCCGGSGAGSGGGGSGASGNNAGG (配列番号: 27)
J	27, 001	SGSGGGGCTGGGGGGSGGZGVGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG HMCWCTTHHCWSCGCCCCCCCHMCMSCSCSCGCCSCCCCCCCCCCSCMCCMYC CSCCCCCCSCGCGGGGGGGGGGGGGSGGSGTGGG (配列番号: 29)

图8

【图 9】

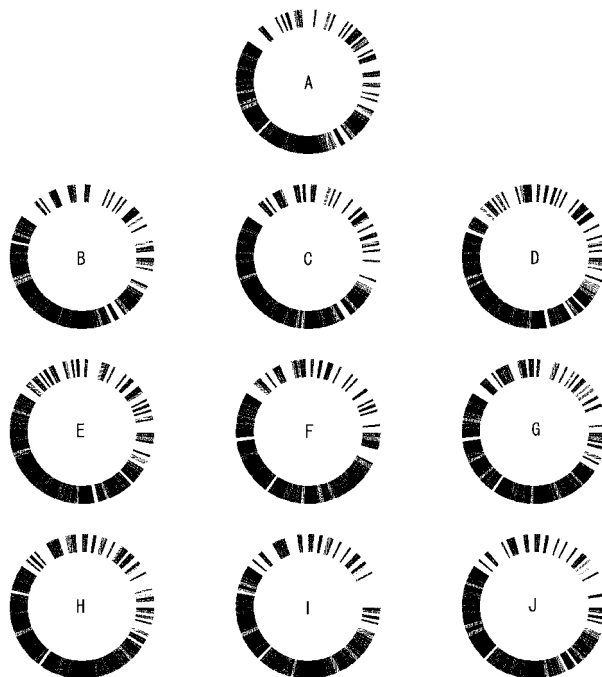


图 9

【 図 1 0 】

[illegible]

图10

【 図 1 1 】

[illegible]

图 11

【 図 1 2 】

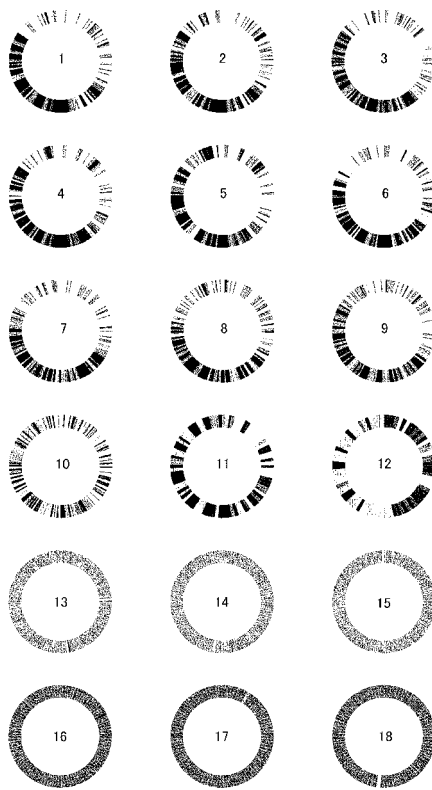


图 12

【図 13】

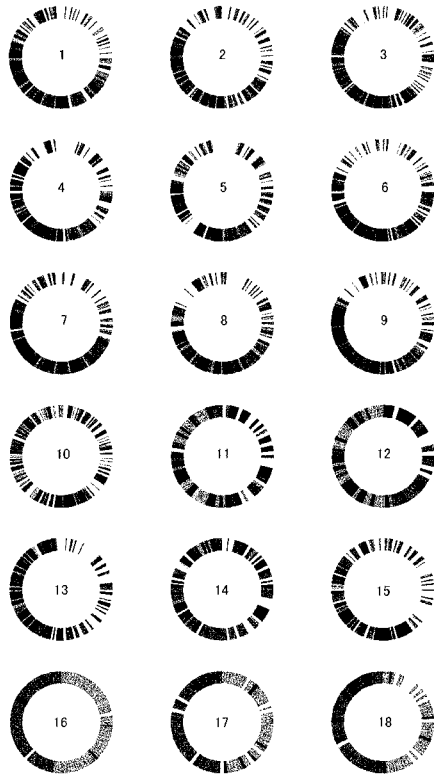


図13

【配列表】

2010033344000001.app