(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2021-80241
(P2021-80241A)
(43)公開日　令和3年5月27日(2021.5.27)

(51)Int.Cl.			F I			テーマコード（参考）
C40B	40/06	(2006.01)	C40B	40/06		4B050
C12Q	1/6869	(2018.01)	C12Q	1/6869	Z	4B063
C12Q	1/6827	(2018.01)	C12Q	1/6827	Z	
C12N	9/16	(2006.01)	C12N	9/16	Z	

審査請求 未請求　請求項の数 23　ＯＬ　（全 59 頁）

(21)出願番号　特願2020-77622(P2020-77622)
(22)出願日　令和2年4月24日(2020.4.24)
(31)優先権主張番号　特願2019-207386(P2019-207386)
(32)優先日　令和1年11月15日(2019.11.15)
(33)優先権主張国・地域又は機関
日本国(JP)

(71)出願人　000000918
花王株式会社
東京都中央区日本橋茅場町１丁目１４番１０号
(74)代理人　110000084
特許業務法人アルガ特許事務所
(72)発明者　大坪　裕紀
栃木県芳賀郡市貝町赤羽２６０６　花王株式会社研究所内
(72)発明者　松村　奨士
栃木県芳賀郡市貝町赤羽２６０６　花王株式会社研究所内

Ｆターム(参考)　4B050 LL03
4B063 QA17 QQ43 QR08 QR14 QR42 QS25 QS34

(54)【発明の名称】シーケンシング用ライブラリの調製方法

(57)【要約】
【課題】シーケンシングエラーを低減させるシーケンシング用ライブラリの提供。
【解決手段】サンプルＤＮＡを断片化すること；及び、調製したサンプルＤＮＡの断片を１本鎖特異的ヌクレアーゼで処理し、該断片から１本鎖部分を除去すること、を含むシーケンシング用ライブラリの調製方法。
【選択図】なし

【特許請求の範囲】
【請求項１】
シーケンシング用ライブラリの調製方法であって、
サンプルＤＮＡを断片化すること；及び、
調製したサンプルＤＮＡの断片を１本鎖特異的ヌクレアーゼで処理し、該断片から１本鎖部分を除去すること、
を含み、
該サンプルＤＮＡが、生細胞から抽出したＤＮＡ、凍結細胞から抽出したＤＮＡ、又はそれらのＤＮＡの保存サンプルである、
方法。
【請求項２】
前記１本鎖特異的ヌクレアーゼが１本鎖特異的エンドヌクレアーゼ、１本鎖特異的エキソヌクレアーゼ、又はそれらの組み合わせである、請求項１記載の方法。
【請求項３】
前記１本鎖特異的ヌクレアーゼでの処理が、前記サンプルＤＮＡの断片を１本鎖特異的エンドヌクレアーゼで処理した後に、さらに１本鎖特異的エキソヌクレアーゼで処理することを含む、請求項２記載の方法。
【請求項４】
前記１本鎖特異的エンドヌクレアーゼがＳ１　ｎｕｃｌｅａｓｅ又はＭｕｎｇ　Ｂｅａｎ　Ｎｕｃｌｅａｓｅである、請求項２又は３記載の方法。
【請求項５】
前記１本鎖特異的エンドヌクレアーゼが、前記サンプルＤＮＡの断片１ｎｇあたり０．０２Ｕ／ｎｇ以上のＳ１　ｎｕｃｌｅａｓｅである、請求項２又は３記載の方法。
【請求項６】
前記１本鎖特異的エンドヌクレアーゼが、前記サンプルＤＮＡの断片１ｎｇあたり０．０２Ｕ／ｎｇ以上のＭｕｎｇ　Ｂｅａｎ　Ｎｕｃｌｅａｓｅである、請求項２又は３記載の方法。
【請求項７】
前記１本鎖特異的エキソヌクレアーゼが$RecJ_f$である、請求項２又は３記載の方法。
【請求項８】
前記１本鎖特異的エキソヌクレアーゼが、前記サンプルＤＮＡの断片１ｎｇあたり０．１０Ｕ／ｎｇ以上の$RecJ_f$である、請求項２又は３記載の方法。
【請求項９】
前記１本鎖特異的ヌクレアーゼで処理した前記サンプルＤＮＡの断片を、末端修復、末端への塩基付加、及び増幅からなる群より選択されるいずれか１つ以上の処理に供することをさらに含む、
請求項１～８のいずれか１項記載の方法。
【請求項１０】
前記増幅がＰＣＲであり、前記１本鎖特異的ヌクレアーゼがＳ１　ｎｕｃｌｅａｓｅであり、かつ
前記サンプルＤＮＡの断片１ｎｇあたりの該Ｓ１　ｎｕｃｌｅａｓｅのユニット数（Ｕ／ｎｇ）が０．０５Ｕ／ｎｇ以下のとき、該ＰＣＲにおける前記サンプルＤＮＡ　１Ｍｂｐあたりの初期ＤＮＡ量が２５０ａｍｏｌ以下であるか、又は
前記サンプルＤＮＡの断片１ｎｇあたりのＳ１　ｎｕｃｌｅａｓｅのユニット数（Ｕ／ｎｇ）が０．０５Ｕ／ｎｇより大きいとき、下記式で算出される指標が６０以下である：
指標＝該ＰＣＲにおける初期ＤＮＡ量（ａｍｏｌ／ＭｂｐサンプルＤＮＡ）×$3^{\log S1\ nuclease\ (U/ng)}$
（式中、Ｓ１　ｎｕｃｌｅａｓｅ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である）、
請求項９記載の方法。

【請求項１１】
前記増幅がＰＣＲであり、前記１本鎖特異的ヌクレアーゼがＭｕｎｇ　Ｂｅａｎ　Ｎｕｃｌｅａｓｅであり、かつ
前記サンプルＤＮＡの断片１ｎｇあたりの該Ｍｕｎｇ　Ｂｅａｎ　Ｎｕｃｌｅａｓｅのユニット数（Ｕ／ｎｇ）が０．０５Ｕ／ｎｇ以下のとき、該ＰＣＲにおける前記サンプルＤＮＡ　１Ｍｂｐあたりの初期ＤＮＡ量が２５０ａｍｏｌ以下であるか、又は
前記サンプルＤＮＡの断片１ｎｇあたりのＭｕｎｇ　Ｂｅａｎ　Ｎｕｃｌｅａｓｅのユニット数（Ｕ／ｎｇ）が０．０５Ｕ／ｎｇより大きいとき、下記式で算出される指標が６０以下である：
指標＝該ＰＣＲにおける初期ＤＮＡ量（ａｍｏｌ／ＭｂｐサンプルＤＮＡ）×$3^{\log \mathrm{Mung\ Bean\ Nuclease\ (U/ng)}}$
（式中、Ｍｕｎｇ　Ｂｅａｎ　Ｎｕｃｌｅａｓｅ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である）、
請求項９記載の方法。
【請求項１２】
請求項１～１１のいずれか１項記載の方法で調製されたシーケンシング用ライブラリをシーケンシングすることを含む、ＤＮＡのシーケンシング方法。
【請求項１３】
前記シーケンシング方法が、以下：
（１）前記ライブラリをシーケンシングし、該ライブラリに含まれる複数の増幅断片の各々について１つ以上のリード配列を作成し、該複数の増幅断片についての複数のリード配列を得ること；
（２）得られた複数のリード配列の中から、該ライブラリの調製に用いたサンプルＤＮＡ上の同一領域の配列情報を有するリード配列を集めてグループ化することにより、リード配列のグループを１つ以上作成すること；及び、
（３）該リード配列のグループに含まれるリード配列の間で配列情報のコンセンサスを取ること、
を含む、請求項１２記載の方法。
【請求項１４】
前記（１）が、前記サンプルＤＮＡの断片を構成する２本の相補鎖の各々に由来する増幅断片に対して１つ以上のリード配列を作成することを含む、請求項１３記載の方法。
【請求項１５】
前記（２）が、参照配列上の同一の位置にマッピングされるリード配列を同じグループに分けることを含む、請求項１４記載の方法。
【請求項１６】
前記（３）が、前記リード配列のグループの中から、前記サンプルＤＮＡ断片の２本の相補鎖の各々に由来するリード配列を少なくとも１つずつ集め、集めたリード配列の間で配列情報のコンセンサスを取ることを含む、請求項１５記載の方法。
【請求項１７】
前記（１）において、前記複数のリード配列が、以下からなるリード配列のペアを複数個含み：
リード１：前記増幅断片を構成する２本の相補鎖のうちの一方の鎖の配列を５'末端側から３'側へ読んだ配列に相当する配列情報を含むリード配列、
リード２：該一方の鎖の配列を３'末端側から５'側へ読んだ配列に相当する配列情報を含むリード配列、
前記（２）が、得られたリード配列のペアの中から、該サンプルＤＮＡ上の同一領域の配列情報を有するリード配列のペアを集めてグループ化することにより、リード配列のペアのグループを１つ以上作成することを含み、
前記（３）が、該リード配列のペアのグループに含まれるリード配列の間で配列情報のコンセンサスを取ることを含む、

請求項１３記載の方法。
【請求項１８】
前記（１）が、前記サンプルＤＮＡの断片を構成する２本の相補鎖の各々に由来する増幅断片に対して１つ以上の前記リード配列のペアを作成することを含む、請求項１７記載の方法。
【請求項１９】
前記（２）が、前記リード配列のペアのリード１とリード２を参照配列に対してマッピングし、リード１の先頭とリード２の先頭とに挟まれる該参照配列の領域が同一であるリード配列のペアを同じグループに分けることを含む、請求項１８記載の方法。
【請求項２０】
前記（２）が、前記リード配列のペアに含まれる一方のリード配列の先頭が前記参照配列上の同じ位置に位置するリード配列のペアを集め、次いで集めたリード配列のペアの中から、該リード配列のペアに含まれるもう一方のリード配列の先頭が該参照配列上の同じ位置に位置するリード配列のペアを集めて、集めたリード配列のペアを同じグループに分けることを含む、請求項１８記載の方法。
【請求項２１】
前記（３）が、前記リード配列のペアのグループの中から、前記サンプルＤＮＡ断片の２本の相補鎖の各々に由来するリード配列のペアを少なくとも１組ずつ集め、集めたリード配列のペアに含まれるリード配列の間で配列情報のコンセンサスを取ることを含む、請求項１９又は２０記載の方法。
【請求項２２】
ゲノムＤＮＡをサンプルＤＮＡとして用いて、請求項１～１１のいずれか１項記載の方法によりシーケンシング用ライブラリを調製すること；及び、
該シーケンシング用ライブラリをシーケンシングすること、
を含む、ゲノムＤＮＡの変異を検出する方法。
【請求項２３】
前記シーケンシングが請求項１３～２１のいずれか１項記載の方法により行われる、請求項２２記載の方法。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明は、シーケンシング用ライブラリの調製方法に関する。
【背景技術】
【０００２】
次世代シーケンシング（ＮＧＳ）技術は、近年著しい発展を遂げ、がん細胞のゲノム変異解析などにおいて盛んに活用されて新しい知見を生み出している。ＮＧＳのためのシーケンサーとしては、イルミナ社のＨｉＳｅｑやＭｉＳｅｑなどのシーケンサーが多用されている。これらのシーケンサーでは、解析対象となる細胞や組織から抽出したサンプルＤＮＡを数百ｂｐの長さに断片化した後、該断片の突出末端を平滑化し、次いで両末端にシーケンシングアダプターを結合してライブラリＤＮＡを調製し、これをシーケンシングする。該末端の平滑化では、一般に、Ｔ４　ＤＮＡポリメラーゼ等の酵素により３'側突出末端が除去され、一方、５'側突出末端は、対となる鎖が合成されて修復される（末端修復）。
【０００３】
シーケンシングからのデータには通常、サンプルＤＮＡの性状やライブラリ調製の工程に起因するエラーが含まれ、これらは解析対象の細胞集団内の変異等の正確な同定への障害となる。例えば、サンプルＤＮＡ自体が保存中の損傷などにより１本鎖になることで、シーケンシングのエラーの原因となることがある。例えば、ホルマリン固定パラフィン包埋（ＦＦＰＥ）サンプルのＤＮＡは、サンプル調製や保存の間に損傷して１本鎖になることがあり、これら１本鎖ＤＮＡ同士は、繰り返し配列などの部分で誤って対形成してキメ

ラ断片を生成する（特許文献１、非特許文献１）。また、ｃｅｌｌ　ｆｒｅｅ　ＤＮＡ（ｃｆＤＮＡ）は、血中で分解されて１本鎖になることがある。１本鎖ＤＮＡのシトシンは、脱アミノ化によりチミジンに変換されやすいため、シーケンシングで高頻度にエラーを引き起こす（特許文献２）。このような１本鎖ＤＮＡに由来するエラーを低減するために、ライブラリ調製の過程でサンプルＤＮＡを１本鎖特異的ヌクレアーゼで処理して１本鎖部分を除去する方法が提案されている（特許文献１、２、及び非特許文献１）。例えば、特許文献１及び非特許文献１には、ＦＦＰＥサンプルのＤＮＡを１本鎖特異的ヌクレアーゼで処理することで、シーケンシングにおけるキメラ断片の検出率及びエラー率が減少したことが報告されている。特許文献２には、ｃｆＤＮＡを１本鎖特異的ヌクレアーゼで処理することで、シトシンの脱アミノ化によるシーケンシングのエラーを低減することが可能であることが記載されている。

【０００４】

あるいは、サンプル調製や保存の過程で解析対象のＤＮＡに生じる酸化修飾等は、シーケンシングのエラーの原因となる（非特許文献２）。また、Ｋｅｎｎｅｄｙらは、シーケンスリードの端部でのエラーの増加が、断片化したＤＮＡの末端修復工程に起因する可能性に触れている（非特許文献３）。

【０００５】

近年、相補鎖情報を活用してシーケンシングのエラーを低減する方法が注目されている。例えば、サンプル調製や保存の過程でＤＮＡに生じる酸化修飾等は、２本鎖のうち片方の鎖だけに起こるため、２本の相補鎖間に固定された変異を同定することで酸化修飾等に起因するエラーを除くことができる（特許文献３、４）。しかしながら、塩基の酸化修飾がＤＮＡ末端の突出部位に存在する場合には、該修飾された塩基が末端修復工程において誤った塩基とペアを形成し、この誤った塩基がＰＣＲ等を経てＤＮＡの２本鎖に固定されることがある。そのため、突出部位に存在する酸化修飾等の起きた塩基は、相補鎖情報を活用したシーケンシングにおいても取り除くことができないエラーとなり得る。非特許文献３には、シーケンスリードの両端から５塩基を削除して解析することで、ＤＮＡの末端修復工程に起因するエラーを低減することを提案している。

【先行技術文献】

【特許文献】

【０００６】

【特許文献１】国際公開広報第２０１５／０５７９８５号
【特許文献２】国際公開広報第２０１９／１２６８０３号
【特許文献３】国際公開広報第２０１３／１４２３８９号
【特許文献４】国際公開広報第２０１９／２０８８２７号

【非特許文献】

【０００７】

【非特許文献１】Nucleic Acids Research, 47(2):e12, 2019
【非特許文献２】Nucleic Acids Research, 41(6):e67, 2013
【非特許文献３】Nature Protocols, 9(11):2586-2606, 2014

【発明の概要】

【発明が解決しようとする課題】

【０００８】

本発明は、シーケンシングエラーを低減させるシーケンシング用ライブラリを調製する方法に関する。

【課題を解決するための手段】

【０００９】

本発明は、シーケンシング用ライブラリの調製方法であって、

サンプルＤＮＡを断片化すること；及び、

調製したサンプルＤＮＡの断片を１本鎖特異的ヌクレアーゼで処理し、該断片から１本鎖部分を除去すること、

を含み、
該サンプルＤＮＡが、生細胞から抽出したＤＮＡ、凍結細胞から抽出したＤＮＡ、又はそれらのＤＮＡの保存サンプルである、
方法を提供する。
【００１０】
また本発明は、前記シーケンシング用ライブラリをシーケンシングすることを含む、ＤＮＡのシーケンシング方法を提供する。
【００１１】
また本発明は、細胞中のゲノムＤＮＡをサンプルＤＮＡとして用いて、前記シーケンシング用ライブラリの調製方法によりシーケンシング用ライブラリを調製すること；及び
該シーケンシング用ライブラリをシーケンシングすること、
を含む、ゲノムＤＮＡの変異を検出する方法を提供する。
【発明の効果】
【００１２】
本発明によれば、サンプル調製や保存の過程で生じる解析対象ＤＮＡの酸化修飾や損傷に起因するシーケンシングエラーを低減することができる。
【図面の簡単な説明】
【００１３】
【図１】ＤＭＳＯ暴露ライブラリにおける６つの変異パターンの変異頻度。データは同一条件で暴露した３サンプルの平均値と標準偏差を示す。
【図２】ＤＭＳＯ暴露ライブラリにおける１２変異パターンの変異頻度に対するリードペア両端の塩基の除去の影響。データは同一条件で暴露した３サンプルの平均値と標準偏差を示す。
【図３】ＤＭＳＯ暴露ライブラリの６つの変異パターンの変異頻度に対するＳ１　ｎｕｃｌｅａｓｅ処理の影響。
【図４】異なるユニット数のＳ１　ｎｕｃｌｅａｓｅで処理したＤＭＳＯ暴露ライブラリにおける１２変異パターンの変異頻度。
【図５】続き。
【図６】ＤＭＳＯ暴露ライブラリの６つの変異パターンの変異頻度に対するＭＢＮ処理の影響。
【図７】異なるユニット数のＭＢＮで処理したＤＭＳＯ暴露ライブラリにおける１２変異パターンの変異頻度。
【図８】続き。
【図９】ＤＭＳＯ暴露ライブラリの６つの変異パターンの変異頻度に対する$RecJ_f$処理の影響。
【図１０】異なるユニット数の$RecJ_f$で処理したＤＭＳＯ暴露ライブラリにおける１２変異パターンの変異頻度。
【図１１】続き。
【図１２】変異原処理したサンプルの変異検出に対するＳ１　ｎｕｃｌｅａｓｅ処理の影響。異なるユニット数のＳ１　ｎｕｃｌｅａｓｅで処理したＤＭＳＯ暴露ライブラリ（ＤＭＳＯ　ｃｏｎｔｒｏｌ）及び３－ＭＣ暴露ライブラリ（３ＭＣ）における変異頻度。
【図１３】続き。
【図１４】変異原処理したサンプルの変異検出に対するＭＢＮ処理の影響。異なるユニット数のＭＢＮで処理したＤＭＳＯ暴露ライブラリ（ＤＭＳＯ　ｃｏｎｔｒｏｌ）及び３－ＭＣ暴露ライブラリ（３ＭＣ）における変異頻度。
【図１５】変異原処理したサンプルの変異検出に対する$RecJ_f$処理の影響。異なるユニット数の$RecJ_f$で処理したＤＭＳＯ暴露ライブラリ（ＤＭＳＯ　ｃｏｎｔｒｏｌ）及び３－ＭＣ暴露ライブラリ（３ＭＣ）における変異頻度。
【図１６】Ｓ１　ｎｕｃｌｅａｓｅ処理したＤＭＳＯ暴露ライブラリのシーケンシングにおけるゲノムに対するカバレッジを示すヒストグラム。横軸はゲノム上の位置、縦軸は約

１００塩基区間のカバレッジを正規化した値。
【図１７】ＭＢＮ処理したＤＭＳＯ暴露ライブラリのシーケンシングにおけるゲノムに対するカバレッジを示すヒストグラム。横軸はゲノム上の位置、縦軸は約１００塩基区間のカバレッジを正規化した値。
【図１８】$ＲｅｃＪ_f$処理したＤＭＳＯ暴露ライブラリのシーケンシングにおけるゲノムに対するカバレッジを示すヒストグラム。横軸はゲノム上の位置、縦軸は約１００塩基区間のカバレッジを正規化した値。
【図１９】断片の誤認識率に対するＳ１　ｎｕｃｌｅａｓｅ処理の影響。縦軸は断片の誤認識率（リードペアのグループに異なるｉｎｄｅｘが含まれる割合（％））を示す。
【図２０】断片の誤認識率に対するＭＢＮ処理の影響。縦軸は断片の誤認識率（リードペアのグループに異なるｉｎｄｅｘが含まれる割合（％））を示す。
【図２１】断片の誤認識率に対する$ＲｅｃＪ_f$処理の影響。縦軸は断片の誤認識率（リードペアのグループに異なるｉｎｄｅｘが含まれる割合（％））を示す。
【図２２】Ｓ１　ｎｕｃｌｅａｓｅ処理を行った断片の誤認識率に対する初期ＤＮＡ量の影響。縦軸は断片の誤認識率（リードペアのグループに異なるｉｎｄｅｘが含まれる割合（％））を示す。
【図２３】断片の誤認識率に対するＳ１　ｎｕｃｌｅａｓｅ＋$ＲｅｃＪ_f$処理の影響。縦軸は断片の誤認識率（リードペアのグループに異なるｉｎｄｅｘが含まれる割合（％））を示す。横軸は使用した$ＲｅｃＪ_f$のユニット数を表す。
【図２４】Ｓ１　ｎｕｃｌｅａｓｅ＋$ＲｅｃＪ_f$処理ＤＭＳＯ暴露ライブラリにおける６変異パターンの変異頻度。凡例は、各バーで示すデータに用いた$ＲｅｃＪ_f$のユニット数を表す。
【発明を実施するための形態】
【００１４】
（１．定義）
本明細書において、「変異（又は突然変異）」（mutation）とは、ＤＮＡに生じる突然変異をいい、例えば、ＤＮＡにおける塩基又は配列の欠失、挿入、置換、付加、逆位、及び転座が挙げられる。本明細書における変異は、１塩基の欠失、挿入、置換、付加、ならびに２以上の塩基からなる配列の欠失、挿入、置換、付加、逆位、及び転座を包含する。また本明細書における変異には、遺伝子のコード領域及び非コード領域における変異が含まれ、また発現するアミノ酸の変化を伴う変異、及び発現するアミノ酸の変化を伴わない変異（サイレント変異）が含まれる。
【００１５】
本発明において評価される物質の「遺伝毒性」とは、該物質が変異を引き起こす性質（いわゆる変異原性）をいう。
【００１６】
本明細書において、「参照配列」とは、解析の対象であるＤＮＡ中に含まれる既知の配列である。当該既知の配列としては、公共のデータベース等に登録されている配列を使用することが好ましいが、予めシーケンサー等で配列決定した解析対象ＤＮＡ中の配列であってもよい。該参照配列の領域や長さ、その数は特に限定されず、解析の目的に応じてＤＮＡ中から適宜選択され得る。
【００１７】
本明細書において、ＰＣＲで得られる「増幅断片」とは、鋳型ＤＮＡのＰＣＲ増幅により得られた２本鎖ＤＮＡ断片をいう。
【００１８】
本明細書において、ＤＮＡ又はその断片に関する「２本の相補鎖」とは、２本鎖のＤＮＡ又はその断片を構成する互いに相補的な２本の１本鎖をいう。
【００１９】
本明細書において、「生リード配列」とは塩基配列のシーケンシングにより読み出された配列情報をいう。また、本明細書において、「リード配列」とは、生リード配列に対し

て、ＰＣＲやシーケンシング反応のために付加したアダプター配列やクオリティの低い塩基等のトリミングなどを行って、生リード配列からシーケンシング対象である塩基配列の情報を取り出したものをいう。ただし、上記のトリミング等の必要がない場合、生リード配列をそのままリード配列として用いることも可能である。また、生リード配列にシーケンシング対象塩基配列の配列情報が複数含まれる場合、それら個々のシーケンシング対象塩基配列の配列情報を個々のリード配列として取り出すことができ、その場合１つの生リード配列から１つ以上のリード配列が作成され得る。したがって基本的には、本明細書におけるリード配列は、サンプルＤＮＡ断片にアダプター配列等が付加される場合でも、該アダプター配列等の配列情報を含まず、サンプルＤＮＡ断片に由来する塩基配列の情報のみを含む。リード配列は、シーケンシング対象である塩基配列（例えば、サンプルＤＮＡ断片の塩基配列）のいずれかの末端の塩基から始まる塩基配列の情報を有する。リード配列の長さは、通常、シーケンサーの性能や仕様に依存する。したがって、リード配列は、場合によっては、シーケンシング対象である塩基配列の一方の末端の塩基から他方の末端の塩基までの配列（全配列）の情報を有していてもよいが、必ずしもその必要はない。

【００２０】

本明細書において、リード配列の「先頭」及び「末尾」とは、それぞれ、該リード配列の作成時に最初に読み取られた末端、及び最後に読み取られた末端をいう。本明細書において、リード配列に関する「配列の向き」とは、該リード配列をマッピングしたＤＮＡ配列における該リード配列の先頭から末尾への方向をいう。

【００２１】

本明細書において、２個以上のリード配列が「サンプルＤＮＡ上の同一領域の配列情報を有する」とは、サンプルＤＮＡの配列（又は参照配列）上においてそれらのリード配列の両末端が配置すると推定される位置が同一であることをいう。該「サンプルＤＮＡ上の同一領域の配列情報を有する」とは、該２個以上のリード配列が１００％配列同一であることを要求しないが、一方、両末端が配置すると推定される位置が１ｂｐでも異なるリード配列は、「サンプルＤＮＡ上の同一領域の配列情報を有する」ものではない。

【００２２】

本明細書において、２個以上のリード配列が「参照配列上の同一の位置にマッピングされる」とは、参照配列にマッピングしたときに、それらのリード配列の先頭と末尾の位置がそれぞれ、参照配列上で同一の位置に配置されることをいう。

【００２３】

本明細書において、「リードペア」とは、１つのシーケンシング対象配列から読み取られた２本のリード配列のペアをいう。リードペアに含まれる該２本のリード配列の一方は、該対象配列を５'末端側から３'側へ読んだ配列に相当する配列情報を含むリード配列（本明細書において「リード１」と称する）であり、他方は、同じ一方の鎖の配列を３'末端側から５'側へ読んだ配列に相当する配列情報を含むリード配列（本明細書において「リード２」と称する）である。

【００２４】

本明細書において、ＤＮＡ、配列又は断片の「リード１の先頭とリード２の先頭とに挟まれる領域」とは、リード１とリード２をマッピングした該ＤＮＡ、配列又は断片における、リード１の先頭が配置される部位からリード２の先頭が配置される部位までの領域（リード１の先頭が配置される部位とリード２の先頭が配置される部位とを含む）をいう。

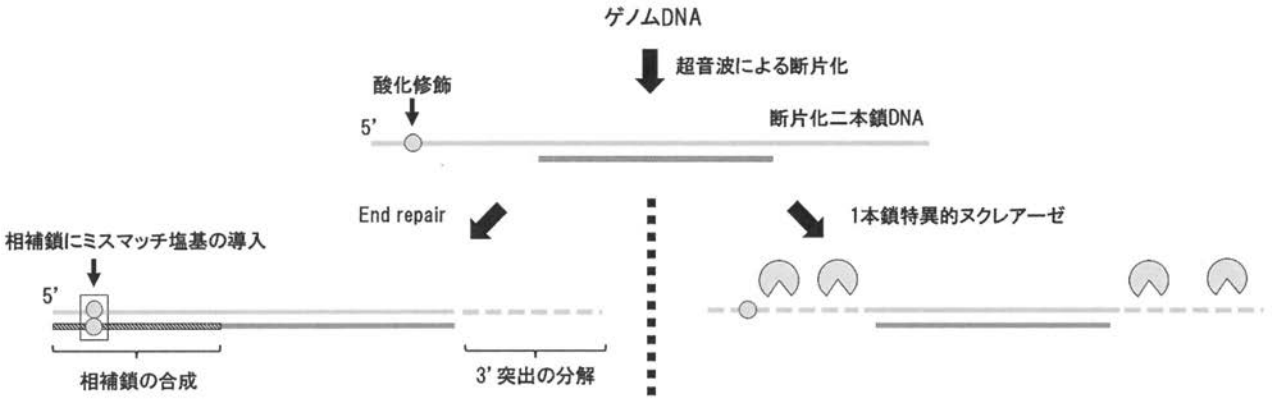
【００２５】

本明細書において、２個以上のリードペアが「サンプルＤＮＡ上の同一領域の配列情報を有する」とは、それらのリードペアの間で、サンプルＤＮＡ配列（又は参照配列）上の「リード１の先頭とリード２の先頭とに挟まれる領域が同一」であることを意味する。２個以上のリードペアが「サンプルＤＮＡ上の同一領域の配列情報を有する」とき、それらのリードペア間でリード配列が１００％配列同一であることは必ずしも要求されない。一方、該「リード１の先頭とリード２の先頭とに挟まれる領域」の末端の位置が１ｂｐでも異なるリードペアは、「サンプルＤＮＡ上の同一領域の配列情報を有する」ものではない

。
【００２６】
本明細書において、２個以上のリードペアが「参照配列上の同一の位置にマッピングされる」とは、参照配列にマッピングしたときに、それらのリードペアの両末端がそれぞれ、参照配列上で同一の位置に配置されることをいう。「リードペアの両末端」は、リード１及びリード２の読み込み開始位置に相当する。
【００２７】
本明細書において、酵素の「ユニット（Ｕ）数」とは、該酵素の活性（触媒活性とも言う）の単位を指しており、酵素ごとにその定義は異なり得る。
【００２８】
本明細書中で引用された全ての特許文献、非特許文献、及びその他の刊行物は、その全体が本明細書中において参考として援用される。
【００２９】
（２．ライブラリ調製方法）
シーケンシング用ライブラリ調製の過程で、ＤＮＡ断片の端部の１本鎖突出部位に酸化修飾等の塩基の修飾が生じた場合、末端修復工程における当該修飾塩基の誤った塩基とのペア形成、及び該誤った塩基を有する鎖のＰＣＲ増幅により、２本の相補鎖に変異が起きた場合と同じ相補鎖情報を有するライブラリが調製される。このようなライブラリは、相補鎖情報を活用したシーケンシングにおいても取り除くことができないエラーをもたらし得る。本発明者は、相補鎖情報を活用したシーケンシングにおいて、ＧＣ→ＴＡ、ＧＣ→ＣＧの変異において、Ｃ→Ａ、Ｃ→Ｇに比べて、Ｇ→Ｔ、Ｇ→Ｃの変異が高頻度に検出されることを確認した（図１）。これら高頻度の変異は、グアニンが酸化修飾されたことに起因するエラーと考えられた。このエラーの原因として、シーケンシング用ライブラリの調製過程で断片化されたサンプルＤＮＡの末端に１本鎖突出が生じ、該１本鎖突出部位のグアニンが酸化修飾されたためと考えられた（下記概念図１左）。
【００３０】
こうした末端突出部位のエラーを除くため、本発明者は、従来のアプローチ（非特許文献３）に従って、シーケンシングで得られたリードペアの両端の１０～２０塩基を除去した。その結果、除去した塩基数に依存して、Ｇ→Ｔ、Ｇ→Ｃの変異頻度が減少した（図２）。この結果は、Ｇ→Ｔ、Ｇ→Ｃの変異がリードペアの両端部に多く存在していることを示し、これらの変異がＤＮＡ断片の末端の１本鎖突出部位のグアニンの酸化修飾に起因するエラーであることを支持した。しかし、この従来のアプローチでは、両端をそれぞれ２０塩基除去したとしてもエラーによるグアニンの変異を十分に低減することはできなかった（図２）。リードペアから除去する塩基数の増加によってエラーをより低減できると予想されるが、リードペアの塩基数の減少は、ＤＮＡ解析の効率や精度を低下させる。
【００３１】
（２－１）概要
本発明においては、ライブラリ調製の過程で、サンプルＤＮＡを超音波等により断片化してＤＮＡ断片を調製したのちに、該サンプルＤＮＡ断片を１本鎖特異的ヌクレアーゼで処理して、その１本鎖部分を除去することにより、酸化修飾等によるシーケンシングのエラーを効率的に低減する（概念図１右）。
【００３２】

【化１】

概念図１



【００３３】

（２－２）サンプルＤＮＡ

本発明によるライブラリの調製方法で用いられる「サンプルＤＮＡ」は、２本鎖ＤＮＡであればよく、その由来は動物、植物、微生物などを含み、特に限定されない。該サンプルＤＮＡの種類としては、ゲノムＤＮＡ、ミトコンドリアゲノムＤＮＡ、葉緑体ゲノムＤＮＡ、プラスミドＤＮＡ、ウイルスゲノムＤＮＡ、合成ＤＮＡなどが挙げられ、限定されないが、ゲノムＤＮＡが好ましい。

【００３４】

好ましくは、該サンプルＤＮＡは、体内で、又は細胞や組織サンプルの調製及び保存過程で、分解や損傷を受けていないか又は分解や損傷が低頻度であるＤＮＡ（以下の本明細書において、「新鮮な」ＤＮＡともいう）である。好ましくは、該「新鮮な」ＤＮＡは、１本鎖に分解された領域をほとんど含まない、ほぼ完全な２本鎖で存在するＤＮＡをいう。例えば、該「新鮮な」ＤＮＡは、生細胞から抽出したＤＮＡ（例えば、生体、器官、組織、又はそれらから採取した細胞から固定処置等を経ることなく直接抽出したＤＮＡ、細菌等の微生物の細胞から直接抽出したＤＮＡ）、凍結細胞から抽出したＤＮＡ（例えば、凍結保存した生体から抽出したＤＮＡ、生体から採取した器官、組織又は細胞を採取後すみやかに凍結保存したものから抽出したＤＮＡ）、及びそれらの抽出したＤＮＡの保存サンプル（例えば、凍結保存、溶剤や不活性ガス中での低温保存をしたサンプル等）、などのＤＮＡの化学修飾や分解を促進する環境に長期間さらされていないＤＮＡであり得る。一方、ＦＦＰＥサンプル由来ＤＮＡのようなホルマリン固定された細胞由来のＤＮＡ、及びｃｆＤＮＡのような一定期間血液中に存在していたＤＮＡは、該「新鮮な」ＤＮＡからは除外され得る。あるいは、「新鮮な」ＤＮＡは、Ａｇｉｌｅｎｔ　４２００　ＴａｐｅＳｔａｔｉｏｎ、又はＡｇｉｌｅｎｔ　２２００　ＴａｐｅＳｔａｔｉｏｎ（いずれもアジレント・テクノロジー社製）で分析したＤＮＡ　Ｉｎｔｅｇｒｉｔｙ　Ｎｕｍｂｅｒ（以下、本明細書において「ＤＩＮ」という）が、好ましくは６以上、より好ましくは７以上、さらに好ましくは７．３以上、さらにより好ましくは７．５以上のＤＮＡとして定義することができる。

【００３５】

サンプルＤＮＡは、細胞から当該分野における通常の方法を用いて抽出又は単離することによって取得することができる。該抽出又は単離には、例えば、市販のＤＮＡ抽出キットなどを用いることができる。あるいは、細胞から抽出又は単離後保存されているＤＮＡを取得し、本発明の方法で使用してもよい。合成ＤＮＡは、公知の化学合成法により合成することができる。

【００３６】

あるいは、本発明の方法では、２本鎖ＤＮＡの代わりに、２本鎖ＲＮＡを用いてもよい。２本鎖ＲＮＡは、それを保有するウイルスや細胞などから、市販のＲＮＡ抽出キットなど、当該分野における通常の方法で抽出又は単離することができる。あるいは、抽出又は単離後保存されている２本鎖ＲＮＡを取得し、本発明の方法で使用してもよい。本発明の方法においてＲＮＡを取得及び解析する場合、取得されたＲＮＡはＰＣＲ前にｃＤＮＡに変換され、該ｃＤＮＡ由来のリード配列の塩基Ｔは塩基Ｕと読み替えられる。

【００３７】

（２－３）ＤＮＡ断片の調製

サンプルＤＮＡの断片化は、超音波処理、酵素処理など、切断箇所がランダムになる当該分野における通常の方法を用いて実施することができる。ＤＮＡの断片化処理の具体的な例としては、コバリス社のＤＮＡ　Ｓｈｅａｒｉｎｇシステムなどを用いた集中超音波処理等が挙げられる。調製する断片の長さは、シーケンサーが精度よく読み取れる長さに応じて適宜選択され得る。一般的には、１００～１０，０００ｂｐが選択され得るが、シーケンサーが精度よく読み取れる限りは１０，０００ｂｐ以上の長さの断片が調製されてもよく、シーケンサーの種類に依存してより適切な範囲が選択され得る。例えば、断片の増幅を行うシーケンシング反応用のシーケンサーにかける場合は、断片の長さは平均長１００～１０００ｂｐが好ましく、平均長２００～５００ｂｐがより好ましい。あるいは、より長い断片を調製し、これを後述するＰＣＲにかけ、シーケンシング反応に適切な長さのＰＣＲ産物を調製してもよい。

【００３８】

（２－４）ヌクレアーゼ処理

本発明の方法では、上述した新鮮なサンプルＤＮＡを断片化した後、得られた断片を１本鎖特異的ヌクレアーゼで処理し、該断片から１本鎖部分を除去する。従来の１本鎖特異的ヌクレアーゼ処理は、ＦＦＰＥサンプルのＤＮＡやｃｆＤＮＡ等の比較的分解や損傷を受けており、既に断片化した状態でサンプル中に存在するＤＮＡを対象としていた。本発明のように分解の程度が低い新鮮なＤＮＡを、ライブラリ調製のために１本鎖特異的ヌクレアーゼで処理したことはこれまで報告されていない。

【００３９】

本発明において、１本鎖特異的ヌクレアーゼによる分解の対象は、主にＤＮＡ断片の末端に存在する１本鎖突出部位であり得るが、この限りではない。例えば、ＤＮＡ断片の両端以外（例えば中心部分）に存在する１本鎖部分も、本発明における１本鎖特異的ヌクレアーゼによる除去の対象であり得、その除去はエラー低減に寄与し得る。例えば、２本鎖ＤＮＡ断片の片方の鎖にニックが存在する場合、後述する末端修復工程（例えば、Ｅｎｄ　Ｒｅｐａｉｒカクテルの酵素での処理）の際にニック以降の鎖が再度合成され、エラー率増加に寄与する可能性がある。当該ニックにおける１本鎖部分を１本鎖特異的ヌクレアーゼで除去することは、エラー低減に寄与し得る。

【００４０】

本発明の方法で使用可能な１本鎖特異的ヌクレアーゼは、１本鎖特異的に作用する限り、エンドヌクレアーゼであってもエキソヌクレアーゼであってもよい。１本鎖特異的エンドヌクレアーゼの例としては、Ｓ１　ｎｕｃｌｅａｓｅ、Ｍｕｎｇ　Ｂｅａｎ　Ｎｕｃｌｅａｓｅ（ＭＢＮ）などが挙げられ、１本鎖特異的エキソヌクレアーゼの例としてはＥｘｏｎｕｃｌｅａｓｅ　VIIなどが挙げられ、１本鎖特異的５'→３'エキソヌクレアーゼの例としてはＲｅｃＪ$_f$などが挙げられる。このうち、１本鎖への特異性が高い点及び２本鎖で挟まれた１本鎖も除去できる点から、Ｓ１　ｎｕｃｌｅａｓｅ及びＭＢＮが好ましく、Ｓ１　ｎｕｃｌｅａｓｅがより好ましい。これらの１本鎖特異的ヌクレアーゼは市販されており、例えば、プロメガ社、タカラバイオ社、Ｎｅｗ　Ｅｎｇｌａｎｄ　Ｂｉｏｌａｂｓ社などから購入することができる。当該１本鎖特異的ヌクレアーゼ処理においては、１種類の酵素のみを用いてもよいが、複数種の酵素を組み合わせて用いてもよい。例えば、サンプルＤＮＡ断片をＳ１　ｎｕｃｌｅａｓｅ、ＭＢＮなどのエンドヌクレアーゼで処理した後、さらにＲｅｃＪ$_f$などのエキソヌクレアーゼで処理することが好ましく、その

逆も同様である。

【００４１】

サンプルＤＮＡ断片の１本鎖特異的ヌクレアーゼによる処理は、通常の手順で、例えば購入元の提供するプロトコルに従って、実施することができる。反応条件は、酵素の至適条件や、基質であるＤＮＡ断片の量に従って適宜決定することができる。例えば、反応液におけるサンプルＤＮＡ断片１ｎｇあたりの酵素活性単位（ユニット数；Ｕ／ｎｇ）は、Ｓ１　ｎｕｃｌｅａｓｅでは、シーケンシングエラー低減の観点からは好ましくは０．０１Ｕ／ｎｇ以上、より好ましくは０．０２Ｕ／ｎｇ以上、さらに好ましくは０．０３Ｕ／ｎｇ以上、さらに好ましくは０．０５Ｕ／ｎｇ以上、さらに好ましくは０．１０Ｕ／ｎｇであり、一方、現実的に実施可能な上限値、及び、高濃度時に非特異的に生じる２本鎖ＤＮＡの分解の観点からは１６．７Ｕ／ｎｇ以下が好ましく、反応効率の観点からは、好ましくは５．００Ｕ／ｎｇ以下、より好ましくは１．６７Ｕ／ｎｇ以下、さらに好ましくは１．００Ｕ／ｎｇ以下、さらに好ましくは０．３０Ｕ／ｎｇ以下である。あるいは、酵素反応の効率の観点から好ましいＳ１　ｎｕｃｌｅａｓｅの酵素量の範囲は、０．０２～５．００Ｕ／ｎｇ、より好ましくは０．０３～１．６７Ｕ／ｎｇ、さらに好ましくは０．０３～１．００Ｕ／ｎｇ、さらに好ましくは０．０５～１．００Ｕ／ｎｇ、さらに好ましくは０．１０～０．３０Ｕ／ｎｇである。また例えば、反応液におけるＭＢＮの酵素量は、シーケンシングエラー低減の観点からは好ましくは０．０１Ｕ／ｎｇ以上、より好ましくは０．０２Ｕ／ｎｇ以上、さらに好ましくは０．０３Ｕ／ｎｇ以上、さらに好ましくは０．０５Ｕ／ｎｇ以上、さらに好ましくは０．１０Ｕ／ｎｇ以上であり、一方、現実的に実施可能な上限値、及び、高濃度時に非特異的に生じる２本鎖ＤＮＡの分解の観点からは１６．７Ｕ／ｎｇ以下が好ましく、反応効率の観点からは、好ましくは、５．００Ｕ／ｎｇ以下、より好ましくは１．６７Ｕ／ｎｇ以下、さらに好ましくは１．００Ｕ／ｎｇ以下、さらに好ましくは０．３０Ｕ／ｎｇ以下である。あるいは、酵素反応の効率の観点から好ましいＭＢＮの酵素量の範囲は、０．０２～５．００Ｕ／ｎｇ、より好ましくは０．０３～１．６７Ｕ／ｎｇ、さらに好ましくは０．０３～１．００Ｕ／ｎｇ、さらに好ましくは０．０５～１．００Ｕ／ｎｇ、さらに好ましくは０．１０～０．３０Ｕ／ｎｇである。また例えば、反応液におけるＲｅｃＪ$_f$の酵素量は、シーケンシングエラー低減の観点からは好ましくは０．１０Ｕ／ｎｇ以上、より好ましくは０．３０Ｕ／ｎｇ以上であり、一方、現実的に実施可能な上限値の観点からは１００Ｕ／ｎｇ以下が好ましく、推奨のＤＮＡの量に近い条件（６０ｎｇ）で酵素反応を行う観点からは１６．７Ｕ／ｎｇ以下が好ましく、反応効率の観点からは１．００Ｕ／ｎｇが好ましい。あるいは、酵素反応の効率の観点から好ましいＲｅｃＪ$_f$の酵素量の範囲は、０．１０～１６．７Ｕ／ｎｇ、より好ましくは０．３０～１．００Ｕ／ｎｇである。なお本明細書において、酵素活性１単位（１Ｕ）は以下の通り定義される：

・Ｓ１　ｎｕｃｌｅａｓｅ：３０ｍＭ酢酸ナトリウム（ｐＨ４．６、２５℃）、５０ｍＭ　ＮａＣｌ、１ｍＭ　ＺｎＣｌ$_2$、５％グリセロール、０．５ｍｇ／ｍＬ変性仔牛胸腺ＤＮＡの混合溶液中において、３７℃で１分間に１μｇの酸可溶性物質を生成する酵素活性。

・ＭＢＮ：熱変性仔牛胸腺ＤＮＡを基質として、３７℃、ｐＨ５．０において、１分間に１μｇの酸可溶性分解物を生成する酵素活性。

・ＲｅｃＪ$_f$：全反応液５０μＬ（１×ＮＥ　Ｂｕｆｆｅｒ　２及び１．５μｇの超音波処理［^{3}Ｈ］標識１本鎖Ｅ．ｃｏｌｉ　ＤＮＡを含む）中、３７℃、１分間で、０．５ｎｇのトリクロロ酢酸可溶性デオキシリボヌクレオチドを生成する酵素活性。

【００４２】

さらに、反応に用いる酵素のユニット数は、後述する増幅（ＰＣＲ）工程の初期ＤＮＡ量と関連し得る。例えばＳ１　ｎｕｃｌｅａｓｅでは、ユニット数が０．０５Ｕ／ｎｇより大きい場合、下記式で算出される指標：

指標＝初期ＤＮＡ量（ａｍｏｌ／ＭｂｐサンプルｄｓＤＮＡ）×$3^{\log S1\ nuclease\ (U/ng)}$

（式中、Ｓ１　ｎｕｃｌｅａｓｅ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である）が

、好ましくは６０以下、より好ましくは３０以下、さらに好ましくは１５以下、さらにより好ましくは７．５以下である。また、例えばＭＢＮでは、ユニット数が０．０５Ｕ／ｎｇより大きい場合、下記式で算出される指標：

指標＝初期ＤＮＡ量（ａｍｏｌ／Ｍｂｐサンプル ＤＮＡ）×$3^{\log MBN\ (U/ng)}$

（式中、ＭＢＮ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である）が、好ましくは６０以下、より好ましくは３０以下、さらに好ましくは１５以下、さらにより好ましくは７．５以下である。一方、０．０５Ｕ／ｎｇ以下のＳ１　ｎｕｃｌｅａｓｅもしくはＭＢＮ、又はユニット数に関わらず$RecJ_f$を用いる場合、前記の式は成立せず、後述する増幅（ＰＣＲ）工程の初期ＤＮＡ量は、サンプルＤＮＡ　１Ｍｂｐあたり、好ましくは２５０ａｍｏｌ以下、より好ましくは１２５ａｍｏｌ以下、さらに好ましくは６２．５ａｍｏｌ以下、さらにより好ましくは３１．３ａｍｏｌ以下、なお好ましくは１５．７ａｍｏｌ以下である。

【００４３】

反応後の酵素は、失活させるか又は洗浄除去することが望ましい。ヌクレアーゼ処理したＤＮＡ断片は、その後のＰＣＲ工程に用いることができるように精製する。ＤＮＡの精製には、エタノール沈殿、電気泳動、カラム精製、ビーズ精製、アフィニティー精製などの通常の手段を用いることができる。

【００４４】

（２－５）追加処理

本発明においては、上記サンプルＤＮＡ断片の１本鎖特異的ヌクレアーゼ処理以降は、通常の手順に従って、シーケンシング用ライブラリを調製することができる。例えば、１本鎖特異的ヌクレアーゼ処理したＤＮＡ断片を、必要に応じて、末端修復、末端への塩基付加、増幅などの処理にかけて、ライブラリを調製する。好ましくは、該末端修復、末端への塩基付加、及び増幅が、この順序で全て行われる。該末端修復、末端への塩基付加、及び増幅の工程は、ＴｒｕＳｅｑ　Ｎａｎｏ　ＤＮＡ　Ｌｉｂｒａｒｙ　Ｐｒｅｐ　Ｋｉｔ（イルミナ社）などの市販の試薬を用いて実施することができる。

【００４５】

（２－５－１）末端修復

サンプルＤＮＡ断片は、１本鎖特異的ヌクレアーゼで処理した後にも、末端に短い１本鎖突出部位が残存することがある。末端修復では、該ヌクレアーゼ処理後のＤＮＡ断片において、該残存する１本鎖突出部位を有する末端を平滑化する。該平滑化処理では、一般に、Ｔ４　ＤＮＡポリメラーゼ等の３'→５'エキソヌクレアーゼにより３'側突出末端が除去され、一方、５'側突出末端は、５'→３'ポリメラーゼにより対となる鎖が合成され、これによりＤＮＡ断片の両端が平滑化される。

【００４６】

（２－５－２）塩基付加

末端への塩基付加は、末端平滑化したＤＮＡ断片に対して、その両端へのシーケンシングに必要な標識配列の付加や、該標識配列を付加するための３'末端へのアデニンの付加を行う処理である。標識配列が付加されたＤＮＡ断片を増幅し、シーケンシングすることで、該ＤＮＡ断片の配列情報と該標識配列の情報とを取得することができ、また該標識配列の情報に従って、リード配列を識別又は分類することができる。例えば、ＤＮＡ断片の両末端に付加した標識配列は、リード配列が該ＤＮＡ断片の全配列の情報を有するかを判断する指標となる。あるいは、ＤＮＡ断片の片方の末端に標識配列を付加し、該標識配列を含まない側からシーケンシングすることで、リード配列が該サンプルＤＮＡ断片の全配列の情報を有するかを判断することができる。

【００４７】

相補鎖情報を活用したシーケンシングに用いるライブラリの調製においては、サンプルＤＮＡの断片の両末端に、リード配列が該断片の２本の相補鎖のいずれに由来するかを識別可能にする標識配列を付加させることが好ましい。例えば、１つのＤＮＡ断片を構成する２本の相補鎖の５'末端側と３'末端側にそれぞれ異なる標識配列を付加させる。一実施

形態においては、１つのＤＮＡ断片の両鎖の間で５'末端側の標識配列は同一であり、両鎖の３'末端側の標識配列も同一であり、かつ両端の標識配列は互いに相補的でない配列を含む（以下の本明細書において、これを「相補鎖標識配列」と呼ぶ；下記概念図２参照）。好ましくは、該相補鎖標識配列においては、標識した各ＤＮＡ断片の間で、５'末端側の標識配列は共通であり、かつ３'末端側の標識配列も共通である。よって、各断片を構成する２つの１本鎖は、それぞれ５'末端側及び３'末端側に異なる標識配列を有するが、該５'末端側の標識配列と該３'末端側の標識配列は各１本鎖間で共通である。一方、該相補鎖標識配列は、該リード配列がサンプルＤＮＡのいずれの個別断片に由来するかを識別する必要はない。このような相補鎖標識配列の例としては、イルミナ社のＴｒｕＳｅｑに付属のアダプター配列が挙げられる。

【００４８】

別の一実施形態においては、サンプルＤＮＡの断片を個別に識別する標識配列（以下の本明細書において、これを「個別断片標識配列」と呼ぶ；例えば、PNAS, 109(36):14508-14513, 2012、又は特許文献１に記載されるような、サンプルＤＮＡ断片固有のタグ配列）をＤＮＡ断片に付加することができる。このような標識は、リード配列がＤＮＡ断片の２本の相補鎖のいずれに由来するかを識別させ、相補鎖情報を活用したシーケンシングを可能にする。ただし、相補鎖情報を活用したシーケンシングの効率の観点からは、特にサンプルＤＮＡのサイズが大きい場合、相補鎖標識配列を用いることが好ましい。

【００４９】

（２－５－３）増幅

ＤＮＡ断片の増幅には、ＰＣＲ等の既存の方法を用いることができる。得られた増幅断片は、必要に応じて通常の手順で精製し、シーケンシング用ライブラリとして用いることができる。ＰＣＲは、市販のＰＣＲ用試薬や機器を用いて、常法に従って実施することができる。あるいは、ＰＣＲ増幅装置を備えたシーケンサーを用いてもよい。サンプルＤＮＡの断片のＰＣＲ増幅をその工程に含む高スループットシーケンサーとしては、ＨｉＳｅｑ（イルミナ社製）、ＭｉＳｅｑ（イルミナ社製）などが上市されている。

【００５０】

好ましくは、当該ＰＣＲにおいては、鋳型として使用されたＤＮＡ断片の各々について、２つ以上の増幅断片がそれぞれ作製される。このとき、鋳型として用いたサンプルＤＮＡの断片の少なくとも一部の各々について２つ以上の増幅断片が調製されればよい。一方、該ＰＣＲで全部の鋳型サンプルＤＮＡの断片について２つ以上の増幅断片を得てもよいが、その必要はない。サンプルＤＮＡの断片のＰＣＲ増幅をその工程に含む高スループットシーケンサーでは、シーケンシング反応に用いるＰＣＲ産物量を一定量用いることがシーケンシング効率の点で推奨されている。そのため、ＰＣＲにかけるサンプルＤＮＡ量（ＰＣＲでの初期ＤＮＡ量）に応じてＰＣＲのサイクル数を変更し、ＰＣＲ産物量を推奨量にあわせることが好ましい。

【００５１】

（３．シーケンシング方法）

上記の手順で得られたライブラリを用いてシーケンシングを実施することができる。本発明で得られたライブラリは、各種シーケンシング方法に適用可能である。好ましくは、本発明で得られたライブラリは、相補鎖情報を活用したシーケンシング（例えば、特許文献４に記載のシーケンシング方法）に用いられる。以下に、特許文献４を参考に、本発明で得られたライブラリを用いた、相補鎖情報を活用したシーケンシング方法（以下、本シーケンシング方法という）の概要を説明する。

【００５２】

（３－１）概要

本シーケンシング方法は、基本的には、本発明で得られたライブラリをシーケンシングし、該ライブラリに含まれる各サンプルＤＮＡの断片由来の複数の増幅断片の各々について１つ以上の読み取り結果（リード配列）を作成し、複数の増幅断片についての複数のリード配列を得ること；該シーケンシングで得られたリード配列の中から、該サンプルＤＮ

Ａ上の同一領域の配列情報を有するリード配列を集めること；集めたリード配列の情報を用いて、該サンプルＤＮＡの配列情報を構築すること、を含む。

【００５３】

（３－２）シーケンシング及びリード配列の作成

ライブラリのシーケンシングは、解析等に必要な部分、例えば後述する変異解析の場合、参照配列との配列比較に使用すべき部分について行えば足りる。例えば、その配列の少なくとも一部、好ましくは全体が、参照配列のＤＮＡ領域に対応する断片をシーケンシングすればよい。哺乳動物細胞等の場合には、エクソン領域等を選択的にシーケンシングしてもよい。領域の選択には、ＳｕｒｅＳｅｌｅｃｔ（アジレント・テクノロジー社製）等のキットが上市されている。

【００５４】

該シーケンシングにより、ライブラリについての生リード配列が取得される。該生リード配列から、ＰＣＲやシーケンシング反応のために付加したアダプター配列やクオリティの低い塩基等のトリミングなどを行ってサンプルＤＮＡの断片に由来する配列を取り出すことで、リード配列が作成される。あるいは、上記トリミング等の必要がない場合、生リード配列をそのままリード配列として用いてもよい。該生リード配列又はリード配列が作成される増幅断片は、該ライブラリに含まれる増幅断片のうちの少なくとも一部である複数の増幅断片であればよい。一方、該ライブラリに含まれる全増幅断片についてリード配列を取得してもよいが、その必要はない。該リード配列は、該複数の増幅断片の各々に対して１つ以上作成される。それらのリード配列は、該増幅断片（すなわちそれが由来するサンプルＤＮＡの断片）の２本の相補鎖のいずれかについての配列情報を有する。したがって、該ライブラリのシーケンシングにより、複数のリード配列が得られる。なお、この段階で得られた該複数のリード配列を含むデータを、本明細書において「シーケンシングデータ」と呼ぶことがある。

【００５５】

（３－３）リード配列のグループ化

次いで、得られた複数のリード配列の中から、各リード配列の配列情報に基づいて、サンプルＤＮＡ上の同一領域の配列情報を有するリード配列を集める。集めたリード配列は、グループ化される。したがって、本発明の方法で作成される「リード配列のグループ」とは、サンプルＤＮＡ上の同一領域の配列情報を有するリード配列の集合であり、言い換えると、同一のサンプルＤＮＡ断片に由来すると推定されるリード配列の集合である。本発明の方法においては、通常、ライブラリ調製の際にＰＣＲにかけたサンプルＤＮＡ断片の数とシーケンシングデータの量に依存して、１つ以上のリード配列のグループが作成され得る。

【００５６】

本発明の方法の一実施形態においては、ライブラリに含まれる１増幅断片に対して、１本以上のリード配列が作成され、得られたリード配列は上述のようにグループ化される。好ましい実施形態においては、上述したリード配列のグループの作成に利用されるリード配列は、元のサンプルＤＮＡの断片（すなわち該リード配列が由来する増幅断片の元となるサンプルＤＮＡの断片）の全配列の情報を有するリード配列である。シーケンシングで得られたリード配列の中から元のサンプルＤＮＡの断片の全配列の情報を有するリード配列を選抜する手順としては、リード配列の末尾の塩基の読み取り精度（クオリティ値）が高いリード配列を選別する方法、末端に標識配列を付加したライブラリを調製し、これをシーケンシングし、該標識配列の情報の有無に基づいてリード配列を選別する方法、などが挙げられる。このうち、標識配列を用いた方法についてより具体的な手順の例を説明する：まず、サンプルＤＮＡの断片の両末端にそれぞれ異なる標識配列を付加し、これをＰＣＲ増幅することにより、両末端に該標識配列を有する増幅断片を含むライブラリを調製する；得られたライブラリをシーケンシングし、該増幅断片由来のリード配列と、それに付随する該標識配列の情報を取得する。該両末端の標識配列の両方の情報が付随するリード配列は、元のサンプルＤＮＡの断片の全配列の情報を有するリード配列とみなされる。

別の例では、サンプルＤＮＡの断片の片方の末端に標識配列を付加し、これをＰＣＲ増幅して該標識配列を含む増幅断片を調製する；得られた増幅断片を、該標識配列のない末端の側からシーケンシングする。該標識配列の情報が付随するリード配列は、元のサンプルＤＮＡ断片の全配列の情報を有するリード配列とみなされる。ここで該標識配列の情報は、生リード配列から取得してもよく、又はシーケンシングプライマーの配列情報から取得してもよい。

【００５７】

集めたリード配列からリード配列のグループを作成する手段としては、例えば、参照配列上の同一の位置にマッピングされるリード配列を集める方法、少なくとも両末端領域の配列が同等であるリード配列を集める方法、などが挙げられる。なお、「少なくとも両末端領域の配列が同等」とは、アラインさせたリード配列が、少なくとも両末端領域において配列同一性が８０％以上、好ましくは９０％以上、より好ましくは９５％以上、さらに好ましくは９７％以上であり、かつ両末端が同じ位置にアラインすることをいう。該「末端領域」の長さは適宜選択することができ、例えば末端を含め、１０塩基以上、好ましくは１０～３０塩基程度であればよい。あるいは、配列全体の同一性が８０％以上、好ましくは９０％以上、より好ましくは９５％以上、さらに好ましくは９７％以上であり、かつ両末端が同じ位置にアラインするリード配列を集めることで、リード配列のグループを作成してもよい。

【００５８】

（３－４）リード配列のグループからのサンプルＤＮＡ配列情報の抽出

次に、得られたリード配列のグループから、サンプルＤＮＡの配列情報を抽出する。詳細には、該リード配列のグループに含まれるリード配列の情報を用いて１つの配列データを導き出す。得られた配列データは、該グループのリード配列が由来する特定のサンプルＤＮＡの断片についてのコンセンサス配列を表す。

【００５９】

例えば、リード配列のグループに含まれるリード配列の間で配列情報のコンセンサスを取ることで、１つの配列データを作成することができる。リード配列間でのコンセンサスを取る具体的な手法としては、以下が挙げられる：リード配列をアライメントし、アライメントした全てのリード配列の対応する塩基が一致した場合にその塩基を"コンセンサス塩基"とする方法；リード配列をアライメントした後、配列上の各位置で最大の頻度で出現する塩基を決定し、"コンセンサス塩基"として抽出する方法；リード配列をアライメントした後、対応する位置にある塩基の中でシーケンサーでの読み取り精度（クオリティ値）の最も高い塩基を"コンセンサス塩基"として採用する方法；リード配列をアライメントした後、クオリティ値や塩基の出現頻度等を基に、確率論的に"コンセンサス塩基"を決定する方法；あるいは、これらを組み合わせた方法、など。

【００６０】

リード配列間でのコンセンサスを取る際には、リード配列のグループに含まれる全てのリード配列が用いられてもよいが、該グループ内の一部のリード配列のみが用いられてもよい。リード配列間でのコンセンサスを取ることにより、シーケンシングにおける読み取りエラーなどのエラーを除外することができるので、高精度な読み取り結果を得ることができる。得られた配列データは、サンプルＤＮＡの一領域の配列を示す最終的な配列データとして取得することができる。

【００６１】

（３－５）相補鎖情報に基づくシーケンシング

シーケンシングエラーを引き起こす、ＤＮＡの酸化修飾等による塩基の置換は、基本的にはＤＮＡ２本鎖のうち片方の鎖だけに起こる。したがって、ＤＮＡの２本の相補鎖それぞれについてのシーケンシング情報を用いることで、片方の鎖にのみ発生した塩基の置換を変異として検出することなく、２本鎖に固定された真の変異のみを同定することが可能となる。ＤＮＡの２本の相補鎖の配列は、相補的であるものの、互いに等価の情報を有する。従って理論上は、シーケンシングで得られたリード配列の中から等価の情報を有する

配列を探すことにより、相補鎖の情報を得ることが可能である。例えば、ある生物種のゲノム配列からサンプルＤＮＡを調製した場合、サンプルＤＮＡの断片を構成する２本の相補鎖それぞれに由来する読み取り領域が同一である２つのリード配列は、解析対象となる生物種の参照配列にマッピングした場合には、ゲノムの同一箇所にマッピングされる。したがって、ゲノムの同一箇所にマップされ得るリード配列を集めて、それらリード配列をその由来する相補鎖によって選抜することで、２本の相補鎖のそれぞれに由来するリード配列を取得することができる。さらにそれら２本の相補鎖に由来するリード配列間でのコンセンサスをとることにより、相補鎖の情報を反映させた高精度なリード情報を得ることが可能である。

【００６２】

本シーケンシング方法では、上記（３－２）で述べたライブラリのシーケンシングの際に、各サンプルＤＮＡの断片を構成する２本の相補鎖の各々に対してリード配列を作成する。より詳細には、ライブラリのシーケンシングにより該ライブラリに含まれる該複数の増幅断片の各々についてのリード配列を作成する際に、各サンプルＤＮＡの断片を構成する２本の相補鎖の各々に由来する増幅断片に対して、それぞれ１つ以上のリード配列が作成される。すなわち、１個のサンプルＤＮＡの断片に対して２つ以上のリード配列が取得され、それらのリード配列はそれらが由来するサンプルＤＮＡの断片の２本の相補鎖の一方及び他方についての配列情報を有する。

【００６３】

次いで、得られた複数のリード配列から、１つ以上のリード配列のグループを作成する。リード配列のグループを作成する手段は、上記（３－３）で述べたとおりである。ここで得られるリード配列のグループには、特定のサンプルＤＮＡの断片についての２本の相補鎖の一方及び他方の配列情報を有するリード配列が含まれている。したがって、該リード配列のグループに含まれるリード配列の間で配列情報のコンセンサスを取ることにより、相補鎖の情報を反映させた配列データを作成することができる。リード配列間でのコンセンサスを取る具体的な手法は、上記（３－４）で述べたとおりである。リード配列間でのコンセンサスを取る際には、リード配列のグループに含まれる全てのリード配列が用いられてもよいが、該グループ内の一部のリード配列のみが用いられてもよい。

【００６４】

好ましくは、上記リード配列間でのコンセンサスを取る工程は、リード配列のグループの中から、サンプルＤＮＡの断片の２本の相補鎖の各々に由来するリード配列を少なくとも１つずつ集め、集めたリード配列の間で配列情報のコンセンサスを取ることを含む。これにより、相補鎖情報を用いたコンセンサスデータ（本明細書において「相補鎖間コンセンサスリード配列」ともいう）を得ることができる。得られた相補鎖間コンセンサスリード配列は、シーケンシングにおける読み取りエラーやＤＮＡ酸化修飾等に起因するエラーなどの片方の鎖にのみ生じるエラーが除外された高精度な読み取り結果であり、サンプルＤＮＡの一断片についての配列を示す最終的な配列データとして取得することができる。

【００６５】

リード配列のグループの中から、サンプルＤＮＡの断片の２本の相補鎖の各々に由来するリード配列を集める手順としては、例えば、以下の手順が挙げられる：予めサンプルＤＮＡの断片に２本の相補鎖を識別できる標識配列を付加することにより、該標識配列を含む増幅断片を調製する；次いで、該増幅断片をシーケンシングし、該増幅断片由来のリード配列と、それに付随する該標識配列の情報を取得する；得られたリード配列から、リード配列のグループを作成する；次いで、リード配列に付随する標識配列の情報を利用して、リード配列のグループの中から互いに相補的な鎖に由来するリード配列を集める。

【００６６】

上記の手順においては、好ましくはサンプルＤＮＡの断片に、上記（２－５－２）で述べたリード配列が該断片の２本の相補鎖のいずれに由来するかを識別可能にする標識配列（例えば、相補鎖標識配列又は個別断片標識配列）を付加する。好ましくは、相補鎖標識配列が用いられる。該標識配列が付加されたサンプルＤＮＡの断片から得られた増幅断片

をシーケンシングすることで、該増幅断片由来のリード配列とそれに付随する該標識配列の情報を取得することができる。

【００６７】

次に、当該標識配列の情報を利用して、リード配列のグループの中から互いに相補的な鎖に由来するリード配列を集める際の好ましい手順を説明する。リード配列のグループに含まれるリード配列を参照配列にマッピングするとき、５'末端側の標識配列の情報が付随し、かつその先頭が末尾に対して参照配列上でより５'側に位置するリード配列と、３'末端側の標識配列の情報が付随し、かつその先頭が末尾に対して参照配列上でより３'側に位置するリード配列は、サンプルＤＮＡの断片の２本の相補鎖のうちの同じ１本鎖に由来する。一方、３'末端側の標識配列の情報が付随し、かつその先頭が末尾に対して参照配列上でより５'側に位置するリード配列と、５'末端側の標識配列の情報が付随し、かつその先頭が末尾に対して参照配列上でより３'側に位置するリード配列は、サンプルＤＮＡの断片の２本の相補鎖のうちのもう一方の１本鎖に由来する。したがって、参照配列にマッピングされたリード配列の参照配列に対する配置と、それに付随する標識配列の情報に基づいて、リード配列のグループ内の各リード配列がサンプルＤＮＡの断片を構成する２本の相補鎖のどちらに由来するかを識別することができる。あるいは、増幅断片の末端に特定の標識配列が付加しているときにのみ開始するシーケンシング反応を行うことにより、標識配列の情報に基づいて、サンプルＤＮＡの断片の特定の１本鎖に由来するリード配列を識別することができる。このようにサンプルＤＮＡの断片の同じ１本鎖に由来するリード配列を予め識別しておくことで、リード配列のグループの中から互いに相補的な鎖に由来するリード配列を集めることができる。

【００６８】

上述したリード配列のグループから相補鎖間コンセンサスリード配列を得る手順の具体的な例としては、リード配列のグループの中から、サンプルＤＮＡの断片の２本の相補鎖の各々に由来する２本のリード配列を選択し、それら２本のリード配列の間で配列情報のコンセンサスを取ることが挙げられる。さらに、該手順を繰り返して複数の相補鎖間コンセンサスリード配列を作成した後、さらにそれらの間でのコンセンサスを取り、１つの相補鎖間コンセンサスリード配列を作成してもよい。あるいは、該相補鎖間コンセンサスリード配列を得る手順の別の具体的な例としては、リード配列のグループに含まれるリード配列を、サンプルＤＮＡの断片の２本の相補鎖の一方に由来する群と他方に由来する群とに分け、各群のリード配列の間でコンセンサスを取り、得られた２つのコンセンサスデータの間でさらにコンセンサスを取り、１つの相補鎖間コンセンサスリード配列を作成することが挙げられる。あるいはサンプルＤＮＡの断片の２本の相補鎖に由来するリード配列を特に区別せず、リード配列のグループに含まれるリード配列の間でコンセンサスを取り、コンセンサスリード配列を作成することが挙げられる。

【００６９】

（３－６）リードペアを用いたサンプルＤＮＡ配列情報の抽出

本シーケンシング方法の一実施形態においては、上記（３－２）で述べたライブラリのシーケンシングの際に、該ライブラリに含まれる該複数の増幅断片の各々に対して１本のリード配列を作成する代わりに、２本のリード配列からなるリード配列のペア（すなわち「リードペア」）が１つ作成される。作成されたリードペアから、上記と同様の原理で、サンプルＤＮＡの配列情報が抽出される。

【００７０】

当該方法においては、ライブラリのシーケンシングにより、各増幅断片に対して１つ以上のリードペアが作成される。また該１つ以上のリードペアの作成は、各サンプルＤＮＡの断片に由来する２つ以上の該増幅断片について行われる。それらのリードペアは、該サンプルＤＮＡの断片の２本の相補鎖のいずれかについての配列情報を有する。したがって、本実施形態においては、上述したライブラリのシーケンシングで得られる複数のリード配列は、複数個のリードペアを含む。

【００７１】

該リードペアを構成する２本のリード配列の一方は、該増幅断片を構成する２本の相補鎖のうちの一方の鎖の配列を５'末端側から３'側へ読んだ配列に相当する配列情報を含むリード配列（すなわち「リード１」）であり、他方は、同じ一方の鎖の配列を３'末端側から５'側へ読んだ配列に相当する配列情報を含むリード配列（すなわち「リード２」）である。リード１とリード２は、オリジナルの鎖（増幅断片を構成する１本鎖）に対して互いに逆向きに配置する。すなわち、該オリジナルの鎖に対してマッピングした場合、リード１の先頭は、その末尾に比べてより該オリジナルの鎖の５'側に配置し、一方、リード２の先頭は、その末尾に比べてより該オリジナルの鎖の３'側に配置する（後出の模式図１参照）。

【００７２】

次いで、得られたシーケンシングデータ中の複数のリードペアの中から、サンプルＤＮＡ上の同一領域の配列情報を有するリードペアを選抜する。集めたリードペアはグループ化される。該リードペアのグループを作成する手段としては、例えば、リードペアのリード１とリード２を参照配列に対してマッピングし、リード１の先頭とリード２の先頭とに挟まれる該参照配列の領域が同一であるリード配列のペアを集めて、同じグループに分ける方法が挙げられる。より詳細な手順の例としては、まず、リードペアに含まれる一方のリード配列（リード１又は２）の先頭が参照配列上の同じ位置に位置するリードペアを集め、次いで集めたリード配列のペアの中から、該リードペアに含まれるもう一方のリード配列（リード２又は１）の先頭が参照配列上の同じ位置に位置するリード配列のペアを集めて、それらを同じグループに分ける方法が挙げられる。

【００７３】

したがって、本発明の方法で作成される「リード配列のペア（リードペア）のグループ」とは、サンプルＤＮＡ上の同一領域の配列情報を有する（すなわち、同一のサンプルＤＮＡ断片に由来する）と推定されるリードペアの集合である。本方法においては、通常、ライブラリ調製に用いたサンプルＤＮＡの断片の数とシーケンシングデータの量に依存して、１つ以上のリードペアのグループが作成され得る。

【００７４】

次いで、得られたリードペアのグループに含まれるリード配列の情報を用いて、サンプルＤＮＡの配列情報を抽出する。例えば、リードペアのグループに含まれるリード配列の間で配列情報のコンセンサスを取ることで、１つの配列データを作成することができる。リード配列間でのコンセンサスを取る具体的な手法は、上記（３－４）で述べたとおりである。リード配列間でのコンセンサスを取る際には、リードペアのグループに含まれる全てのリードペアのリード配列が用いられてもよいが、該グループ内の一部のリードペアのリード配列のみが用いられてもよい。得られた配列データは、サンプルＤＮＡの断片についての配列を示す最終的な配列データとして取得することができる。

【００７５】

（３－７）リードペアを用いた相補鎖情報に基づくシーケンシング

上述したリードペアを用いて、相補鎖情報を用いたＤＮＡのシーケンシング方法を行うことができる。当該方法では、上記（３－６）で述べたライブラリのシーケンシングの際に、各サンプルＤＮＡの断片を構成する２本の相補鎖の各々に由来する増幅断片に対して、１つ以上のリードペアが作成される。すなわち、１個のサンプルＤＮＡの断片に対して２つ以上のリードペアが取得され、それらのリードペアは、該サンプルＤＮＡの断片の２本の相補鎖の一方及び他方についての配列情報を有する。したがって、本実施形態においては、上述したシーケンシングで得られる複数のリード配列は、複数個のリードペアを含む。

【００７６】

次いで、得られた複数のリードペアから、１つ以上のリードペアのグループを作成する。リードペアのグループを作成する手段は、上記（３－５）で述べたとおりである。ここで得られるリードペアのグループには、特定のサンプルＤＮＡの断片についての２本の相補鎖の一方及び他方の配列情報を有するリードペアが含まれている。したがって、該リー

ドペアのグループに含まれるリード配列の間で配列情報のコンセンサスを取ることにより、相補鎖の情報を反映させた配列データを作成することができる。リード配列間でのコンセンサスを取る具体的な手法は、上記（３－４）で述べたとおりである。リード配列間でのコンセンサスを取る際には、リードペアのグループに含まれる全てのリードペアのリード配列が用いられてもよいが、該グループ内の一部のリードペアのリード配列のみが用いられてもよい。

【００７７】

次いで、得られたリードペアのグループに含まれるリード配列の情報を用いて、１つの配列データを導き出す。例えば、リードペアのグループに含まれるリード配列の間で配列情報のコンセンサスを取ることで、１つの配列データを作成することができる。得られた配列データは、該グループのリード配列が由来する特定のサンプルＤＮＡの断片についての配列である。リードペアのグループに、サンプルＤＮＡの断片についての２本の相補鎖の配列情報を有するリード配列が含まれている場合には、それらの間でのコンセンサスを取ることにより、シーケンシングにおける読み取りエラーやＤＮＡ酸化修飾等に起因するエラーなどの片方の鎖にのみ生じるエラーを除外することができる。

【００７８】

好ましくは、リードペアのグループに含まれるリード配列間でのコンセンサスを取る工程は、リードペアのグループの中から、サンプルＤＮＡの断片の２本の相補鎖の各々に由来するリードペアを少なくとも１組ずつ集め、集めたリードペアに含まれるリード配列の間で配列情報のコンセンサスを取ることを含む。これにより、相補鎖間コンセンサスリード配列を得ることができる。得られた相補鎖間コンセンサスリード配列は、サンプルＤＮＡの断片についての配列を示す最終的な配列データとして取得することができる。

【００７９】

リードペアのグループの中からサンプルＤＮＡの断片の２本の相補鎖の各々に由来するリードペアを集める手順としては、例えば、以下の手順が挙げられる：予めサンプルＤＮＡの断片に２本の相補鎖を識別できる標識配列を付加しておくことにより、該標識配列を含む増幅断片を調製する；次いで、該増幅断片をシーケンシングし、該増幅断片由来のリードペアと、それに付随する該標識配列の情報を取得する；得られたリードペアから、リードペアのグループを作成する；次いで、リードペアに付随する標識配列の情報を利用して、リードペアのグループの中から、互いに相補的な鎖に由来するリードペアを集める。

【００８０】

上記の手順においては、好ましくはサンプルＤＮＡの断片に、上記（２－５－２）で述べたリード配列が該断片の２本の相補鎖のいずれに由来するかを識別可能にする標識配列（例えば、相補鎖標識配列又は個別断片標識配列）を付加する。好ましくは、相補鎖標識配列が用いられる。該標識配列が付加されたサンプルＤＮＡの断片から得られた増幅断片をシーケンシングすることで、該増幅断片由来のリードペアと、それに含まれる各リード配列に付随する該標識配列の情報を取得することができる。この場合、各リードペアのリード１とリード２には、いずれか一方に５'末端側の標識配列の情報が、他方に３'末端側の標識配列の情報が、それぞれ付随する。

【００８１】

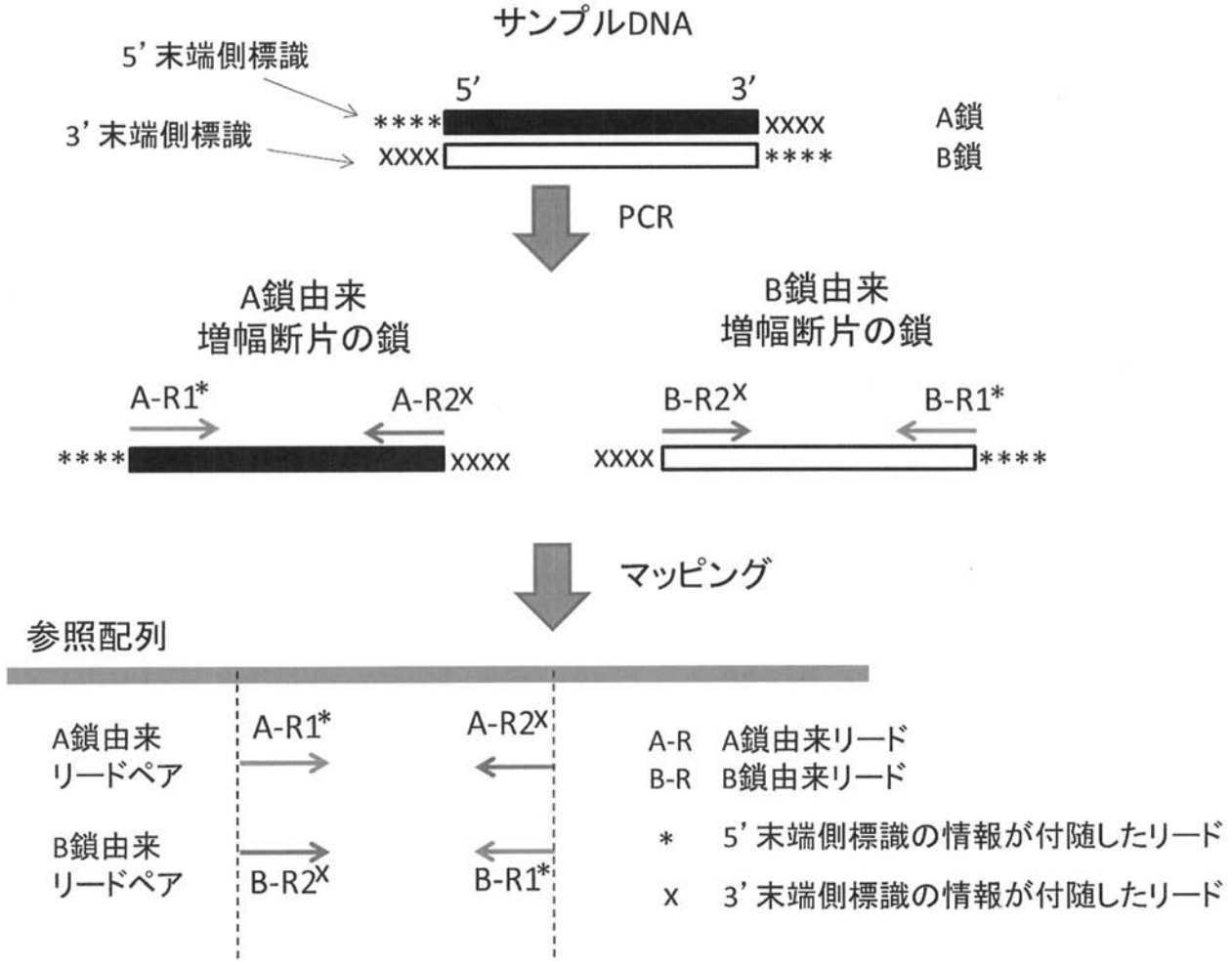
次に、当該標識配列の情報を利用して、リードペアのグループの中から互いに相補的な鎖に由来するリードペアを集める際の好ましい手順を説明する。リードペアのグループに含まれるリードペアを参照配列にマッピングするとき、５'末端側の標識配列の情報が付随するリード配列の先頭が、参照配列上で、リードペアのもう一方のリード配列の先頭よりも５'側に位置する（すなわち、３'末端側の標識配列の情報が付随するリード配列の先頭が、参照配列上で、もう一方のリード配列の先頭よりも３'側に位置する）リードペアと、５'末端側の標識配列の情報が付随するリード配列の先頭が、参照配列上で、リードペアのもう一方のリード配列の先頭よりも３'側に位置する（すなわち、３'末端側の標識配列の情報が付随するリード配列の先頭が、参照配列上で、もう一方のリード配列の先頭よりも５'側に位置する）リードペアに分かれる。前者のリードペアと後者のリードペア

は、サンプルＤＮＡ断片の２本の相補鎖のそれぞれに由来する（下記概念図２参照）。したがって、リードペアに含まれる２本のリード配列に付随する標識配列の情報と、該２本のリード配列の参照配列上での互いの位置関係に基づいて、リードペアのグループ内の各リードペアがサンプルＤＮＡの断片を構成する２本の相補鎖のどちらに由来するかを識別することができる。あるいは、増幅断片の末端に特定の標識配列が付加しているときにのみ開始するシーケンシング反応を行うことにより、標識配列の情報に基づいて、サンプルＤＮＡの断片の特定の１本鎖に由来するリードペアを識別することができる。このようにサンプルＤＮＡ断片の同じ１本鎖に由来するリードペアを予め識別しておくことで、リードペアのグループの中から互いに相補的な鎖に由来するリードペアを集めることができる。

【００８２】

【化２】

概念図 2



【００８３】

上述したリードペアのグループから相補鎖間コンセンサスリード配列を得る手順の具体的な例としては、リードペアのグループの中から、サンプルＤＮＡの断片の２本の相補鎖の各々に由来する２組のリードペアを選択し、それらのリードペアに含まれるリード配列の間で配列情報のコンセンサスを取ることが挙げられる。さらに、該手順を繰り返して複数の相補鎖間コンセンサスリード配列を作成した後、さらにそれらの間でのコンセンサスを取り、１つの相補鎖間コンセンサスリード配列を作成してもよい。あるいは、該相補鎖

間コンセンサスリード配列を得る手順の別の具体的な例としては、リードペアのグループに含まれるリードペアを、サンプルＤＮＡの断片の２本の相補鎖の一方に由来する群と他方に由来する群とに分け、各群のリードペアに含まれるリード配列の間でコンセンサスを取り、得られた２つのコンセンサスデータの間でさらにコンセンサスを取り、１つの相補鎖間コンセンサスリード配列を作成することが挙げられる。あるいはサンプルＤＮＡの断片の２本の相補鎖に由来するリード配列を特に区別せず、リードペアのグループに含まれるリード配列の間でコンセンサスを取り、コンセンサスリード配列を作成することが挙げられる。リードペアのグループから相補鎖間コンセンサスリード配列を作成する手順のより具体的な例は、以下の実施例１（模式図３）に説明されている。

【００８４】

上述したリード配列又はリードペアのグループ分けは、リード配列に含まれるサンプルＤＮＡ自体の配列情報に基づいて行うことができる。互いに相補的な鎖の配列情報を有するリード配列は、サンプルＤＮＡの断片に付加した標識配列の配列情報等に基づいて識別され得る。

【００８５】

（３－８）個別断片標識配列を用いた相補鎖情報の抽出

あるいは、上述した個別断片標識配列を用いることで、サンプルＤＮＡの断片の２本の相補鎖にそれぞれ由来するリード配列を識別することができる。この場合、必ずしもリード配列又はリードペアのグループを作成する必要はなく、個別の標識配列の情報に基づいて、１つのＤＮＡ断片の２本の相補鎖に由来するリード配列を抽出することができる。抽出したリード配列間でのコンセンサスをとることにより、相補鎖の情報を反映させた高精度なリード情報を得ることが可能である。

【００８６】

（４．シーケンシングのための最適条件）

本シーケンシング方法では、サンプルＤＮＡの個別の断片を識別するための標識（個別断片標識配列）を用いない場合、本来異なるＤＮＡ断片に由来する配列を誤って同一断片として誤認識する可能性があり、そのため本来変異として検出されるべきものがエラーと見なされて見逃される可能性がある。

【００８７】

シーケンシングデータからサンプルＤＮＡの配列データ（例えば、リード配列のグループ内でのコンセンサスデータや、相補鎖間コンセンサスリード配列）が得られる効率（解析効率）は、ライブラリ中に同一ＤＮＡ断片由来の増幅産物（順鎖及び相補鎖を含む）が含まれている割合と、該ライブラリを用いたシーケンシングデータの量（リード数又はｂｐ）に依存する。例えば、ライブラリ調製での増幅工程（例えばＰＣＲ）での初期ＤＮＡ量と、シーケンシングデータ量は、上述したリード配列又はリードペアのグループに含まれるリード配列又はリードペアの数、ひいては解析効率に影響する。

【００８８】

断片の誤認識は、ライブラリ中に、異なるサンプルＤＮＡ断片に由来する配列の重複する断片が含まれており、かつそれら断片の双方がシーケンシングされている場合に発生し得る。したがって、断片の誤認識には、ライブラリ中のＤＮＡ配列の多様度（増幅工程での初期ＤＮＡ量）が関係する。また、ライブラリ中のＤＮＡ配列の多様度には、サンプルＤＮＡの配列の多様度が影響し、サンプルＤＮＡの配列の多様度はサンプルＤＮＡのサイズ（総ｂｐ）に概ね依存する。よって、サンプルＤＮＡのサイズも断片の誤認識に影響する。また、１本鎖特異的ヌクレアーゼ処理も断片の誤認識に影響し得る。１本鎖特異的ヌクレアーゼの配列特異性に起因して、該ヌクレアーゼ処理後のＤＮＡ断片には、該ヌクレアーゼが除去しにくい配列が残ることがある。すなわち、該ヌクレアーゼ処理後のＤＮＡ断片では、断片の両端の配列が同一となる確率が高まる可能性があり、これにより断片の誤認識の割合が増加し得る。

【００８９】

したがって、シーケンシングの効率及び精度に影響を与え得る因子としては、主にライ

ブラリ調製の増幅工程（例えばＰＣＲ）での初期ＤＮＡ量、シーケンシングデータ量、そしてサンプルＤＮＡのサイズが重要である。さらに、１本鎖特異的ヌクレアーゼ処理の反応液中におけるＤＮＡの単位重量当たりの１本鎖特異的ヌクレアーゼのユニット数（Ｕ／ｎｇ）も、シーケンシングの効率及び精度に影響を与え得る因子として考慮することが望まれる。さらに、これらの因子に依存するリード配列又はリードペアのグループに含まれるリード配列又はリードペアの数は、シーケンシングの効率及び精度を判断する指標となり得る。

【００９０】

ライブラリ調製の増幅（例えばＰＣＲ）工程における初期ＤＮＡ量（以下、単に初期ＤＮＡ量という）の適切な範囲は、サンプルＤＮＡのサイズに依存し得るが、例えばサンプルＤＮＡの１Ｍｂｐあたり、好ましくは２５０ａｍｏｌ以下、より好ましくは１２５ａｍｏｌ以下、さらに好ましくは６２．５ａｍｏｌ以下、なお好ましくは３１．３ａｍｏｌ以下、さらになお好ましくは１５．６ａｍｏｌ以下、さらになお好ましくは７．８ａｍｏｌ以下、さらになお好ましくは３．９ａｍｏｌ以下、さらになお好ましくは１．７ａｍｏｌ以下、さらになお好ましくは０．８３ａｍｏｌ以下、さらになお好ましくは０．４２ａｍｏｌ以下、さらになお好ましくは０．２１ａｍｏｌ以下である。一方で、該初期ＤＮＡ量は、ゲノムの網羅性を担保する観点から、サンプルＤＮＡの１Ｍｂｐあたり、好ましくは０．０００３ａｍｏｌ以上、より好ましくは０．０００７ａｍｏｌ以上、さらに好ましくは０．００２ａｍｏｌ以上、なお好ましくは０．００５ａｍｏｌ以上、さらになお好ましくは０．０１ａｍｏｌ以上、さらになお好ましくは０．０３ａｍｏｌ以上、さらになお好ましくは０．０５ａｍｏｌ以上、さらになお好ましくは０．１ａｍｏｌ以上、さらになお好ましくは０．３ａｍｏｌ以上、さらになお好ましくは１ａｍｏｌ以上、さらになお好ましくは２ａｍｏｌ以上、さらになお好ましくは３．９ａｍｏｌ以上、さらになお好ましくは７．８ａｍｏｌ以上である。

【００９１】

一例において、本シーケンシング方法における初期ＤＮＡ量は、以下のとおりである：ゲノムサイズ約５Ｍｂｐの細菌の場合、サンプルＤＮＡの１Ｍｂｐあたり、好ましくは０．１～２５０ａｍｏｌ、より好ましくは０．３～２５０ａｍｏｌ、さらに好ましくは１～２５０ａｍｏｌ、なお好ましくは２～１２５ａｍｏｌ、さらになお好ましくは３．９～６２．５ａｍｏｌ、さらになお好ましくは７．８～３１．３ａｍｏｌである；ゲノムサイズ約１０Ｍｂｐの酵母の場合、サンプルＤＮＡの１Ｍｂｐあたり、好ましくは０．０５～２５０ａｍｏｌ、より好ましくは０．１～２５０ａｍｏｌ、さらに好ましくは０．３～１２５ａｍｏｌ、なお好ましくは１～６２．５ａｍｏｌ、さらになお好ましくは２～３１．３ａｍｏｌである；ゲノムサイズ約１００Ｍｂｐの線虫の場合、サンプルＤＮＡの１Ｍｂｐあたり、好ましくは０．００５～３１．３ａｍｏｌ、より好ましくは０．０１～３１．３ａｍｏｌ、さらに好ましくは０．０３～１５．６ａｍｏｌ、なお好ましくは０．１～７．８ａｍｏｌ、さらになお好ましくは０．３～３．９ａｍｏｌである；ゲノムサイズ約３Ｇｂｐのマウスの場合、サンプルＤＮＡの１Ｍｂｐあたり、好ましくは０．０００３～１．７ａｍｏｌ、より好ましくは０．０００７～１．７ａｍｏｌ、さらに好ましくは０．００２～１．７ａｍｏｌ、なお好ましくは０．００５～０．８３ａｍｏｌ、さらになお好ましくは０．０１～０．４２ａｍｏｌ、さらになお好ましくは０．０３～０．２１ａｍｏｌ、である。なお、本願明細書において、初期ＤＮＡ量は増幅工程に使用するＤＮＡサンプル中のＤＮＡ量であり、プライマーなどのＤＮＡ量は含まない。

【００９２】

ただし、上記のサンプルＤＮＡ　１Ｍｂｐあたりの初期ＤＮＡ量の範囲は、上述したライブラリ調製での１本鎖特異的ヌクレアーゼ処理における該ヌクレアーゼのユニット数に依存し得る。例えば、該１本鎖特異的ヌクレアーゼ処理で０．０５Ｕ／ｎｇ以下のＳ１ｎｕｃｌｅａｓｅを用いる場合、断片の誤認識への影響が十分に小さいので、初期ＤＮＡ量の適切な範囲は上記のとおりである。

一方、該１本鎖特異的ヌクレアーゼ処理で０．０５Ｕ／ｎｇより大きいユニット数でＳ

１　ｎｕｃｌｅａｓｅを用いる場合、ユニット数の増加に伴い断片の誤認識率が増加し得る。そのため、該ヌクレアーゼ処理での反応液中におけるＳ１　ｎｕｃｌｅａｓｅのユニット数（Ｕ／ｎｇ）に応じて初期ＤＮＡ量を設定することが望ましい。Ｓ１　ｎｕｃｌｅａｓｅのユニット数（＞０．０５Ｕ／ｎｇ）と初期ＤＮＡ量の適切な条件は、下記の式より算出される指標で表され、

指標＝初期ＤＮＡ量（ａｍｏｌ／Ｍｂｐ）×$3^{\log S1\ nuclease\ (U/ng)}$

（式中、Ｓ１　ｎｕｃｌｅａｓｅ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である。）

当該指標は、好ましくは６０以下、より好ましくは３０以下、さらに好ましくは１５以下、さらにより好ましくは７．５以下である。

【００９３】

例えば、該１本鎖特異的ヌクレアーゼ処理で０．０５Ｕ／ｎｇ以下のＭＢＮを用いる場合、断片の誤認識への影響が十分に小さいので、初期ＤＮＡ量の適切な範囲は上記のとおりである。

一方、該１本鎖特異的ヌクレアーゼ処理で０．０５Ｕ／ｎｇより大きいユニット数でＭＢＮを用いる場合、ユニット数の増加に伴い断片の誤認識率が増加し得る。そのため、該ヌクレアーゼ処理での反応液中におけるＭＢＮのユニット数（Ｕ／ｎｇ）に応じて初期ＤＮＡ量を設定することが望ましい。ＭＢＮのユニット数（＞０．０５Ｕ／ｎｇ）と初期ＤＮＡ量の適切な条件は、下記の式より算出される指標で表され、

指標＝初期ＤＮＡ量（ａｍｏｌ／Ｍｂｐ）×$3^{\log MBN\ (U/ng)}$

（式中、ＭＢＮ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である。）

当該指標は、好ましく６０以下、より好ましくは３０以下、さらに好ましくは１５以下、さらにより好ましくは７．５以下である。

【００９４】

一方、該１本鎖特異的ヌクレアーゼ処理で$RecJ_f$を用いる場合、ユニット数に関わらず断片の誤認識への影響は十分に小さいので、初期ＤＮＡ量の適切な範囲は上記のとおりである。

【００９５】

一方、解析に十分な量のデータ（配列情報）を確保するためには、本シーケンシング方法における初期ＤＮＡ量は、好ましくは０．１ａｍｏｌ以上、より好ましくは１ａｍｏｌ以上、さらに好ましくは５ａｍｏｌ以上、なお好ましくは２０ａｍｏｌ以上、さらになお好ましくは３９ａｍｏｌ以上、さらになお好ましくは７８ａｍｏｌ以上である。解析効率の観点からは、初期ＤＮＡ量は、好ましくは１０００００ａｍｏｌ以下、より好ましくは２００００ａｍｏｌ以下、さらに好ましくは５０００ａｍｏｌ以下である。例えば、本シーケンシング方法における初期ＤＮＡ量は、好ましくは０．１～１０００００ａｍｏｌ、より好ましくは１～１０００００ａｍｏｌ、さらに好ましくは５～１０００００ａｍｏｌ、なお好ましくは２０～１０００００ａｍｏｌ、さらになお好ましくは２０～２００００ａｍｏｌ、さらになお好ましくは３９～２００００ａｍｏｌ、さらになお好ましくは７８～２００００ａｍｏｌ、さらになお好ましくは２０～５０００ａｍｏｌ、さらになお好ましくは３９～５０００ａｍｏｌ、さらになお好ましくは７８～５０００ａｍｏｌである。

【００９６】

初期ＤＮＡ量に対して大きすぎる又は少なすぎるシーケンシングデータ量は、解析効率を低下させ得る。本シーケンシング方法におけるシーケンシングデータ量は、初期ＤＮＡ量１ａｍｏｌあたりのリードペア数又はリード配列数で、好ましくは０．０２×10^6個（リード配列又はリードペアの塩基対量で４Ｍｂｐ、これはリード配列の長さ、又はリードペアに含まれるリード配列の合計長の平均が２００ｂｐの場合の値であり、該リード配列長さ又はリードペアに含まれるリード配列の合計長の平均値に合わせて変化し得る値である、以下同）以上、より好ましくは０．０４×10^6個（８Ｍｂｐ）以上、さらに好ましくは０．０８×10^6個（１６Ｍｂｐ）以上、なお好ましくは０．１６×10^6個（３２Ｍｂｐ）以上であり、かつ、好ましくは１０×10^6個（２０００Ｍｂｐ）以下、より好ましくは５×10^6個（１０００Ｍｂｐ）以下、さらに好ましくは２．５×10^6個（５０

０Ｍｂｐ）以下、なお好ましくは２×１０^{6}個（４００Ｍｂｐ）以下である。例えば、本シーケンシング方法におけるシーケンシングデータ量は、初期ＤＮＡ量１ａｍｏｌあたりのリードペア数又はリード配列数で、好ましくは０．０２～１０×１０^{6}個（４～２０００Ｍｂｐ）、より好ましくは０．０４～５×１０^{6}個（８～１０００Ｍｂｐ）、さらに好ましくは０．０８～２．５×１０^{6}個（１６～５００Ｍｂｐ）、なお好ましくは０．１６～２×１０^{6}個（３２～４００Ｍｂｐ）である。

【００９７】

本シーケンシング方法において、最大の解析効率をもたらすリード配列又はリードペアのグループに含まれるリード配列又はリードペアの数の平均値は、初期ＤＮＡ量及びシーケンシングデータ量にかかわらずほぼ一定である（特許文献４参照）。本シーケンシング方法における、リード配列のグループに含まれるリード配列数、又はリードペアのグループに含まれるリードペア数は、該グループ間の平均で、好ましくは１．０５以上、より好ましくは１．１以上、さらに好ましくは１．２以上であり、なお好ましくは１．４以上であり、かつ、好ましくは３０以下、より好ましくは２０以下、さらに好ましくは１０以下、なお好ましくは５以下である。例えば、本シーケンシング方法において、リード配列又はリードペアのグループに含まれるリード配列又はリードペアの数は、該グループ間の平均で、好ましくは１．０５～３０、より好ましくは１．１～２０、さらに好ましくは１．２～１０、なお好ましくは１．４～５である。

【００９８】

適切なシーケンシングデータ量は、サンプルＤＮＡのサイズに依存し得る。より大きなサイズのサンプルＤＮＡに対しては、より多くの初期ＤＮＡ量が必要となる。一方、サンプルＤＮＡのサイズに対してシーケンシングデータ量が多すぎる場合、解析効率が低下する。本シーケンシング方法におけるシーケンシングデータ量は、サンプルＤＮＡの１Ｍｂｐあたりのリード配列又はリードペア数で、好ましくは０．０５×１０^{6}個（１０Ｍｂｐ）以上、より好ましくは０．１×１０^{6}個（２０Ｍｂｐ）以上、さらに好ましくは０．２×１０^{6}個（４０Ｍｂｐ）以上、なお好ましくは０．５×１０^{6}個（１００Ｍｂｐ）以上、さらになお好ましくは１×１０^{6}個（２００Ｍｂｐ）以上、さらになお好ましくは２×１０^{6}個（０．４Ｇｂｐ）以上であり、かつ、好ましくは１６００×１０^{6}個（３２０Ｇｂｐ）以下、より好ましくは８００×１０^{6}個（１６０Ｇｂｐ）以下、さらに好ましくは４００×１０^{6}個（８０Ｇｂｐ）以下、なお好ましくは２００×１０^{6}個（４０Ｇｂｐ）以下、さらになお好ましくは１００×１０^{6}個（２０Ｇｂｐ）以下、さらになお好ましくは５０×１０^{6}個（１０Ｇｂｐ）以下である。例えば、本シーケンシング方法におけるシーケンシングデータ量は、サンプルＤＮＡの１Ｍｂｐあたりリード配列又はリードペア数で、好ましくは０．０５～１６００×１０^{6}個（０．０１～３２０Ｇｂｐ）、より好ましくは０．１～８００×１０^{6}個（０．０２～１６０Ｇｂｐ）、さらに好ましくは０．２～４００×１０^{6}個（０．０４～８０Ｇｂｐ）、なお好ましくは０．５～２００×１０^{6}個（０．１～４０Ｇｂｐ）、さらになお好ましくは１～１００×１０^{6}個（０．２～２０Ｇｂｐ）、さらになお好ましくは２～５０×１０^{6}個（０．４～１０Ｇｂｐ）である。なお、哺乳動物由来ゲノムＤＮＡなどのようにサンプルＤＮＡのサイズが大きい場合で、かつサンプルＤＮＡの配列全体に対しての配列データの網羅性が特に問題とならない場合、本シーケンシング方法におけるシーケンシングデータ量は、サンプルＤＮＡの１Ｍｂｐあたりのリード配列又はリードペア数で、０．０５×１０^{6}個（１０Ｍｂｐ）未満であってもよい。例えば、ゲノムサイズ約３Ｇｂｐのマウスの場合のシーケンシングデータ量は、サンプルＤＮＡの１Ｍｂｐあたりのリード配列又はリードペア数で、好ましくは０．００００３～１６×１０^{6}個（０．００６～３２００Ｍｂｐ）、より好ましくは０．００００７～８×１０^{6}個（０．０１４～１６００Ｍｂｐ）、さらに好ましくは０．０００１～４×１０^{6}個（０．０２～８００Ｍｂｐ）、なお好ましくは０．０００３～２×１０^{6}個（０．０６～４００Ｍｂｐ）、さらになお好ましくは０．０００５～１×１０^{6}個（０．１～２００Ｍｂｐ）、さらになお好ましくは０．００１～０．５×１０^{6}個（０．２～１００Ｍｂｐ）である。

【００９９】
サンプルＤＮＡのサイズが小さすぎると、シーケンシング用のライブラリ中の配列の多様性が低下して断片の誤認識の確率が高くなることがある。本シーケンシング方法におけるサンプルＤＮＡのサイズは、好ましくは１０ｋｂｐ以上、より好ましくは１００ｋｂｐ以上、さらに好ましくは１Ｍｂｐ以上、なお好ましくは４Ｍｂｐ以上であるが、サンプルＤＮＡの由来する生物のゲノムＤＮＡのサイズ等に依存し得る。
【０１００】
本シーケンシング方法の好ましい一実施形態においては、サンプルＤＮＡのサイズは約５Ｍｂｐであり、ＰＣＲ初期ＤＮＡ量は、好ましくは１０～１２５０ａｍｏｌであり、シーケンシングデータ量は、リード配列又はリードペア数で０．２～１２５００×10^6個（０．０４～２５００Ｇｂｐ）、好ましくは０．４～６２５０×10^6個（０．０８～１２５０Ｇｂｐ）、より好ましくは０．８～３１２５×10^6個（０．１６～６２５Ｇｂｐ）、さらに好ましくは１．６～２５００×10^6個（０．３２～５００Ｇｂｐ）である。
より好ましくは、サンプルＤＮＡのサイズは約５Ｍｂｐであり、ＰＣＲ初期ＤＮＡ量は２０～６２５ａｍｏｌであり、シーケンシングデータ量は、リード配列又はリードペア数で０．４～６２５０×10^6個（０．０８～１２５０Ｇｂｐ）、好ましくは０．８～３１２５×10^6個（０．１６～６２５Ｇｂｐ）、より好ましくは１．６～１５６３×10^6個（０．３２～３１３Ｇｂｐ）、さらに好ましくは３．２～１２５０×10^6個（０．６４～２５０Ｇｂｐ）である。
さらに好ましくは、サンプルＤＮＡのサイズは約５Ｍｂｐであり、ＰＣＲ初期ＤＮＡ量は３９～３１３ａｍｏｌであり、シーケンシングデータ量は、リード配列又はリードペア数で０．７８～３１３０×10^6個（０．１５６～６２６Ｇｂｐ）、好ましくは１．５６～１５６５×10^6個（０．３１２～３１３Ｇｂｐ）、より好ましくは３．１２～７８３×10^6個（０．６２４～１５７Ｇｂｐ）、さらに好ましくは６．２４～６２６×10^6個（１．２４８～１２５Ｇｂｐ）である。
本シーケンシング方法の別の好ましい一実施形態においては、サンプルＤＮＡのサイズは約５Ｍｂｐであり、リード配列又はリードペアのグループあたりのリード配列又はリードペアの数は、該グループ間の平均で、１．０５～３０、好ましくは１．１～２０、さらに好ましくは１．２～１０、なお好ましくは１．４～５である。
上述したとおり、上記ＰＣＲ初期ＤＮＡ量は、ライブラリ調製での１本鎖特異的ヌクレアーゼ処理における該ヌクレアーゼのユニット数に依存し得る。
【０１０１】
本シーケンシング方法のさらに別の好ましい一実施形態においては、サンプルＤＮＡのサイズは約３Ｇｂｐであり、ＰＣＲ初期ＤＮＡ量は、好ましくは１０～５０００ａｍｏｌであり、シーケンシングデータ量は、リード配列又はリードペア数で０．２～５００００×10^6個（０．０４～１００００Ｇｂｐ）、好ましくは０．４～２５０００×10^6個（０．０８～５０００Ｇｂｐ）、より好ましくは０．８～１２５００×10^6個（０．１６～２５００Ｇｂｐ）、さらに好ましくは１．６～１００００×10^6個（０．３２～２０００Ｇｂｐ）である。
より好ましくは、サンプルＤＮＡのサイズは約３Ｇｂｐであり、ＰＣＲ初期ＤＮＡ量は２０～２５００ａｍｏｌであり、シーケンシングデータ量は、リード配列又はリードペア数で０．４～２５０００×10^6個（０．０８～５０００Ｇｂｐ）、好ましくは０．８～１２５００×10^6個（０．１６～２５００Ｇｂｐ）、より好ましくは１．６～６２５０×10^6個（０．３２～１２５０Ｇｂｐ）、さらに好ましくは３．２～５０００×10^6個（０．６４～１０００Ｇｂｐ）である。
さらに好ましくは、サンプルＤＮＡのサイズは約３Ｇｂｐであり、ＰＣＲ初期ＤＮＡ量は３９～１２５０ａｍｏｌであり、シーケンシングデータ量は、リード配列又はリードペア数で０．７８～１２５００×10^6個（０．１５６～２５００Ｇｂｐ）、好ましくは１．５６～６２５０×10^6個（０．３１２～１２５０Ｇｂｐ）、より好ましくは３．１２～３１２５×10^6個（０．６２４～６２５Ｇｂｐ）、さらに好ましくは６．２４～２５

００×１０^{6}個（１．２４８～５００Ｇｂｐ）である。
本シーケンシング方法のなお別の好ましい一実施形態においては、サンプルＤＮＡのサイズは約３Ｇｂｐであり、リード配列又はリードペアのグループあたりのリード配列又はリードペアの数は、該グループ間の平均で、１．０５～３０、好ましくは１．１～２０、さらに好ましくは１．２～１０、なお好ましくは１．４～５である。
上述したとおり、上記ＰＣＲ初期ＤＮＡ量は、ライブラリ調製での１本鎖特異的ヌクレアーゼ処理における該ヌクレアーゼのユニット数に依存し得る。
【０１０２】
サイズ約５Ｍｂｐのサンプルの例としては、サルモネラ属細菌のゲノム（約４．８６Ｍｂｐ）が挙げられる。サルモネラ属細菌の好ましい例としては、Ａｍｅｓ試験に使用されるＳ．ｔｙｐｈｉｍｕｒｉｕｍ　ＬＴ－２株、ＴＡ１００株、ＴＡ９８株、ＴＡ１５３５株、ＴＡ１５３８株、ＴＡ１５３７株等が挙げられる。
【０１０３】
（５．シーケンシング方法の応用）
本発明のライブラリを用いたシーケンシングで得られた配列データは、ＤＮＡ断片の１本鎖部分の酸化修飾等に起因するシーケンシングエラーが除外された高精度な配列データである。したがって、本発明のライブラリを用いたシーケンシングは、これに限定されないが、変異解析に応用することができる。より詳細には、例えば、ゲノムＤＮＡの変異解析による、試験物質の遺伝毒性の評価や、生殖発生毒性等のその他毒性の評価、ゲノムＤＮＡに対する経時変化、生活環境、遺伝的要素などの影響の評価、培養細胞の品質評価などに応用することができる。これらの応用においては、変異解析の対象であるゲノムＤＮＡから本発明のライブラリを調製し、これをシーケンシングして配列データを取得する。次いで、得られた配列データを用いて変異解析を行い、解析対象ゲノムＤＮＡの変異を検出する。
【０１０４】
したがって、本発明はまた、ゲノムＤＮＡの変異を検出する方法を提供する。当該方法は、細胞中のゲノムＤＮＡをサンプルＤＮＡとして用いて、本発明によるシーケンシング用ライブラリの調製方法によりシーケンシング用ライブラリを調製すること、該シーケンシング用ライブラリをシーケンシングすること、を含む。該シーケンシングにより、該ゲノムＤＮＡについての配列データが作成される。該配列データを参照配列と比較して、該配列データと該参照配列とで塩基がマッチしない部位を変異部位として検出することで、該ゲノムＤＮＡの変異を検出することができる。
【０１０５】
一実施形態において、本発明によるゲノムＤＮＡの変異を検出する方法は、試験物質の遺伝毒性の評価に利用される。本実施形態では、該ゲノムＤＮＡは、試験物質に暴露した細胞のゲノムＤＮＡである。好ましくは、該ゲノムＤＮＡは、試験物質に暴露した細胞（被験細胞）のゲノムＤＮＡと、該試験物質に暴露していない細胞（対照細胞）のゲノムＤＮＡである。好ましくは、これらのゲノムＤＮＡは新鮮なＤＮＡである。該新鮮なＤＮＡは、好ましくはＤＩＮが６以上のＤＮＡであり、より好ましくはＤＩＮが７以上のＤＮＡであり、より好ましくはＤＩＮが７．３以上のＤＮＡ、さらにより好ましくはＤＩＮが７．５以上のＤＮＡである。本実施形態では、該被験細胞のゲノムＤＮＡについて検出した変異と、該対照細胞のゲノムＤＮＡについて検出した変異とが比較される。例えば、該被験細胞でのみ検出された変異を、試験物質の暴露により生じた変異として同定することができる。本実施形態において使用される細胞は、特に限定されず、微生物細胞、動物細胞、植物細胞を含み得る。動物の例としては、好ましくは哺乳動物、鳥類、カイコ、線虫などが挙げられ、微生物の例としては、大腸菌、サルモネラ菌、酵母などが挙げられるが、これらに限定されない。本実施形態において使用される細胞の好ましい例としては、サルモネラ属細胞、及び大腸菌細胞が挙げられるが、これらに限定されない。サルモネラ属細胞の好ましい例としては、Ａｍｅｓ試験に使用されるＳａｌｍｏｎｅｌｌａ　ｔｙｐｈｉｍｕｒｉｕｍ　ＬＴ－２株、ＴＡ１００株、ＴＡ９８株、ＴＡ１５３５株、ＴＡ１５３８

株、ＴＡ１５３７株等が挙げられる。大腸菌の好ましい例としては、分子生物学研究で汎用されるＫ－１２株や、Ａｍｅｓ試験に使用されるＷＰ２株、ＷＰ２　ｕｖｒＡ株等が挙げられる。本実施形態において使用される細胞の別の好ましい例としては、生体から採取した哺乳動物細胞、及び哺乳動物由来培養細胞が挙げられる。哺乳動物の好ましい例としては、マウス、ラット、ハムスター、チャイニーズハムスター、ウサギ、ヒトなどが挙げられ、このうちマウス及びヒトが好ましい。本実施形態において使用される細胞の別の好ましい例としては、生体から採取した鳥類細胞、及び鳥類由来培養細胞が挙げられる。鳥類の好ましい例としては、ニワトリが挙げられ、鳥類由来培養細胞の例としてはＤＴ４０などが挙げられる。

【０１０６】

該試験物質の例としては、その遺伝毒性を評価したい物質であれば特に制限されない。例えば、遺伝毒性を有すると疑われる物質、又は遺伝毒性の有無を確認したい物質、どのような変異を誘発するかを調べたい物質などが挙げられる。試験物質は、天然に存在する物質であっても、化学的もしくは生物学的方法等で人工的に合成した物質であってもよく、又は化合物であっても、組成物もしくは混合物であってもよい。あるいは、該試験物質は、紫外線や放射線などであってもよい。細胞を試験物質に暴露する手段は、試験物質の種類に応じて適宜選択すればよく、特に限定されない。例えば、細胞を含む培地に試験物質を添加する方法、細胞を試験物質の存在する雰囲気下に置く方法などが挙げられる。

【０１０７】

別の一実施形態において、本発明によるゲノムＤＮＡの変異を検出する方法は、ゲノムＤＮＡに対する経時変化、生活環境、遺伝的要素などの影響の評価に利用される。経時変化としては、細胞や個体の成長、加齢、老化、継代培養などが挙げられ、生活環境としては、食生活、運動などの生活習慣、居住地などが挙げられ、遺伝的要素としては、性別、種、特定の遺伝子の欠損や塩基対置換などが挙げられるが、これらに限定されない。本実施形態の好適な例は、ゲノムＤＮＡに対する経時変化の影響の評価であり、該ゲノムＤＮＡには、経時変化した細胞のゲノムＤＮＡが用いられる。より好ましくは、該ゲノムＤＮＡは、経時変化した細胞（被験細胞）のゲノムＤＮＡと、より経時変化していない細胞（対照細胞）のゲノムＤＮＡである。好ましくは、これらのゲノムＤＮＡは新鮮なＤＮＡである。該新鮮なＤＮＡは、好ましくはＤＩＮが６以上のＤＮＡであり、より好ましくはＤＩＮが７以上のＤＮＡであり、より好ましくはＤＩＮが７．３以上のＤＮＡ、さらにより好ましくはＤＩＮが７．５以上のＤＮＡである。本実施形態では、該被験細胞のゲノムＤＮＡについて検出した変異と、該対照細胞のゲノムＤＮＡについて検出した変異とが比較される。該対照細胞として用いられる、より経時変化していない細胞としては、成長、加齢、老化又は継代培養の程度が被験細胞より少ない細胞（例えば、より若い細胞、老化処理していない細胞、継代していないか継代数の少ない細胞など）が挙げられる。例えば、該被験細胞でのみ検出された変異を、経時変化により生じた変異として同定することができる。本実施形態において使用される細胞の好ましい例としては、生体から採取した哺乳動物細胞、及び哺乳動物由来培養細胞が挙げられる。哺乳動物の好ましい例としては、上述したとおりである。

【０１０８】

別の一実施形態において、本発明によるゲノムＤＮＡの変異を検出する方法は、培養細胞の品質評価に利用される。本実施形態で用いられる該ゲノムＤＮＡは、変異の有無を調べたい培養細胞のゲノムＤＮＡであればよい。該変異の有無を調べたい培養細胞の例としては、ある一定期間培養した細胞であって、その変異の傾向を確認したいものが挙げられる。好ましくは、該ゲノムＤＮＡは、該変異の有無を調べたい培養細胞（被験細胞）のゲノムＤＮＡと、対照細胞のゲノムＤＮＡである。対照細胞としては、例えば、同じ種類の培養細胞であって、遺伝情報既知の（例えば変異の有無及びその変異タイプが確認されている）細胞が用いられる。好ましくは、これらのゲノムＤＮＡは新鮮なＤＮＡである。該新鮮なＤＮＡは、好ましくはＤＩＮが６以上のＤＮＡであり、より好ましくはＤＩＮが７以上のＤＮＡであり、より好ましくはＤＩＮが７．３以上のＤＮＡ、さらにより好ましく

はＤＩＮが７．５以上のＤＮＡである。本実施形態では、該被験細胞のゲノムＤＮＡについて検出した変異と、該対照細胞のゲノムＤＮＡについて検出した変異とが比較される。例えば、該被験細胞でのみ検出された変異を、培養中に生じた変異として同定することができる。

【０１０９】

本発明によるゲノムＤＮＡの変異を検出する方法で検出される変異としては、塩基対置換型変異、及び短い挿入／欠失変異が挙げられる。塩基対置換型変異とは、ＤＮＡの塩基対情報を別の塩基対に変化させる変異であり、例えば、１塩基対置換型変異、及び２塩基対又は３塩基対以上が置換した多塩基対置換型変異を含む。本発明では、好ましくは１塩基対置換型変異が検出される。一方、短い挿入／欠失変異とは、ＤＮＡの配列中に短い塩基配列の挿入又は欠失を引き起こす変異であり、好ましくは挿入又は欠失した塩基の長さが１０ｂｐ以下、より好ましくは１～５ｂｐの挿入又は欠失変異をいう。

【０１１０】

塩基対置換型変異、及び短い挿入／欠失変異の検出は、ＷＯ／２０１８／１５０５１３（その全体を本明細書に援用する）に記載の手順に従って実施することができる。その例として、以下に、解析対象ゲノムＤＮＡにおける１塩基対置換型変異のパターンを検出する場合の好ましい手順を記載する。塩基対置換型変異の検出においては、シーケンシングで取得された配列データが参照配列と比較され、該配列データと該参照配列とで塩基がマッチしない部位が変異部位として検出される。検出された部位は、塩基対置換型変異を有する変異部位として取得される。本発明においては、変異解析の目的に応じて、該参照配列との比較に、取得された配列データの一部を用いてもよく、又は全部を用いてもよい。

【０１１１】

次いで、検出した変異部位の塩基と変異前の塩基の種類に基づいて、各変異を塩基の変異パターンに従って分類する。さらに、該塩基の変異パターンの各々について、出現頻度を決定することができる。これらの手順は、Ｐｙｔｈｏｎ等のプログラミング言語を用いて作成したプログラム等を用いて実施することができる。

【０１１２】

より詳細な例においては、配列データに含まれる各塩基を、下記(i)～(iv)に分ける。

(i) 参照配列上の塩基がＡである位置に存在する塩基
(ii) 参照配列上の塩基がＴである位置に存在する塩基
(iii)参照配列上の塩基がＧである位置に存在する塩基
(iv) 参照配列上の塩基がＣである位置に存在する塩基

上記(i)及び(ii)は、参照配列の塩基対がＡＴであった部位に存在する塩基であり、上記(iii)及び(iv)は、参照配列の塩基対がＧＣであった部位に存在する塩基である。これらの塩基の中から、参照配列と塩基がマッチしない（すなわち塩基対置換変異している）ものを検出する。次いで、検出された変異部位の各々について、参照配列と配列データの配列情報に基づいて変異前及び後の塩基対を求める。これらのデータから、各変異を、変異前の塩基対がＡＴであった場合について[ＡＴ→ＴＡ、ＡＴ→ＣＧ、及びＡＴ→ＧＣ]の３パターン、変異前の塩基対がＧＣであった場合について[ＧＣ→ＴＡ、ＧＣ→ＣＧ、及びＧＣ→ＡＴ]の３パターンの、全部で６つの塩基対の変異パターンに分類することができる。さらに、各変異パターンに属する変異の総数、及び解析した塩基の総数に基づいて、各変異パターンの出現頻度を決定することができる。例えば、ＡＴ、ＧＣ塩基対それぞれについての解析した塩基の総数に基づいて、各々の塩基対ごとに３種類の変異パターンの出現頻度を算出することができる。

【０１１３】

さらに、上記の各変異パターンを、変異検出の際にリード配列がマッピングされた参照配列上の塩基によってさらに２パターンに分類することができる。例えば、変異パターンがＧＣ→ＴＡの変異であれば、参照配列上のＧ上でＴが検出される場合とＣ上でＡが検出される場合に分けられる。これらをそれぞれＧからＴへの変異（Ｇ→Ｔ）、ＣからＡへの変異（Ｃ→Ａ）と定義する。したがって、Ｇ→Ｔ及びＣ→Ａに分けて変異頻度を算出する

ことができる。ＡＴ→ＴＡ、ＡＴ→ＣＧ、ＡＴ→ＧＣ、ＧＣ→ＣＧ、及びＧＣ→ＡＴについても同様である。２本鎖ＤＮＡに固定された真の変異ならば、これら２パターンの変異頻度は同等になる。一方、これら２パターンの間で変異頻度に偏りが認められる場合、リード配列の由来するサンプルＤＮＡの２本の鎖の間で変異頻度が異なることを意味し、この変異は、酸化修飾等による塩基の変異に起因するエラーである可能性が高い。したがって、上記のような２パターンへの分類は、シーケンシングエラーの検出に利用することができる。

【０１１４】

本発明においては、多塩基対置換型変異を解析することもできる。多塩基対置換型変異としては、例えば、２塩基対置換型変異及び３塩基対置換型変異が挙げられる。多塩基対置換型変異の解析の場合には、例えば、変異前の塩基配列に応じて変異パターンを分類し（例えば２塩基対置換型においては４×４＝１６通り）、次いで、各変異パターンに属する変異の総数、及び解析した変異の総数に基づいて、各変異パターンの出現頻度を決定することができる。

【０１１５】

本発明においては、１塩基対置換型変異のシーケンスコンテクスト解析を行うこともできる。この解析では、上記手順で１塩基対置換型変異を検出した後、検出した各変異について、参照配列に基づいて、変異前の塩基と、該変異前の塩基の上流及び下流に隣接する塩基とを含む配列（いわゆるコンテクスト）を決定する。続いて、各変異を、塩基対の変異パターン及び該コンテクストに従ってタイプ分けする。すなわち、検出した変異を、上述した手順で６つの塩基対の変異パターン［ＡＴ→ＴＡ、ＡＴ→ＣＧ、ＡＴ→ＧＣ、ＧＣ→ＴＡ、ＧＣ→ＣＧ、及びＧＣ→ＡＴ］に分ける。一方で、検出した各変異を、コンテクストに従って分類する。例えば、変異部位の両隣の１塩基ずつを含めた３塩基長のコンテクストは、４×４の１６群［例えば、Ｃからの変異の場合、ＡＣＡ、ＡＣＣ、ＡＣＧ、ＡＣＴ、ＣＣＡ、ＣＣＣ、ＣＣＧ、ＣＣＴ、ＧＣＡ、ＧＣＣ、ＧＣＧ、ＧＣＴ、ＴＣＡ、ＴＣＣ、ＴＣＧ、及びＴＣＴ］に分類される。結果、各変異は、塩基対の変異パターンとコンテクストに従って、全部で９６（４×６×４）のタイプに分類される。さらに長いコンテクストを解析することも可能である。例えば、変異部位の両隣の２塩基ずつを含めた５塩基長のコンテクストに従うと、各変異は２５６群（４×４×４×４）に分類され、この分類と６つの塩基対パターンにより、各変異は最終的に全部で１５３６（４×４×６×４×４）のタイプに分類される。さらに変異部位の両隣のｎ塩基ずつを含めた２ｎ＋１塩基長のコンテクストに従うと、各変異は4^{2n}群に分類され、この分類と６つの塩基対パターンにより、各変異は最終的に全部で4^{2n}×６個のタイプに分類される。次いで、各変異タイプに属する変異の総数、及び解析した塩基の総数に基づいて、上記変異タイプの各々の変異頻度を決定することができる。

【０１１６】

次に、解析対象ゲノムＤＮＡにおける短い挿入／欠失変異を検出する場合の好ましい手順を記載する。短い挿入／欠失変異の検出においては、配列データをそれぞれ参照配列と比較することによって、各配列データにおける該参照配列に対して塩基が挿入又は欠失されている部位を検出する。該参照配列との比較には、取得された配列データの一部を用いてもよく、又は全部を用いてもよい。検出される挿入又は欠失部位としては、好ましくは挿入又は欠失した塩基の長さが１０ｂｐ以下、より好ましくは１～５ｂｐである部位がよいが、これに限定されない。検出された部位は、挿入又は欠失変異を有する変異部位として取得される。

【０１１７】

さらに、取得された各変異について、変異のタイプ（挿入変異か又は欠失変異か）、該挿入又は欠失部位の塩基長、あるいは挿入又は欠失した塩基の種類を決定することができる。特定の塩基長の挿入又は欠失部位を検出する手順は、上述したＰｙｔｈｏｎ等のプログラミング言語を用いて作成したプログラムを用いて行うことができる。さらに、各配列データと参照配列との比較によって、挿入又は欠失した塩基の種類を同定することができ

る。これらにより、各配列データにおける挿入又は欠失部位の塩基長、あるいは挿入又は欠失部位の塩基の種類を決定することができる。さらに、挿入又は欠失の頻度を、塩基長及び／又は塩基の種類ごとに決定してもよい。例えば、各リード配列について取得した挿入又は欠失変異を塩基長ごとに分類し、それぞれの頻度を決定することができる。また例えば、挿入又は欠失した塩基をその種類（Ａ、Ｔ、Ｇ、及びＣ）ごとに分類し、それぞれの頻度を決定することができる。さらに、該塩基長及び塩基の種類による分類を組み合わせたより細かい変異の分類を行い、それぞれの頻度を決定することができる。

【０１１８】

本発明の例示的実施形態として、さらに以下の物質、製造方法、用途、方法等を本明細書に開示する。ただし、本発明はこれらの実施形態に限定されない。

【０１１９】

〔１〕シーケンシング用ライブラリの調製方法であって、

サンプルＤＮＡを断片化すること；及び、

調製したサンプルＤＮＡの断片を１本鎖特異的ヌクレアーゼで処理し、該断片から１本鎖部分を除去すること、

を含む、

方法。

〔２〕前記サンプルＤＮＡが、

好ましくはホルマリン固定細胞のＤＮＡ又はｃｆＤＮＡではなく、より好ましくは、生細胞から抽出したＤＮＡ、凍結細胞から抽出したＤＮＡ、又はそれらのＤＮＡの保存サンプルであり、かつ

好ましくは、ＤＩＮが６以上、さらに好ましくは７以上、さらに好ましくは７．３以上、さらにより好ましくは７．５以上である、

〔１〕記載の方法。

〔３〕前記１本鎖特異的ヌクレアーゼが、

好ましくは、１本鎖特異的エンドヌクレアーゼ、１本鎖特異的エキソヌクレアーゼ、又はそれらの組み合わせであり、

より好ましくは、Ｓ１ ｎｕｃｌｅａｓｅ、Ｍｕｎｇ Ｂｅａｎ Ｎｕｃｌｅａｓｅ（ＭＢＮ）、ＲｅｃＪ$_f$、及びＥｘｏｎｕｃｌｅａｓｅ VIIからなる群より選択される少なくとも１種である、

〔１〕又は〔２〕記載の方法。

〔４〕好ましくは、前記１本鎖特異的ヌクレアーゼでの処理が、前記サンプルＤＮＡの断片を１本鎖特異的エンドヌクレアーゼで処理した後に、さらに１本鎖特異的エキソヌクレアーゼで処理することを含むか、又は１本鎖特異的エキソヌクレアーゼで処理した後に、さらに１本鎖特異的エンドヌクレアーゼで処理することを含む、〔３〕記載の方法。

〔５〕好ましくは、前記１本鎖特異的エンドヌクレアーゼがＳ１ ｎｕｃｌｅａｓｅであり、

前記サンプルＤＮＡの断片１ｎｇ当たりのＳ１ ｎｕｃｌｅａｓｅのユニット数（Ｕ／ｎｇ）が、

好ましくは０．０１Ｕ／ｎｇ以上、より好ましくは０．０２Ｕ／ｎｇ以上、さらに好ましくは０．０５Ｕ／ｎｇ以上であり、かつ好ましくは１６．７Ｕ／ｎｇ以下、より好ましくは５．００Ｕ／ｎｇ以下、さらに好ましくは１．６７Ｕ／ｎｇ以下であるか、又は、

好ましくは０．０２～５．００Ｕ／ｎｇ、より好ましくは０．０５～１．６７Ｕ／ｎｇである、〔３〕又は〔４〕記載の方法。

〔６〕好ましくは、前記１本鎖特異的エンドヌクレアーゼがＭＢＮであり、

前記サンプルＤＮＡの断片１ｎｇ当たりのＭＢＮのユニット数（Ｕ／ｎｇ）が、

好ましくは０．０１Ｕ／ｎｇ以上、より好ましくは０．０２Ｕ／ｎｇ以上、さらに好ましくは０．０３Ｕ／ｎｇ以上、さらに好ましくは０．０５Ｕ／ｎｇ以上、さらに好ましくは０．１０Ｕ／ｎｇ以上であり、かつ好ましくは１６．７Ｕ／ｎｇ以下、より好ましくは５．００Ｕ／ｎｇ以下、さらに好ましくは１．６７Ｕ／ｎｇ以下、さらに好ましくは１

．００Ｕ／ｎｇ以下、さらに好ましくは０．３０Ｕ／ｎｇ以下であるか、又は、
好ましくは０．０２～５．００Ｕ／ｎｇ、より好ましくは０．０３～１．６７Ｕ／ｎｇ、さらに好ましくは０．０３～１．００Ｕ／ｎｇ、さらに好ましくは０．０５～１．００Ｕ／ｎｇ、さらに好ましくは０．１０～０．３０Ｕ／ｎｇである、
〔３〕又は〔４〕記載の方法。
〔７〕好ましくは、前記１本鎖特異的エキソヌクレアーゼが$RecJ_f$であり、
前記サンプルＤＮＡの断片１ｎｇ当たりの$RecJ_f$のユニット数（Ｕ／ｎｇ）が、
好ましくは０．１０Ｕ／ｎｇ以上、より好ましくは０．３０Ｕ／ｎｇ以上であり、かつ好ましくは１００Ｕ／ｎｇ以下、より好ましくは１６．７Ｕ／ｎｇ以下、さらに好ましくは１．００Ｕ／ｎｇ以下であるか、又は、
好ましくは０．１０～１６．７Ｕ／ｎｇ、より好ましくは０．３０～１．００Ｕ／ｎｇである、
〔３〕又は〔４〕記載の方法。
〔８〕好ましくは、前記１本鎖特異的ヌクレアーゼで処理した前記サンプルＤＮＡの断片を、末端修復、末端への塩基付加、及び増幅からなる群より選択されるいずれか１つ以上の処理に供することをさらに含み、
より好ましくは、前記１本鎖特異的ヌクレアーゼで処理した前記サンプルＤＮＡの断片を、末端修復、末端への塩基付加、及び増幅に供することをさらに含む、
〔１〕～〔７〕のいずれか１項記載の方法。
〔９〕好ましくは、前記末端への塩基付加が、前記サンプルＤＮＡの断片の両末端への標識配列の付加である、〔８〕記載の方法。
〔１０〕好ましくは、前記増幅がＰＣＲである、〔８〕又は〔９〕記載の方法。
〔１１〕前記１本鎖特異的ヌクレアーゼがＳ１　ｎｕｃｌｅａｓｅであり、前記サンプルＤＮＡの断片１ｎｇあたりの該ヌクレアーゼのユニット数（Ｕ／ｎｇ）が０．０５Ｕ／ｎｇ以下のとき、前記ＰＣＲにおける該サンプルＤＮＡ　１Ｍｂｐあたりの初期ＤＮＡ量が、好ましくは２５０ａｍｏｌ以下、より好ましくは１２５ａｍｏｌ以下、さらに好ましくは６２．５ａｍｏｌ以下、さらにより好ましくは３１．３ａｍｏｌ以下、なお好ましくは１５．７ａｍｏｌであるか；
前記１本鎖特異的ヌクレアーゼがＳ１　ｎｕｃｌｅａｓｅであり、前記サンプルＤＮＡの断片１ｎｇあたりの該ヌクレアーゼのユニット数（Ｕ／ｎｇ）が０．０５Ｕ／ｎｇより大きいとき、下記式で算出される指標：
指標＝ＰＣＲにおける初期ＤＮＡ量（ａｍｏｌ／ＭｂｐサンプルＤＮＡ）×$3^{\log S1\ nuclease\ (U/ng)}$
（式中、Ｓ１　ｎｕｃｌｅａｓｅ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である）
が、好ましくは６０以下、より好ましくは３０以下、さらに好ましくは１５以下、さらにより好ましくは７．５以下であるか；
前記１本鎖特異的ヌクレアーゼがＭＢＮであり、前記サンプルＤＮＡの断片１ｎｇあたりの該ヌクレアーゼのユニット数（Ｕ／ｎｇ）が０．０５Ｕ／ｎｇ以下のとき、前記ＰＣＲにおける該サンプルＤＮＡ　１Ｍｂｐあたりの初期ＤＮＡ量が、好ましくは２５０ａｍｏｌ以下、より好ましくは１２５ａｍｏｌ以下、さらに好ましくは６２．５ａｍｏｌ以下、さらにより好ましくは３１．３ａｍｏｌ以下、なお好ましくは１５．７ａｍｏｌであるか；
前記１本鎖特異的ヌクレアーゼがＭＢＮであり、前記サンプルＤＮＡの断片１ｎｇあたりの該ヌクレアーゼのユニット数（Ｕ／ｎｇ）が０．０５Ｕ／ｎｇより大きいとき、下記式で算出される指標：
指標＝ＰＣＲにおける初期ＤＮＡ量（ａｍｏｌ／ＭｂｐサンプルＤＮＡ）×$3^{\log MBN\ (U/ng)}$
（式中、ＭＢＮ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である）
が、好ましく６０以下、より好ましくは３０以下、さらに好ましくは１５以下、さらにより好ましくは７．５以下である、

〔１０〕記載の方法。
【０１２０】
〔１２〕前記〔１〕～〔１１〕のいずれか１項記載の方法で調製されたシーケンシング用ライブラリをシーケンシングすることを含む、ＤＮＡのシーケンシング方法。
〔１３〕好ましくは、前記シーケンシング方法が、以下：
（１）前記ライブラリをシーケンシングし、該ライブラリに含まれる複数の増幅断片の各々について１つ以上のリード配列を作成し、該複数の増幅断片についての複数のリード配列を得ること；
（２）得られた複数のリード配列の中から、該ライブラリの調製に用いたサンプルＤＮＡ上の同一領域の配列情報を有するリード配列を集めてグループ化することにより、リード配列のグループを１つ以上作成すること；及び、
（３）該リード配列のグループに含まれるリード配列の間で配列情報のコンセンサスを取ること、
を含む、〔１２〕記載の方法。
〔１４〕好ましくは、前記（１）が、前記サンプルＤＮＡの断片を構成する２本の相補鎖の各々に由来する増幅断片に対して１つ以上のリード配列を作成することを含む、〔１３〕記載の方法。
〔１５〕好ましくは、前記（２）が、参照配列上の同一の位置にマッピングされるリード配列を同じグループに分けることを含む、〔１４〕記載の方法。
〔１６〕好ましくは、前記（３）が、前記リード配列のグループの中から、前記サンプルＤＮＡ断片の２本の相補鎖の各々に由来するリード配列を少なくとも１つずつ集め、集めたリード配列の間で配列情報のコンセンサスを取ることを含む、〔１５〕記載の方法。
〔１７〕好ましくは、
前記（１）において、前記複数のリード配列が、以下からなるリード配列のペアを複数個含み：
リード１：前記増幅断片を構成する２本の相補鎖のうちの一方の鎖の配列を５'末端側から３'側へ読んだ配列に相当する配列情報を含むリード配列、
リード２：該一方の鎖の配列を３'末端側から５'側へ読んだ配列に相当する配列情報を含むリード配列、
前記（２）が、得られたリード配列のペアの中から、該サンプルＤＮＡ上の同一領域の配列情報を有するリード配列のペアを集めてグループ化することにより、リード配列のペアのグループを１つ以上作成することを含み、
前記（３）が、該リード配列のペアのグループに含まれるリード配列の間で配列情報のコンセンサスを取ることを含む、
〔１３〕記載の方法。
〔１８〕好ましくは、前記（１）が、前記サンプルＤＮＡの断片を構成する２本の相補鎖の各々に由来する増幅断片に対して１つ以上の前記リード配列のペアを作成することを含む、〔１７〕記載の方法。
〔１９〕好ましくは、前記（２）が、前記リード配列のペアのリード１とリード２を参照配列に対してマッピングし、リード１の先頭とリード２の先頭とに挟まれる該参照配列の領域が同一であるリード配列のペアを同じグループに分けることを含む、〔１８〕記載の方法。
〔２０〕好ましくは、前記（２）が、前記リード配列のペアに含まれる一方のリード配列の先頭が前記参照配列上の同じ位置に位置するリード配列のペアを集め、次いで集めたリード配列のペアの中から、該リード配列のペアに含まれるもう一方のリード配列の先頭が該参照配列上の同じ位置に位置するリード配列のペアを集めて、集めたリード配列のペアを同じグループに分けることを含む、〔１８〕記載の方法。
〔２１〕好ましくは、前記（３）が、前記リード配列のペアのグループの中から、前記サンプルＤＮＡ断片の２本の相補鎖の各々に由来するリード配列のペアを少なくとも１組ずつ集め、集めたリード配列のペアに含まれるリード配列の間で配列情報のコンセンサスを

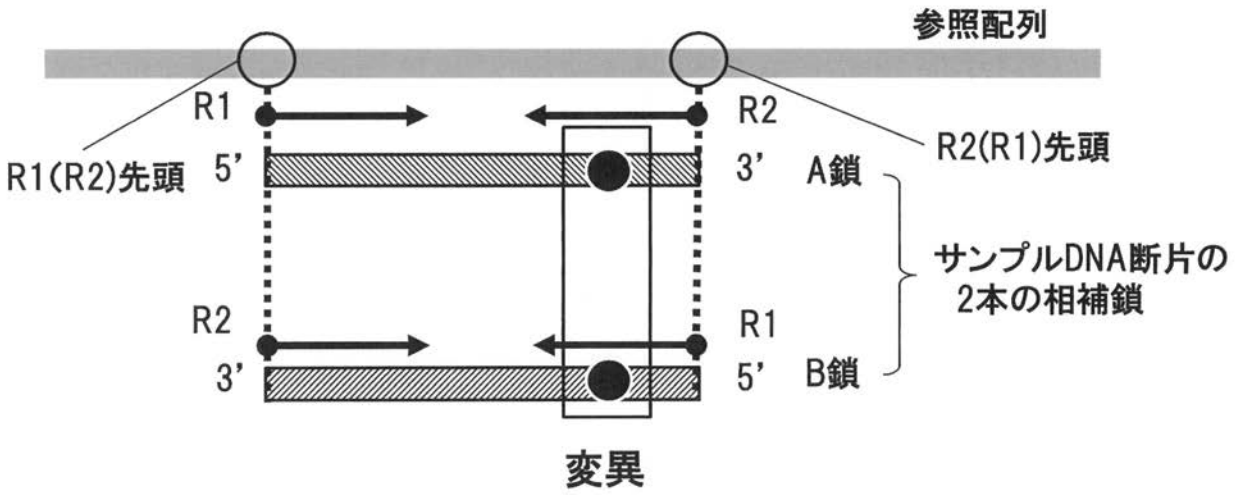
取ることを含む、〔１９〕又は〔２０〕記載の方法。
【０１２１】
〔２２〕ゲノムＤＮＡをサンプルＤＮＡとして用いて、前記〔１〕～〔１１〕のいずれか１項記載の方法によりシーケンシング用ライブラリを調製すること；及び、
該シーケンシング用ライブラリをシーケンシングすること、
を含む、ゲノムＤＮＡの変異を検出する方法。
〔２３〕好ましくは、前記シーケンシングが前記〔１３〕～〔２１〕のいずれか１項記載の方法により行われる、〔２２〕記載の方法。
〔２４〕好ましくは、前記変異が塩基対置換型変異である、〔２２〕又は〔２３〕記載の方法。
【実施例】
【０１２２】
以下、実施例を示し、本発明をより具体的に説明する。
【０１２３】
参考例１　シーケンシング及び変異解析
後述の比較例及び実施例で用いたシーケンシング方法及び変異解析のフローを以下に説明する。基本的には、特許文献４に記載される相補鎖情報を活用した高精度シーケンシング法を用いた。具体的には、ライブラリをシーケンシングし、同一のＤＮＡ断片に由来すると推定されるリードペアを集めた。次いで、該ＤＮＡ断片の２本の相補鎖（以下、Ａ鎖及びＢ鎖と称する）のそれぞれに由来すると推定されるリード配列間でのコンセンサスリード配列（相補鎖間コンセンサスリード配列）を作成した。得られた相補鎖間コンセンサスリード配列は変異解析に使用した。
【０１２４】
１）相補鎖情報を活用したシーケンシング
シーケンサーにはイルミナ社のＨｉＳｅｑを用いた。ＨｉＳｅｑシーケンサー用のライブラリには、サンプルＤＮＡ断片の２本の相補鎖の双方に由来するＰＣＲ産物が含まれる。したがって、このライブラリをシーケンシングすることで、該２本の相補鎖のそれぞれについてリード１とリード２を作成した。
【０１２５】
互いに相補的な鎖のリード配列を識別するため、ＰＣＲの前に、サンプルＤＮＡ断片の両末端に、相補鎖標識配列（イルミナ社のＴｒｕＳｅｑに付属のアダプター配列）を連結した。次いで、該アダプター配列に特異的に結合するプライマーを用いたＰＣＲにより、該アダプター配列を含むＰＣＲ産物を生成し、シーケンシング用のライブラリとして用いた。ＨｉＳｅｑシーケンサーにおいては、該アダプター配列がシーケンシングに使用されるフローセル上のオリゴＤＮＡ断片とアニーリングすることで、該フローセル上に増幅産物が結合され、シーケンシングされる。
【０１２６】
シーケンシングでは、ライブラリ中の各ＰＣＲ産物に含まれる個々の増幅断片（サンプルＤＮＡ断片に由来する）に対して、それぞれ２本のリード配列（リード１、リード２）のペアが取得された。このとき、該増幅断片の一方の鎖の配列を５'側から３'側へ読んだ配列情報を含むリード配列がリード１（Ｒ１）であり、同じ鎖の配列を３'側から５'側へ読んだ配列に相当する配列情報を含むリード配列がリード２（Ｒ２）であった。
【０１２７】
２）リード配列の編集、及び相補鎖情報の抽出
１）で得られたリード配列を、アダプター配列及びクオリティの低い塩基等のトリミングを行った後、参照配列へマッピングした。サンプルＤＮＡ断片の２本の相補鎖由来のリードペアを参照配列上にマッピングしたときの、参照配列に対する各リードペアの配置の概念図を模式図１に示す。参考のため、模式図１には、各リードペアが由来するサンプルＤＮＡ断片の２本の相補鎖を図示する。互いに相補的な鎖に由来するリードペアの間では、リード１の先頭とリード２の先頭とに挟まれる参照配列の領域は同一である。したがっ

て参照配列上でのリードペアのマッピング位置に基づいて、同じサンプルＤＮＡ断片に由来すると考えられるリードペアを集めた。

【０１２８】

【化３】

模式図1: 参照配列に対する相補鎖由来リードペアのマッピングの概念図。



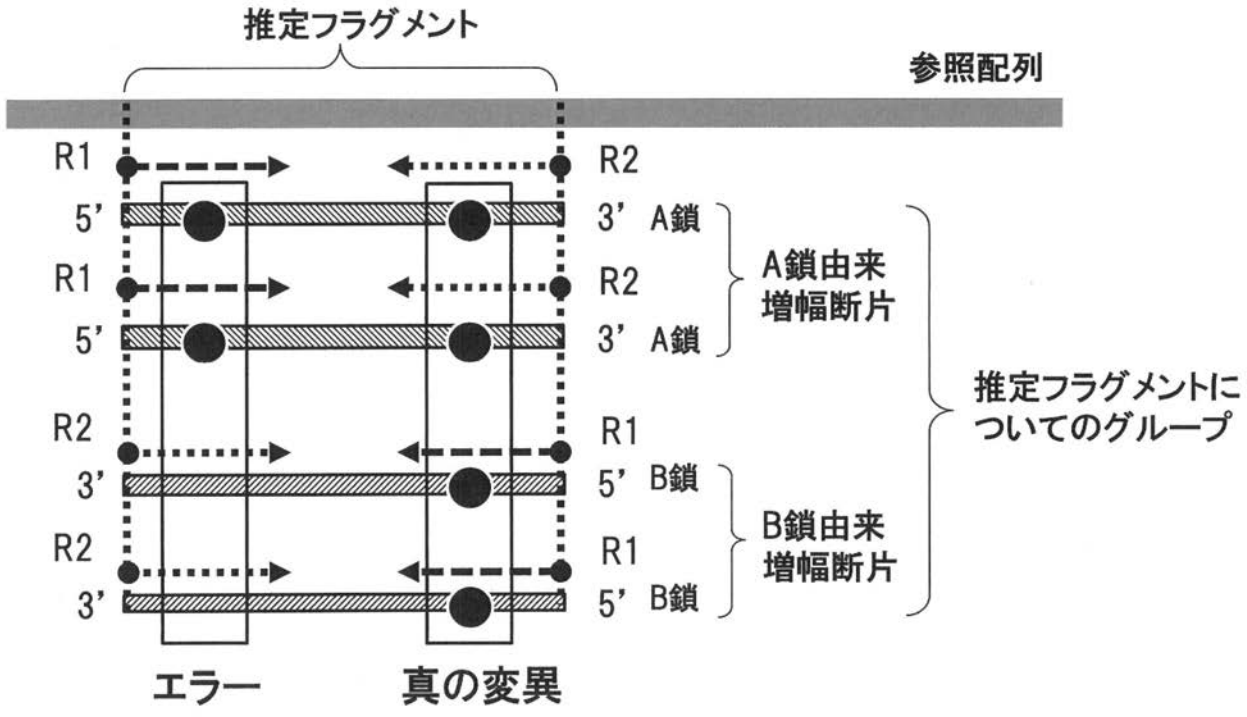
【０１２９】

なお本方法に関し、参照配列上における、マッピングしたリード１（リード２）の先頭からリード２（リード１）の先頭までの領域、言い換えると、リードペア（リード１、リード２）を参照配列上にマッピングしたときに、リード１の先頭とリード２の先頭とに挟まれる該参照配列の領域を、「推定フラグメント」と称する。推定フラグメントが共通するリードペアの群を、推定フラグメントについての「グループ」と称する（模式図２）。

【０１３０】

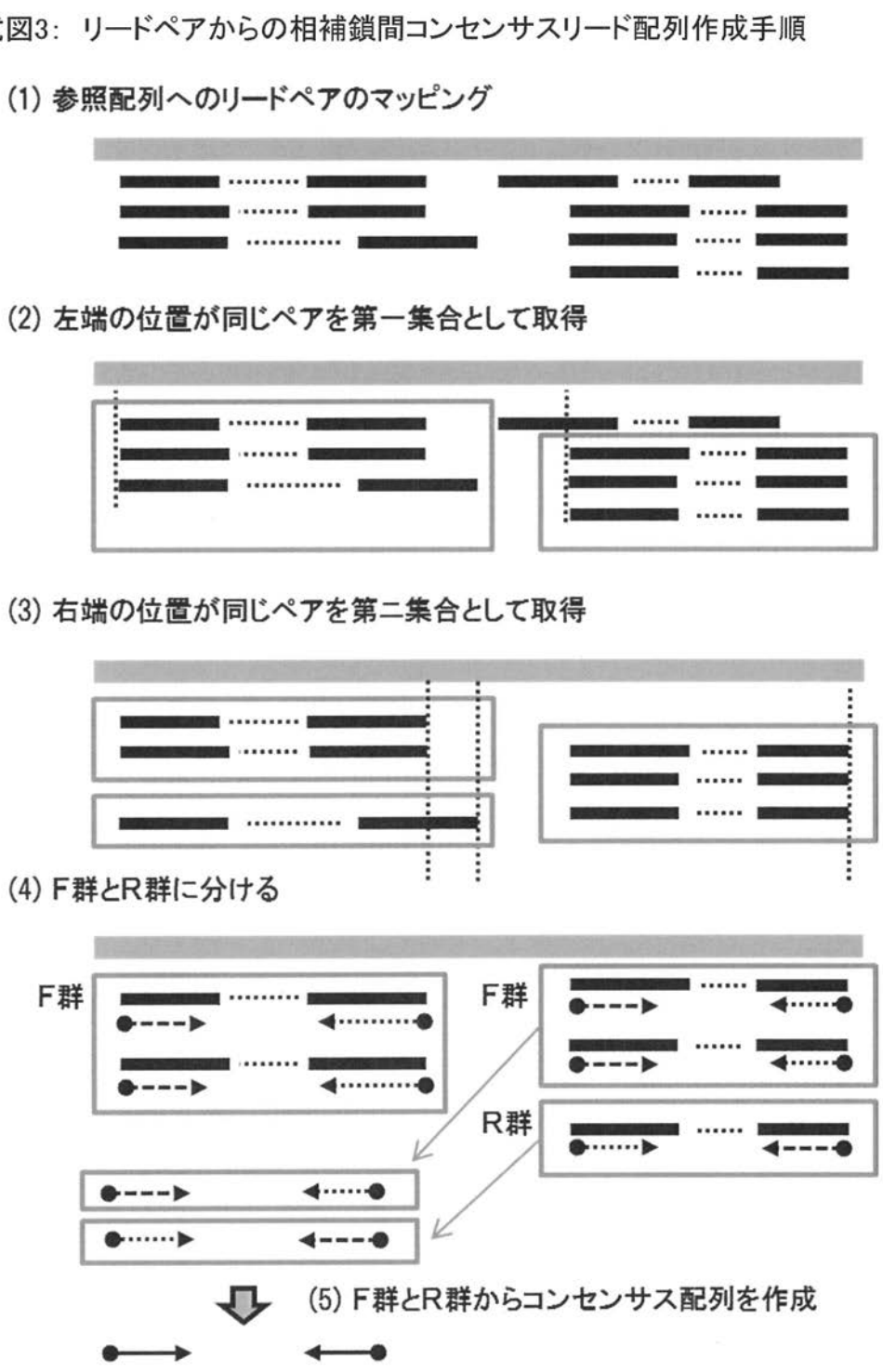
【化４】

模式図2: 推定フラグメント



【０１３１】
次いで、推定フラグメントについてのグループから、互いに相補的な２本の鎖のそれぞれに由来するリードペアの組み合わせを、リードペアのセットとして取得した。
【０１３２】
サンプルＤＮＡ断片から得られた増幅断片は、サンプルＤＮＡ断片に元々含まれる変異を両鎖に保有するのに加えて、片方の鎖のみに、サンプルＤＮＡ断片の酸化修飾などに起因する塩基の置換を有することがある。このようなケースを模式図１、２に例示する。模式図１に示すサンプルＤＮＡ断片は、変異による塩基の置換（真の変異）を両鎖に１つずつ保有する。一方、模式図２に示した該サンプルＤＮＡ断片由来の増幅断片は、変異による塩基の置換（真の変異）を両鎖に保有するのに加え、片方の鎖のみにサンプル調製過程で生じた塩基の置換（エラー）を有する。これらの真の変異及びエラーは、各リードペアのリード１とリード２に読み取られている。本方法では、相補鎖に由来するリードペアのセットの有する配列情報から、両鎖に固定された真の変異と片方の鎖のみに生じたエラーとを区別し、真の変異を抽出した。
【０１３３】
本方法では、集めたリードペアのセットから相補鎖間コンセンサスリード配列を作成した。相補鎖間コンセンサスリード配列の作成においては、まず、推定フラグメントの共通するリードペアを集め、それらをＡ鎖由来のリードペアとＢ鎖由来のリードペアとに分けた。次いで、１つ以上のＡ鎖由来のリードペアと１つ以上のＢ鎖由来のリードペアとの組み合わせをリードペアのセットとして取得し、それらを用いて相補鎖間コンセンサスリード配列を作成した。リードペアのセットに含まれるＡ鎖由来又はＢ鎖由来のリードペアの数は特に限定されず、Ａ鎖由来とＢ鎖由来双方のリードペアが少なくとも１つ以上含まれていれば良いとした。例えば、Ａ鎖由来のリードペアが２つで、Ｂ鎖由来のリードペアが２つの場合や、Ａ鎖由来のリードペアが３つで、Ｂ鎖由来のリードペアが１つの場合でも、それらの間でコンセンサスを取ることで相補鎖間コンセンサスリード配列を作成した。
【０１３４】
リードペアの集合化から相補鎖間コンセンサスリード配列作成までのより具体的な手順の例を、以下の模式図３に示す。模式図３のとおり、本方法では、まず、各相補鎖由来のリードペアを参照配列にマッピングした（１）。このとき、参照配列上で左端（参照配列上の最も５'側に配置する端）が同じ位置にあるリードペアの群を第一集合として取得した（２）。次いで、該第一集合から、参照配列上で右端（参照配列上の最も３'側に配置する端）が同じ位置にあるリードペアの群を分け、第二集合として取得した（３）。この第二集合は、推定フラグメントの共通するリードペアの集合であった。次いで、第二集合を、Ａ鎖に由来する群（Ｆ群）と、Ｂ鎖に由来する群（Ｒ群）とに分けた（４）。このとき、Ａ鎖に由来する群であるかＢ鎖に由来する群であるかは、シーケンシングの際に取得される標識配列の情報に基づいて識別することができた。本方法においては、サンプルＤＮＡ断片に付加されたアダプター配列中の標識配列を認識し結合するフローセルを用いてシーケンシング反応を行った。フローセル内での断片の増幅後、５'側に付加されたアダプター配列中の標識配列を特異的に切断することにより、各増幅断片のリード１、リード２のシーケンシングの方向性を統一することで、標識配列の情報に基づいてリードペアをＦ群とＲ群とに分けた。該Ｆ群とＲ群は、それぞれ、ＤＮＡ断片を構成する２本の相補鎖のいずれか一方に由来するリードペアの集合であった。したがって、該Ｆ群とＲ群との間でコンセンサスを取ることにより、相補鎖間コンセンサスリード配列を作成した（５）。
【０１３５】

【化５】



【０１３６】

相補鎖間コンセンサスリード配列を作成することにより、片方の鎖にのみ生じた置換はエラーとして除外し、両方の鎖に共通して存在する置換を真の変異として取得した。

【０１３７】

３）変異解析

２）で得られた相補鎖間コンセンサスリード配列を参照配列上に再度マッピングすることで、解析対象ゲノムの変異を検出した。参照配列に再マッピングした相補鎖間コンセンサスリード配列から変異した塩基を検出するための具体的な手順は、ＰＣＴ／ＪＰ２０１７／００５７００に記載された手順に従った。

【０１３８】

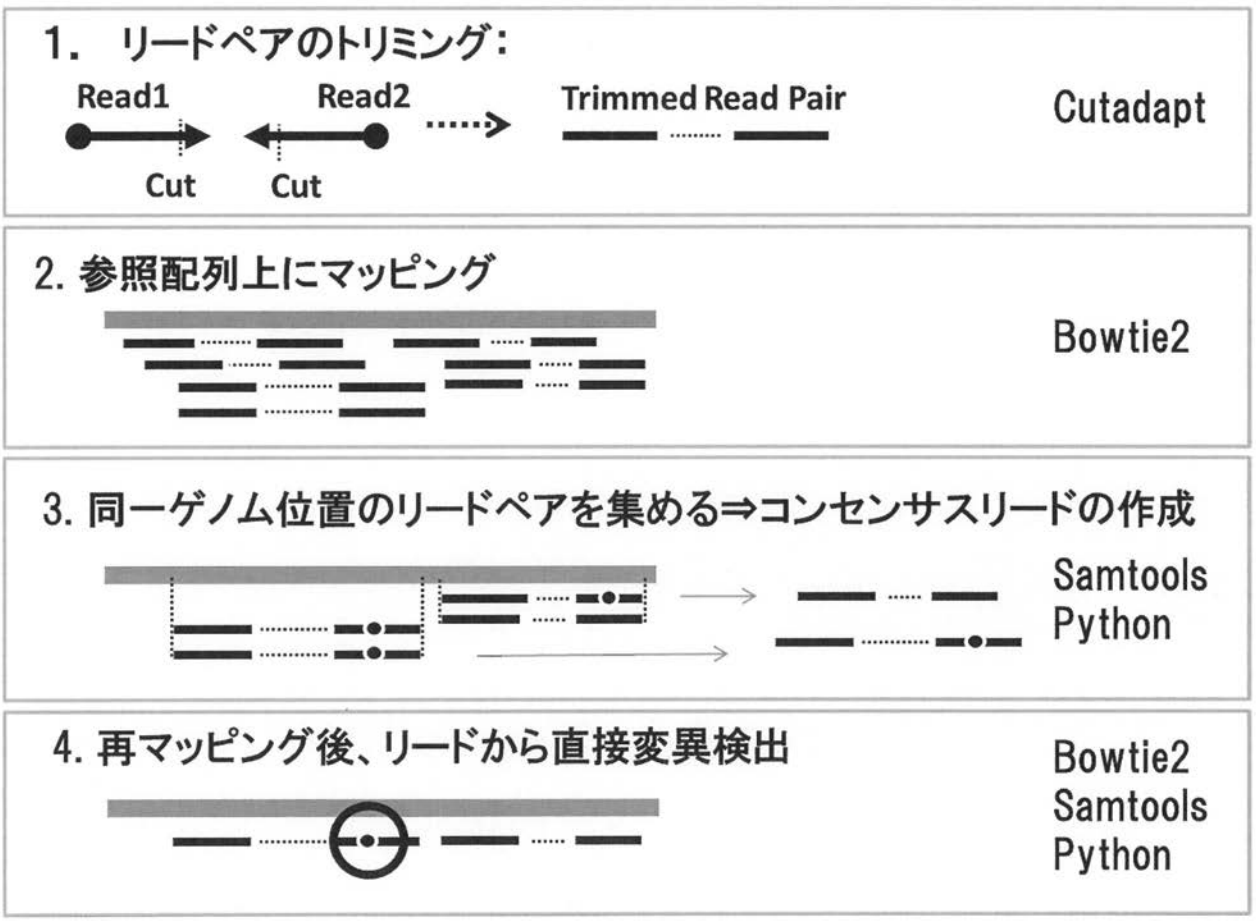
４）ソフトウェア、プログラム

リード配列の編集、相補鎖情報の抽出、及び変異解析のフローを模式図４に示す。解析には、Ｃｕｔａｄａｐｔソフトウェア、Ｂｏｗｔｉｅ２ソフトウェア、Ｓａｍｔｏｏｌｓソフトウェア、及びプログラミング言語Ｐｙｔｈｏｎを用いて作成したプログラムを用いた。まず、各ライブラリ由来のＦａｓｔｑファイル（リード１、及びリード２）に対して、Ｃｕｔａｄａｐｔソフトウェアを用いて、アダプター配列及びクオリティの低い塩基等のトリミングを行った。その後、各ライブラリ由来のＦａｓｔｑファイルを、Ｂｏｗｔｉｅ２ソフトウェアを用いて参照配列へマッピングし、Ｓａｍフォーマットのファイルを得た。Ｓａｍｔｏｏｌｓソフトウェアを用いてＳａｍフォーマットのファイルのリードの並び替えを行い、次いで、プログラミング言語Ｐｙｔｈｏｎで作成したプログラムを用いて、推定フラグメントについてのグループを作成し、その中からリードペアのセットを集め、相補鎖間コンセンサスリード配列を作成した。得られた相補鎖間コンセンサスリード配列を、再度Ｂｏｗｔｉｅ２ソフトウェアで参照配列にマッピングし、Ｓａｍｔｏｏｌｓソフトウェア、及び、プログラミング言語Ｐｙｔｈｏｎで作成したプログラムを用いて、変異解析を行った。

【０１３９】

【化６】

模式図4: 解析フロー



【０１４０】

比較例１　リードペアの両端の塩基の除去によるエラー低減

参考例１のシーケンシング法を用いて、新鮮なゲノムＤＮＡの断片における末端１本鎖突出部位に由来するエラーの存在を検証した。また、末端部のエラーに対する既存の改善法であるリードペアの両端から塩基を除く方法によるエラーの低減効果を検討した。

【０１４１】

１）サンプルＤＮＡ

サンプルＤＮＡとして、ジメチルスルホキシド（ＤＭＳＯ；和光純薬工業製）を暴露したＳａｌｍｏｎｅｌｌａ ｔｙｐｈｉｍｕｒｉｕｍ ＬＴ－２ ＴＡ１００株（以下、単に「ＴＡ１００株」とも称する）のゲノムＤＮＡを用いた。

【０１４２】

ＴＡ１００細胞株のＤＭＳＯへの暴露は、Ａｍｅｓ試験のプレインキュベーション法に準拠して実施した（Ｍｏｌ．Ｍｅｃｈ．Ｍｕｔａｇｅｎ．，４５５：２９－６０，２０００，Ｓｃｉ．Ｒｅｐ．８（１）：９５８３）。２ｍＬのニュートリエントブイヨンＮｏ．２（Ｏｘｏｉｄ社製）にＴＡ１００株を植菌し、３７℃、１８０ｒｐｍで４時間振とう培養し、ＯＤ６６０値が１．０以上の前培養液を得た。試験管内に、ＤＭＳＯ１００μＬ、Ｓ９ ｍｉｘ（家田貿易社製）５００μＬ、及び前培養液１００μＬを添加し、３７℃のウォーターバス中で２０分間、１００ｒｐｍで振とう培養した（ＤＭＳＯ暴露細胞）。２０分間の振とう培養後、培養液を含む試験管をウォーターバスから取り出し、予め分注しておいた２ｍＬのＮｕｔｒｉｅｎｔ Ｂｒｏｔｈ溶液（Ｓ９ ｍｉｘを１８．５％含む）に培養液５０μＬを添加し、インキュベーター内で３７℃、１８０ｒｐｍで１４時間追培養した。培養後、菌懸濁液を回収し、７５００ｒｐｍで５分間遠心し、上清を除去して細胞を回収した。ＤＭＳＯ暴露細胞から、ＤＮｅａｓｙ Ｂｌｏｏｄ ＆ Ｔｉｓｓｕｅ Ｋｉｔ（キアゲン社製）を用い、推奨プロトコルに従って、Ｔｏｔａｌ ＤＮＡを回収した。得られたＤＮＡサンプルの２本鎖ＤＮＡの濃度を、Ｑｕｂｉｔ３．０ Ｆｌｕｏｒｏｍｅｔｅｒ（Ｔｈｅｒｍｏ Ｆｉｓｈｅｒ Ｓｃｉｅｎｔｉｆｉｃ社製）を用いて、付属のＱｕｂｉｔ™ ｄｓＤＮＡ ＢＲ Ａｓｓａｙ Ｋｉｔで測定した。

【０１４３】

２）シーケンシング用ライブラリの調製

サンプルＤＮＡからのライブラリ調製には、ＴｒｕＳｅｑ Ｎａｎｏ ＤＮＡ Ｌｉｂｒａｒｙ Ｐｒｅｐ Ｋｉｔ（イルミナ社製、以下ＴｒｕＳｅｑと略記する）を用いた。ＴｒｕＳｅｑの推奨プロトコルは、ＤＮＡの断片化、Ｅｎｄ Ｒｅｐａｉｒ（２本鎖ＤＮＡ断片の１本鎖突出末端の平滑化）、Ａ－ｔａｉｌｉｎｇ（２本鎖ＤＮＡ断片の３'末端へのアデニンの付加）、Ａｄａｐｔｅｒ ｌｉｇａｔｉｏｎ（２本鎖ＤＮＡ断片両末端へのアダプターの付加）、及びＰＣＲ ｅｎｒｉｃｈｍｅｎｔ（ＰＣＲ増幅によるライブラリＤＮＡの濃縮）から構成される。１）で得たＤＭＳＯ暴露細胞由来ＤＮＡの１２０ｎｇ相当量を複数サンプル用意し、それらをＤＮＡ Ｓｈｅａｒｉｎｇシステム ＭＥ２２０（コバリス社製）で推奨プロトコルに従って平均約３５０ｂｐの長さに断片化した。得られた断片化ＤＮＡに、Ｅｎｄ Ｒｅｐａｉｒ、Ａ－ｔａｉｌｉｎｇ、Ａｄａｐｔｏｒ Ｌｉｇａｔｉｏｎを実施した。得られたＡｄａｐｔｏｒ Ｌｉｇａｔｉｏｎの反応液を推奨プロトコルに従って精製し、２本鎖ＤＮＡ断片の両末端にアダプターが付加されたＤＮＡ（アダプター付加ＤＮＡ）を得た。Ａｇｉｌｅｎｔ ４２００ ＴａｐｅＳｔａｔｉｏｎ（アジレント・テクノロジー社製）のＨｉｇｈ Ｓｅｎｓｉｔｉｖｉｔｙ Ｄ５０００キットを用いてアダプター付加ＤＮＡの濃度を測定した。

【０１４４】

特許文献４に基づいて、ＰＣＲに用いるアダプター付加ＤＮＡの初期量（初期ＤＮＡ量）の最適条件を７８ａｍｏｌ（１５．６ａｍｏｌ／Ｍｂｐ）と推定した。これに従い、ＴｒｕＳｅｑに付属のＲｅｓｕｓｐｅｎｓｉｏｎ ｂｕｆｆｅｒで段階的に希釈し、７８ａｍｏｌのアダプター付加ＤＮＡを含む希釈液２５μＬを得た。得られた希釈液を、推奨プロトコルに従いＰＣＲ ｅｎｒｉｃｈｍｅｎｔに供した。７８ａｍｏｌの初期ＤＮＡ量とシーケンシングに必要なＤＮＡ量を考慮して、１５サイクルのＰＣＲを実施した。反応液から推奨プロトコルに従ってＤＮＡを精製し、ライブラリとした。Ａｇｉｌｅｎｔ ４２００ ＴａｐｅＳｔａｔｉｏｎのＨｉｇｈ Ｓｅｎｓｉｔｉｖｉｔｙ Ｄ１０００キットを用いてライブラリＤＮＡの濃度を測定した。

【０１４５】

３）シーケンシング及び変異解析

２）で調製したライブラリを、２×１００ｂｐのリード長でシーケンシングし、ライブラリあたり、平均で約１０Ｇｂｐ（約５０Ｍリードペア）のシーケンシングデータを得た。得られたシーケンシングデータから相補鎖間コンセンサスリード配列を作成し、参照配列にマッピングした後、変異した塩基を検出した。シーケンシング、相補鎖間コンセンサスリード配列の作成、及び変異解析は参考例１の手順に従って実施した。なお、参照配列には、ＧｅｎＢａｎｋ（www.ncbi.nlm.nih.gov/genbank/）から取得したＳ．ｔｙｐｈｉｍｕｒｉｕｍ　ＬＴ－２株（以下、単にＬＴ－２株とも略記する。）のゲノム配列を用いた（ＧｅｎＢａｎｋ　ａｓｓｅｍｂｌｙ　ａｃｃｅｓｓｉｏｎ：ＧＣＡ＿０００００６９４５．２）。

【０１４６】

４）変異頻度の算出

Ｐｙｔｈｏｎで作成したプログラムを用いて、各ライブラリについて、参照配列にマッピングされた全相補鎖間コンセンサスリード配列中の全解析対象塩基を、対応する参照配列の塩基（Ａ、Ｔ、Ｇ、及びＣ）によって４群に分けた。そして、各群の塩基の総数と参照配列に対して変異した塩基を検出した。検出された変異を、６つの変異パターン（ＡＴ→ＴＡ、ＡＴ→ＣＧ、ＡＴ→ＧＣ、及びＧＣ→ＴＡ、ＧＣ→ＣＧ、ＧＣ→ＡＴ）に分類し、各変異パターンにおける変異頻度を算出した。さらに、各変異パターンを、リード配列がマッピングされた参照配列上の塩基によって、さらに２パターンの変異に分類して、各々の変異頻度を算出した。すなわち、ＡＴ→ＴＡはＡ→Ｔ及びＴ→Ａに、ＡＴ→ＣＧはＡ→Ｃ及びＴ→Ｇに、ＡＴ→ＧＣはＡ→Ｇ及びＴ→Ｃに、ＧＣ→ＴＡはＧ→Ｔ及びＣ→Ａに、ＧＣ→ＣＧはＧ→Ｃ及びＣ→Ｇに、ＧＣ→ＡＴはＧ→Ａ及びＣ→Ｔに分類して、これら１２種の変異パターンそれぞれについて変異頻度を算出した。

【０１４７】

５）リードペアの両端からの塩基の除去によるエラー低減

参考例１の手順に従って、３）で得た相補鎖間コンセンサスリード配列を再度参照配列にマッピングしてＳａｍフォーマットのファイルを作成した。該Ｓａｍフォーマットファイル中で、リードペアの両端の０塩基（ｃｏｎｔｒｏｌ）、１０塩基、又は２０塩基を、Ｐｙｔｈｏｎで作成したプログラムを用いてクオリティ値を下げることで、変異解析の対象から除外した。その後、参考例１の手順に従って変異解析を行った。変異頻度は、４）に示した１２種の変異パターンについて算出した。

【０１４８】

６）結果と考察

４）で算出したサンプルＤＮＡにおける６つの変異パターンについての変異頻度を図１に示す。ＡＴ塩基対の変異頻度に比べてＧＣ塩基対の変異頻度が大きいことから、グアニンの酸化修飾によるエラーの存在が推測された。また、５）で算出した両末端を除去したリードペアから求めた１２種の変異パターンについての変異頻度を図２に示す。ＧＣ塩基対の変異（ＧＣ→ＴＡ、ＧＣ→ＣＧ）において、Ｃ→Ａ、Ｃ→Ｇに比べて、Ｇ→Ｔ、Ｇ→Ｃの変異が高頻度に検出された。真の変異は、Ｇ、Ｃの両塩基で同等の頻度で検出されるはずである。高頻度のグアニンの変異が検出されたことは、これが真の変異ではなく、酸化修飾等による塩基の変異に起因するエラーであることを示唆する。また、Ｇ→Ｔ、Ｇ→Ｃの変異頻度は、リードペアの両端から除去した塩基数に依存して減少した。この結果は、該グアニンの変異によるエラーがリードペアの両端部に多く存在していることを示した。したがって、ＤＮＡ断片の末端１本鎖部位における酸化修飾等によるグアニンの変異が、該エラーの主な原因となっていると考えられた。

【０１４９】

次いでＧＣ→ＴＡ及びＧＣ→ＣＧの変異について、Ｇの変異とＣの変異の間での変異頻度の差を算出し、下記式に基づいて、リードペアの両端除去によるエラーの減少率を求めた。

エラー減少率（％）＝（Ａ－Ｂ）／Ａ×１００

Ａ：両端の塩基を除去しないとき（ｃｏｎｔｒｏｌ）のＧＣ間の変異頻度の差

Ｂ：両端から塩基を除いたときのＧＣ間の変異頻度の差

エラー減少率を表１に示す。エラー減少率は、１０塩基の除去で＜３０％であり、２０塩基の除去でも４０％程度であった。なおＫｅｎｎｅｄｙら（非特許文献３）が報告した両端から５塩基除く方法は、１０塩基除くよりもさらにエラー低減効果が小さいと推測された。これらの結果は、両端２０塩基の除去ではＤＮＡ断片の末端１本鎖突出部分を十分に削除できなかったことを表す。除去する塩基数を増加することによりエラーをより低減できると予想されるが、リードペアからの多数の塩基の削除は、変異解析に充てられる塩基数が減少するため解析効率を低下させる。結果、ＤＮＡ断片の末端１本鎖部位における酸化修飾等に起因するエラーの改善にとって、リードペアの両端の変異解析対象からの除去は有効なアプローチとは言えない。

【０１５０】

【表１】

変異パターン	エラー減少率（%）	
	両端部10塩基除去	両端部20塩基除去
GC to TA	27.5	40.2
GC to CG	26.2	40.0

【０１５１】

実施例１　１本鎖特異的ヌクレアーゼを用いたライブラリ調製法によるエラー低減

ＤＮＡ断片の１本鎖特異的ヌクレアーゼ処理によるエラー低減効果を評価した。

【０１５２】

１）サンプルＤＮＡ

比較例１の１）と同様の手順で、ＤＭＳＯ暴露細胞を調製した、また同様の手順で、ＴＡ１００株を３－Ｍｅｔｈｙｌｃｈｏｌａｎｔｈｒｅｎｅ（３－ＭＣ）に暴露した。３－ＭＣ（シグマアルドリッチ社製、ＣＡＳＲＮ．５６－４９－５）は、ＤＭＳＯに溶解した。試験管内に、３－ＭＣ溶液１００μＬ、Ｓ９ ｍｉｘ（家田貿易社製）５００μＬ、及びＴＡ１００株の前培養液１００μＬを添加し（３－ＭＣ量：１０００μｇ／ｔｕｂｅ）、３７℃のウォーターバス中で２０分間、１００ｒｐｍで振とう培養した（３－ＭＣ暴露細胞）。比較例１の１）と同様の手順で菌懸濁液から細胞を回収し、ＤＮＡを抽出した。

【０１５３】

２）Ａｍｅｓ試験

Ａｍｅｓ試験用に、上記と同様の条件で３－ＭＣを暴露した菌懸濁液を調製した。これに、４５℃に加温した２ｍＬのｔｏｐ ａｇａｒ（１％ＮａＣｌ、１％ａｇａｒ、０．０５ｍＭ Ｈｉｓｔｉｄｉｎｅ及び０．０５ｍＭ Ｂｉｏｔｉｎを含む）を添加し、ボルテックスで撹拌した後、最小グルコース寒天培地（テスメディア（登録商標）ＡＮ；オリエンタル酵母工業製）の上に重層した。得られたプレートを３７℃で４８時間培養後、観察されたコロニーを計数した。

【０１５４】

３）シーケンシング用ライブラリの調製

Ｉ）サンプルＤＮＡの断片化

ＤＭＳＯ暴露細胞又は３－ＭＣ暴露細胞由来ＤＮＡの６０ｎｇ又は１００ｎｇ相当量を複数サンプル用意し、それらをＤＮＡ Ｓｈｅａｒｉｎｇシステム ＭＥ２２０で平均約３５０ｂｐの長さに断片化した。各サンプルの断片を２群に分けた。ヌクレアーゼで処理しない群（非処理群）については、次の工程のＥｎｄ Ｒｅｐａｉｒを行うために、推奨プロトコルに従って、ＴｒｕＳｅｑに付属のＲｅｓｕｓｐｅｎｓｉｏｎ ｂｕｆｆｅｒでＤＮＡ断片を懸濁し、６０μＬの溶出液を得た。ヌクレアーゼで処理する群（処理群）については、ＤＮＡ断片をＴｒｕＳｅｑに付属のＳａｍｐｌｅ Ｐｕｒｉｆｉｃａｔｉｏｎ Ｂｅａｄｓ（以下、単にビーズとも略記する）に吸着させ、８０％エタノール水で２回洗浄し、乾燥させるステップを推奨プロトコルに従って行い、精製した。その後の溶出操

作では、Ｄｉｓｔｉｌｌｅｄ ｗａｔｅｒ（ＤＷ、ニッポンジーン社製）でビーズを懸濁し、３０μＬのＤＮＡ断片を含むＤＮＡ溶出液を得た。

【０１５５】

II）ヌクレアーゼ処理

１本鎖特異的ヌクレアーゼには、Ｓ１ ｎｕｃｌｅａｓｅ（プロメガ社、カタログ番号：Ｍ５７６１）、Ｍｕｎｇ Ｂｅａｎ Ｎｕｃｌｅａｓｅ（ＭＢＮ）（タカラバイオ社、カタログ番号：２４２０Ａ）、又は$RecJ_f$（Ｎｅｗ Ｅｎｇｌａｎｄ Ｂｉｏｌａｂｓ社、カタログ番号：Ｍ０２６４Ｌ）を用いた。各酵素の活性値（ユニット数）は以下の通り定義した。

・Ｓ１ ｎｕｃｌｅａｓｅ：３０ｍＭ酢酸ナトリウム（ｐＨ４．６、２５℃）、５０ｍＭ ＮａＣｌ、１ｍＭ $ZnCl_2$、５％グリセロール、０．５ｍｇ／ｍＬ変性仔牛胸腺ＤＮＡの混合溶液中において、３７℃で１分間に１μｇの酸可溶性物質を生成する酵素活性を１Ｕとした。

・ＭＢＮ：熱変性仔牛胸腺ＤＮＡを基質として、３７℃、ｐＨ５．０において、１分間に１μｇの酸可溶性分解物を生成する酵素活性を１Ｕとした。

・$RecJ_f$：全反応液５０μＬ（１×ＮＥ Ｂｕｆｆｅｒ ２及び１．５μｇの超音波処理［^3H］標識１本鎖Ｅ．ｃｏｌｉ ＤＮＡを含む）中、３７℃、１分間で、０．５ｎｇのトリクロロ酢酸可溶性デオキシリボヌクレオチドを生成するために必要な酵素量を１Ｕとした。

【０１５６】

II－１）Ｓ１ ｎｕｃｌｅａｓｅ処理

Ｉ）で得られたＤＮＡ溶出液に、Ｓ１ ｎｕｃｌｅａｓｅに付属の１０×Ｒｅａｃｔｉｏｎ Ｂｕｆｆｅｒを４μＬ添加した。１×Ｒｅａｃｔｉｏｎ ＢｕｆｆｅｒでＳ１ ｎｕｃｌｅａｓｅを適宜希釈し、１、３、１０、３０、１００、３００ＵをＤＮＡ溶出液に添加し、ＤＷを添加し、全量を４０μＬとした。Ｓ１ ｎｕｃｌｅａｓｅを１０００Ｕ添加するサンプルについては、ビーズ精製で得られた３０μＬのＤＮＡ溶出液に４．６μＬの１０×Ｒｅａｃｔｉｏｎ Ｂｕｆｆｅｒを添加し、Ｓ１ ｎｕｃｌｅａｓｅの原液を１２μＬ加えて全量を４６μＬとした。Ｓ１ ｎｕｃｌｅａｓｅを加えた反応液を撹拌し、３０℃で３０分間インキュベートした。反応液中にライブラリ調製開始時のＤＮＡ全量（６０ｎｇ）が存在すると考えると、１ｎｇ当たりのＤＮＡに対するＳ１ ｎｕｃｌｅａｓｅのユニット数はそれぞれ、０．０２、０．０５、０．１７、０．５０、１．６７、５．００、１６．７Ｕ／ｎｇであった。反応液中のＳ１ ｎｕｃｌｅａｓｅの失活のために、０．５Ｍ ＥＤＴＡ（ｐＨ８．０）（ニッポンジーン社製）を３μＬ添加し、７０℃で１０分間インキュベートした。失活させた反応液からＤＮＡを精製するため、反応液と等量のＴｒｕＳｅｑに付属のビーズを添加し、推奨プロトコルに従って精製操作を進め、ＴｒｕＳｅｑに付属のＲｅｓｕｓｐｅｎｓｉｏｎ ｂｕｆｆｅｒで懸濁し、６０μＬの溶出液を得た（Ｓ１ ｎｕｃｌｅａｓｅ処理群）。

【０１５７】

II－２）ＭＢＮ処理

Ｉ）で得られたＤＮＡ溶出液に、ＭＢＮに付属の１０×Ｍｕｎｇ Ｂｅａｎ Ｎｕｃｌｅａｓｅ Ｂｕｆｆｅｒを５μＬ添加した。１×Ｍｕｎｇ Ｂｅａｎ Ｎｕｃｌｅａｓｅ ＢｕｆｆｅｒでＭＢＮを適宜希釈し、３、１０、３０、１００ＵをＤＮＡ溶出液に添加し、全量を５０μＬとした。ＭＢＮを加えた反応液を撹拌し、３７℃で１０分間インキュベートした。酵素反応液中のＭＢＮの失活のために、０．５Ｍ ＥＤＴＡ（ｐＨ８．０）を３μＬ添加し、６５℃で１０分間インキュベートした。反応液中に１００ｎｇのＤＮＡ断片が存在すると考えると、１ｎｇ当たりのＤＮＡに対するユニット数はそれぞれ、０．０３、０．１、０．３、１．０Ｕ／ｎｇであった。失活させた反応液からＤＮＡを精製するため、反応液と等量のＴｒｕＳｅｑに付属のビーズを添加し、推奨プロトコルに従って精製操作を進め、ＴｒｕＳｅｑに付属のＲｅｓｕｓｐｅｎｓｉｏｎ ｂｕｆｆｅｒで懸濁し、６０μＬの溶出液を得た（ＭＢＮ処理群）。

【０１５８】
II－３）ＲｅｃＪ$_f$処理
Ｉ）で得られたＤＮＡ溶出液に、ＲｅｃＪ$_f$に付属の１０× ＮＥ Ｂｕｆｆｅｒ ２を５μＬ添加した。１×ＮＥ Ｂｕｆｆｅｒ ２でＲｅｃＪ$_f$を適宜希釈し、３、１０、３０、１００ユニットをＤＮＡ溶出液に添加し、全量を５０μＬとした。ＲｅｃＪ$_f$を加えた反応液を撹拌し、３７℃で６０分間インキュベートした。反応液中に１００ｎｇのＤＮＡ断片が存在すると考えると、１ｎｇ当たりのＤＮＡに対するユニット数はそれぞれ、０．０３、０．１、０．３、１．０Ｕ／ｎｇであった。酵素反応液中のＲｅｃＪ$_f$の失活のために、６５℃で２０分間インキュベートした。失活させた反応液からＤＮＡを精製するため、反応液と等量のＴｒｕＳｅｑに付属のビーズを添加し、推奨プロトコルに従って精製操作を進め、ＴｒｕＳｅｑに付属のＲｅｓｕｓｐｅｎｓｉｏｎ ｂｕｆｆｅｒで懸濁し、６０μＬの溶出液を得た（ＲｅｃＪ$_f$処理群）。
【０１５９】
III）Ｅｎｄ Ｒｅｐａｉｒ、Ａ－ｔａｉｌｉｎｇ、Ａｄａｐｔｏｒ Ｌｉｇａｔｉｏｎ及びＰＣＲ ｅｎｒｉｃｈｍｅｎｔ
II）で得られた非処理群、Ｓ１ ｎｕｃｌｅａｓｅ処理群、ＭＢＮ処理群、及びＲｅｃＪ$_f$処理群に、比較例１の２）と同様の手順で、ＴｒｕＳｅｑの推奨プロトコルに従ってＥｎｄ Ｒｅｐａｉｒ、Ａ－ｔａｉｌｉｎｇ、Ａｄａｐｔｏｒ Ｌｉｇａｔｉｏｎを実施した。得られたＡｄａｐｔｏｒ Ｌｉｇａｔｉｏｎの反応液を推奨プロトコルに従って精製し、２本鎖ＤＮＡ断片の両末端にアダプターが付加されたＤＮＡ（アダプター付加ＤＮＡ）を得た。Ａｇｉｌｅｎｔ ４２００ ＴａｐｅＳｔａｔｉｏｎ（アジレント・テクノロジー社製）のＨｉｇｈ Ｓｅｎｓｉｔｉｖｉｔｙ Ｄ５０００キットを用いてアダプター付加ＤＮＡの濃度を測定した。次いで、比較例１の２）と同様の手順でＰＣＲ ｅｎｒｉｃｈｍｅｎｔを実施し、ライブラリを得た。
【０１６０】
４）シーケンシング及び変異解析
３）で調製したライブラリを、２×１５０ｂｐのリード長でシーケンシングし、ライブラリあたり、平均で約１５Ｇｂｐ（約５０Ｍリードペア）のシーケンシングデータを得た。得られたシーケンシングデータから、相補鎖間コンセンサスリード配列の作成、及び変異検出を実施した。シーケンシング、相補鎖間コンセンサスリード配列の作成、及び変異解析は参考例１の手順に従って実施した。
【０１６１】
５）変異頻度の算出
比較例１の４）と同様の手順で、６つの変異パターン及び１２種の変異パターンについて変異頻度を算出した。次いでＧＣ→ＴＡ及びＧＣ→ＣＧの変異について、Ｇの置換とＣの置換の間での変異頻度の差を算出し、下記式に基づいて、ヌクレアーゼ処理群でのエラーの減少率を求めた。
エラー減少率（％）＝（Ａ－Ｂ）／Ａ×１００
Ａ：非処理群（０Ｕ／ｎｇ）でのＧＣ間の変異頻度の差
Ｂ：各ユニット数でのヌクレアーゼ処理群でのＧＣ間の変異頻度の差
【０１６２】
６）解析効率の算出
変異解析の際に用いた各ライブラリの相補鎖間コンセンサスリード配列中のリードペア数（本）と各ライブラリのシーケンシングで読み取ったリードペアの総数（シーケンシングデータ量）（本）から、各ライブラリの解析効率を算出した。
解析効率（％）＝（相補鎖間コンセンサスリード配列中のリードペア数）／（シーケンシングデータ量）×１００
【０１６３】
７）グループあたりの平均リードペア数
４）で作成した相補鎖間コンセンサスリード配列について、推定フラグメントについて

のグループあたりのリードペア数を計数し、リードペア数が等しいグループの数を集計して、平均リードペア数を算出した。

平均リードペア数 ＝ $\{\Sigma_i$（i×（i本のリードペアを含むグループ数））$\}$ ／（グループの総数）

（ｉはグループに含まれるリードペアの本数を指す。）

【０１６４】

８）結果と考察

Ｉ）Ａｍｅｓ試験の復帰突然変異体数

表２に３－ＭＣ暴露後の復帰突然変異体コロニー数を示す。データは３枚のプレートでの測定値と、その平均値を示す。３－ＭＣ暴露により復帰突然変異体コロニー数の増加が認められたことから、３－ＭＣ暴露によりＴＡ１００株のゲノム中に変異が導入されたことが確認された。

【０１６５】

【表２】

Ames試験における復帰変異体コロニー数

3-MC濃度 (μg/plate)	復帰突然変異体数			
	1	2	3	平均
0 (DMSO)	112	93	97	101
10	887	881	1279	1016
20	1539	1484	1509	1511
100	1519	1539	1797	1618

【０１６６】

II）１本鎖特異的ヌクレアーゼによるエラー低減効果

II－１）Ｓ１ ｎｕｃｌｅａｓｅ

ＤＭＳＯ暴露ライブラリにおける６変異パターンの変異頻度を図３に示す。非処理群（Ｓ１ ｎｕｃｌｅａｓｅ ０Ｕ／ｎｇ）では、比較例１と同じようにＧＣ塩基対の変異頻度が高かった。一方で、Ｓ１ ｎｕｃｌｅａｓｅ処理群（Ｓ１ ｎｕｃｌｅａｓｅ ０．２～１６．７Ｕ／ｎｇ）では、ユニット数依存的に変異頻度が減少し、０．１７Ｕ／ｎｇでエラー低減効果が飽和した。続いて、同じライブラリでの１２種類の変異パターンの頻度を図４～５に示す。非処理群では比較例１と同じように、Ｃ→Ａ、Ｃ→Ｇに比べて、Ｇ→Ｔ、Ｇ→Ｃの変異を高頻度に検出した。そして、Ｓ１ ｎｕｃｌｅａｓｅ処理群では、ユニット数の増加に伴ってＧ→Ｔ、Ｇ→Ｃの変異頻度が減少した。ＧＣ→ＴＡ、ＧＣ→ＣＧについてのエラー減少率を表３に示す。０．１７Ｕ／ｎｇ以上で変異頻度の減少が飽和し、ＧＣ間の変異頻度の偏りが大きく改善された。これは、Ｓ１ ｎｕｃｌｅａｓｅがサンプルＤＮＡの断片中の１本鎖部位を特異的に分解し、該１本鎖部位に存在していた酸化修飾されたグアニンを除去したためと考えられた。０．１７Ｕ／ｎｇ以上のＳ１ ｎｕｃｌｅａｓｅ処理により、ＤＮＡ断片の末端１本鎖部位の塩基の酸化修飾に起因するエラーを取り除くことができることが確認された。

【０１６７】

【表３】

エラー減少率 (%)

S1 nuclease (U / ng反応液中DNA)		0.02	0.05	0.17	0.50	1.67	5.00	16.7
変異パターン	GC to TA	52.8	35.3	83.1	69.9	81.7	84.5	79.6
	GC to CG	30.4	50.6	84.9	87.5	86.0	98.3	98.1

【０１６８】

II－２）Ｍｕｎｇ Ｂｅａｎ Ｎｕｃｌｅａｓｅ

図６にＭＢＮ処理時のＤＭＳＯ暴露ライブラリにおける６変異パターンの変異頻度を示

した。ＭＢＮ処理群（０．０３～１．００Ｕ／ｎｇ）において、ユニット数依存的に変異頻度が減少した。続いて、同じライブラリでの１２種類の変異パターンの頻度を図７～８に示す。ＭＢＮ処理群では、Ｇ→Ｃの変異頻度が大きく減少し、Ｇ→Ｃ、Ｃ→Ｇ間の変異頻度の差が大きく減少した。Ｇ→Ｔの変異頻度の減少は認められたが、Ｓ１　ｎｕｃｌｅａｓｅと比較すると小さく、Ｇ→Ｔ、Ｃ→Ａ間の変異頻度の差はユニット数が大きくなっても残っていた。ＧＣ→ＴＡ、ＧＣ→ＣＧについてのエラー減少率を表４に示す。ＧＣ→ＣＧに関しては、０．０３Ｕ／ｎｇ以上でエラー低減効果があり、０．１０Ｕ／ｎｇ以上でＧＣ間の変異頻度の偏りが大きく改善された。一方、ＧＣ→ＴＡに関しては、ＧＣ間の変異頻度の差は低減したものの、効果は小さかった。これは、ＤＭＳＯ暴露ライブラリにおけるＧＣ→ＴＡの変異頻度がII－１で示した結果よりも低かったことが一因と考えられた。同一条件でＤＭＳＯを暴露して調製したＤＮＡ（ｎ＝３）におけるＧ→Ｔ及びＣ→Ａの変異頻度の平均値はそれぞれ０．１７７×１０$^{-6}$及び０．０４２×１０$^{-6}$であった。該平均値に対するエラー減少率は１１．４％（０．０３Ｕ／ｎｇ）、４０．２％（０．１０Ｕ／ｎｇ）、１５．６％（０．３０Ｕ／ｎｇ）、５７．８％（１．００Ｕ／ｎｇ）となった。したがって、Ｓ１　ｎｕｃｌｅａｓｅと比較すると小さいが、ＭＢＮのエラー低減効果は認められた。

【０１６９】

【表４】

エラー減少率（%）

MBN (U / ng反応液中DNA)		0.03	0.10	0.30	1.00
変異パターン	GC to TA	-33.5	9.8	-27.2	36.4
	GC to CG	63.3	90.6	81.2	97.7

【０１７０】

II－３）ＲｅｃＪ$_f$

図９にＲｅｃＪ$_f$処理時のＤＭＳＯ暴露ライブラリにおける６変異パターンの変異頻度を示した。なお、非処理群の結果はＭＢＮ処理群と共通である。ＲｅｃＪ$_f$処理群（０．０３～１．００Ｕ／ｎｇ）において、ユニット数依存的に変異頻度が減少した。続いて、同じライブラリでの１２種類の変異パターンの頻度を図１０～１１に示す。ＲｅｃＪ$_f$処理群では、Ｇ→Ｔ、Ｇ→Ｃの変異頻度の減少が認められ、Ｇ→Ｔ、Ｃ→Ａ間及びＧ→Ｃ、Ｃ→Ｇ間の変異頻度の差も減少したが、Ｓ１　ｎｕｃｌｅａｓｅと比較するとその効果は小さかった。ＧＣ→ＴＡ、ＧＣ→ＣＧについてのエラー減少率を表５に示す。また、II－２の時と同様、ＤＭＳＯ暴露ライブラリのＧＣ→ＴＡの変異頻度が低いことを考慮し、同一条件でＤＭＳＯを暴露して調製したＤＮＡ（ｎ＝３）におけるＧ→Ｔ及びＣ→Ａの変異頻度の平均値と比較した。これらの平均値を用いて算出したエラー減少率は－１０．８％（０．０３Ｕ／ｎｇ）、３５．２％（０．１０Ｕ／ｎｇ）、５４．１％（０．３０Ｕ／ｎｇ）、６２．３％（１．００Ｕ／ｎｇ）となった。したがって、ＲｅｃＪ$_f$は、ＧＣ→ＴＡに関してはＭＢＮと同等のエラー低減効果を示し、ＧＣ→ＣＧに関しては、Ｓ１　ｎｕｃｌｅａｓｅ、ＭＢＮと比較すると小さいが、エラー低減効果は認められた。また、ＧＣ→ＴＡ、ＧＣ→ＣＧともに０．１０Ｕ／ｎｇ以上でエラー低減効果があると考えられた。

【０１７１】

【表５】

エラー減少率（%）

$RecJ_f$ (U / ng反応液中DNA)		0.03	0.10	0.30	1.00
変異パターン	GC to TA	-66.9	2.3	30.8	43.3
	GC to CG	-30.6	8.4	51.5	60.0

【０１７２】
III）３－ＭＣの変異頻度の上昇率の改善
III－１）Ｓ１　ｎｕｃｌｅａｓｅ
５）の方法でＤＭＳＯ暴露ライブラリ（ＤＭＳＯ　ｃｏｎｔｒｏｌ）、及び、３－ＭＣ暴露ライブラリ（３ＭＣ）における６変異パターンの変異頻度をＳ１　ｎｕｃｌｅａｓｅのユニット数ごとに算出した結果を図１２～１３に示す。非処理群（ｃｏｎｔｒｏｌ、０Ｕ／ｎｇ）では、ＤＭＳＯ　ｃｏｎｔｒｏｌと比較した３－ＭＣにおける変異頻度の明確な上昇はいずれの変異パターンにおいても検出されなかったが、Ｓ１　ｎｕｃｌｅａｓｅ処理群では、３－ＭＣでＧＣ→ＴＡの変異頻度の明確な増加が見られた。この変異パターンは、３－ＭＣに暴露された遺伝子組換えマウスの肝臓で検出された変異パターンと一致していた（Ｅｎｖｉｒｏｎ．Ｍｏｌ．Ｍｕｔａｇｅｎ．，２０００，３６：２６６－２７３）。これらの結果は、Ｓ１　ｎｕｃｌｅａｓｅ処理により１本鎖上のグアニン由来のシーケンシングエラーが減少した一方、真の変異は検出されたためと考えられた。表６に、ＤＭＳＯ　ｃｏｎｔｒｏｌに対する３－ＭＣでのＧＣ→ＴＡ変異頻度の上昇率（ＳＮ　ｒａｔｉｏ）を示す。０．１７Ｕ／ｎｇ以上のＳ１　ｎｕｃｌｅａｓｅ処理により、シーケンシングエラーが低減することで、変異原処理により誘発される低頻度な変異が検出可能になることが示唆された。
【０１７３】
【表６】
GC→TA変異頻度の上昇率 (3MC/DMSO control)

S1 nuclease (U / ng反応液中DNA)	0	0.02	0.05	0.17	0.50	1.67	5.00	16.7
SN ratio	1.27	2.1	1.97	2.72	3.75	3.35	1.63	3.59

【０１７４】
III－２）Ｍｕｎｇ　Ｂｅａｎ　Ｎｕｃｌｅａｓｅ
III－１同様、ＭＢＮ処理群における結果を図１４に示す。ＭＢＮ処理群では、Ｓ１　ｎｕｃｌｅａｓｅ処理群と同様に、３－ＭＣにおいてＧＣ→ＴＡの変異頻度の増加が見られた。表７に、ＤＭＳＯ　ｃｏｎｔｒｏｌに対する３－ＭＣでのＧＣ→ＴＡ変異頻度の上昇率（ＳＮ　ｒａｔｉｏ）を示す。本実験での非処理群（０Ｕ／ｎｇ）におけるＳＮ　ｒａｔｉｏは、III－１に比べて高かった。これは、III－１に比べてＤＭＳＯ　ｃｏｎｔｒｏｌのＧＣ→ＴＡの変異頻度が低く、３－ＭＣでのＧＣ→ＴＡの変異頻度が大きいためであった。そこで、同一条件でＭＢＮ処理なしのＤＭＳＯ　ｃｏｎｔｒｏｌ及び３－ＭＣ（それぞれｎ＝３）を調製し、各々についてＧＣ→ＴＡの変異頻度の平均値を算出し、それらの平均値からＳＮ　ｒａｔｉｏを求めた。その結果、ＤＭＳＯ　ｃｏｎｔｒｏｌ、及び３－ＭＣのＧＣ→ＴＡの平均値はそれぞれ、０．１０９×１０$^{-6}$、０．１７６×１０$^{-6}$となり、ＳＮ　ｒａｔｉｏは１．６１となった。したがって、０．１０Ｕ／ｎｇ以上のＭＢＮでＳＮ　ｒａｔｉｏが改善することが推測された。
【０１７５】
【表７】
GC→TA変異頻度の上昇率 (3MC/DMSO control)

MBN (U / ng反応液中DNA)	0	0.03	0.10	0.30	1.00
SN ratio	2.04	1.60	1.81	1.34	1.92

【０１７６】
III－３）ＲｅｃＪ$_f$
III－１同様、ＲｅｃＪ$_f$処理群における結果を図１５に示す。Ｓ１　ｎｕｃｌｅａｓｅ及びＭＢＮ処理群と同様に、ＲｅｃＪ$_f$処理群でも、３－ＭＣにおいてＧＣ→ＴＡの変異頻度の増加が見られた。表８に、ＤＭＳＯ　ｃｏｎｔｒｏｌに対する３－ＭＣでのＧＣ→

ＴＡ変異頻度の上昇率（ＳＮ　ｒａｔｉｏ）を示す。III－２）で算出した変異頻度の平均値のＳＮ　ｒａｔｉｏ（１．６１）を考慮すると、０．１０Ｕ／ｎｇ以上のＲｅｃＪ$_f$にエラー低減効果があると考えられた。

【０１７７】

【表８】

GC→TA変異頻度の上昇率 (3MC/DMSO control)

$RecJ_f$ (U / ng反応液中DNA)	0	0.03	0.10	0.30	1.00
SN ratio	2.04	1.48	2.14	1.68	1.94

【０１７８】

IV）解析効率と平均リードペア数

本実施例でのシーケンシングは最適条件と推定される初期ＤＮＡ量７８ａｍｏｌの条件（特許文献４参照）で実施されたが、ヌクレアーゼ処理がシーケンシング最適条件に影響を及ぼしている可能性がある。そこで、シーケンシングの解析効率と平均リードペア数（特許文献４）に基づいて、本実施例でのシーケンシングが最適条件下でなされたか否かを評価した。表９～１１に各ユニット数のＳ１　ｎｕｃｌｅａｓｅ、ＭＢＮ、及びＲｅｃＪ$_f$で処理したライブラリにおける解析効率と平均リードペア数の算出結果を示す。特許文献４で算出されたシーケンシングの最適条件は、解析効率が５～１０％程度、平均リードペア数が約２本であり、本実施例でも近い結果が得られた。したがって、ヌクレアーゼ処理によるシーケンシング条件への影響は小さく、本実施例でもほぼ最適条件でシーケンシングが行われたと考えられた。

【０１７９】

【表９】

DMSO暴露ライブラリ

S1 nuclease (U / ng反応液中DNA)	0	0.02	0.05	0.17	0.50	1.67	5.00	16.7
解析効率 (%)	8.0	8.3	8.3	8.0	8.5	8.8	8.9	8.9
平均リードペア数 (本)	1.9	2.0	2.0	2.2	1.9	2.0	2.1	1.9

3-MC暴露ライブラリ

S1 nuclease (U / ng反応液中DNA)	0	0.02	0.05	0.17	0.50	1.67	5.00	16.7
解析効率 (%)	8.3	7.9	7.7	8.2	8.5	8.5	8.8	8.0
平均リードペア数 (本)	2.4	2.2	2.0	2.0	2.1	2.2	2.3	2.2

【０１８０】

【表１０】

DMSO 暴露ライブラリ

MBN (U / ng反応液中DNA)	0	0.03	0.10	0.30	1.00
解析効率 (%)	8.2	8.8	8.6	8.9	9.0
平均リードペア数 (本)	2.4	2.4	2.1	2.2	2.3

3-MC 暴露ライブラリ

MBN (U / ng反応液中DNA)	0	0.03	0.10	0.30	1.00
解析効率 (%)	8.3	8.6	8.5	8.8	8.7
平均リードペア数 (本)	2.7	2.5	2.1	2.2	2.2

【０１８１】

【表１１】

DMSO 暴露ライブラリ

$RecJ_f$ (U / ng反応液中DNA)	0	0.03	0.10	0.30	1.00
解析効率 (%)	8.2	8.6	7.4	8.6	8.8
平均リードペア数 (本)	2.4	2.4	2.8	2.6	2.3

3-MC 暴露ライブラリ

$RecJ_f$ (U / ng反応液中DNA)	0	0.03	0.10	0.30	1.00
解析効率 (%)	8.2	8.4	7.4	8.4	8.5
平均リードペア数 (本)	2.7	2.4	2.5	2.4	2.3

【０１８２】

実施例２　変異解析に対する影響の評価

本実施例では、ＤＮＡ断片の１本鎖特異的ヌクレアーゼ処理が変異解析に与える影響を評価するため、１）相補鎖間コンセンサスリード配列のＬＴ－２株のゲノムに対する網羅性、及び、２）異なるＤＮＡ断片の同一断片としての誤認識（断片の誤認識）を調べた。実施例１で得られた各サンプルのリードペア、相補鎖間コンセンサスリード配列を用いた。１本鎖特異的ヌクレアーゼにはＳ１　ｎｕｃｌｅａｓｅ、ＭＢＮ、及び$RecJ_f$を用いた。

【０１８３】

１）相補鎖間コンセンサスリード配列のＬＴ－２株のゲノムに対する網羅性

シーケンシングでのゲノム全体のカバレッジを調べ、ゲノムの特定の部位が特異的にシーケンシングされていないか評価した。ＤＭＳＯ暴露ライブラリの非処理群及びＳ１　ｎｕｃｌｅａｓｅ処理群、ＭＢＮ処理群、及び$RecJ_f$処理群の相補鎖間コンセンサスリード配列から各ゲノム位置におけるカバレッジの情報を抽出し、プログラミング言語Ｐｙｔｈｏｎで作成したプログラムにより、ゲノム領域をおよそ１００塩基ごとに区切り、各領域におけるカバレッジを求め、正規化し（カバレッジの総和が１となる）、ヒストグラムを作成した。さらに、ＬＴ－２株のゲノムにマッピングした際のｃｏｖｅｒｅｄ　ｒａｔｅ（カバレッジが１以上になったゲノム位置の割合）、平均カバレッジ（ｍｅａｎ　ｃｏｖｅｒａｇｅ）、カバレッジの標準偏差（ＳＤ　ｏｆ　ｃｏｖｅｒａｇｅ）、及び変動係数（ＣＶ）を算出した。

変動係数（ＣＶ）（％）＝（カバレッジの標準偏差）／（平均カバレッジ）×１００

【０１８４】
非処理群、及びシーケンシングエラー低減効果が明確に現れた０．１７Ｕ／ｎｇ以上でのＳ１　ｎｕｃｌｅａｓｅ処理群でのカバレッジのヒストグラムを図１６に示した。全データに共通してみられるゲノム位置８０００００から９０００００番目あたりのカバレッジがない部分は、ＴＡ１００株におけるｕｖｒＢ遺伝子の欠損部位である（Ｊ．Ａｐｐｌ．Ｔｏｘｉｃｏｌ．，２０１７，３７：１１２５－１１２８）。いずれのユニット数のＳ１　ｎｕｃｌｅａｓｅ処理群においても、非処理群と比べてヒストグラムに大きな変化は見られなかった。表１２上には、非処理群及びＳ１　ｎｕｃｌｅａｓｅ処理群の相補鎖間コンセンサスリード配列をＬＴ－２株のゲノムにマッピングした際の、各群でのｃｏｖｅｒｅｄ　ｒａｔｅ、ｍｅａｎ　ｃｏｖｅｒａｇｅ、ＳＤ　ｏｆ　ｃｏｖｅｒａｇｅ、及びＣＶを示す。Ｓ１　ｎｕｃｌｅａｓｅのユニット数が増加しても、ｃｏｖｅｒｅｄ　ｒａｔｅやＣＶは非処理群と大きく変わらないことが確認された。また、非処理群、及び１．００Ｕ／ｎｇでのMBN処理群及び$RecJ_f$処理群でのカバレッジのヒストグラムを図１７～１８に示した。MBN処理群、及び$RecJ_f$処理群のどちらにおいても、非処理群と比べてヒストグラムに大きな変化は見られなかった。表１２下には、非処理群、MBN処理群及び$RecJ_f$処理群の相補鎖間コンセンサスリード配列をＬＴ－２株のゲノムにマッピングした際の、各群でのｃｏｖｅｒｅｄ　ｒａｔｅ、ｍｅａｎ　ｃｏｖｅｒａｇｅ、ＳＤ　ｏｆ　ｃｏｖｅｒａｇｅ、及びＣＶを示す。MBN処理群、及び$RecJ_f$処理群のどちらも、ｃｏｖｅｒｅｄ　ｒａｔｅやＣＶは非処理群と大きく変わらないことが確認された。以上の結果より、サンプルＤＮＡ断片の１本鎖特異的ヌクレアーゼ処理によりシーケンシングされるゲノム領域が偏ることはおおむねないものと考えられた。

【０１８５】
【表１２】

S1 nuclease (U / ng 反応液中DNA)	Covered rate (%)	Mean coverage	SD of coverage	CV (%)
0	97.51	140.5	37.07	26.38
0.17	97.52	171.8	43.97	25.59
0.50	97.51	176.9	45.97	25.99
1.67	97.50	184.8	48.63	26.31
5.00	97.52	218.4	58.62	26.84
16.7	97.50	187.8	57.56	30.64

Nuclease (U / ng 反応液中DNA)	Covered rate (%)	Mean coverage	SD of coverage	CV (%)
0U	97.54	263.1	66.81	25.40
1.00 (MBN)	97.51	261.6	67.55	25.82
1.00 ($RecJ_f$)	97.51	231.7	59.45	25.66

【０１８６】
２）断片の誤認識
相補鎖間コンセンサスリード配列の作成の際、異なる細胞由来のリードペアが偶然に参照配列上の同一の位置にマッピングされると、同じ２本鎖ＤＮＡ断片由来のリードペアとして誤認識される。このとき、ある細胞のＤＮＡから変異の入ったリードペアが得られ、別の細胞のＤＮＡから変異のないリードペアが得られていた場合、真の変異がエラーとして除かれてしまう。こうした異なるＤＮＡ断片の同一断片としての誤認識（断片の誤認識）は、ライブラリ調製でのＤＮＡ断片増幅過程で初期ＤＮＡ量を解析対象のゲノムサイズに応じて調整し、ライブラリ中のアダプター付加ＤＮＡの多様性を調整することで最小限に抑えられる。実施例１のライブラリは、全て初期ＤＮＡ量が７８ａｍｏｌであることから、断片の誤認識は通常無視できるレベルである。本実施例では、断片の誤認識が１本鎖特異的ヌクレアーゼでの処理により増加しないか調べた。

【０１８７】
本解析では、サンプルＤＮＡの識別のため、アダプター配列内のｉｎｄｅｘ情報を利用した。異なるｉｎｄｅｘ情報を持つアダプター配列を用いてＤＭＳＯ暴露ライブラリと３－ＭＣ暴露ライブラリを調製し、シーケンシングデータを得た。それぞれのライブラリのＦａｓｔｑファイル（リード１、リード２）の先頭から２５Ｍリードずつを抽出し、リード１同士、及び、リード２同士で１つにまとめ、２種類のｉｎｄｅｘ情報を含む５０Ｍリードのファストｑファイルをリード１、リード２それぞれ作成した。このようにして、１本鎖特異的ヌクレアーゼのユニット数ごとに、ゲノムＤＮＡの由来の異なるリードペアが混合されたシーケンシングデータを作成した。このデータを参照配列にマッピングし、参考例１の方法に従ってリードペアのグループを作成した。これらのグループのうち、２つ以上のリードペアが含まれるグループを抽出し、各グループ中のリードペアのｉｎｄｅｘ情報をもとに、ゲノムＤＮＡの由来の異なるリードペアが含まれる割合（異なるｉｎｄｅｘが含まれる割合＝断片の誤認識率）を算出した。
異なるｉｎｄｅｘが含まれる割合（％）＝（異なるｉｎｄｅｘ情報が含まれるグループ数）／（２つ以上のリードペアが含まれるグループ数）×１００
【０１８８】
Ｉ）Ｓ１　ｎｕｃｌｅａｓｅ
各ユニット数のＳ１　ｎｕｃｌｅａｓｅ処理群での異なるｉｎｄｅｘが含まれる割合、即ち断片の誤認識率を図１９及び表１３に示す。Ｓ１　ｎｕｃｌｅａｓｅのユニット数の増加に伴い、異なるｉｎｄｅｘが含まれる割合は増加していた。本実施例では、２種類のｉｎｄｅｘ情報を用いたことから、実際に起こった断片の誤認識のうちのおよそ半分が検出されたと推定され、したがって、算出された異なるｉｎｄｅｘが含まれる割合の約２倍の値が、実際の誤認識率と推定された。シーケンシングエラーが大きく低減される０．１７Ｕ／ｎｇ以上でのＳ１　ｎｕｃｌｅａｓｅ処理では、断片の誤認識率はおよそ７％以上で、変異頻度への影響が懸念されるレベルであった。
【０１８９】
【表１３】

S1 nuclease (U / ng 反応液中DNA)	0	0.02	0.05	0.17	0.50	1.67
異なるindexが含まれる割合 (%)	1.53	1.60	1.77	3.56	7.44	10.32

【０１９０】
II）ＭＢＮ
各ユニット数のＭＢＮ処理群での断片の誤認識率を図２０及び表１４に示す。ＭＢＮにおいても、ユニット数の増加に伴い、異なるｉｎｄｅｘが含まれる割合は増加した。０．１０Ｕ／ｎｇ以上では断片の推定誤認識率（異なるｉｎｄｅｘが含まれる割合の約２倍の値）はおよそ６％以上で、変異検出への影響が懸念されるレベルであった。
【０１９１】
【表１４】

MBN (U / ng 反応液中 DNA)	0	0.03	0.10	0.30	1.00
異なる index が含まれる割合 (%)	1.58	1.83	3.16	6.30	12.65

【０１９２】
III）ＲｅｃＪ$_f$
各ユニット数のＲｅｃＪ$_f$処理群での断片の誤認識率を図２１及び表１５に示す。ユニット数が増加に伴い、異なるｉｎｄｅｘが含まれる割合が僅かに増加したが、変異検出に影響するほどではなかった。

【０１９３】
【表１５】

$RecJ_f$ (U / ng 反応液中 DNA)	0	0.03	0.10	0.30	1.00
異なる index が含まれる割合 (%)	1.58	1.42	1.29	1.52	1.81

【０１９４】
３）結果と考察
実施例１の結果から、末端修復の前にＤＮＡをＳ１　ｎｕｃｌｅａｓｅ、ＭＢＮ又は$RecJ_f$で処理することで、シーケンシングにおけるエラーを低減できることが確認できた。したがって、エラー低減効果は１本鎖特異的ヌクレアーゼに共通することが示された。エラー低減効果は、Ｓ１　ｎｕｃｌｅａｓｅ＞ＭＢＮ＞$RecJ_f$の順で大きかった。この理由の１つとして、両側が２本鎖である１本鎖部分は、１本鎖特異的エキソヌクレアーゼ（$RecJ_f$）では分解できないが、１本鎖特異的エンドヌクレアーゼ（Ｓ１　ｎｕｃｌｅａｓｅ、及びＭＢＮ）では分解できること考えられた。一方、実施例２の結果から、Ｓ１　ｎｕｃｌｅａｓｅ及びＭＢＮにおいては高精度シーケンシング法を併用すると、断片の誤認識率が増え、変異頻度に影響があることが明らかとなった。誤認識率が増えた原因は、Ｓ１　ｎｕｃｌｅａｓｅ及びＭＢＮ活性の配列特異性によるものと考えられた。すなわち、ＤＮＡ断片の末端にＳ１　ｎｕｃｌｅａｓｅ及びＭＢＮで分解されにくい１本鎖配列が残ったことで、リードペアの両末端部位が偶然に一致し、参照配列上の同一の位置にマッピングされる可能性が上昇することにより誤認識率が増加したと推測された。この問題を解決するためには、（ｉ）初期ＤＮＡ量をさらに減少させる、又は（ｉｉ）Ｓ１　ｎｕｃｌｅａｓｅもしくはＭＢＮ処理後、断片を特異性の異なる１本鎖特異的ヌクレアーゼでさらに処理する、という２つの手段が考えられた。これらの手段の有効性について、Ｓ１　ｎｕｃｌｅａｓｅを用いて、この後の実施例で検討した。一方、$RecJ_f$に関しては、誤認識率が大きく増えなかったことから、変異検出への影響を受けずに使用することができると考えられた。
【０１９５】
実施例３　断片の誤認識率に対する初期ＤＮＡ量の影響
断片の誤認識率、すなわちリードペアが偶然に参照配列上の同一の位置にマッピングされる可能性は、ライブラリ中のサンプルＤＮＡの多様性を減少させることで抑えられる。そこで本実施例では、ライブラリ調製における初期ＤＮＡ量を７８ａｍｏｌよりもさらに減少させることでリードの偶然の重なりを低下させることができるか検討した。
【０１９６】
１）シーケンシング用ライブラリの調製
比較例１及び実施例１で調製したＤＭＳＯ暴露細胞及び３－ＭＣ暴露細胞由来のゲノムＤＮＡをサンプルＤＮＡとした。１２０ｎｇ相当量のＤＮＡをそれぞれ複数サンプル用意し、実施例１に記載の方法でＳ１　ｎｕｃｌｅａｓｅ処理したライブラリを調製した。Ｓ１　ｎｕｃｌｅａｓｅのユニット数は、シーケンシングエラーの低減と断片の誤認識率を考慮して０．０８Ｕ／ｎｇ（ＤＮＡ）及び０．２５Ｕ／ｎｇ（ＤＮＡ）とした。アダプター付加ＤＮＡのＰＣＲ　ｅｎｒｉｃｈｍｅｎｔの過程では、初期ＤＮＡ量を３９及び２０ａｍｏｌとし、ＰＣＲ産物のＤＮＡ量を考慮して、３９ａｍｏｌのＤＮＡは１６サイクル、２０ａｍｏｌのＤＮＡは１７サイクルでＰＣＲ増幅してライブラリを調製した。
【０１９７】
２）シーケンシング及び断片の誤認識率の算出
実施例１と同様にライブラリをシーケンシングした。次いで実施例２と同様の手順で断片の誤認識率（異なるｉｎｄｅｘが含まれる割合）を算出した。
【０１９８】
３）結果と考察

断片の誤認識率を図２２及び表１６に示す。実施例２の２）と同様、表１６の値の約２倍の値が、実際の誤認識率と推定された。特許文献４の実施例に記載のとおり、初期ＤＮＡ量を減少させることで断片の誤認識率を減少させることができた。そして、０．０８Ｕ／ｎｇのＳ１　ｎｕｃｌｅａｓｅで処理する場合は、初期ＤＮＡ量を３９ａｍｏｌ以下にすれば、実際の誤認識率はおよそ５％以下となり、変異を見逃す懸念をできる限り小さくすることができた。同様に、０．２５Ｕ／ｎｇのＳ１　ｎｕｃｌｅａｓｅの場合は、初期ＤＮＡ量を２０ａｍｏｌ以下にすると誤認識率が５％以下となった。

【０１９９】

【表１６】

異なるindexが含まれる割合(%)

S1 nuclease (U / ng 反応液中DNA)		0	0.08	0.25
初期 DNA 量	39 amol	0.68	2.69	4.32
	20 amol	0.31	1.29	2.05

【０２００】

実施例２の断片の誤認識と本実施例の結果から、酵素量（Ｕ／ｎｇ）に応じて断片の誤認識が増加するが、適切な初期ＤＮＡ量を選択することで断片の誤認識を減少させることが可能であることが示された。Ｓ１　ｎｕｃｌｅａｓｅ処理下での断片の誤認識の増加率は、下記式で定義され、かつ実施例２の結果に基づいて酵素量ごとに表１７のように算出された。

断片の誤認識の増加率＝[Ｓ１　ｎｕｃｌｅａｓｅ（Ｕ／ｎｇ）処理時の断片の誤認識率（％）]／[Ｓ１　ｎｕｃｌｅａｓｅ非処理時の断片の誤認識率（％）]

【０２０１】

【表１７】

S1 nuclease (U/ng 反応液中 DNA)	0.017	0.050	0.167	0.500	1.67
断片の誤認識の増加率	1.04	1.16	2.32	4.85	6.72

【０２０２】

表１７に示すとおり、酵素量０．０５Ｕ／ｎｇ以下では断片の誤認識率への影響は無視できるレベルであったが、酵素量が０．０５Ｕ／ｎｇより大きい条件では、断片の誤認識が増加した。例えば酵素量が０．０５Ｕ／ｎｇより大きく０．１６７Ｕ／ｎｇ以下の範囲では、断片の誤認識の増加率は、Ｓ１　ｎｕｃｌｅａｓｅ非処理時の２倍程度と推測された。上記のとおり断片の誤認識率は初期ＤＮＡ量に依存するため、酵素量が０．０５Ｕ／ｎｇより大きい場合、適切な初期ＤＮＡ量の範囲はＳ１　ｎｕｃｌｅａｓｅ非処理時の２分の１程度、すなわち、２５０ａｍｏｌ／Ｍｂｐの２分の１である１２５ａｍｏｌ／Ｍｂｐ以下と考えられた。同様に、酵素量が０．１６７Ｕ／ｎｇより大きく０．５Ｕ／ｎｇ以下の場合、断片の誤認識の増加率はＳ１　ｎｕｃｌｅａｓｅ非処理時の４倍程度と推測でき、適切な初期ＤＮＡ量の範囲は、６２．５ａｍｏｌ／Ｍｂｐ以下と考えられた。酵素量が０．５Ｕ／ｎｇより大きい場合、断片の誤認識の増加率はＳ１　ｎｕｃｌｅａｓｅ非処理時の８倍以下と推測でき、適切な初期ＤＮＡ量の範囲は、３１．３ａｍｏｌ／Ｍｂｐ以下と見積もることができた。

【０２０３】

ライブラリ調製とシーケンシングにおける適切な条件は、Ｓ１　ｎｕｃｌｅａｓｅの処理濃度の増加率と断片の誤認識の増加率の関係、及び、初期ＤＮＡ量と断片の誤認識の増加率の関係を組み合わせることでも導出することができた。例えば、表１７に示すとおり、Ｓ１　ｎｕｃｌｅａｓｅ処理における酵素量が、０．１７Ｕ／ｎｇから１．６７Ｕ／ｎｇに１０倍増えたとき、断片の誤認識率はおよそ３倍増えた。したがって、Ｓ１　ｎｕｃｌｅａｓｅの酵素量の増加による断片の誤認識率は[$3^{\log \text{S1 nuclease (U/ng)}}$]（式中、Ｓ１　ｎｕｃｌｅａｓｅ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である）で

表すことができる。一方、本実施例の結果から、初期ＤＮＡ量が２倍に増えると、断片の誤認識率も２倍に増える傾向があった。以上の２つの結果を考慮して、Ｓ１　ｎｕｃｌｅａｓｅの酵素量が０．０５Ｕ／ｎｇより大きい場合のライブラリ調製とシーケンシングにおける条件は下記の式で表される指標に反映される：

指標＝ＰＣＲにおける初期ＤＮＡ量（ａｍｏｌ／Ｍｂｐサンプル ＤＮＡ）×３$^{\log S1\ nuclease\ (U/ng)}$

（式中、Ｓ１　ｎｕｃｌｅａｓｅ（Ｕ／ｎｇ）＞０．０５、ｌｏｇは常用対数である）。各条件における上記の指標の数値を表１８に示す。上記実施例で調べた適切な条件範囲を考慮すると、好ましい条件でのシーケンシングを可能にする指標の値は６０以下、より好ましくは３０以下、さらに好ましくは１５以下、さらにより好ましくは７．５以下であると考えられた。

【０２０４】

【表１８】

S1 nuclease (U/ng 反応液中 DNA)		0.08	0.17	0.25	0.50	1.67
初期DNA量 (amol) ／サンプル DNA (Mbp)	500	152.8	212.7	258.1	359.2	638.0
	250	76.4	106.3	129.0	179.6	319.0
	125	38.2	53.2	64.5	89.8	159.5
	62.5	19.1	26.6	32.3	44.9	79.7
	31.3	9.55	13.3	16.1	22.5	39.9
	15.6	4.77	6.65	8.06	11.2	19.9
	7.81	2.39	3.32	4.03	5.61	9.97
	3.91	1.19	1.66	2.02	2.81	4.98
	1.95	0.60	0.83	1.01	1.40	2.49

【０２０５】

また、実施例２の結果から、ＭＢＮにおいても、Ｓ１　ｎｕｃｌｅａｓｅと同等のユニット数で同等の誤認識率を示したので、上記で導出した関係式と適切な条件範囲を、そのまま適用できると考えられた。一方、実施例２で示したとおり、ＲｅｃＪ$_f$のユニット数が断片の誤認識率に及ぼす影響は無視できるレベルであった。

【０２０６】

実施例４　断片の誤認識率に対する異なるヌクレアーゼ処理の影響

Ｓ１　ｎｕｃｌｅａｓｅ処理による断片の誤認識率の増加は、ＤＮＡ断片の末端におけるＳ１　ｎｕｃｌｅａｓｅで分解されにくい１本鎖の残存が原因と推測された。このため、Ｓ１　ｎｕｃｌｅａｓｅで処理後、ＤＮＡ断片を特異性の異なる１本鎖特異的ヌクレアーゼでさらに処理することで、誤認識率が改善されると考えられた。エンドヌクレアーゼであるＳ１　ｎｕｃｌｅａｓｅと異なり、ＲｅｃＪ$_f$は１本鎖の５’末端から分解する５’→３’エキソヌクレアーゼ活性を有する。本実施例では、Ｓ１　ｎｕｃｌｅａｓｅ処理後にＤＮＡ断片をＲｅｃＪ$_f$でさらに処理することによる断片の誤認識率への影響を調べた。

【０２０７】

１）シーケンシング用ライブラリの調製

比較例１及び実施例１で調製したＤＭＳＯ暴露細胞及び３－ＭＣ暴露細胞由来のゲノムＤＮＡをサンプルＤＮＡとした。１００ｎｇ相当量のＤＮＡをそれぞれ複数サンプル用意し、実施例１の３）Ｉ）に記載の方法で、３０μＬのサンプルＤＮＡの断片を含むＤＮＡ溶出液を得た。次いで、実施例１の３）II－１）に記載の方法で断片を３０Ｕ（０．３Ｕ／ｎｇ）のＳ１　ｎｕｃｌｅａｓｅで処理した。ＥＤＴＡの添加と熱失活の後、ビーズを添加し、反応液からＤＮＡを精製し、２群に分けた。ＲｅｃＪ$_f$非処理群は、ＴｒｕＳｅｑに付属のＲｅｓｕｓｐｅｎｓｉｏｎ　ｂｕｆｆｅｒで懸濁し、６０μＬの溶出液に調製

した。ＲｅｃＪ$_f$処理群は、Ｄｉｓｔｉｌｌｅｄ　ｗａｔｅｒでビーズを懸濁して３０μＬの溶出液を得た後、実施例１の３）II－３）に記載の方法でＲｅｃＪ$_f$（３（０．０３）、１０（０．１）、３０（０．３）、１００（１．０）Ｕ（Ｕ／ｎｇ））処理した。熱失活の後、ＤＮＡの精製のため、反応液にビーズを添加し、ＴｒｕＳｅｑに付属のＲｅｓｕｓｐｅｎｓｉｏｎ　ｂｕｆｆｅｒで懸濁し、６０μＬの溶出液を得た。得られた溶出液からＴｒｕＳｅｑの推奨プロトコルに基づいてライブラリを調製した。アダプター付加ＤＮＡのＰＣＲ　ｅｎｒｉｃｈｍｅｎｔの過程では、初期ＤＮＡ量を７８ａｍｏｌとし、１５サイクルで増幅した。

【０２０８】

２）シーケンシング及び断片の誤認識率の算出

実施例１と同様にライブラリをシーケンシングした。次いで実施例２と同様の手順で断片の誤認識率（異なるｉｎｄｅｘが含まれる割合）を算出した。

【０２０９】

３）結果と考察

０．３０Ｕ／ｎｇのＳ１　ｎｕｃｌｅａｓｅで処理後、各ユニット数のＲｅｃＪ$_f$で処理した断片での誤認識率を図２３及び表１９に示す。ＲｅｃＪ$_f$のユニット数の増加に伴い、断片の誤認識率が僅かだが減少した。これは、Ｓ１　ｎｕｃｌｅａｓｅが分解しきれなかった１本鎖部分を配列特異性の異なるＲｅｃＪ$_f$が分解したことによるものと考えられた。したがって、配列特異性の異なる１本鎖特異的ヌクレアーゼの組合せ処理により、断片の誤認識率を低減できると考えられた。

【０２１０】

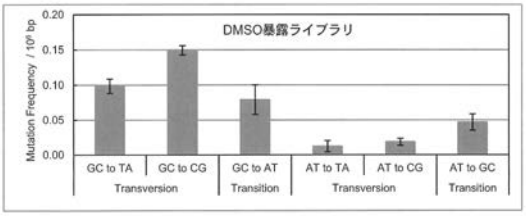
【表１９】

$RecJ_f$ (U / ng 反応液中 DNA)	0	0.03	0.10	0.30	1.00
異なる index が含まれる割合 (%)	6.03	6.14	5.85	5.80	5.43

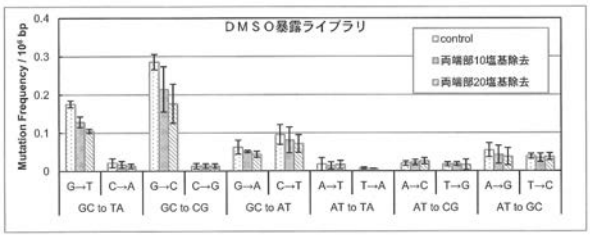
【０２１１】

図２４に、ＤＭＳＯ暴露ライブラリにおける６変異パターンの変異頻度を示す。本実験では０．３０Ｕ／ｎｇのＳ１　ｎｕｃｌｅａｓｅ処理のみ（ＲｅｃＪ$_f$　０Ｕ／ｍｇ）でもエラーが十分に低減しており、このため、ＲｅｃＪ$_f$の追加処理によって誤認識率がそれほど低減しなかったのだと考えられた。

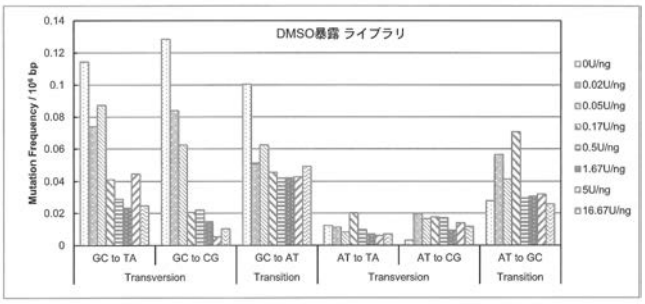
【図１】


DMSO暴露ライブラリ
Mutation Frequency / 10⁶ bp
0.20
0.15
0.10
0.05
0.00
GC to TA
GC to CG
GC to AT
AT to TA
AT to CG
AT to GC
Transversion
Transition
Transversion
Transition

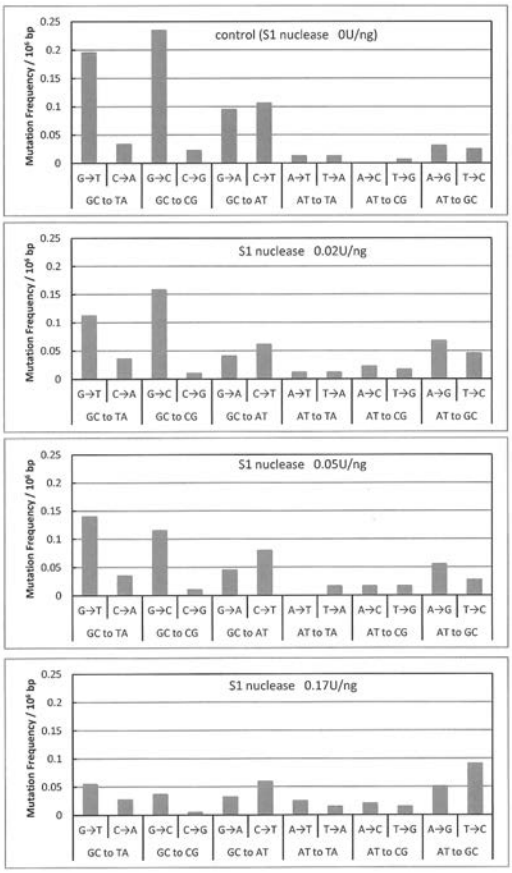
【図２】


DMSO暴露ライブラリ
Mutation Frequency / 10⁶ bp
0.4
0.3
0.2
0.1
0
control
両端部10塩基除去
両端部20塩基除去
G→T
C→A
G→C
C→G
G→A
C→T
A→T
T→A
A→C
T→G
A→G
T→C
GC to TA
GC to CG
GC to AT
AT to TA
AT to CG
AT to GC

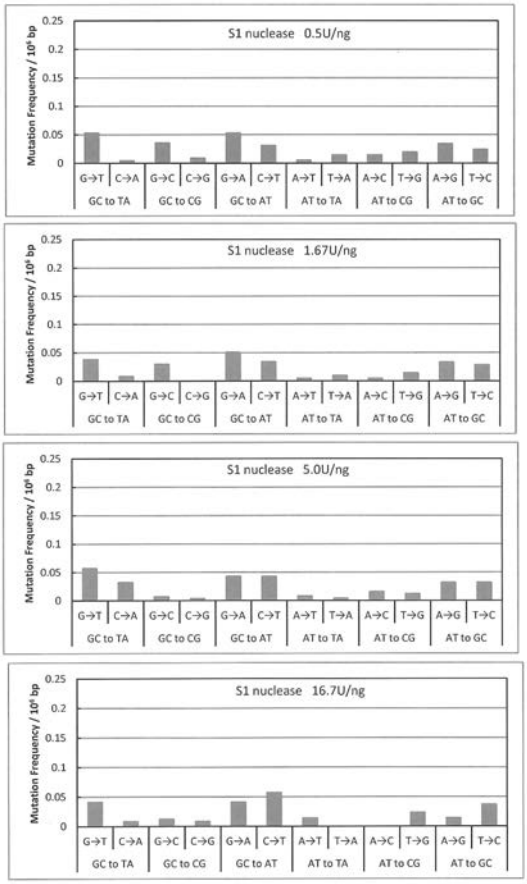
【図３】


DMSO暴露 ライブラリ
Mutation Frequency / 10⁶ bp
0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
0U/ng
0.02U/ng
0.05U/ng
0.17U/ng
0.5U/ng
1.67U/ng
5U/ng
16.67U/ng
GC to TA
GC to CG
GC to AT
AT to TA
AT to CG
AT to GC
Transversion
Transition
Transversion
Transition

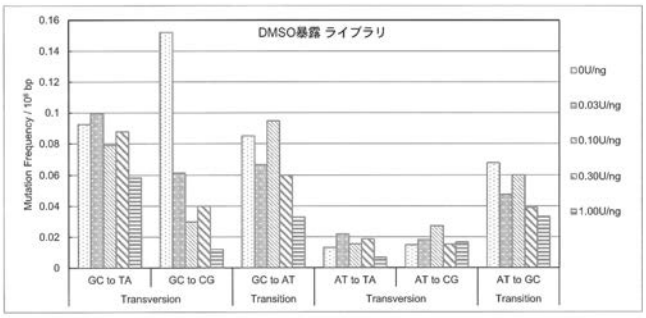
【図４】


control (S1 nuclease 0U/ng)
S1 nuclease 0.02U/ng
S1 nuclease 0.05U/ng
S1 nuclease 0.17U/ng
Mutation Frequency / 10⁶ bp
0.25
0.2
0.15
0.1
0.05
0
G→T
C→A
G→C
C→G
G→A
C→T
A→T
T→A
A→C
T→G
A→G
T→C
GC to TA
GC to CG
GC to AT
AT to TA
AT to CG
AT to GC

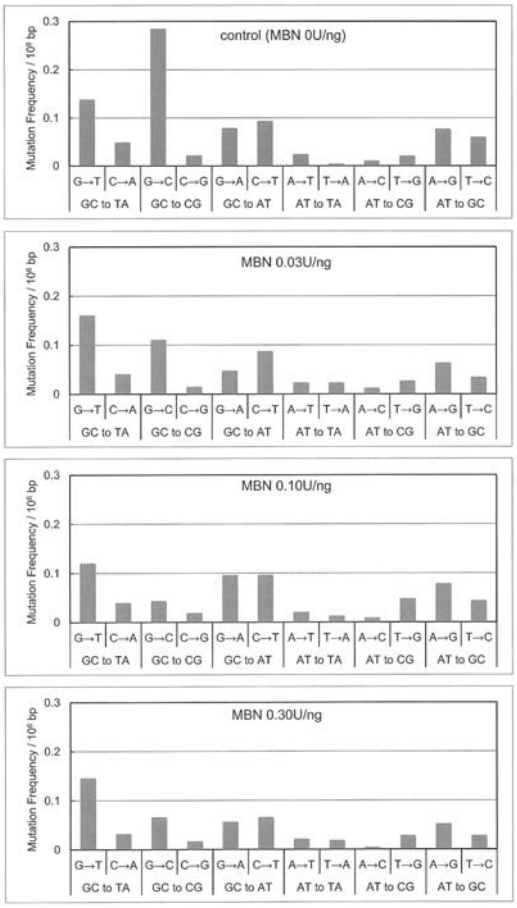
【図５】


S1 nuclease 0.5U/ng
S1 nuclease 1.67U/ng
S1 nuclease 5.0U/ng
S1 nuclease 16.7U/ng
Mutation Frequency / 10⁶ bp
0.25
0.2
0.15
0.1
0.05
0
G→T
C→A
G→C
C→G
G→A
C→T
A→T
T→A
A→C
T→G
A→G
T→C
GC to TA
GC to CG
GC to AT
AT to TA
AT to CG
AT to GC

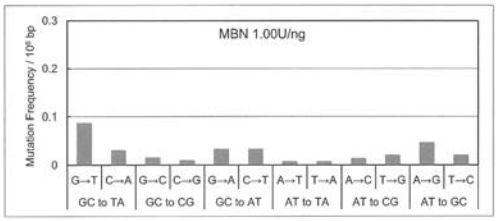
【図６】



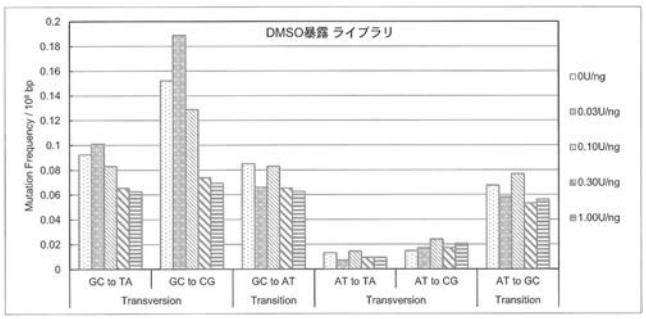
【図７】



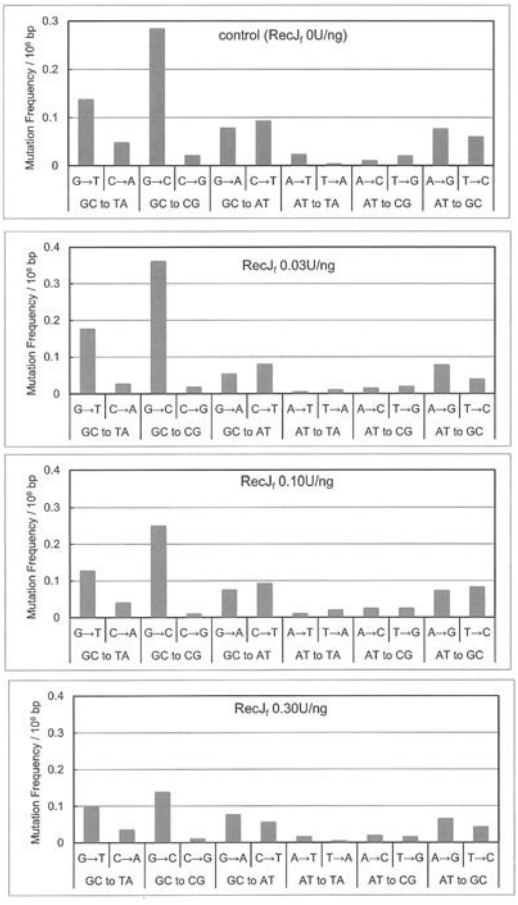
【図８】



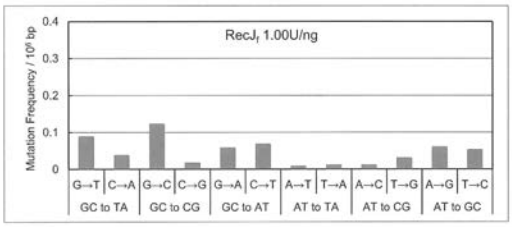
【図９】



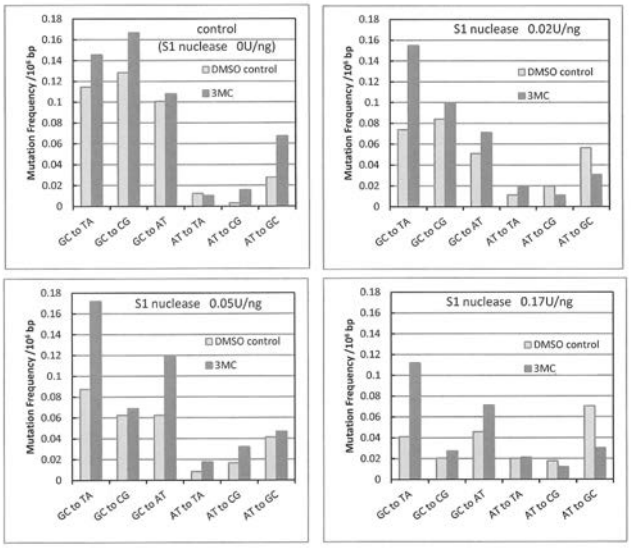
【図１０】



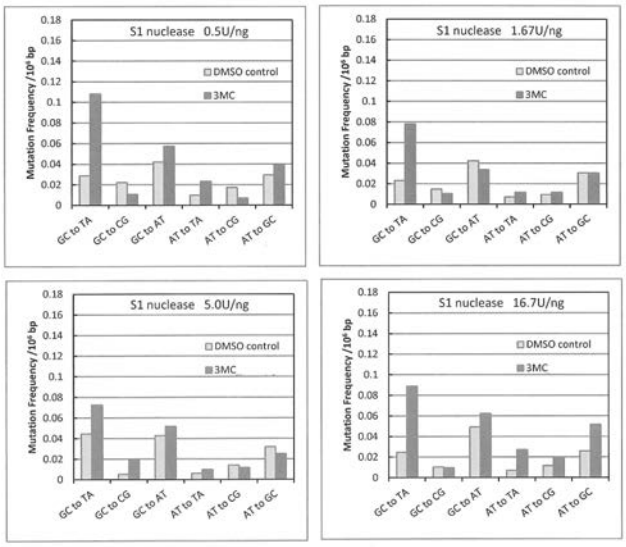
【図１１】



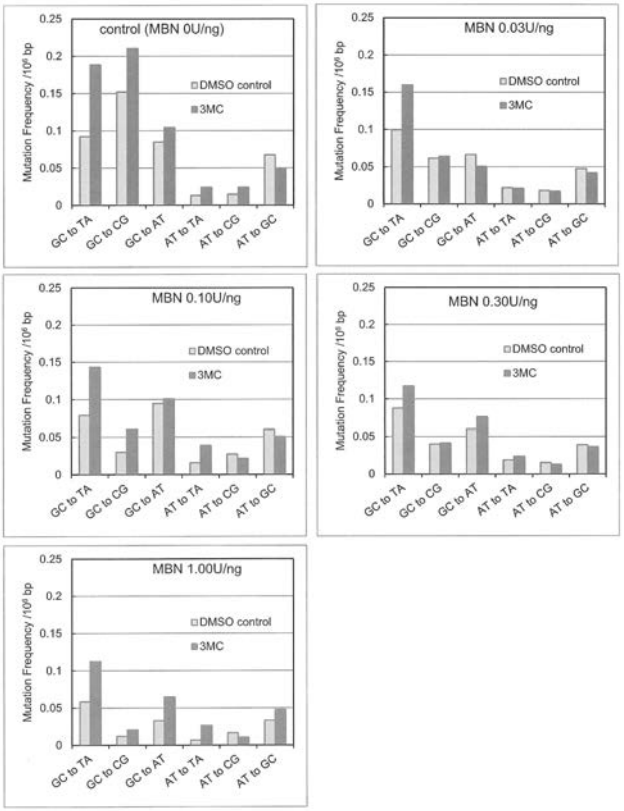
【図１２】



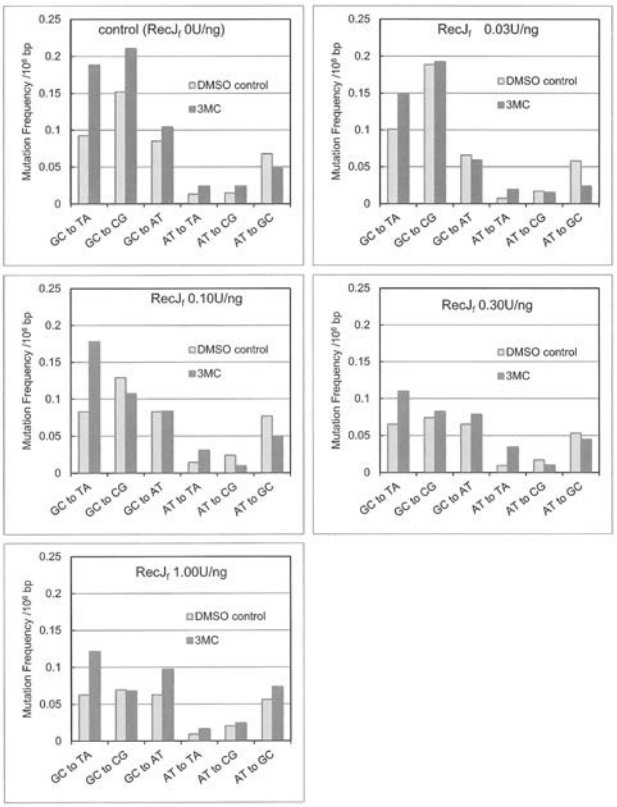
【図１３】



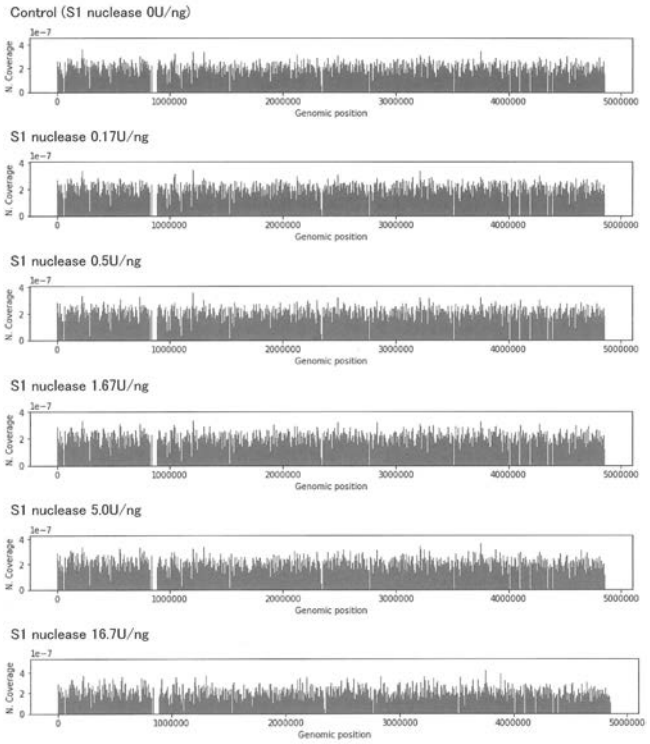
【図１４】



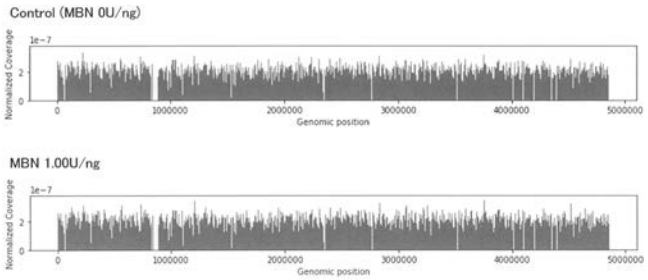
【図１５】



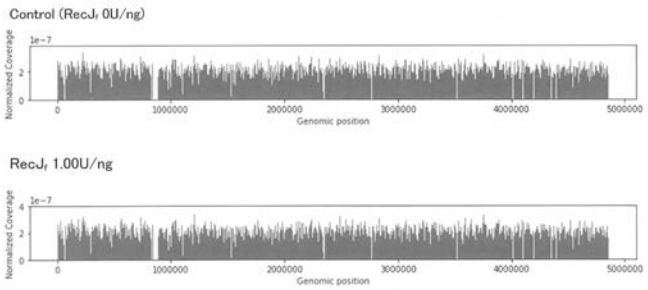
【図１６】



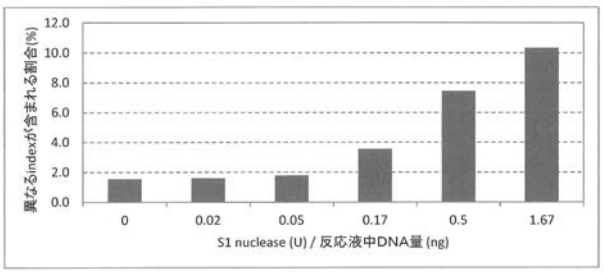
【図１７】



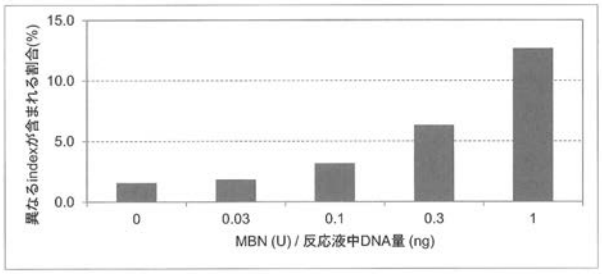
【図１８】



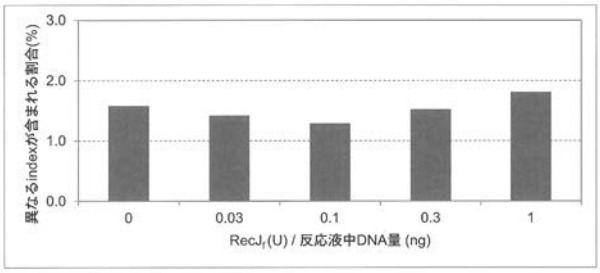
【図１９】



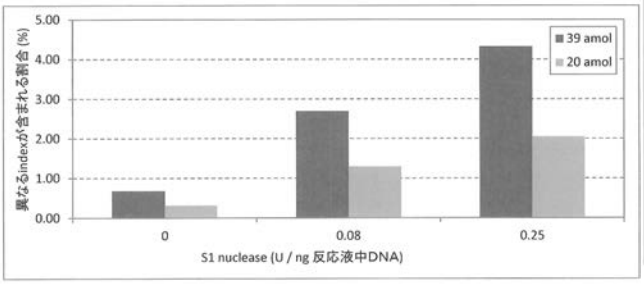
【図２０】



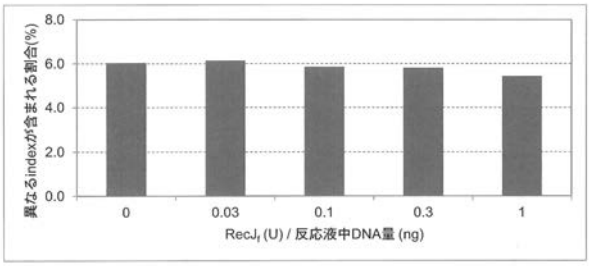
【図２１】



【図２２】



【図２３】



【図２４】