

(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2013-223502

(P2013-223502A)

(43) 公開日 平成25年10月31日(2013.10.31)

(51) Int.Cl.	F I	テーマコード (参考)
C 1 2 Q 1/68 (2006.01)	C 1 2 Q 1/68 A	4 B 0 6 3
C 1 2 Q 1/34 (2006.01)	C 1 2 Q 1/34	

審査請求 有 請求項の数 17 O L (全 24 頁)

(21) 出願番号	特願2013-124744 (P2013-124744)	(71) 出願人	508186875 キージーン ナムローゼ フェンノートシ ャップ オランダ王国, エヌエル-6700 アー エー ワーゲニンゲン, ポストブス 21 6
(22) 出願日	平成25年6月13日 (2013.6.13)	(74) 代理人	100077838 弁理士 池田 憲保
(62) 分割の表示	特願2009-519391 (P2009-519391) の分割	(74) 代理人	100082924 弁理士 福田 修一
原出願日	平成19年7月10日 (2007.7.10)	(72) 発明者	ヴァン アイク, ミカエル ヨセフス テ レシア オランダ王国, エヌエル-5373 エー イ ヘルベン, パストール ストレイボス ストラート 12
(31) 優先権主張番号	60/830, 121		
(32) 優先日	平成18年7月12日 (2006.7.12)		
(33) 優先権主張国	米国 (US)		

最終頁に続く

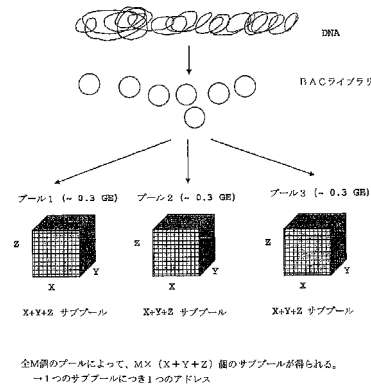
(54) 【発明の名称】 制限断片のクローン源を識別するための方法

(57) 【要約】 (修正有)

【課題】 制限断片のクローン源を識別するための方法の提供。

【解決手段】 (a) 試料DNAを準備する工程、(b) 各々が前記試料DNAの一部を含有する人工染色体のクローンのバンクを作製する工程、(c) 2つ以上のプールに存在する前記人工染色体のクローンを複数のプールで組合せてライブラリを作製する工程、(d) 前記複数のプールを1つ又は複数の制限エンドヌクラーゼで消化し、各プールに対して制限断片の組を得る消化工程、(e) 前記制限断片の片側又は両側にアダプタをライゲートし、前記アダプタの少なくとも1つが、プール特異的識別子区間または縮重識別子区間を含有することで、夫々アダプタとライゲートした制限断片を得るライゲートする工程、(f) プール特異的区間及び前記制限断片の一部をシーケンシング工程、(g) 前記プール特異的識別子を使用して工程(f)より前記制限断片配列を割当てする工程を含む方法。

【選択図】 図1



【特許請求の範囲】

【請求項 1】

制限断片のクローン源を識別するための方法であって、

(a) 試料 D N A を準備する工程と、

(b) 各々が前記試料 D N A の一部を含有する人工染色体のクローンのバンクを作製する工程と、

(c) 前記人工染色体のクローンであって各クローンは、2つ以上のプールに存在する前記クローンを複数のプールで組合せる工程であって、ライブラリを作製する、組合せる工程と、

(d) 前記複数のプールを1つ又は複数の制限エンドヌクレアーゼで消化し、各プールに対して制限断片の組が得られる、消化する工程と、

(e) 前記制限断片の片側又は両側にアダプタをライゲートする工程であって、前記アダプタの少なくとも1つが、プール特異的識別子区間または縮重識別子区間を含有することで、夫々アダプタとライゲートした制限断片を得るライゲートする工程と、

(f) 少なくともプール特異的区間及び前記制限断片の一部でシーケンシングする工程と、

(g) 前記プール特異的識別子を使用して、対応するクローンに対して、工程 (f) の前記アダプタとライゲートした制限断片で決定される前記制限断片配列を割り当てる工程とを含むことを特徴とする方法。

10

【請求項 2】

請求項 1 に記載の方法において、前記制限断片の一部で同一の配列を含有するが、異なるプール特異的識別子を担持するアダプタとライゲートした制限断片をクラスタリングすることによって、該制限断片が前記対応するクローンに割り当てられることを特徴とする方法。

20

【請求項 3】

請求項 1 に記載の方法において、前記アダプタとライゲートした制限断片は、増幅されて前記シーケンシングの前に単位複製配列を得ることを特徴とする方法。

【請求項 4】

請求項 1 に記載の方法において、前記シーケンシングが、ハイスループットシーケンシングによって行われることを特徴とする方法。

30

【請求項 5】

請求項 4 に記載の方法において、前記ハイスループットシーケンシングが固体支持体上で行われることを特徴とする方法。

【請求項 6】

請求項 4 に記載の方法において、前記ハイスループットシーケンシングが合成時解読 (S e q u e n c i n g - b y - S y n t h e s i s) に基づくことを特徴とする方法。

【請求項 7】

請求項 4 に記載の方法において、前記ハイスループットシーケンシングが、

前記単位複製配列又は前記アダプタとライゲートした制限断片を、ビーズにアニーリングする工程であって、各ビーズは、単一のアダプタとライゲートした制限断片又は単位複製配列とアニーリングする工程と、

40

前記ビーズを、各油中水型マイクロリアクタが単一のビーズを含むように前記油中水型マイクロリアクタ中で乳化させる工程と、

エマルション P C R を実施し、前記アダプタとライゲートした制限断片又は前記単位複製配列を前記ビーズ表面上で増幅する、実施する工程と、

前記ビーズを夫々のウェルが単一のビーズを含むように前記ウェルに充填する工程と、

ピロリン酸シグナルを発生させる工程とを含むことを特徴とする方法。

【請求項 8】

請求項 7 に記載の方法において、前記方法は、さらに、増幅した単位複製配列を含有するビーズを選択 / 濃縮する工程を有することを特徴とする方法。

50

【請求項 9】

請求項 4 に記載の方法において、前記ハイスルーブットシーケンシングが、前記アダプタとライゲートした制限断片又は前記単位複製配列を、第 1 のプライマー及び第 2 のプライマー又は第 1 のプライマー結合配列及び第 2 のプライマー結合配列をそれぞれ含有する表面にアニーリングする工程と、

架橋増幅を実施し、増幅したアダプタとライゲートした制限断片又は増幅した単位複製配列のクラスタが得られる、実施する工程と、

標識した可逆性終止ヌクレオチドを使用して、前記増幅したアダプタとライゲートした制限断片又は前記増幅した単位複製配列のヌクレオチド配列を決定する工程とを含むことを特徴とする方法。

10

【請求項 10】

請求項 1 に記載の方法において、前記識別子が、4 b p ~ 16 b pであることを特徴とする方法。

【請求項 11】

請求項 10 に記載の方法において、前記識別子が、4 b p ~ 10 b pであることを特徴とする方法。

【請求項 12】

請求項 11 に記載の方法において、前記識別子が、4 b p ~ 8 b pであることを特徴とする方法。

【請求項 13】

請求項 12 に記載の方法において、前記識別子が、4 b p ~ 6 b pであることを特徴とする方法。

20

【請求項 14】

請求項 9 に記載の方法において、前記識別子が 2 つ以上の同一の連続塩基を含有しないことを特徴とする方法。

【請求項 15】

請求項 10 乃至 14 の内のいずれか一項に記載の方法において、2 つ以上のクローンに対して、前記対応する識別子が、少なくとも 2 つの異なるヌクレオチドを含有することを特徴とする方法。

【請求項 16】

請求項 10 乃至 14 の内のいずれか一項に記載の方法において、前記少なくとも 1 つのプライマーが、その 3' 末端で 1 個 ~ 10 個の選択的ヌクレオチドを担持し、単位複製配列のランダムサブセットが得られることを特徴とする方法。

30

【請求項 17】

請求項 10 乃至 14 の内のいずれか一項に記載の方法において、前記少なくとも 1 つのプライマーが、その 3' 末端で 1 個 ~ 4 個の選択的ヌクレオチドを担持し、単位複製配列のランダムサブセットが得られることを特徴とする方法。

【発明の詳細な説明】**【技術分野】****【0001】**

本発明は、分子生物学及びバイオテクノロジーの分野に関する。特に、本発明は、核酸の検出及び同定の分野に関する。より詳細には、本発明は、ハイスルーブットシーケンシング技術を使用した、ゲノム又はその一部の物理地図の作製に関する。

40

【背景技術】**【0002】**

統合された遺伝的且つ物理的なゲノム地図は、地図に基づく遺伝子単離、比較ゲノム解析に、及びゲノムシーケンシングプロジェクトのための配列対応 (sequence-ready) クローンの供給源として非常に貴重である。ゲノム研究用の種の物理的且つ遺伝的マーカーの統合地図が利用可能であることの影響は非常に大きい。統合地図によって、正確且つ迅速な遺伝子マッピング、並びにマイクロサテライト遺伝子座及び SNP マーカーの正確な

50

マッピングが可能になる。複雑性が様々なゲノムの物理地図をアセンブリするのに様々な方法が開発されている。より良好な特徴を示すアプローチの1つは、制限酵素を使用してゲノムのサブクローンから大量のDNA断片を生成することである（非特許文献1、非特許文献2、非特許文献3）。これらのフィンガープリントを比較して、関連クローンを特定し、コンティグにおいて重複するクローンをアセンブリする。しかし、ゲル間のDNA移動の変動、反復DNAの存在、制限部位の異常な分布及びクローン提示の歪みのために、複雑なゲノムの大きい挿入クローンの順番の決定（ordering）に対するフィンガープリントの有用性は制限されている。したがって、フィンガープリントと、PCRベース又はハイブリダイゼーションベースの方法との組合せを用いて、複雑なゲノムの最も高品質な物理地図を構築している。しかし、フィンガープリント技術の利用の欠点の1つは、間接的な方法である断片パターンマッチング法に基づいていることである。

10

【0003】

実際の配列データに基づきコンティグを作製すること、即ちより直接的な方法によって、物理地図を作製することが好ましい。配列を基にした物理地図はより正確であるだけでなく、同時に対象の種の全ゲノム配列の決定にも寄与する。近年、より効率的且つ費用効果的にクローンの完全ヌクレオチド配列の決定を可能にするハイスループットシーケンシング法が利用可能になっている。

【0004】

しかしながら、制限断片全体のシーケンシングによる検出は、依然として比較的不経済である。さらに、例えば本明細書中の他の部分で開示される、現行技術水準のシーケンシング技術（454 Life Sciences、www.454.com、Solexa、www.solexa.com及びHelicos、www.helicosbio.comによる）は、その圧倒的なシーケンシング能にも関わらず、制限長断片のシーケンシングしか提供することができない。また、現行方法は1回の操作で多くの試料を同時に処理することができない。

20

【先行技術文献】

【特許文献】

【0005】

【特許文献1】欧州特許第534858号

【特許文献2】米国特許第6045994号

【非特許文献】

30

【0006】

【非特許文献1】Brenner et al., Proc. Natl. Acad. Sci., (1989), 86, 8902-8906

【非特許文献2】Gregory et al., Genome Res. (1997), 7, 1162-1168

【非特許文献3】Marra et al., Genome Res. (1997), 7, 1072-1084

【非特許文献4】Albert L. Lehninger著「生化学の原理（Principles of Biochemistry）」、793-800（Worth Pub. 1982）

【非特許文献5】Vos et al., Nucleic Acid Research, 1995, 23, 21, 4407-4414

【非特許文献6】Iannone et al. Cytometry 39:131-140, 2000

【発明の概要】

【発明が解決しようとする課題】

40

【0007】

本発明の目的は、制限断片のクローン源を識別するための方法を提供することにある。

【課題を解決するための手段】

【0008】

本発明者らは、ライブラリにおけるクローンの制限酵素による消化、得られた配列のアダプタ-ライゲーション、（選択的）増幅、ハイスループットシーケンシング及びデコンプレッションの組合せを使用することによって、大きくて複雑なゲノムであっても物理地図をアセンブリするのに使用することができるコンティグが得られることを見出し、本発明をなすにいたったものである。

【0009】

50

本発明によれば、制限断片のクローン源を識別するための方法であって、

(a) 試料DNAを準備する工程と、

(b) 各々が前記試料DNAの一部を含有する人工染色体のクローンのバンクを作製する工程と、

(c) 前記人工染色体のクローンであって各クローンは、2つ以上のプールに存在する前記クローンを複数のプールで組合せる工程であって、ライブラリを作製する、組合せる工程と、

(d) 前記複数のプールを1つ又は複数の制限エンドヌクレアーゼで消化し、各プールに対して制限断片の組が得られる、消化する工程と、

(e) 前記制限断片の片側又は両側にアダプタをライゲートする工程であって、前記アダプタの少なくとも1つが、プール特異的識別子区間または縮重識別子区間を含有することで、夫々アダプタとライゲートした制限断片を得るライゲートする工程と、

(f) 少なくともプール特異的区間及び前記制限断片の一部でシーケンシングする工程と、

前記プール特異的識別子を使用して、対応するクローンに対して、工程(f)の前記アダプタとライゲートした制限断片で決定される前記制限断片配列を割り当てる工程とを含むことを特徴とする方法が得られる。

【0010】

定義

以下の説明及び実施例において、多くの用語が使用される。かかる用語に与えられる範囲を含む、明細書及び特許請求の範囲についての明確で且つ一貫した理解を提供するために、以下の定義が提供される。本明細書中で特に規定しない限り、使用されるすべての技術用語及び科学用語は、本発明が属する技術分野の当業者によって共通に理解されるものと同一の意味を有する。すべての刊行物、特許出願、特許及び他の参考文献の開示内容全体が、参照により本明細書中に援用される。

【0011】

核酸：本発明に記載の核酸は、ピリミジン塩基及びプリン塩基、好ましくはシトシン、チミン及びウラシル、並びにアデニン及びグアニンの任意のポリマー又はオリゴマーをそれぞれ含み得る（非特許文献4（この全体がすべての目的のために参照により本明細書中に援用される）を参照されたい）。本発明では、任意のデオキシリボヌクレオチド、リボヌクレオチド又はペプチド核酸成分、及びそれらの任意の化学的な変形（例えばこれらの塩基のメチル化形態、ヒドロキシメチル化形態又はグリコシル化形態）等が考えられる。ポリマー又はオリゴマーは、組成物中で不均質であっても又は均質であってもよく、天然の供給源から単離してもよく、又は人為的若しくは合成的に生成してもよい。また核酸は、DNA若しくはRNA、又はそれらの混合物であってもよく、恒久的に又は一時的に、一本鎖形態、又はホモ二重鎖、ヘテロ二重鎖、及びハイブリッド状態を含む、二本鎖形態で存在してもよい。

【0012】

AFLP：AFLPとは、1つ又は複数の制限エンドヌクレアーゼにより核酸を消化して制限断片を生成すること、アダプタを制限断片にライゲートすること、及びアダプタと（一部）相補的であり、制限エンドヌクレアーゼの残部と（一部）相補的であり、A、C、T又はG（又は場合によってはU）の中からランダムに選択される少なくとも1つのヌクレオチドをさらに含有する少なくとも1つのプライマーを用いて、アダプタとライゲートした制限断片を増幅することに基づく核酸の選択的増幅法を指す。AFLPには事前配列情報は全く必要なく、AFLPは任意の開始DNAで実施し得る。概してAFLPは、

(a) 核酸、特にDNA又はcDNAを1つ又は複数の特異的な制限エンドヌクレアーゼで消化する工程であって、対応する一連の制限断片へとDNAを断片化する、消化する工程と、

(b) こうして得られた制限断片を、一端が制限断片の一端又は両端と適合する二本鎖合成オリゴヌクレオチドアダプタとライゲートする工程であって、それによりアダプタ

10

20

30

40

50

とライゲートした（好ましくは、タグ付けした）開始DNAの制限断片を生成する、ライゲートする工程と、

（c）アダプタとライゲートした（好ましくは、タグ付けした）制限断片を、ハイブリダイズ条件下で、その3'末端に選択ヌクレオチドを含有する1つ又は複数のオリゴヌクレオチドプライマーと接触させる工程と、

（d）プライマーとハイブリダイズした、アダプタとライゲートした（好ましくは、タグ付けした）制限断片を、PCR又は同様の技法によって増幅する工程であって、プライマーがハイブリダイズする開始DNAの制限断片に沿って、ハイブリダイズしたプライマーのさらなる伸長を引き起こす、増幅する工程と、

（e）こうして得られた増幅又は伸長したDNA断片を、検出、同定、又は回収する工程とを含む。

10

【0013】

したがって、AFLPはアダプタとライゲートした断片の再現可能なサブセットを提供する。AFLPはとりわけ特許文献1、特許文献2、及び非特許文献5に記載されている。AFLPに関するさらなる詳細に関しては、これらの刊行物が参照される。AFLPは複雑度低減技法及びDNAフィンガープリント技術として一般に使用される。AFLPをフィンガープリント技術として使用する状況の中で、AFLPマーカ-の概念が発展してきた。

【0014】

選択塩基：選択塩基は、アダプタと相補的な部分、及び制限部位の残部と相補的な部分を含有するプライマーの3'末端に位置し、A、C、T又はGの中からランダムに選択される。選択塩基を用いてプライマーを伸長することにより、続く増幅ではアダプタとライゲートした制限断片の再現可能なサブセットのみ、即ち、選択塩基を担持するプライマーを使用して増幅し得る断片のみが生成される。選択ヌクレオチドは1個～10個の数でプライマーの3'末端に付加され得る。典型的には、1個～4個で十分であり、且つ好ましい。両プライマーは様々な数の選択塩基を含有し得る。それぞれ付加された選択塩基によって、サブセット中の増幅したアダプタとライゲートした制限断片（単位複製配列）の量を約4分の1に低減させる。典型的には、AFLPで使用される選択塩基の数は $+N+M$ で示される（この場合、1つのプライマーがN個の選択ヌクレオチドを担持するとともに他のプライマーがM個の選択ヌクレオチドを担持する）。したがって、 $EcoRI/MseI+1/+2$ AFLPとは、開始DNAのEcoRI及びMseIによる消化、適切なアダプタのライゲーション、並びに1つの選択塩基を担持するEcoRI制限位置を対象とする1つのプライマー、及び2つの選択ヌクレオチドを担持するMseI制限部位を対象とするもう1つのプライマーを用いる増幅の省略表現である。少なくとも1つの選択ヌクレオチドをその3'末端に担持する、AFLPで使用されるプライマーもAFLPプライマーとして示される。選択ヌクレオチドをその3'末端に担持せず、且つ実際はアダプタ及び制限部位の残部と相補的なプライマーは、AFLP+0プライマーとして示されることもある。

20

30

【0015】

クラスタリング：「クラスタリング」という用語は、短く又は長く続く同一又は類似のヌクレオチドの存在に基づいて、2つ以上のヌクレオチド配列を比較すること及び短く（又は長く）続く同一又は類似の配列に基づいて、或る特定の最小レベルの配列相同性を有する配列をグループ化することを意味する。

40

【0016】

アセンブリ：（部分）重複配列の一群の順番決定に基づくコンティグの構築であり、「コンティグ構築」とも呼ばれる。

【0017】

アライメント：例えばギャップを導入することによって、アライメント中の様々な配列にわたって配列同一性を有する領域を得る可能性を最大にするための図面における複数の配列の位置決め。以下でさらに説明されるように、幾つかのヌクレオチド配列のアライメ

50

ント法が当該技術分野で既知である。

【0018】

識別子：アダプタ若しくはプライマーに付加され得るか、又はその配列内に含まれ得るか、又はそうでなければ標識として使用されて一意の識別子を提供し得る短い配列。かかる配列識別子（タグ）は、特異的な核酸試料を同定するために一意的に使用される、多様であるが規定された長さを有する一意の塩基配列であり得る。例えば、4 bpのタグにより、 $4^4 = 256$ 個の異なるタグができる。典型例は、ハイブリダイゼーションによる一意的な検出に一般に使用されるタグとして当該技術分野において既知のZIP配列である（非特許文献6）。かかる識別子を使用することによって、さらなる処理時にPCR試料の起源が求められ得る。異なる核酸試料に由来する処理産物を組合せる場合には、概して、異なる核酸試料は異なる識別子を使用して同定される。

10

【0019】

シーケンシング：「シーケンシング」という用語は、核酸試料、例えばDNA又はRNA中のヌクレオチドの順番（塩基配列）を決定することを指す。

【0020】

ハイスループットスクリーニング：HTSと略されることの多いハイスループットスクリーニングは、特に生物学及び化学の分野に関連した科学実験用の方法である。現代のロボット工学及び他の専門化した実験ハードウェアの組合せによって、研究者は大量の試料を同時に効率的にスクリーニングすることが可能となる。

【0021】

制限エンドヌクレアーゼ：制限エンドヌクレアーゼ又は制限酵素は、二本鎖DNA分子中の特異的なヌクレオチド配列（標的部位）を認識し、すべての標的部位又はその近くでDNA分子の両鎖を切断する酵素である。

20

【0022】

制限断片：制限エンドヌクレアーゼによる消化によって生成するDNA分子は制限断片と称される。任意の所定のゲノム（又はその起源に関わらず核酸）は、特定の制限エンドヌクレアーゼによって制限断片の別個の組へと消化される。制限エンドヌクレアーゼ切断に由来するDNA断片は、多様な技法でさらに使用することができ、例えばゲル電気泳動によって検出することができる。

【0023】

ライゲーション：2つの二本鎖DNA分子を相互に共有結合させるリガーゼ酵素によって触媒される酵素反応は、ライゲーションと称される。概して、両方のDNA鎖は相互に共有結合するが、鎖の一端の化学的修飾又は酵素的修飾を通じて、2つの鎖のうちの1つの鎖のライゲーションを防ぐことも可能である。その場合には、共有結合が2つのDNA鎖のうち1つのみで生じる。

30

【0024】

合成オリゴヌクレオチド：化学的に合成することができる、好ましくは約10塩基～約50塩基を有する一本鎖DNA分子は、合成オリゴヌクレオチドと称される。概して、これらの合成DNA分子は、一意又は所望のヌクレオチド配列を有するように設計されるが、関連配列を有し、且つヌクレオチド配列内の特定位置に異なるヌクレオチド組成を有する分子のファミリーを合成することが可能である。「合成オリゴヌクレオチド」という用語は、設計した又は所望のヌクレオチド配列を有するDNA分子を指すために使用される。

40

【0025】

アダプタ：制限断片末端にライゲートすることができるように設計されている、限られた数の塩基対、例えば約10塩基対長～約50塩基対長を有する短い二本鎖DNA分子。アダプタは概して、互いに一部が相補的なヌクレオチド配列を有する2つの合成オリゴヌクレオチドから構成される。適切な条件下、溶液中で2つの合成オリゴヌクレオチドを混合すると、これらは互いにアニーリングして二本鎖構造を形成する。アニーリング後、アダプタ分子の一端は、制限断片の末端と適合し、それとライゲートすることができるよう

50

に設計される。アダプタのもう一方の端は、ライゲートすることができないように設計することができるが、そうである必要はない(二重にライゲートしたアダプタ)。

【0026】

アダプタとライゲートした制限断片：アダプタでキャッピングした制限断片。

【0027】

プライマー：概して、「プライマー」という用語は、DNAの合成をプライムすることができるDNA鎖を指す。DNAポリメラーゼは、プライマーなしではDNAをdenovo合成することができない。即ち、アセンブリするヌクレオチドの順番を指示するための鋳型として相補鎖が使用される反応において、DNAポリメラーゼは既存のDNA鎖しか伸長することができない。本発明者らは、プライマーとして、ポリメラーゼ連鎖反応(PCR)で使用される合成オリゴヌクレオチド分子を指す。

10

【0028】

DNAの増幅：「DNAの増幅」という用語は、典型的には、PCRを使用する二本鎖DNA分子のin vitroでの合成を指し示すために使用される。他の増幅方法が存在し、それらが趣旨を逸脱することなく、本発明で使用され得ることに留意されたい。

【図面の簡単な説明】

【0029】

【図1】プーリング戦略の概略図である。

【図2】シロイヌナズナ染色体4に対する4つの連続したBACコンティグ-プーリング戦略を示す図である。

20

【図3】グループ内で重複しない代替的な最小タイリングパス(tiling path)を示す図である。

【図4】BACプールで扱われる認識配列-アガロースゲル上の増幅産物を示す図である。

【図5】再アセンブリした最小タイリングパス-1.9Mbのコンティグ部分の拡大を示す図である。

【発明を実施するための形態】

【0030】

本発明の実施するための形態について説明する。

【0031】

本発明は、一態様において、制限断片のクローン源を識別するための方法であって、

30

(a) 試料DNAを準備する工程と、

(b) 各々が前記試料DNAの一部を含有する人工染色体のクローンのバンクを作製する工程と、

(c) 前記人工染色体のクローンであって各クローンは、2つ以上のプールに存在する前記クローンを複数のプールで組合せる工程であって、ライブラリを作製する、組合せる工程と、

(d) 前記複数のプールを1つ又は複数の制限エンドヌクレアーゼで消化し、各プールに対して制限断片の組が得られる、消化する工程と、

(e) 前記制限断片の片側又は両側にアダプタをライゲートする工程であって、前記アダプタの少なくとも1つが、プール特異的識別子区間または縮重識別子区間を含有することで、夫々アダプタとライゲートした制限断片を得るライゲートする工程と、

40

(f) 少なくともプール特異的区間及び前記制限断片の一部でシーケンシングする工程と、

前記プール特異的識別子を使用して、対応するクローンに対して、工程(f)の前記アダプタとライゲートした制限断片で決定される前記制限断片配列を割り当てる工程とを含むことを特徴とする方法に関する。

【0032】

この方法の工程(a)では、試料DNAを準備する。例えばSambrook et al (Sambrook and Russell (2001) "Molecular Cloning: A Laboratory Manual (3rd edition)", Cold

50

Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press) によって開示されたような当該技術分野における任意の方法でこれを達成することができる。試料DNAは、任意の種、特にヒト、植物又は動物由来であり得る。本明細書の他の部分に記載されるように、本発明はまた、例えばAFLPに基づく選択的増幅を介した再現可能なサブセットの作製によって、任意のサイズのゲノムに合わせた方法を提供するので、ゲノムの一部のみを使用することが可能であるが、これは必須という訳ではない。したがって典型的には、本発明は全ゲノムを使用する。

【0033】

工程(b)では、人工クローンバンクを作製する。ライブラリは、細菌人工染色体ライブラリであり得るか(BAC)、又は酵母に基づき得る(YAC)。コスミド、PAC、TAC又はMAC等に基づく他のライブラリも可能である。BACライブラリが好ましい。ライブラリは、高品質であることが好ましく、挿入サイズが大きいゲノムライブラリが好ましい。これは、個々のBACが、研究中のゲノムDNAに大きい挿入を含有することを意味する(典型的には、125kbp超)。好ましい大きい挿入のサイズは、種によって異なる。本願を通して、人工染色体の例として、BACに対して言及が為されている。しかし、本発明はそれらに限定されず、本発明の要旨を逸脱することなく、他の人工染色体を使用することができることに留意されたい。好ましくは、ライブラリは、少なくとも5ゲノム当量、より好ましくは少なくとも7ゲノム当量、最も好ましくは少なくとも8ゲノム当量を含有する。特に好ましくは、少なくとも10ゲノム当量である。ライブラリにおけるゲノム当量数が高くなれば、得られるコンティグ及び物理地図の信頼性が高くなる。

10

20

【0034】

ライブラリにおける個々のクローンをプールし、多くの人工染色体又はクローンを含有するプールを形成する。プーリング(pooling)は、1つの試料への多くの個々のクローン(例えば、10個のプールへの100個のクローン(それぞれが10個のクローンを含有する))の単純な組合せであり得るが、より精密なプーリング戦略を用いてもよい。プールにわたるクローン分布は、各クローンが少なくとも2つ以上のプールに存在するようなものが好ましい。プールは、1つのプール当たり10個~10000個、好ましくは100個~1000個、より好ましくは250個~750個のクローンを含有することが好ましい。1つのプール当たりのクローンの数は大いに異なり得ることが観察され、このばらつきは、例えば研究中のゲノムのサイズに関係がある。典型的には、プール又はサブプールの最大サイズは、識別子の組によってプール中のクローンを一意的に同定する能力によって支配される。以下でさらに詳述されるように、プールにおけるゲノム当量の典型的な範囲は、約0.2~0.3であり、これも1つのゲノム当たりで異なり得る。プールは、当該技術分野で既知のプーリング戦略に基づいて作製される。当業者は、ゲノムサイズ等の要因に基づき、最適なプーリング戦略を選択することができる。得られるプーリング戦略は環境に依存し、その例は、プレートプーリング、N次元プーリング(例えば2次元プーリング、3次元プーリング、6次元プーリング)、又は複雑なプーリングである。多くのプールの操作を容易にするために、3次元プーリングを示す添付の図1で例示されるように、プールは順番にスーパープール(即ちスーパープールは、クローンプールのプールである)に組合せられるか、又はサブプールに分けられ得る。プーリング戦略及びそのデコンポリューション(即ち1つ又は複数のプール又はサブプールにおけるクローンの既知の関連指標(即ち標識又は識別子)の存在検出による、ライブラリにおける個々のクローンの正確な同定)の他の例は、例えば米国特許第6975943号又はKlein et al. in Genome Research, (2000), 10, 798-807に記載される。プーリング戦略は、ライブラリにおけるあらゆるクローンがプールの一意的な組合せがあらゆるクローンに適合するプール全体に分布するようなものが好ましい。その結果、或る特定の(サブ)プールの組合せがクローンを一意的に同定する。

30

40

【0035】

制限エンドヌクレアーゼでプールを消化して、制限断片を得る。各プールが別々にエン

50

ドヌクレアーゼ消化に供されるのが好ましい。各プールを同じ（組合せ）のエンドヌクレアーゼ（複数可）で処理する。原則として、あらゆる制限エンドヌクレアーゼを使用することができる。制限エンドヌクレアーゼは、フリークエントカッター（MseI又はPstI等の4又は5カッター）又はレアカッター（EcoRI、HindIII等の6以上のカッター）であり得る。典型的には、平均して或る量で存在するか、又はその後の工程に適している或る特定の長さ分布を有する制限断片が得られるように、制限エンドヌクレアーゼが選択される。或る特定の実施の形態では、2つ以上の制限エンドヌクレアーゼを使用することができ、また或る特定の実施の形態では、レアカッターと、フリークエントカッターとの組合せを使用することができる。大きいゲノムには、例えば3つ以上の制限エンドヌクレアーゼを有利に使用することができる。

10

【0036】

制限断片の一端又は両端に対して、工程（e）でアダプタをライゲートし、アダプタとライゲートした制限断片を得る。典型的に、本明細書の他の部分で規定されるように、アダプタは合成オリゴヌクレオチドである。本発明で使用されるアダプタは、本明細書の他の部分で規定されるように、本質的に識別子区間を含有することが好ましい。或る特定の実施の形態では、アダプタは、プール特異的識別子を含有する、即ち各プールに対して、プールを明確に示す一意の識別子を含有するアダプタを使用する。或る特定の実施の形態では、アダプタは、プール特異的識別子を含有するプライマーと組合せて使用される縮重識別子区間を含有する。

20

【0037】

或る特定の実施の形態では、特にアダプタがプール特異的識別子を含有するときに、アダプタとライゲートした制限断片をより大きいグループに組合せることができる。より大きいグループへのこの組合せは、プールから得られるアダプタとライゲートした制限断片の各組の並行増幅数を低減させるのに役立つ。

【0038】

プライマーの組（少なくとも1つのプライマーが、アダプタにおけるプール特異的な位置又は縮重識別子の位置にプール特異的識別子を含有する）を使用して、アダプタとライゲートした制限断片を増幅することができる。この実施の形態は、上記で概説したように、増幅前にアダプタとライゲートした制限断片のグループ化も可能にする。代替的な実施の形態では、別々にプライマーの組（少なくとも1つのプライマーがプール特異的な区間を含有する）を使用することによって一意的にプールを同定して、各プールのアダプタとライゲートした制限断片（アダプタが縮重識別子区画を含有していた）を増幅する。

30

【0039】

どちらにしても、結果として、それが由来するプールと結び付く、アダプタとライゲートした制限断片（単位複製配列とも表される）の組が単位複製配列におけるプール特異的識別子の存在により増幅される。或る特定の実施の形態では、本質的に本明細書の他の部分で記載されるように、その3'末端で選択的ヌクレオチドを担持するプライマーを用いて、選択的増幅によって、単位複製配列のサブセットを作製してもよい。

【0040】

或る特定の実施の形態では、単位複製配列を、組合せた単位複製配列又はいわゆる配列ライブラリと組合せてもよい。

40

【0041】

この方法の工程（i）では、単位複製配列は、本明細書中の下記に記載されるようにシーケンシング、好ましくはハイスループットシーケンシングにかけられる。シーケンシング中に、単位複製配列のヌクレオチド配列の少なくとも一部が決定される。好ましくは、少なくともプール特異的識別子の配列及び単位複製配列の制限断片の一部が決定される。好ましくは、制限断片の少なくとも10ヌクレオチドの配列が決定される。或る特定の実施の形態では、制限断片の少なくとも11、12、13、14又は15ヌクレオチドが決定される。また、決定する最小のヌクレオチド数はゲノムに依存する。例えば、植物にはより反復性のある配列が存在し、これによってより長い配列（25bp～30bp

50

)が決定される。例えば、既知のシロイヌナズナゲノムの計算では、シーケンシング工程で6bpの制限部位を含む場合、1つの制限断片当たり約20bpを決定する必要があることが示されている。全制限断片の配列を決定することが可能であるが、これはBACクロンのコンティグ構築に絶対必要なものではない。

【0042】

シーケンシング工程では、精度の増大を得るために、少なくとも5の包括度で、配列ライブラリをシーケンシングしてもよい。これは、1つの特異的なアダプタとライゲートした制限断片の単位複製配列の増幅から得られる少なくとも5つの単位複製配列を有する配列を決定することを意味する。言い換えれば、各制限断片は、(統計的に)少なくとも5回シーケンシングされる。さらに精度が高まるので、包括度の増大は好ましく、好ましくは包括度は少なくとも7であり、より好ましくは少なくとも10である。包括度の増大を利用して、「サンプリング変動」として知られる現象を補う。

10

【0043】

以下の工程では、コンピュータ制御方法によって、典型的にコンピュータ内では、(一部)シーケンシングされた単位複製配列は、対応するクローンと相関がある。制限断片由来の部分においてヌクレオチドの同一区間を含有する単位複製配列が選択される。その後、これらの単位複製配列に存在する様々なプール特異的識別子を同定する。様々なプール特異的識別子の組合せ、及びしたがって制限断片の配列は、特定のクローンに一意的に割り当てることができる(上記で「デコンボリューション」として記載されたプロセス)。例えば、3次元プーリング戦略(X、Y、Z)の場合、3つのプールに特異的な識別子の組合せによって、ライブラリにおける各プールが一意的に扱われる。各クローンがライブラリにおいて2回以上発生するので、ライブラリにおけるクローンの各発生に対して、同じ制限断片由来の区画と組合せて、3つのプールに特異的な識別子を組合せることができる。言い換えれば、クローンに起因する制限断片由来の区画は、3つの異なる識別子でタグ付けされる。3つの識別子と組合せて観察される場合、一意の制限断片由来の区画を単一のBACクロンに割り当てることができる。制限断片由来部分で他の一意のヌクレオチド区画を含有する各単位複製配列に対して、これを繰り返すことができる。1つのプール当たりのゲノム当量を比較的low(0.3未満、好ましくは0.2未満)維持し、それにより同じ断片が、異なるクローンに由来する同じプールで2回存在する確率を減らすことによって、このデコンボリューションプロセスを簡易にすることができる。

20

30

【0044】

プーリング概念の例示的な図面が図1に与えられる。試料DNAを、BACライブラリに変換する。BACライブラリをプールの組(M)にプールする(3つのプールが示され、各プールが約0.3GEを含有する)。各プールは、(X+Y+Z)サブプール(典型的にはマイクロタイタープレート束)に分けられる。

【0045】

ライブラリにおいて特定のクローンと結びつくシーケンシングされた単位複製配列が、制限断片由来の区画の配列マッチングに基づき、コンティグを構築するのに使用される。次いで、各クローンのコンティグをアライメントして、物理地図を作製する。

【0046】

本方法の利点は、とりわけBACコンティグ構築に関する従来技術に比べて、BACコンティグ構築に関する精度の向上にある。さらに、配列情報に基づく物理地図の構築はより正確である。これは、本方法が物理地図の構築の直接的な方法であるとともに、ゲノム配列の決定に寄与し、さらにSTS開発及び比較マッピング目的に好適な配列情報を与えるためである。

40

【0047】

本発明で使用されるハイスループットシーケンシングは、特に生物学及び化学の分野に関連した科学実験用の方法である。現代のロボット工学及び他の専門化した実験ハードウェアの組合せによって、研究者は大量の試料を同時に効率的にスクリーニングすることが可能となる。

50

【 0 0 4 8 】

シーケンシングは、国際公開第W O 0 3 / 0 0 4 6 9 0号、同第W O 0 3 / 0 5 4 1 4 2号、同第W O 2 0 0 4 / 0 6 9 8 4 9号、同第W O 2 0 0 4 / 0 7 0 0 0 5号、同第W O 2 0 0 4 / 0 7 0 0 0 7号、及び同第W O 2 0 0 5 / 0 0 3 3 7 5号（すべて454 Life Sciences名義）、Seo et al. (2004)による、Proc. Natl. Acad. Sci. USA 101:5488-93、並びにHelicos、Solexa、US Genomics等の技術（これらは参照により本明細書中に援用される）で開示されている方法等のハイスループットシーケンシング法を使用して実施されることが好ましい。

【 0 0 4 9 】

454 Life sciences技術

或る特定の実施の形態において、シーケンシングは、国際公開第W O 0 3 / 0 0 4 6 9 0号、同第W O 0 3 / 0 5 4 1 4 2号、同第W O 2 0 0 4 / 0 6 9 8 4 9号、同第W O 2 0 0 4 / 0 7 0 0 0 5号、同第W O 2 0 0 4 / 0 7 0 0 0 7号、及び同第W O 2 0 0 5 / 0 0 3 3 7 5号（すべて454 Life Sciences名義）（これらは参照により本明細書中に援用される）に開示されている装置及び/又は方法を使用して実施されることが好ましい。記載されている技術は、1回の操作での2000万～4000万塩基のシーケンシングを可能にし、競合技術よりも100倍速く、且つ安価である。シーケンシング技術は、大まかには5工程：1）一本鎖DNA（ssDNA）のライブラリを作製するための、DNAの断片化及び特定のアダプタのライゲーション、2）ssDNAのビーズへのアニーリング、油中水型マイクロリアクタにおけるビーズの乳化、及び個々のssDNA分子をビーズ上で増幅するためのエマルジョンPCRの実施、3）表面上に増幅したssDNA分子を含有するビーズの選択/濃縮、4）PicoTiter（商標）プレートにおける、DNA担持ビーズの沈着、並びに5）ピロリン酸光シグナルの発生による10000ウェルでの同時シーケンシングから構成される。以下、本方法をより詳細に説明する。

【 0 0 5 0 】

好ましい実施の形態において、シーケンシングは、

- (a) アダプタと結合した（adapted）断片をビーズにアニーリングする（各ビーズは単一のアダプタと結合した断片とアニーリングする）工程と、
- (b) ビーズ上にアニーリングした断片を油中水型マイクロリアクタ中で乳化及び増幅させる（各油中水型マイクロリアクタは単一のビーズを含む）工程と、
- (c) ビーズをウェルに充填する（各ウェルは単一のビーズを含む）工程と、ピロリン酸シグナルを発生する工程とを含む。

【 0 0 5 1 】

第1の工程（a）において、シーケンシングアダプタは、組合せライブラリ内で断片とライゲートする。上記シーケンシングアダプタは少なくとも、ビーズと結合した相補的なオリゴヌクレオチドにアニーリングするための領域、シーケンシングプライマー領域、及びPCRプライマー領域を含む。したがって、アダプタと結合した断片が得られる。

【 0 0 5 2 】

第1の工程では、アダプタと結合した断片をビーズにアニーリングする（各ビーズは単一のアダプタと結合した断片とアニーリングする）。アダプタと結合した断片のプールに過剰のビーズを加えることにより、大部分のビーズに関して、1つのビーズ当たり1つのアダプタと結合した単一断片のアニーリングが保証される（ポワソン分布）。本発明では、クローンから得られた制限断片にライゲートするアダプタは、ビーズとアニーリング可能な区画を含み得る。

【 0 0 5 3 】

次の工程では、ビーズを油中水型マイクロリアクタ中で乳化させる（各油中水型マイクロリアクタは単一のビーズを含む）。PCR試薬が油中水型マイクロリアクタ内に存在することにより、PCR反応をマイクロリアクタ内部で起こすことが可能となる。続いて、

マイクロリアクタを破壊して、DNAを含むビーズ（DNAポジティブビーズ）を濃縮する、即ち増幅断片を含有しないビーズから分離する。

【0054】

続く工程では、濃縮ビーズをウェルに充填する（各ウェルは単一のビーズを含む）。ウェルは、好ましくは大量の断片の同時シーケンシングを可能にするPicoTiter（商標）プレートの一部である。

【0055】

酵素担持ビーズの添加後、ピロシーケンシングを使用して断片の配列を決定する。引き続き工程では、従来のシーケンシング試薬の存在下、PicoTiter（商標）プレート及びビーズ並びにその中の酵素ビーズを異なるデオキシリボヌクレオチドに曝し、デオキシリボヌクレオチドを取り込む際に、発生する光シグナルを記録する。正確なヌクレオチドの取込みにより、検出し得るピロシーケンシングシグナルが発生される。

【0056】

ピロシーケンシング自体は当該技術分野において既知であり、とりわけwww.biotagebio.com; www.pyrosequencing.comの技術欄に記載されている。本技術は、例えば、国際公開第WO03/004690号、同第WO03/054142号、同第WO2004/069849号、同第WO2004/070005号、同第WO2004/070007号、及び同第WO2005/003375号（すべて454 Life Sciences名義）及びMargulies et al., nature 2005, 437, 376-380（これらは参照により本明細書中に援用される）でさらに適用される。

【0057】

本発明において、ビーズは好ましくは、ビーズ結合単位複製配列を得るために重合による伸長が可能なプライマー配列又はその一部を備える。他の実施の形態において、増幅で使用されるプライマーは、続くエマルジョン重合、その後のシーケンシングを可能にするために、例えばその5'末端に単位複製配列をビーズに結合させる配列を備える。代替的には、単位複製配列は、ビーズ又は表面とのライゲーション前にシーケンシングアダプタとライゲートし得る。単位複製配列をシーケンシングすることにより、識別子の同一性が明らかになり、ひいては識別子の組合せにより、クローンの同一性が明らかになる。

【0058】

Solexa技術

ハイスループットシーケンシング法の1つは、Solexa（英国）（www.solexa.co.uk）が利用可能であり、とりわけ、国際公開第WO0006770号、同第WO0027521号、同第WO0058507号、同第WO0123610号、同第WO0157248号、同第WO0157249号、同第WO02061127号、同第WO03016565号、同第WO03048387号、同第WO2004018497号、同第WO2004018493号、同第WO2004050915号、同第WO2004076692号、同第WO2005021786号、同第WO2005047301号、同第WO2005065814号、同第WO2005068656号、同第WO2005068089号、同第WO2005078130号に記載されている。本質的に特に本明細書の他の部分で記載されるように人工染色体プールのアダプタとライゲートした断片の場合に、本方法はDNAのアダプタとライゲートした断片で開始する。アダプタとライゲートしたDNAは、典型的にはフローセルで、固体表面と結合したプライマーの濃密な叢（lawn）とランダムに結合する。アダプタとライゲートした断片の他端は、表面上で相補的なプライマーとハイブリダイズする。いわゆる固相架橋増幅において、ヌクレオチド及びポリメラーゼの存在下でプライマーを伸長することにより、二本鎖の断片が得られる。本固相架橋増幅は選択的増幅であり得る。変性及び固相架橋増幅の反復の結果、増幅した断片の濃密なクラスタが表面全体に分布して得られる。4つの異なる標識をした可逆性終止ヌクレオチド、プライマー及びポリメラーゼをフローセルに加えることによって、シーケンシングを開始する。1回目のプライマー伸長の後、標識を検出し、第1の取込み塩基の同一性を記

10

20

30

40

50

録し、ブロックした3'末端及び蛍光体を取込み塩基から除去する。その後、同様に、第2の塩基の同一性を判定し、このようにしてシーケンシングを継続する。

【0059】

本発明において、アダプタとライゲートした制限断片又は単位複製配列は、プライマー結合配列又はプライマー配列を介して表面に結合する。識別子配列及び制限断片(の一部)を含む配列は概説したように決定する。現在利用可能なSolexa技術により、約25塩基対の断片のシーケンシングが可能である。アダプタ及び表面に結合したプライマーの経済的設計によって、シーケンシング工程は、試料識別子、制限エンドヌクレアーゼの認識配列の残部、及びいずれかの任意選択的な選択塩基を読み取る。6bpの試料識別子を使用すると、残部はレアカッターEcoRI由来であり(AACCT)、2つの選択塩基を使用すると、12bpの制限断片の内部配列が得られ、これを使用して、試料中の制限断片を一意的に同定することができる。

10

【0060】

上記のSolexaシーケンシング技術に基づく、好ましい実施の形態において、アダプタとライゲートした制限断片の増幅は、その3'末端に最大で1つの選択ヌクレオチドを含有するプライマー、好ましくはその3'末端に選択ヌクレオチドを含有しないプライマー(即ち、プライマーがアダプタのみと相補的である)(+0プライマー)を用いて実施される。

【0061】

本明細書中に記載されるシーケンシング法を対象とする代替的な実施の形態において、増幅で使用されるプライマーは、(本明細書中に記載されるプライマー又はプライマー結合配列の代用として)特定の区間を含有し得る。これは、次のシーケンシング工程で使用されて、アダプタでキャッピングした制限断片又は単位複製配列を表面に結合させる。これらは、概して鍵領域又は5'-プライマー適合配列として示される。

20

【0062】

本発明は、プール特異的な区間又は縮重識別子区間を含有するアダプタ、及び/又はプール特異的識別子を含有するプライマーそれぞれにおいて具体化される。

【実施例】

【0063】

以下、実施例に基づいて説明する。

30

【0064】

(実施例1)

合成時解読(Sequencing By Synthesis)(SBS)アプローチに基づくシロイヌナズナ(Arabidopsis thaliana)のdenovoのBACに基づく物理地図構築

本実施例は、以下の概括に基づく。

【0065】

全シロイヌナズナゲノムは約125Mbpである。細菌人工染色体(BAC)は、平均約100kbのゲノム挿入を有する。シロイヌナズナゲノムの1×物理的包括度に対して、BACの1ゲノム当量(GE)は、約1250個のBACを含む。至適結果を得るには、1つのBACプールが0.34GE(約384個のBAC)以下を含有するように、BACプールを構築することが好ましい。統計分析によって、0.34GEでは、2つの同一のBAC(即ち全く同じ物理的位置にマッピングされる2つのBAC)を見出す確率は5%未満であることが予測される。BACプールのGEが低くなると、同じ位置にマッピングされる2つのBACを見出す確率がさらに低くなる。直接的な3次元プーリングシステムを計算に使用する。2つの異なる高品質BACライブラリ(2つの異なるクローニング酵素、例えばEcoRI及びHindIII)の合計10GEのBACは、高品質物理地図の構築には十分である。シロイヌナズナに関する10GEのBACは、約12500個のBACである。

40

【0066】

配列タグ(制限断片の一部と識別子との組合せ)、例えばEcoRI/MseI若しく

50

は HindIII / MseI 等の AFLP 断片、又は幾つかの複合酵素 (EC) の組合せが、レアカッター制限部位から生成される。

【0067】

本実施例では、HindIII / MseI の複合酵素を使用する。シロイヌナズナゲノムにおける HindIII / MseI 断片の分布は、100 kb 当たり 50 ~ 120 個の断片であると推測される。

【0068】

ハイスループットシーケンシングの設定

図1も参照されたい。0.3 GE は、384 個の BAC に対応する。X + Y + Z 次元を用いた 384 個の BAC の 3 次元プーリングによって、8 + 12 + 4 = 24 個のサブプールが得られる。10 GE に関して、 $M \times (X + Y + Z) = 30 \times (8 + 12 + 4) = 720$ 個のサブプール。

10

【0069】

各サブプールに関して、

1つの BAC 当たり 100 個のシーケンシングされたタグ、

1つのタグ当たり 10 倍の配列冗長性 (sequence redundancy)、

3次元プーリング (各 BAC 断片は、各 (X、Y、Z) 次元でシーケンシングされる) を生成することが目的である。

【0070】

このことは、0.34 GE のプールの架橋増幅に基づくハイスループットシーケンシングに関して、一連のシーケンシングによって、8つのサブプール $\times (12 \times 4 \times 100 \times 10) + 12$ 個のサブプール $\times (8 \times 4 \times 100 \times 10) + 4$ つのサブプール $\times (12 \times 8 \times 100 \times 10) = 1152000$ 個のサブプールが読み取られる必要があることを意味する。このことは、1 GE に対して、1 GE 当たり $3 \times 1152000 = 3456000$ 個の読み取り、及び 10 GE 当たり $10 \times 3456000 = 34560000$ 個の読み取りが必要であることを意味する。

20

【0071】

単一の BAC は、潜在的に約 20 bp の一意の配列タグ (制限部位を含む) を約 100 個生成する。配列数は、複合酵素の選択及び / 又は組合せに依存する。

【0072】

個々の BAC が調整され、「デコンボリューション」工程によって扱われるサブプールの配列から添付の配列タグを推測することができる。結果として、デコンボリューションを介して、各配列タグを対応する個々の BAC に割り当てることが可能である。反復配列タグは無視する。デコンボリューションプロセスによって、1つの BAC 当たり 100 個の一連のタグが得られ、続いてアガロースゲルで分析された BAC 断片に関して Cari Soderlund によって説明されたように、FPC (フィンガープリント化コンティグ) 型プロセスによって、denovo 物理地図のアセンブリが得られる (Soderlund et al. 2000 - Genome Research 10; 1772-1787)。最終的に、物理地図の遺伝子地図への固定がコンピュータ内で行われる。より大きいゲノムに対しては、他のプーリング戦略が必要な場合がある。

30

40

【0073】

合成時解読 (SBS) アプローチに基づくキュウリ (Cucumis sativus) の denovo の BAC に基づく物理地図構築

本実施例は、以下の概括に基づく。

【0074】

キュウリの全ゲノムは、約 350 Mb p である。細菌人工染色体 (BAC) は、平均して約 100 kb のゲノム挿入を有する。シロイヌナズナゲノムの 1x 物理的包括度に対して、BAC の 1ゲノム当量 (GE) は、約 3500 個の BAC を含む。至適結果を得るには、1つの BAC プールが 0.34 GE (約 384 個の BAC) 以下を含有するように、BAC プールを構築することが好ましい。統計分析によって、0.34 GE では、2つの

50

同一の B A C (即ち全く同じ物理的位置にマッピングされる 2 つの B A C) を見出す確率は 5 % 未満であることが予測される。B A C プールの G E が低くなると、同じ位置にマッピングされる 2 つの B A C を見出す確率がさらに低くなる。直接的な 3 次元プーリングシステムを計算に使用する。2 つの異なる高品質 B A C ライブラリ (2 つの異なるクローニング酵素、例えば E c o R I 及び H i n d I I I) の合計 1 0 G E の B A C は、高品質物理地図の構築には十分である。キュウリに関する 1 0 G E の B A C は、約 3 5 0 0 0 個の B A C である。

【 0 0 7 5 】

配列タグ (制限断片の一部と識別子との組合せ)、例えば E c o R I / M s e I 若しくは H i n d I I I / M s e I 等の A F L P 断片、又は幾つかの複合酵素 (E C) の組合せが、レアカッター制限部位から生成される。

10

【 0 0 7 6 】

本実施例では、複合酵素 H i n d I I I / M s e I を使用する。キュウリゲノム中の H i n d I I I / M s e I 断片の分布は、1 0 0 k b 当たり 5 0 個 ~ 1 2 0 個の断片であると推測される。

【 0 0 7 7 】

ハイスループットシーケンシングの設定

図 1 を参照されたい。0 . 3 G E は 1 1 5 2 個の B A C に対応する。X + Y + Z 次元による 1 1 5 2 個の B A C の 3 次元プーリングによって、 $8 + 1 2 + 1 2 = 3 2$ 個のサブプールが得られる。1 0 G E に対して、 $M \times (X + Y + Z) = 3 0 \times (8 + 1 2 + 1 2) = 9 6 0$ 個のサブプール。

20

【 0 0 7 8 】

各サブプールに関して、

1 つの B A C 当たり 1 0 0 個のシーケンシングされたタグ、

1 つのタグ当たり 1 0 倍の配列冗長性、

3 次元プーリング (各 B A C 断片は、各 (X、Y、Z) 次元でシーケンシングされる) を生成することが目的である。

【 0 0 7 9 】

このことは、0 . 3 4 G E のプールの架橋増幅に基づくハイスループットシーケンシングに関して、一連のシーケンシングによって、8 個のサブプール $\times (1 2 \times 1 2 \times 1 0 0 \times 1 0) + 1 2$ 個のサブプール $\times (8 \times 1 2 \times 1 0 0 \times 1 0) + 1 2$ 個のサブプール $\times (1 2 \times 8 \times 1 0 0 \times 1 0) = 3 4 5 6 0 0 0$ 個のサブプールが読み取られる必要があることを意味する。このことは、1 G E に対して、1 G E 当たり $3 \times 3 4 5 6 0 0 0 = 1 0 3 6 8 0 0 0$ 個の読み取り、及び 1 0 G E 当たり $1 0 \times 1 0 3 6 8 0 0 0 = 1 0 3 6 8 0 0 0 0$ 個の読み取りが必要であることを意味する。

30

【 0 0 8 0 】

単一の B A C は、潜在的に約 2 0 b p の一意の配列タグ (制限部位を含む) を約 1 0 0 個生成する。配列数は、複合酵素の選択及び / 又は組合せに依存する。

【 0 0 8 1 】

個々の B A C が調整され、「デコンボリューション」工程によって扱われるサブプールの配列から添付の配列タグを推測することができる。結果として、デコンボリューションを介して、各配列タグを対応する個々の B A C に割り当てることが可能である。反復配列タグは無視する。デコンボリューションプロセスによって、1 つの B A C 当たり 1 0 0 個の一連のタグが得られ、続いてアガロースゲルで分析された B A C 断片に関して Cari Soderlund によって説明されたように、F P C (フィンガープリント化コンティグ) 型プロセスによって、d e n o v o 物理地図のアセンブリが得られる (Soderlund et al. 2000 - Genome Research 10; 1772-1787)。最終的に、物理地図の遺伝子地図への固定がコンピュータ内で行われる。より大きいゲノムに対しては、他のプーリング戦略が必要な場合がある。

40

【 0 0 8 2 】

50

プールしたBACからAFLP鑄型(EcoRI/MseI又はHindIII/MseI)を調製する。2個のHindIII+1プライマーと、1個のMseI+0プライマーとの組合せを用いて、AFLP増幅を行う(EcoRIに関しても同じ)。2個の+1プライマーの使用によって、プールからの約50%のH/M(又はE/M)断片の増幅、即ち各複合酵素に対して平均して $70/2=35$ 個の制限断片が増幅されるのを確実にする。各BACプールの5'末端に一意の識別子タグを含有するAFLPプライマーで、AFLP増幅反応を行う。したがって、少なくとも74個の識別子配列が必要である。4つの塩基タグ($4^4=256$ 個の可能性)によって、これを達成することができる。本実施例では一方向性のシーケンシングを行うので、HindIIIプライマーには、識別子配列だけが必要である。

10

【0083】

すべてのプールのAFLP反応混合物を等量で混合し、断片ライブラリを作製する。断片ライブラリを用いて、配列ライブラリを構築する。

【0084】

3次元プーリング戦略では、このことは、各次元で断片毎に平均して複数回サンプリングすることを意味する。結果は、制限断片のHindIII(又はEcoRI)部位に由来する100bpの配列である。上記のように、1つのBACクローン当たり平均して35個の配列が得られる。この配列は、FPC(<http://www.agcol.arizona.edu/software/fpc/>から入手可能なSoderlund製のソフトウェアパッケージ)と同様の手順を用いるが、配列マッチングに基づいて(より詳細には)、コンティグアセンブリの基礎となる。

20

【0085】

再現性のある複雑性の低減を利用することの利点は、物理地図の構築に必要な断片が少なくなることである。上記のキュウリの例における複雑性の50%の低減は、読み取りを103680000個から51840000個にする。本明細書に記載されるような複雑性の低減を利用した本発明のさらなる利点は、制御可能な品質で物理地図を作製することができることである。このことは、+1のAFLP増幅、例えば+Cをプライマーに組合せることで、BACプールの複雑性を低減させることによって、+1の増幅を4つすべてのプライマーの組合せ(A、C、T、G)と比べて、物理地図の品質(包括度)が約25%になることを意味する。しかしながら、2つ又は3つのプライマーの組合せを用いると、+1の増幅を4つすべてのプライマーの組合せ(A、C、T、G)で得られる包括度と比べると、包括度の増大、即ち例えばそれぞれ55%又は90%が得られる。

30

【0086】

BACクローンの役割

3つの異なるタグ付けしたプライマーで同じBACクローン由来の断片を増幅する。したがって、3つのタグとの組合せで観察された一意の配列をライブラリ中の単一のBACクローンに割り当てる。反復配列は、複数のタグとの組合せで観察されるので、単一のBACクローンには結び付けることはできない。このことは、かなりの割合の断片に影響を与えるが、1つのBACクローン当たり35個の断片の中で少なくとも1つのサブセットが一意的なものである。

【0087】

BACプールの10倍の配列包括度(3.3倍/次元)は、(個々のクローンの濃度差及びサンプリングの変動等のために)すべての予測される断片が観察される訳ではないことを意味する。したがって、(一意の)配列分画は、1つ又は2つのタグとの組合せで観察されるにすぎない(又は全く観察されない)。このことが、配列を単一のBACクローンに割り当てるのを邪魔している。しかしながら、これが同じクローンに由来する制限断片間のサンプリングの変動によるものであれば、35個の断片をサンプリングすることは、タグの組合せが、BACに対する正確なアドレスを提供することを意味する:下記を参照されたい。

40

【0088】

【表 1】

	タグ1	タグ2	タグ3
断片1	X	X	
断片2		X	X
断片3		X	X
断片4	X		X
断片5	X	X	X
等			
断片35	X	X	

10

【0089】

上記のスキームは、コンティグ構築が、コンティグにおいて相互に断片をグループ化する、即ち断片5（一意の配列を有し、3つのタグを組合せてサンプリングした）が、同様に断片1～断片4（+35）が由来すると考えられる、ライブラリ中でBACアドレスを規定することを示している。

【0090】

したがって、このアプローチの強みは、十分多い制限断片（上記実施例では35個）における配列情報を正確なコンティグを構築するのに使用し、（実験の変動により）たとえBACアドレスが各個々の断片配列から得ることができなくとも、3次元タグ付けシステムを利用することで大部分のBACに対する直接的な同定を可能にすることである。しかしながら、同じBAC由来の断片からのタグの組合せによってBACアドレスが提供される。

20

【0091】

このように、配列ベースのBACコンティグ作製に由来する情報は、従来のアプローチ（即ちコンティグ+BACアドレス）と同じである。個々のクローンフィンガープリント化アプローチに関するBACアドレスは定義により既知であることが観察される。

【0092】

（実施例2）

配列タグBACマッピングによるハイスループット物理地図作製の手順

シロイヌナズナの第4染色体に対してマッピングし、4つのBACコンティグ（1.8 Mb、1.2 Mb、0.5 Mb及び1.9 Mb）における全体の物理的な広がりが5.4 Mbに及ぶ合計72個のBAC（BAC＝細菌人工染色体）をTAIR及び他のデータベースから選択した。BACライブラリのドナー植物は、コロンビア型のシロイヌナズナである。サイズが70 kb～150 kbに及ぶ72個のBACを36個のBACの2つのグループ、「AB」グループ及び「XY」グループに分けた。2つのグループ内の36個のBACには内部重複はないが、ABグループとXYグループとを組合せたBACは、ABグループ及びXYグループのBACを交互に並べた4つの連続した最小のタイリングパスコンティグにアセンブリすることができる（図2～図5を参照されたい）。

30

【0093】

72個のシロイヌナズナのBAC、ABグループ36個及びXYグループ36個のプーリング戦略

40

【0094】

【表 2】

ABグループ

	B1	B2	B3	B4	B5	B6
A1	F23J03	T30A10	T25P22	T09A04	T05L19	F07L13
A2	T12H20	T22B04	F25E04	T26M18	T04C09	F07K02
A3	F07K19	F16G20	T32A16	T22A06	F06i07	F24A06
A4	F08F16	F28M20	F10N07	F08B04	T16i18	F04i10
A5	T16L01	F17i05	F28A23	T04L20	T12J05	F23E12
A6	F14H08	T19P05	T10C14	F06D23	T03E09	T06O13

10

XYグループ

	Y1	Y2	Y3	Y4	Y5	Y6
X1	T03H13	T08A17	T15G18	F17A08	F28M11	F24G24
X2	T04F09	F25i24	F08L21	T05C23	F16J13	T01P17
X3	T12H17	F21P08	F09D16	T19F09	F22K18	F13M23
X4	T30C03	F03L17	F11C18	F10M06	F04D11	F26P21
X5	F17M05	T09O24	T04G07	F10M10	F11i11	T04K12
X6	F05M05	T19K04	F23E13	T02G10	F07O06	T08H13

20

T=TAMU BACライブラリ - 12.5 μ gクロラムフェニコール/mlF=IGF BACライブラリ - 50 μ gカナマイシン/ml

【0095】

クロラムフェニコール (TAMU BACクローン) 又はカナマイシン (IGF BACクローン) を含む標準TY培地 200 μ l で個々のクローンとして72個のBACを一晚培養した。すべてのクローンを6 \times 6フォーマットで培養し、プーリング手順を容易にした。翌朝、1つのグループ当たり12個のプールを作製するように、2次元 (6 \times 6) で液体培養物をプールした。各プールは、培養BACを有する培地600 μ l (個々のBAC当たり100 μ l) を含有していた。Sambrook et al. (2001) による標準的なアルカリミニプレップ法に従って24個すべてのBACプールからDNAを単離した。

30

【0096】

各BACプールのDNA 50 ngを制限酵素EcoRI及びMseIで消化し、続いてVos et al. (1995) によって記載された標準AFLP法に従って、EcoRI及びMseIのAFLPアダプタをライゲートした。制限/ライゲーションミックスをミリQ水で10倍に希釈し、5 μ lを増幅工程に使用した。各プールがプール特異的な4つのヌクレオチドアドレス配列でタグ付けされるように、増幅工程で使用したプライマーを4つのヌクレオチド認識配列を用いて設計した。この認識配列は、個々のBAC座標に対するすべての配列のデコンボリューションを容易にするのに必要である。

40

【0097】

使用したEcoRI + 0プライマー及びMseI + 0プライマーの両方は、5'認識配列を担持するアダプタ適合性の5'リン酸化プライマーであり、各プール座標で異なる (図4を参照されたい)。5'リン酸化は、ピロシーケンシングアダプタのライゲーションに必要である。94 (30秒)、56 (60秒)、72 (60秒)のプロファイル30サイクルで増幅を行った。増幅後、産物をアガロースゲル上で確認し (図4)、各グループの増幅したプール産物12個をグループプール (AB c q, XY) にプールするとともに定量した。すぐに各グループプールのDNA 5 μ gを454 Life Sciencesのシーケンシングのためにさらなる調製工程で処理した。Margulies et al. (2005) に従ってGS20プラットフォームで454 Life Sciencesのピロシーケンシングを行った。

50

【 0 0 9 8 】

B A Cコンティグのデータセット及びアセンブリの解析

G S 2 0 ピロシークエンシング機によって得られる D N A 配列の読み取り表は、3つの工程で解析した：

工程 1) プール試料コードから成る最初の 4 つのヌクレオチドを同定し、対応するプール標識を割り当てた。コードが未知であった場合、このセットから、この読み取りを除いた。

【 0 0 9 9 】

工程 2) 次のプライマー配列を含有する 1 6 個又は 1 7 個のヌクレオチド (制限酵素に依存する) を同定した。プライマー配列と 1 0 0 % 同一であった場合、読み取りを承認して、データセットに加え、そうでなければ除いた。

10

【 0 1 0 0 】

工程 3) 工程 2 のすべての読み取りをプライマー配列後の 1 4 個のヌクレオチドに対してトリミングした。

【 0 1 0 1 】

その後、正確にトリミングした配列の読み取りをすべてグループ化した。1 0 0 % 同一な読み取りをすべて同定し、それらに対応するプールに割り当てた。各一意の読み取りグループを「タグ」と呼ぶ。ちょうど 2 つのプールで見出されたタグ (X 座標に関するタグ及び Y 座標に関するタグの両方) を特異的な B A C と結び付けた。この手順は デコンボリユーション と呼ぶ。

20

【 0 1 0 2 】

両方の B A C グループに関して、デコンボリユーションした B A C に対する一意のタグをすべて列挙した。1 つ又は複数の共通のタグを有する B A C 対を同定した。その後、B A C コンティグを表 3 に示されるようにアセンブリすることができた。

【 0 1 0 3 】

【表 3】

表 3. BAC対 (例えばX1Y1及びA1B1) 間で共通で、且つ各プールにおいて少なくとも2回起こるすべての配列タグに対するBACリンク。コンティグに数字を付ける。

BACリンク	タグ数	コンティグ	BACリンク	タグ数	コンティグ
X1Y1_A1B1	8	コンティグ1	X4Y2_A4B1	12	コンティグ4
X1Y2_A1B1	18	コンティグ1	X4Y2_A4B2	4	コンティグ4
X1Y2_A1B2	6	コンティグ1	X4Y3_A4B2	5	コンティグ4
X1Y3_A1B2	19	コンティグ1	X4Y3_A4B3	20	コンティグ4
X1Y3_A1B3	3	コンティグ1	X4Y4_A4B3	5	コンティグ4
X1Y4_A1B3	10	コンティグ1	X4Y4_A4B4	11	コンティグ4
X1Y4_A1B5	10	コンティグ1	X4Y5_A4B5	9	コンティグ5
X1Y5_A1B4	16	コンティグ1	X4Y6_A4B5	7	コンティグ5
X1Y5_A1B5	12	コンティグ1	X4Y6_A4B6	6	コンティグ5
X1Y6_A1B4	13	コンティグ1	X5Y1_A5B1	6	コンティグ6
X1Y6_A1B6	4	コンティグ1	X5Y2_A5B1	5	コンティグ6
X2Y1_A1B6	1	コンティグ1	X5Y2_A5B2	28	コンティグ6
X2Y1_A2B1	3	コンティグ1	X5Y3_A5B2	4	コンティグ6
X2Y2_A2B1	4	コンティグ1	X5Y3_A5B3	26	コンティグ6
X2Y2_A2B2	2	コンティグ1	X5Y4_A5B4	4	コンティグ7
X2Y3_A2B2	5	コンティグ1	X5Y5_A5B4	3	コンティグ7
X2Y4_A2B3	4	コンティグ2	X5Y5_A5B5	1	コンティグ7
X2Y4_A2B4	2	コンティグ2	X5Y6_A5B5	16	コンティグ7
X2Y5_A2B4	1	コンティグ2	X5Y6_A5B6	19	コンティグ7
X2Y5_A2B5	1	コンティグ2	X6Y1_A5B6	7	コンティグ7
X2Y6_A2B5	4	コンティグ2	X6Y1_A6B1	14	コンティグ7
X3Y1_A2B6	3	コンティグ3	X6Y2_A6B1	3	コンティグ7
X3Y1_A3B1	5	コンティグ3	X6Y2_A6B2	14	コンティグ7
X3Y2_A3B1	4	コンティグ3	X6Y3_A6B2	14	コンティグ7
X3Y2_A3B2	2	コンティグ3	X6Y3_A6B3	8	コンティグ7
X3Y3_A3B2	1	コンティグ3	X6Y4_A6B3	14	コンティグ7
X3Y3_A3B3	5	コンティグ3	X6Y5_A6B5	13	コンティグ8
X3Y4_A3B3	15	コンティグ3	X6Y6_A6B5	8	コンティグ8
X3Y4_A3B4	1	コンティグ3	X6Y6_A6B6	14	コンティグ8
X3Y5_A3B4	2	コンティグ3			
X3Y5_A3B5	13	コンティグ3			
X3Y6_A3B5	7	コンティグ3			
X3Y6_A3B6	7	コンティグ3			
X4Y1_A3B6	10	コンティグ3			

10

20

30

【0104】

1.8 Mb、1.2 Mb、0.5 Mb及び1.9 Mbの4つのBACの最小タイリングパスが、個々のBACに対する配列タグのデコンボリューション後に直接的な方法で再アセンブリすることができることが実証された(表1及び図4)。作製したGS20タグと、コンピュータ内で予測した断片との比較は、EcoRI/MseI断片の70%~80%がシーケンシングされたことを示した。したがって、4つのBACコンティグを再アセンブリする際に、2つのBAC間で幾つかの小さい物理的重複は検出することができなかった。

40

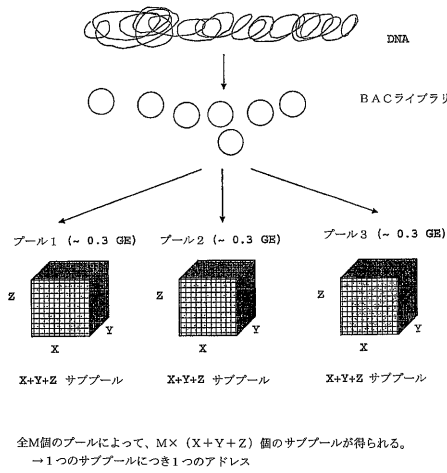
【0105】

短い読み取り(14bp)が、BACタイリングパスを再アセンブリするには十分であるということは、短い読み取り長を有するハイスループットシーケンシングプラットフォーム(例えばIlluminaのゲノム分析器及びSOLiD(ABI))が、提唱された方法

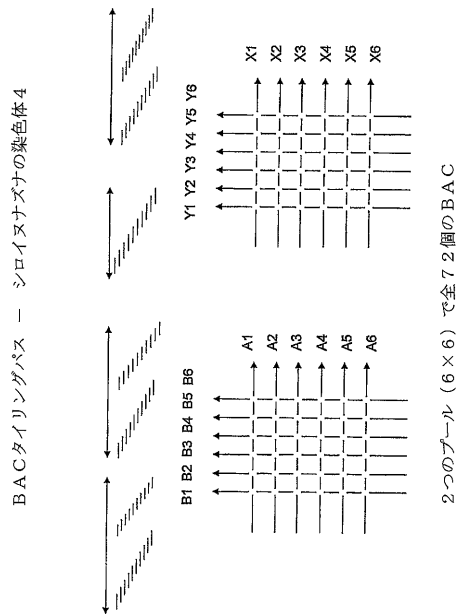
50

に従って、ハイスループットな物理地図のアセンブリを可能にすることを示す。

【 図 1 】

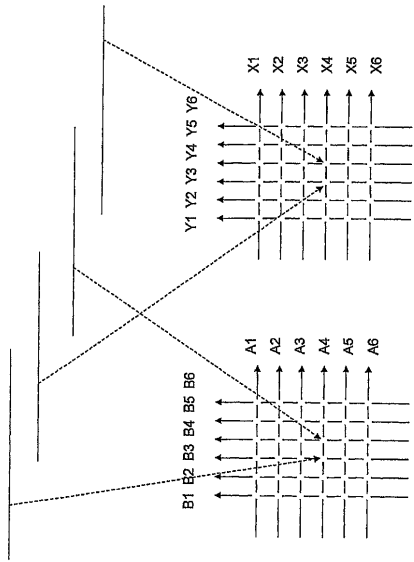


【 図 2 】



【 図 3 】

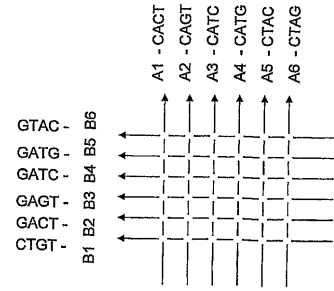
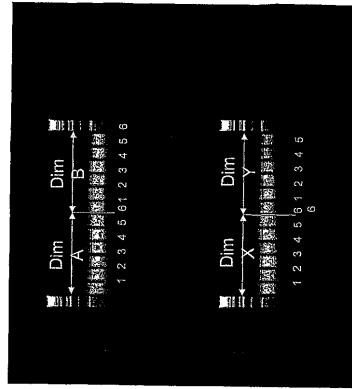
BACタイリングパス - シロイヌナズナの染色体4



【 図 4 】

BACプールのDNAの増幅

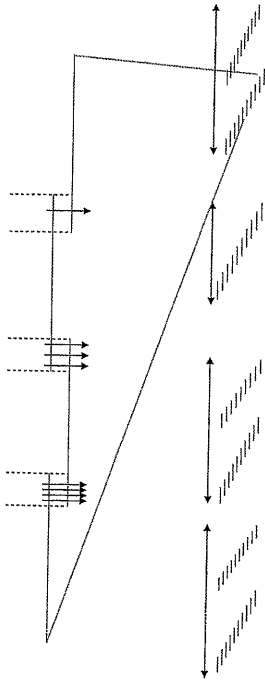
各プール盛盤に対する一意のアドレス



【 図 5 】

再アセンブリした最小タイリングパス

EcoRIタグのみの使用



フロントページの続き

(72)発明者 イェッセ, タコ ベテル

オランダ王国, エヌエル - 6 7 0 2 セーハー ワーゲニンゲン, トルベッケストラート 3 7 2

Fターム(参考) 4B063 QA01 QA13 QA18 QQ42 QQ52 QR14 QR32 QR35 QR55 QR62

QS25