

①⑨ RÉPUBLIQUE FRANÇAISE  
INSTITUT NATIONAL  
DE LA PROPRIÉTÉ INDUSTRIELLE  
PARIS

①① N° de publication : 2 772 048

(à n'utiliser que pour les  
commandes de reproduction)

②① N° d'enregistrement national : 97 15319

⑤① Int Cl<sup>6</sup> : C 12 Q 1/68, C 12 N 5/10, 15/79

①②

DEMANDE DE BREVET D'INVENTION

A1

②② Date de dépôt : 04.12.97.

③① Priorité :

④③ Date de mise à la disposition du public de la  
demande : 11.06.99 Bulletin 99/23.

⑤⑥ Liste des documents cités dans le rapport de  
recherche préliminaire : *Se reporter à la fin du  
présent fascicule*

⑥① Références à d'autres documents nationaux  
apparentés :

⑦① Demandeur(s) : *BIOMETHODES Societe a responsa-  
bilite limitee* — FR.

⑦② Inventeur(s) : DELCOURT MARC.

⑦③ Titulaire(s) :

⑦④ Mandataire(s) : BREESE MAJEROWICZ.

⑤④ PROCÉDE DE CLONAGE PAR DIGESTION MULTIPLE, VECTEURS POUR SA MISE EN OEUVRE ET SES  
APPLICATIONS.

⑤⑦ La présente invention concerne une nouvelle métho-  
de de clonage fondée sur l'utilisation systématique, combi-  
née, et préparative des sites de restriction présents sur les  
inserts constituant une banque d'ADN complémentaire ou  
génomique. L'invention concerne également les applica-  
tions de ce procédé ainsi que les vecteurs pour sa mise en  
oeuvre et les banques obtenues.

FR 2 772 048 - A1



PROCÉDÉ DE CLONAGE PAR DIGESTION MULTIPLE,  
VECTEURS POUR SA MISE EN OEUVRE ET SES APPLICATIONS.

5 La présente invention concerne le domaine de  
la biologie moléculaire et se rapporte plus  
particulièrement au clonage de gènes.

Le clonage de gènes est un domaine en pleine  
expansion qui vise notamment à associer fonctions et  
gènes. Ce domaine se développe principalement selon deux  
10 grands axes, d'une part la biologie moléculaire inverse,  
qui consiste à identifier la fonction de séquences en  
leur faisant passer des batteries de tests, et d'autre  
part la biologie moléculaire directe, qui consiste à  
trouver la séquence responsable d'une activité observée.

15 L'avènement de la technique dite "PCR" pour  
"Polymerase Chain Reaction" a considérablement simplifié  
le clonage par homologie. Aussi un grand nombre de  
laboratoires s'emploient-ils à cloner de nouveaux gènes  
en utilisant cette technique qui relève alors de la  
20 biologie moléculaire inverse. Ces travaux sont  
relativement simples et efficaces, mais demandent  
néanmoins un grand investissement en temps.

A l'inverse, le clonage stratégique n'a pas  
fait de progrès majeur, et la biologie moléculaire  
25 attend toujours un moyen simple, systématique et rapide  
de cloner des gènes de cette façon.

La présente invention vise précisément à  
offrir une nouvelle méthode de clonage désignée aussi  
ci-après DMD pour "Digestion Multiple Différentielle",  
30 qui repose sur l'utilisation systématique, combinée, et  
préparative des sites de restriction présents sur les  
inserts constituant une banque d'ADN complémentaire ou  
génomique.

35 Outre son application dans le clonage  
stratégique, le procédé de clonage de l'invention permet  
également de faciliter grandement le clonage par

homologie, et offre une solution alternative et/ou complémentaire aux techniques actuelles que sont la PCR ou le balayage d'une banque avec une sonde.

La présente invention trouve également des applications dans le domaine fastidieux du séquençage quand celui-ci n'a qu'une activité d'identification des inserts, ainsi que dans l'étude du polymorphisme humain, notamment dans le cadre de la recherche des prédispositions génétiques.

On désignera ci-après par Caractéristique Enzymatique (CE) la résistance, notée "r" ou la sensibilité, notée "s" d'un insert à une enzyme de restriction. Ce qui signifie :

- qu'un insert qualifié "s" pour une enzyme de restriction donnée contient le site de clivage de cette enzyme, et

- qu'un insert qualifié "r" pour une enzyme de restriction ne contient pas le site de clivage de cette enzyme.

La Caractéristique Enzymatique Multiple (CEM) est alors l'ensemble des CE obtenues pour plusieurs enzymes. On peut alors représenter la CEM d'un insert contenant un site Eco, un site Bam, deux sites Sca, mais pas de site Hind ni Stu, de la façon suivante: Eco<sup>s</sup>Bam<sup>s</sup>Sca<sup>s</sup>Hind<sup>r</sup>Stu<sup>r</sup>.

Le procédé de l'invention propose donc d'associer à l'insert recherché au sein d'une banque sa caractéristique enzymatique multiple (CEM). La recherche dudit insert peut être réalisée par tout moyen connu de l'homme du métier, comme sa capacité à hybrider à une sonde nucléique, l'activité enzymatique de son produit protéique, l'expression d'une protéine que l'on pourra détecter, etc...

Le procédé de clonage selon l'invention est fondé sur la démonstration que quand le nombre d'enzymes est suffisamment grand, chaque insert de la banque possède une CEM originale, et qu'en conséquence,  
5 l'invention offre une stratégie simple de clonage des gènes à partir de leur CEM. Le procédé selon l'invention repose donc sur le balayage de banques d'ADN en utilisant de façon combinatoire la répartition des sites de restriction sur les inserts constituant cette banque.

10

Le principe du procédé selon l'invention repose également sur l'utilisation d'un vecteur de type nouveau car substantiellement dépourvu de tous les sites de clivage par des enzymes de restrictions pour n'en  
15 garder que trois :

- un site A nécessaire pour construire la banque, et

- deux sites B, identiques et flanquant le site A, utile pour sous-cloner le gène une fois celui-ci  
20 identifié et cloné.

L'ensemble pourra être désigné ci-après "trilinker", dont la fabrication schématique est donnée à la figure 1 en annexe.

Avantageusement les sites B sont des sites  
25 octonucléotidiques, de façon à minimiser le risque que des sites B soient présents dans les inserts clonés. Il sera ainsi possible de sous-cloner facilement en un seul morceau.

Environ 100 enzymes de restriction à site  
30 hexanucléotidique ont été découvertes à ce jour. 70 d'entre-elles ont un site de reconnaissance de type palyndrome continu ou discontinu.

Un vecteur de l'invention, avantageusement  
35 un plasmide, ne contient plus, mis à part ces trois sites, de sites de restriction hexa ou penta-

nucléotidiques correspondant à des enzymes de restriction déjà identifiées ou qui le seront dans le futur. On entend donc par "substantiellement", que cette destruction peut être partielle en ce sens qu'elle ne  
5 concerne que certains des sites connus; seules les enzymes correpondantes seront alors utilisées dans le procédé de l'invention. Dans l'exposé de l'invention qui suit, on considère, comme indiqué précédemment que 60 à 70 types de sites ont été détruits.

10 Un vecteur de l'invention peut être construit à partir d'un plasmide déjà existant et possédant toutes les fonctions nécessaires pour permettre d'y créer et d'y manipuler une banque d'ADNC ou génomique. Il est cependant essentiel, dans le  
15 procédé de l'invention, que la banque ne contienne plus de vecteur seul refermé sur lui-même. Il est donc avantageux que le vecteur de l'invention contienne un système qui élimine tout vecteur seul refermé sur lui-même, comme un gène suicide, un système de disruption de  
20 la proximité d'un promoteur  $\lambda$  ou tout autre système connu de l'homme du métier.

La technique utilisée, dans le cadre de l'invention pour détruire tous les sites de restriction  
25 est une des techniques de mutagenèse dirigée simple ou multiple déjà décrites dans l'art antérieur, ou tout autre technique connue par l'homme du métier, comme le remplacement de segments du plasmide par oligonucléotides. On obtient ainsi un vecteur résistant  
30 à 70 enzymes de restriction, numérotées de I à LXX, et sensible à 2, désignées ci-avant A et B. L'idée de détruire simultanément la majorité des sites de restriction présents sur un plasmide a été envisagée par D. H. Jones et al. (BioTechniques 1994, 16, 4 : 694)  
35 mais dans un contexte tout différent. En effet, cet article décrit la destruction par mutagenèse multiple de

31 des 37 sites d'un très petit vecteur, de façon à créer un nouvel outil facilitant certaines manipulation de l'ADN.

5 Le procédé de l'invention est par ailleurs basé sur la création d'une banque d'ADN, laquelle selon l'application envisagée contient de 1 à  $10^8$  et de préférence de l'ordre de  $10^5$  à  $4 \cdot 10^6$  fragments différents de l'ordre de 0,1 kb à 5 kb chacun, et de  
10 préférence, selon les applications du procédé de l'invention, de 1 à 2 kb. Dans le mode particulier de mise en oeuvre du procédé de l'invention décrit à l'exemple 6 ci-après concernant l'étude du polymorphisme, la banque peut ne comprendre qu'un seul  
15 fragment.

Ainsi, pour l'application du procédé de l'invention à la fabrication de banques d'expression ou au clonage par homologie, il a été préparé, comme représenté schématiquement à la figure 2 en annexe, une  
20 banque d'ADNc de  $10^5$  fragments différents de 1 kb chacun.

Toutefois, ce modèle constitue une approximation et est donc légèrement faux, car la taille des fragments est hétérogène. La taille moyenne des  
25 inserts étant sous-estimée et la taille de la banque étant surestimée, ce modèle a été choisi afin de disposer d'une base simple (homogénéité de la taille des fragments) et d'un système d'étude qui désavantage l'analyse (taille de la banque surestimée, taille des  
30 fragments sous-estimée) de façon à ce que le procédé de l'invention soit reproductible dans tous les cas.

Pour l'application du procédé de l'invention au Southern Blot d'identification et à l'étude du polymorphisme, il a été préparé, comme représenté  
35 schématiquement à la figure 3 en annexe, une banque d'ADN génomique de  $4 \cdot 10^6$  fragments différents de 1kb

chacun. Cette banque a été obtenue par l'utilisation d'une enzyme correspondant à un site de fréquence 1/1024 (de type AT(ACGT)(TGCA)TA).

5 Tous types de banques d'ADN, tels que celles obtenus par amplification au hasard par PCR utilisant des oligonucléotides dégénérés ou non, entrent dans le cadre de la présente invention.

10 En conséquence, l'invention a pour objet un procédé de clonage d'un fragment d'acide nucléique comprenant les étapes suivantes :

- on prépare une banque d'ADN susceptible de contenir ledit fragment,

15 - on crible ladite banque en utilisant de façon combinatoire au moins 10 et de préférence 60 à 70 enzymes de restriction, pour isoler par tout moyen approprié le clone contenant ledit fragment.

20 Dans ce procédé, la préparation de la banque d'ADN susceptible de contenir le fragment d'acide nucléique, consiste à insérer chacun des fragments d'ADN d'un échantillon dans un vecteur dépourvu de tout site de clivage par des enzymes de restriction à l'exception :

25 - d'un site pour la construction de la banque, et

- éventuellement de deux autres sites identiques, différents du premier site et le flanquant, utiles pour sous-cloner la séquence d'acide nucléique une fois celle-ci identifiée et clonée.

30 Le procédé comprend plus particulièrement les étapes suivantes :

35 a) on prépare une banque d'ADN susceptible de contenir ledit fragment d'acide nucléique, consistant à insérer chacun des fragments d'ADN d'un échantillon dans un vecteur substantiellement dépourvu de tout site

de clivage par des enzymes de restriction à l'exception :

- d'un site pour la construction de la banque, et

5                   - éventuellement de deux autres sites identiques et différent du premier site et le flanquant, utile pour sous-cloner la séquence d'acide nucléique une fois celle-ci identifiée et clonée.

10                   L'échantillon à partir duquel sont issus les fragments constituant la banque peut être toute cellule eucaryote (mammifère, plante, levure, etc...), ou tout organisme procaryote (virus, bactérie, etc...). Il peut s'agir d'ADN génomique, d'ADNc, de fragments d'amplification par PCR ou de tout autre banque d'ADN

15                   susceptible d'être préparée par l'homme du métier.

                  b) On réalise la digestion en parallèle de la banque par plusieurs enzymes de restriction, au moins 10 et de préférence environ 60 à 70 enzymes de restriction de façon à obtenir autant de banques mono-digérés que d'enzymes utilisées.

20

                  c) On transfecte indépendamment les banques mono-digérées dans des hôtes cellulaires appropriés de façon à obtenir des lots correspondant d'hôtes cellulaires.

25                   d) On teste par tout moyen approprié chacun des lots obtenus à l'étape (c) pour évaluer l'intégrité de la séquence d'acide nucléique à cloner et ainsi établir sa CEM.

30                   En effet, si la digestion par une enzyme désignée "I" n'altère pas l'intégrité de l'insert présent dans la banque, celui-ci est considéré I<sup>r</sup>, au contraire, si elle l'altère, celui-ci est considéré I<sup>s</sup>.

                  Les étapes (a) à (d) ci-dessus permettent l'analyse du fragment à cloner selon l'invention.

Le procédé de l'invention comprend en outre les étapes suivantes permettant de purifier le fragment à cloner :

5 e) On reprend la banque totale de l'étape (a) et on la digère sensiblement simultanément par les enzymes qui n'affectent pas l'intégrité du fragment à cloner. Pour lesquelles il a donc été considéré "r".

10 f) On isole le clone résistant contenant le fragment d'acide nucléique à cloner par tout moyen approprié et éventuellement on le sous-clone en utilisant les deux sites ménagés dans le vecteur à cet effet.

15 g) On séquence éventuellement le fragment d'acide nucléique à cloner.

Eventuellement, par sécurité on peut effectuer des étapes de vérification entre les étape (a) et (b) d'une part, et (e) et (f) d'autre part consistant :

20 a') A vérifier la présence du fragment d'acide nucléique à cloner dans la banque en transfectant dans un hôte cellulaire qui n'a pas ledit fragment et en testant par tout moyen approprié la présence du fragment dans ledit hôte. Avantageusement, on utilise dans cette étape (a') des cellules COS, classiquement mise en oeuvre pour les transfections.

25 e') A transformer la banque multi-digérée de l'étape (e) dans des hôtes compétents de façon à vérifier la nature des fragments clonés. Par exemple, cette étape consiste à étaler sur petri et vérifier par minipréparations d'ADN plasmidique (miniprep) que les inserts sont bien sensibles aux enzymes notées "s" lors de l'établissement de la CEM.

35 Avantageusement, la banque d'ADN préparée à l'étape (a) contient de 1 à  $10^8$  et de préférence de

l'ordre de  $10^5$  à  $4.10^6$  fragments différents de l'ordre de 0,1 kb à 5 kb chacun, et de préférence de l'ordre de 1 à 2 kb.

5 Dans le procédé de l'invention, on préfère que les deux sites de sous-clonage soient des sites octonucléotidiques, de façon à minimiser le risque que des sites B soient présents dans les inserts clonés.

10 Il est essentiel, dans le procédé de l'invention, que la banque ne contienne plus de vecteur seul refermé sur lui-même. Il est donc avantageux que le vecteur de l'invention contienne un système qui élimine tout vecteur seul refermé sur lui-même, comme un gène suicide, un système de disruption de la proximité d'un promoteur  $\lambda$  ou tout autre système connu de l'homme du  
15 métier.

Les tests réalisés aux étapes (d) et (a') pour vérifier l'intégrité de la séquence d'acide nucléique à cloner peuvent être tout moyen de mise en  
20 évidence soit de la séquence elle-même, comme une sonde, soit de la protéine codée par ladite séquence, tel qu'un ligand comme un anticorps, ou encore de l'activité de cette protéine, comme un marqueur enzymatique, qui peuvent être détectées par tout moyen connu de l'homme  
25 du métier, comme un marquage fluorescent ou radioactif.

Les applications du procédé de clonage selon l'invention sont très nombreux et l'on peut citer notamment celles détaillées dans les exemples ci-après :

- 30 - Le clonage d'un gène par banque d'expression.
- Le clonage par homologie.
  - Le Southern Blot d'identification dénommé par l'Inventeur "identiblot".
  - 35 - L'étude du polymorphisme.

Aussi, on entendra dans ce qui suit par gène, insert, séquence, le fragment d'acide nucléique à cloner selon le procédé de l'invention décrit ci-dessus.

5 Un procédé de clonage d'un gène par banque d'expression selon l'invention comprend les étapes suivantes :

10 a) On prépare une banque d'ADNc susceptible de contenir ledit gène en insérant dans un vecteur substantiellement dépourvu de tout site de clivage par des enzymes de restriction à l'exception d'un site pour la construction de la banque et de deux autres sites identiques et différent du premier site et le flanquant, utile pour sous-cloner le gène une fois celui-ci  
15 identifié et cloné.

b) On vérifie la présence du gène recherché dans la banque en transfectant dans une lignée cellulaire qui n'a pas l'activité ou le phénotype recherché et en mesurant sa restauration de façon à  
20 distinguer les cellules transfectées des cellules non transfectées.

c) On digère la banque indépendamment par au moins 10 et de préférence environ 60 à 70 enzymes de restriction.

25 d) On transfecte indépendamment les banques mono-digérées de l'étape (c).

e) On teste par tout moyen approprié chacun des lots obtenus à l'étape (d) pour la présence de l'activité associée au gène à cloner et on évalue  
30 l'intégrité de la séquence dudit gène afin d'établir la CEM de l'activité associée audit gène.

On entend par activité associée au gène la détection par tout moyen de la protéine codée par ledit gène ou de l'activité de cette protéine quelle qu'elle  
35 soit (ligand, enzyme, induction de tumeur, etc...).

f) On reprend la banque totale de l'étape (a) et on la digère sensiblement simultanément par les environ 50 à 55 enzymes qui, en moyenne, n'affectent pas l'activité mesurée à l'étape (e).

5 g) On transforme la banque multi-digérée dans des bactéries compétentes.

h) On sous-clone en utilisant l'enzyme correspondant au site de sous-clonage ménagés dans le vecteur, puis éventuellement on séquence le gène.

10

L'invention concerne également, les banques mono- ou multi-digérées obtenues dans le procédé de l'invention et notamment aux étapes (c) et (f) des procédés ci-dessus, ainsi que les supports, tels que des tubes, des membranes, des plaques, des gels, etc... qui peuvent contenir lesdites banques, leur produit d'expression ou des hôtes les contenant.

15

L'invention se rapporte aussi aux hôtes cellulaires ou bactériens contenant les banques ci-dessus.

20

Un procédé de clonage par homologie selon l'invention comprend les étapes suivantes :

25

a) On prépare une banque d'ADNc comme décrit à l'étape (a) précédente.

b) On digère la banque indépendamment par chacune d'au moins 10 et de préférence environ 70 enzymes de restriction.

30

c) On transforme les produits de la digestion de l'étape (b) dans des bactéries compétentes.

d) On fait pousser les bactéries transformées en milieu sélectif, de façon à produire des banques digérées débarrassées des produits clivés.

35

e) On clive séparément chacune de ces banques par l'enzyme correspondant au site de sous-clonage ménagé dans le vecteur et l'on dépose séparément

chacun de ces produits de digestion dans un puits de gel d'agarose ou d'acrylamide.

5 f) On fait migrer les produits de digestion de l'étape (e), puis l'on transfère sur membrane par exemple de nitro-cellulose, et l'on hybride avec une sonde spécifique du gène à cloner par homologie, ou bien l'on dépose directement les produits de l'étape (d) sur une membrane de nitrocellulose.

10 g) On analyse la CEM du signal.  
e) On fait les multi-digestions correspondantes, de façon à ce que le seul clone résistant soit le vecteur portant le gène à cloner.

15 Le procédé de clonage par homologie ci-dessus peut être appliqué à l'identification :

a) d'allèles de différentes souches d'animaux de la même espèce, ou de différents individus chez l'homme (souvent très homologues).

20 b) d'équivalents géniques présents chez différentes espèces (modérément homologues).

c) d'épissages alternatifs d'un même gène au sein d'un même tissu ou entre différents tissus (homologie totale par sections).

25 d) de différents membres d'une famille génique, distribuée au sein d'un même tissus ou dans différents tissus. (homologie imprévisible, souvent très forte dans certains domaines).

30 La version simplifiée de ce procédé consiste à effectuer des dots au lieu de Blots, c'est-à-dire à s'abstenir de cliver les banques mono-digérées par l'enzyme correspondant au site de sous-clonage ménagé dans le vecteur, et de les déposer directement en un point sur une membrane de nitro-cellulose.

35 Un procédé de Southern Blot d'identification d'inserts selon l'invention permet d'identifier un

fragment d'ADN sans avoir à le séquencer même partiellement. Ce procédé comprend les étapes suivantes :

5 (a) on prépare une banque d'ADN susceptible de contenir ledit insert, consistant à insérer chacun des fragments d'ADN d'un échantillon dans un vecteur substantiellement dépourvu de tout site de clivage par des enzymes de restriction à l'exception d'un premier site pour la construction de la banque, et de deux 10 autres sites identiques et différent du premier site et le flanquant, utile pour sous-cloner la séquence d'acide nucléique une fois celle-ci identifiée et clonée.

15 b) On digère cette banque par chacune d'au moins 10 et de préférence environ 60 à 70 enzymes de restriction.

c) On transforme les banques monodigérés obtenues à l'étape (b) ci-dessus dans des bactéries compétentes, ou des hôtes équivalents.

20 d) On fait pousser les bactéries en milieu sélectif, de façon à produire des banques monodigérées débarassées des produits clivés.

25 e) On clive séparément chacune de ces banques par l'enzyme correspondant aux deux autres sites identiques et différent du premier site et le flanquant, et on dépose les produits de digestion dans les puits d'un gel d'agarose ou d'acrylamide.

f) On fait migrer ce gel et on le transfère sur une membrane par exemple de nitrocellulose.

30 g) On utilise les inserts à identifier comme sondes marquées, soit un par un, soit plusieurs par plusieurs.

35 h) On teste par tout moyen approprié chacun des lots obtenus à l'étape (g) pour associer les inserts à identifier à une CEM. Cette CEM correspond à l'action des enzymes de l'étape (b).

Un procédé d'étude du polymorphisme selon l'invention est identique au procédé de Southern Blot d'identification d'inserts mais est caractérisé en ce que :

5                   - La banque d'ADN génomique de l'étape (a) est issue du sujet étudié, par exemple d'un malade, ou de sujets de la famille du malade.

                  - Les inserts utilisés comme sondes sont des marqueurs de polymorphisme déjà décrits.

10                   Le procédé d'étude du polymorphisme selon l'invention trouve son application soit dans le cadre de la recherche de marqueurs du polymorphisme associés à une pathologie, soit dans le cadre d'un diagnostic de cette pathologie.

15                   Une variante du procédé de clonage précédent pour l'étude du polymorphisme d'un individu comprend les étapes suivantes :

20                   a) On définit la CEM de chacun des marqueurs connus de façon à permettre leur identification.

                  b) On constitue une banque d'ADN génomique du sujet étudié consistant à insérer chacun des fragments d'ADN d'un échantillon dans un vecteur substantiellement dépourvu de tout site de clivage par des enzymes de restriction à l'exception :

25                   - d'un site pour la construction de la banque, et

                  - éventuellement de deux autres sites identiques et différent du premier site et le flanquant, utile pour sous-cloner la séquence d'acide nucléique une fois celle-ci identifiée et clonée.

30                   c) On effectue la digestion de la banque par des batteries d'enzymes correspondant aux CEM attribuées aux marqueurs étudiés.

35                   d) On transforme les banques multidigérées dans des bactéries compétentes.

e) On cultive ces bactéries en milieu liquide ou solide contenant l'agent sélectif du plasmide. Si l'allèle recherché existe, les bactéries poussent, par contre, si cet allèle n'existe pas, elles ne poussent pas, et l'on dispose ainsi d'un profil pour chaque allèle.

Avantageusement, à l'étape (b), on prépare une banque dont les fragments ont une longueur moyenne de 1000 à 4000 et de préférence de 2000 nucléotides.

Un mode de réalisation de la variante ci-dessus permet également la détection des nombreux allèles de segments polymorphes, comme par exemple la gp120 des virus HIV ou la p53 d'oncogènes cellulaires, dans le cadre de travaux de recherche ou d'un diagnostic. Le segment concerné est avantageusement amplifié par PCR et cloné dans le plasmide, puis le procédé est identique à celui décrit précédemment et exposé en détail dans l'exemple 6 ci-après.

L'invention a aussi pour objet un mélange d'au moins 10 et de préférence de 60 à 70 enzymes de restriction susceptible d'être utilisé dans un procédé selon l'invention.

L'invention concerne encore les vecteurs susceptibles d'être utilisés dans les procédés de l'invention. Ces vecteurs sont caractérisés par le fait qu'ils sont substantiellement dépourvus de tout site de clivage par des enzymes de restriction à l'exception :

- d'un site pour la construction de la banque, et

- éventuellement de deux autres sites identiques, différents du premier site et le flanquant, utiles pour sous-cloner une séquence d'acide nucléique clonée dans ledit vecteur.

Un tel vecteur est avantageusement un plasmide obtenu par mutagenèse dirigée simple ou multiple ou tout autre technique connue de l'homme du métier.

5 Un vecteur de l'invention porte avantageusement un système qui élimine tout vecteur seul refermé sur lui-même, comme un gène suicide, un système de disruption de la proximité d'un promoteur  $\lambda$  ou tout autre système connu de l'homme du métier.

10 D'autres avantages et caractéristiques du procédé de l'invention apparaîtront dans la description qui suit se rapportant à des exemples détaillés de mise en oeuvre du procédé de l'invention dans diverses applications qui ne sauraient être considérés comme une limitation quelconque de l'invention.

15 Exemple 1 : La préparation et l'exploitation d'une banque d'expression.

20 Une première application du procédé de l'invention se rapporte à la préparation et à l'exploitation d'une banque d'expression.

25 Cette technique suscite depuis le milieu des années 1980 un grand intérêt. De nombreux gènes, comme ceux codant pour des récepteurs aux cytokines, des marqueurs de surface des lymphocytes, des protéines liant l'ADN, etc... ont été identifiés par cette technique (U. Gubler et al., Annals of the N. Y. Academy of Science 795 : 36-40, 1996 ; D. Pennica et al. PNAS 92(4) : 1142, 1995 ; R. M. O'Brien et al., Biochemical Journal 312 : 17 - 21, 1995). Toutefois, les techniques de clonage par banque d'expression sont toujours fastidieuses et difficilement reproductibles à ce jour.

35 Les applications actuelles des banques d'expression concernent l'identification du gène codant

pour une protéine, dont les moyens de mise en évidence peuvent être regroupé en quatre catégories principales :

- les anticorps,
- les liaisons protéine/protéine autre  
5 qu'anticorps/antigène,
- les oligonucléotides marqués par exemple  
par fluorescence dans le cas où la protéine recherchée  
est une protéine se liant à l'ADN,
- les tests d'activité de la protéine,  
10 quelle qu'elle soit.

La banque contenant le gène codant pour la protéine recherchée peut être transformée dans des bactéries ou des levures, et l'on utilise l'anticorps, la protéine ou l'oligonucléotide marqué pour balayer la  
15 banque à la recherche de la colonie qui l'exprime. Ces systèmes marchent souvent très mal, car les protéines n'ont pas la même conformation et les mêmes modifications post-traductionnelles chez les bactéries ou levures et chez les cellules mammifères.

20 Le problème de la transfection de banques dans les cellules mammifères réside dans le fait que, à la différence des systèmes de bactéries ou de levures, plusieurs plasmides sont intégrés dans chaque cellule. Pour contenir cette difficulté, il est nécessaire  
25 d'utiliser des techniques de fractionnement successif des banques, ou de tris cytofluorimétriques répétés (T. Kitamura et al., PNAS 92 (20) : 9146, 1995 ; D. R. Gehlert et al., Molecular Pharmacology, 49 (2) : 224, 1996). Chacune de ces deux techniques représente un  
30 travail important, très long (plusieurs mois), qui n'est pas toujours conclusif, et qui n'est pas adapté au clonage simultané deux inserts ou plus.

35 Le procédé de l'invention est beaucoup plus simple et moins coûteux que tous ceux de l'art antérieur cité ci-dessus. De plus, il peut être appliqué au

clonage simultané de plusieurs inserts. Ce dernier point revêt une importance majeure lorsqu'on sait que beaucoup de protéines de surface sont composées de plusieurs chaînes et n'atteignent la surface que lorsque toutes les chaînes sont produites, par exemple dans le cas des complexes majeurs d'histocompatibilité. Or aucune des techniques de l'art antérieur basée sur les banques d'expression ne permet d'avoir accès à ces protéines.

5  
10 Un procédé de l'invention pour la constitution de banques d'expression comprend les étapes suivantes :

15 a) On prépare une banque en insérant dans le site A du vecteur l'ADN complémentaire du tissu ou de la lignée cellulaire d'intérêt. La figure 2 en annexe est une représentation schématique de cette étape.

20 b) On vérifie la présence du gène recherché dans la banque en transfectant dans une lignée cellulaire qui n'a pas l'activité ou le phénotype recherché et en mesurant sa restauration, par exemple avec un anticorps, de façon à distinguer les cellules transfectées (+) des cellules non transfectées (-). La figure 4 et 5 en annexe sont une représentation schématique de cette étape.

25 c) Avantageusement, on utilise des cellules COS, classiquement mise en oeuvre pour les transfections.

30 d) On digère la banque indépendamment par chacune des 60 à 70 enzymes de restriction disponibles. On obtient ainsi 60 à 70 tubes. La figure 6 en annexe est une représentation schématique de cette étape.

35 e) On transfecte indépendamment les 60 à 70 banques mono-digérées, et tester la présence de l'activité recherchée dans chacun de ces 60 à 70 lots de cellules transfectées. La figure 7 en annexe est une représentation schématique de cette étape.

e) On établit ainsi la CEM de l'activité recherchée. Si la digestion par l'enzyme I n'altère pas l'activité de l'insert présent dans la banque, on notera  $I^r$ . Si elle l'altère, on notera  $I^s$ .

5 On obtient, dans l'exemple de la figure 7 en annexe, la CEM suivante:  $I^s II^r \dots LXX^r$ .

En moyenne, on estime que  $55 \pm 4$  enzymes sur 70 auront un r, et  $15 \pm 4$  auront un s pour un insert de 1 kb. En effet, la probabilité de clivage par une enzyme de restriction à site hexanucléotidique, à une position donnée prise au hasard, est  $p = 1/4^6 = 1/4096$ . Sur un gène dont la taille est n nucléotides, la probabilité théorique d'avoir 1, 2, 3, ... clivages suit une loi binomiale de probabilité p et de nombre d'évènements n.

10 La probabilité de n'avoir aucun clivage est  $C_n^0 p^n (1-p)^0$ . Dans le cas où  $n = 1000$ , on obtient une probabilité de 78,3%. La probabilité d'avoir un clivage ou plus est d'environ 21,7 %. Le nombre moyen d'enzymes ne clivant pas est donc de  $0,783 \times 70 = 55$ , et l'écart type est

15 proche de 4.

20

f) On reprend la banque totale et on la digère sensiblement simultanément par les 55 enzymes qui n'affectent pas l'activité mesurée qui est associée à "r" pour l'insert recherché. La figure 8 en annexe est

25 une représentation schématique de cette étape.

En pratique, cette digestion ne peut être totalement simultanée pour des raisons de compatibilité de tampons, et l'on est obligé de 2 ou 3 multi-digestions successives correspondant aux deux ou trois

30 tampons choisis.

La probabilité pour un insert pris au hasard d'être clivé par x de ces 55 enzymes suit aussi une loi binomiale, de probabilité 0,783 et de nombre d'évènements 55. La probabilité pour un insert de n'être clivé par aucune des 55 enzymes est donc  $C_{55}^0 \times (0,783)^{55} \times (1 - 0,783)^0 = (0,783)^{55} = 1,4 \cdot 10^{-6}$ .

35

Il ne reste donc, en moyenne, que l'insert recherché, plus  $1,4 \cdot 10^{-6} \times 10^5 = 0,14$  insert parasite.

5 L'utilisation partielle de la CEM, correspondant uniquement aux enzymes associées à un r, est suffisante pour "pêcher" le gène, lequel est déjà pur à plus de 85%.

10 g) On transforme la banque multi-digérée dans des bactéries compétentes. Avantagement on étale sur petri et l'on vérifie par miniprep que les inserts sont bien sensibles aux enzymes notées "s" lors de l'établissement de la CEM. La figure 9 en annexe est une représentation schématique de cette étape.

15 h) On sous-clone en utilisant l'enzyme B dans un vecteur d'étude, tel que Bluescript. Puis avantagement on séquence.

20 Chacune de ces 8 étapes ne demande que très peu de temps, puisque le procédé ci-dessus devrait peut être réalisé en 16 jours dont 10 de travail.

25 Le procédé de l'invention permet notamment de générer des banques d'expression pour de nombreux tissus et lignées, puis en faire les 60 à 70 simples digestions, les transfecter, faire l'extrait cellulaire de ces cellules transfectées, et les déposer sur une membrane comme représenté à la figure 10.

30 Il ne reste alors plus qu'à révéler la membrane par exemple avec un anticorps, afin d'obtenir immédiatement la CEM de l'insert recherché comme représenté sur la figure 11. La banque est alors multi-digérée en fonction de la CEM obtenue, puis on transforme des bactéries compétentes avec ce produit multi-digéré, et le gène est cloné en 3 jours au lieu de plusieurs mois par les techniques de l'art antérieur.

On peut aussi garder ces membranes ad vitam aeternam à 4°C ou congelées de façon à disposer de ces gènes "préclonés".

5 L'exemple ci-dessus est basé sur l'utilisation d'un anticorps spécifique mais d'autres systèmes de révélation du phénotype associé à la CEM peut être utilisé.

Exemple 2 : Le clonage par homologie.

10

Le clonage par homologie est intensivement utilisé par de très nombreux laboratoires de biologie moléculaire pour l'identification :

15 a) d'allèles de différentes souches d'animaux de la même espèce, ou de différents individus chez l'homme (souvent très homologues).

b) d'équivalents géniques présents chez différentes espèces (modérément homologues).

20 c) d'épissages alternatifs d'un même gène au sein d'un même tissu ou entre différents tissus (homologie totale par sections).

25 d) de différents membres d'une famille génique, distribuée au sein d'un même tissu ou dans différents tissus. (homologie imprévisible, souvent très forte dans certains domaines).

Les stratégies utilisées dans l'art antérieur sont essentiellement les deux suivantes (M. Parmentier et al., Nature, 355 : 453, 1992) :

30 - La PCR par homologie et ses dérivés, lesquels soulèvent le problème du choix des amorces alors que l'on ne connaît pas les parties conservées. La fenêtre entre le bruit de fond aspécifique et les vraies amplifications homologues est étroite. De plus, la  
35 partie amplifiée ne représente le plus souvent pas tout le gène et on est alors obligé d'aller chercher les

morceaux manquants, et notamment la partie en 5', par des techniques laborieuses comme la PCR ancrée.

5 - L'hybridation de banques existant actuellement. Cette méthode est efficace, mais nécessite beaucoup de travail. En outre, elle ne s'applique qu'à l'identification d'allèles (a) ou d'équivalents géniques (b). Pour l'identification des épissages alternatifs (c) ou des différents membres d'une famille génique (d), on reclone toujours l'ADNc majoritaire, donc dans la  
10 plupart des cas, le gène dont on dispose déjà.

Le procédé de l'invention permet une fois qu'on a cloné un gène, de produire une sonde qui s'hybride sur des gènes homologues. On utilise donc  
15 cette sonde pour pêcher ses homologues par hybridation à faible stringence.

Le procédé de clonage par homologie selon l'invention comprend les étapes suivantes :

20 a) On prépare une banque d'ADNc comme décrit à l'étape (a) de l'exemple 1.

b) On digère la banque indépendamment par chacune des 70 enzymes de restriction comme décrit à l'étape (c) de l'exemple 1.

25 c) On transforme ces 60 à 70 digestions dans des bactéries compétentes. La figure 12 en annexe est une représentation schématique de cette étape.

30 d) On fait pousser les bactéries transformées en milieu sélectif, de façon à produire des grandes quantités des banques digérées débarrassées des produits clivés (non transformants en raison de leur linéarisation). Les éléments clivés sont désormais absents des banques. La figure 13 en annexe est une représentation schématique de cette étape.

e) On clive séparément chacune de ces banques par l'enzyme B. La figure 14 en annexe est une représentation schématique de cette étape.

5 On dépose alors séparément chacun de ces 60 à 70 produits de digestion dans un puits de gel d'agarose, comme représenté à la figure 15 en annexe.

Puis, on fait migrer, l'on transfère sur nitro-cellulose, et l'on hybride avec la sonde, comme représenté à la figure 16 en annexe.

10 f) On analyse la CEM du signal. Ainsi, si une bande est présente dans la piste "banque non digérée", mais absente dans la piste "banque digérée par I", c'est que l'insert hybridé est sensible à l'enzyme I, et ainsi de suite.

15 g) On fait les multi-digestions correspondantes, et le seul plasmide résistant est le plasmide recherché. Avantageusement, on peut ensuite faire des minipreps avec ce plasmide en utilisant les enzymes "s" pour confirmer.

20 Le procédé décrit ci-dessus relève à la fois du Southern Blot, car l'on fait migrer de l'ADN, et du Northern Blot, car ce que l'on fait migrer correspond aux produits transcrits de la cellule, et pourrait donc être intitulé le "Blot équatorial".

25 Ce procédé peut être utilisé industriellement, en proposant les Blots équatoriaux correspondant à de nombreuses banques, comme celles réalisées dans l'exemple 1. Ainsi, en disposant de la sonde et des Blots équatoriaux judicieusement choisis, il devient quasi immédiat de trouver tous les allèles, correspondants d'espèce, épissages alternatifs et isotypes (deux jours de travail, sans compter le séquençage).

30

35

Une version simplifiée de ce procédé consiste à effectuer des Dots au lieu de Blots, c'est-à-dire à s'abstenir de cliver les banques mono-digérées par l'enzyme B, et à les déposer directement en un point sur une membrane de nitro-cellulose. Lors de l'hybridation avec la sonde, les points générant un signal, par exemple radioactif, correspondraient aux enzymes pour lesquelles le plasmide serait résistant. Ceux ne retenant plus de signal correspondraient aux enzymes pour lesquelles l'insert est sensible. L'analyse est alors proche de celle développée dans l'exemple 1, comme représentée aux figures 17 et 18 en annexe.

Cette version simplifiée est cependant insuffisante pour cloner des isotypes ou des épissages alternatifs exprimés dans une même cellule.

Exemple 3 : Southern Blot d'identification "identiblot".

Les applications précédentes du procédé de l'invention visent à simplifier et rendre plus rapide le clonage par expression ou par homologie et en offrant de nouvelles possibilités.

Le Southern Blot d'identification propose un raccourci qui faciliterait les travaux du chercheur impliqué dans les techniques de clonage moléculaire.

En effet, Le cas est fréquent dans l'art antérieur où une stratégie de clonage aboutit à l'obtention de nombreux inserts, parmi lesquels se trouve le gène recherché et de nombreux parasites. Or dans l'art antérieur, pour identifier un insert parmi tous ces parasites, il est indispensable de le séquencer, au moins partiellement, ce qui représente un travail considérable. Le séquençage est une technique fastidieuse et trop puissante pour ce simple usage

d'identification d'inserts, en effet la lecture de 10 nucléotides suffit le plus souvent à identifier un insert. Une technique moins puissante, mais moins fastidieuse, utilisant la DMD, pourrait lui être substituée.

5

Ainsi, comme décrit précédemment, il est possible de faire des Southern Blots de banques d'ADN génomique de façon industrielle. Une représentation schématique de la préparation d'une telle banque d'ADN génomique est donnée à la figure 3 en annexe. Ce Southern Blot, rebaptisé "identiblot" dans le cadre de la présente invention est suffisamment informatif pour permettre l'identification du fragment d'ADN homologue de la sonde.

10

15

Ce southern blot est même cent millions de fois trop informatif. En effet, la banque génomique est constituée de 4 millions d'inserts différents de 1 kb de long. Pour les mêmes raisons que celles exposées précédemment, un insert pris au hasard sera résistant à 55 enzymes sur 70 et sensible à 15. Le nombre de combinaisons enzymatique possible est donc  $C_{70}^{15}$ , soit  $70!/(55!15!) = 7,2 \cdot 10^{14}$ . Ce nombre est plus de cent millions de fois supérieur à la taille de la banque. Le fait de considérer que les inserts ont tous une taille différente ajoute encore des possibilités. Chaque insert de la banque est donc associé à une CEM originale.

20

25

Ce procédé présente un intérêt majeur dans le cas de stratégies de clonages aboutissant à de forts taux de clones faux positifs, par exemple dans les cas de l'utilisation de banques soustractives ou de stratégies de clonage par insertion.

30

35

En effet, à partir des cinquante clones étudiés, il suffit de préparer une "sonde multiplex", c'est à dire un marquage commun de cinquante inserts (dans un seul tube), et une seule hybridation du filtre

de nitro-cellulose. On obtient l'identité des cinquante inserts d'un seul coup, en comparant leurs CEM avec celles préalablement entrées dans une base de données informatique.

5                    La figure 19 en annexe représente un exemple où trois sondes A, B, et C ont été utilisées simultanément. Les inserts génomiques correspondant à des CEM non décrites, et eux seuls, seront clonés par exemple par multi-digestion, puis séquencés.

10                    A l'inverse du cas des banques d'ADNc, dont la variété est insondable, il est suffisant de produire un seul type de Blot, pour chacune des 10 espèces étudiées couramment en biologie : humain, souris, rat, 15 drosophile, tabac, levure, etc...

#### Exemple 4 : Etude du polymorphisme humain.

20                    Les travaux de recherche menés dans le cadre de l'étude du polymorphisme nécessitent l'utilisation d'un nombre croissant de marqueurs génétiques. Leur utilisation dans le cadre du diagnostic des maladies génétiques est également en forte augmentation. Il est probable que dans un futur proche des cartes 25 personnelles de prédisposition génétique pourront être établies, de façon à ce que chacun évite de s'exposer à certains risques environnementaux, comme le tabac en cas de prédisposition au cancer du poumon, le sucre pour les terrains diabétiques, etc....

30                    Les techniques disponibles dans l'art antérieur permettant d'étudier le polymorphisme sur un allèle sont principalement la PCR, le Southern Blot et l'étude des marqueurs de l'ADN satellite (potentiellement utilisés en combinaison). La méthode de 35 l'invention constitue donc une alternative, plus efficaces, à ces techniques.

La mise en oeuvre du procédé de clonage selon l'invention pour l'étude du polymorphisme humain est proche de celle de l'exemple 3 ci-dessus. La différence majeure réside dans le fait qu'il est  
5 nécessaire de réaliser la banque, ses digestions, la migration sur gel et le transfert sur membrane, pour chaque sujet.

Ce procédé permet également de localiser très rapidement l'origine d'une maladie génétique par  
10 l'analyse des différents membres de la famille hébergeant la pathologie. Dans une famille dite "à risque" pour une pathologie, il convient de faire les Identiblots correspondant aux différents membres de la famille, et de les tester au moyen de polysondes faites  
15 à partir des marqueurs génétiques déjà existants. En un jour, on peut hybrider au moins vingt membranes, soit obtenir l'allèle de 1000 marqueurs. Les facteurs génétiques multiples pourraient être définis avec cette technique.

Cette application du procédé de l'invention implique un investissement en temps de l'ordre de plusieurs heures de travail, mais est rapidement rentabilisé par la possibilité de tester les marqueurs de prédisposition génétique 50 par 50, au moyen de  
25 sondes multiplex préparées en routine et constituées du nombre équivalent de marqueurs.

Exemple 5 : Deuxième approche de l'étude du polymorphisme humain selon l'invention.

30 Cette seconde mise en oeuvre du procédé de l'invention à l'étude du polymorphisme concerne l'application de la DMD à l'étude des marqueurs génétiques par RFLP. Le procédé selon l'invention  
35 comprend alors les étapes suivantes :

a) On a préalablement définis la CEM de chacun des marqueurs connus de façon à permettre leur identification. Cette identification est faite une fois pour toutes et un polymorphisme est donc caractérisé par une variation de la CEM.

b) On constitue une banque d'ADN génomique du sujet étudié dans le vecteur précédemment décrit. Comme indiqué précédemment, il est impératif que les vecteurs seuls ligasés sur eux-même soient éliminés. Avantageusement, on prépare une banque dont les fragments ont une longueur moyenne de 2000 nucléotides.

c) On effectue la digestion de la banque par des batteries d'enzymes correspondant aux CEM attribuées aux marqueurs étudiés.

Ainsi, par exemple, dans le premier puits, on met les cinquantes enzymes qui vont digérer toute la banque sauf un premier marqueur. Dans le second puits, on met une autre batterie d'enzymes pour un second marqueur, etc...

d) On transforme les banques multidigérées dans des bactéries compétentes.

e) On cultive en milieu solide (boite de petri) contenant l'agent sélectif du plasmide de façon à ce que si l'allèle recherché existe, les bactéries poussent. Par contre, si cet allèle n'existe pas, elles ne poussent pas. On dispose ainsi d'un profil pour chaque allèle, et l'on peut étudier un nombre illimité de marqueurs en une seule fois.

Ce procédé peut être automatisé en préparant des plaques de 96 puits contenant tous les mélanges enzymatiques, ces plaques sont stockées en congélateur. Différents types de plaques pourraient être produites:

- Marqueurs répartis sur tout le génome.
- Marqueurs répartis sur un seul chromosome.
- Marqueurs répartis sur une région précise.

- Marqueurs reliés aux risques de pathologies.

- Etc...

5 On distribue dans chaque puits une quantité fixe d'ADN sous forme de banque, puis l'on incube à 37°C pour les digestions. On rajoute alors les bactéries compétentes, et on suit le mécanisme classique de transformation (choc thermique, incubation sans agent de sélection, etc...). On prélève à la pipette 96 canaux et  
10 l'on dépose sur un pétri. La lecture est effectuée à l'oeil ou à l'aide d'un spectrophotomètre.

Cette seconde approche offre l'avantage d'être facilement automatisable, et confère à grande échelle un gain de temps très important. En effet, il  
15 faut tester un grand nombre de marqueurs pour chaque individu, car la création de la banque représente un investissement en travail. En outre, elle permet de s'exonérer de toute radioactivité, coûteuse et nocive.

20 Exemple 6 : Troisième approche de l'étude du polymorphisme humain selon l'invention.

Cette variante du procédé d'étude du polymorphisme humain selon l'invention se rapporte à  
25 l'application de la DMD à l'étude des différents allèles d'un seul marqueur.

Dans l'art antérieur, ces différentiations se font essentiellement par séquençage. Dans un futur proche, les puces d'ADN (T. Pastinen et al., Genome  
30 Research 7 : 606-14, 1997 ; J. G. Hacia et al., Nature Genetics, 14 : 441, 1996), permettront d'automatiser ces travaux. Ceux-ci sont particulièrement utiles dans le cas de l'examen de la gp120 du virus HIV ou d'oncogènes cellulaires tels p53, présents sous de nombreux allèles.

35 Conformément à la présente invention, cette variante consiste à préalablement définir la CEM de

chacun des allèles du segment étudié. Cette  
identification est réalisée une fois pour toute et se  
limite aux allèles comportant au moins un site de  
restriction de différence. Puis on amplifie par PCR le  
5 fragment étudié. On clone ledit fragment dans le vecteur  
de l'invention. Ce procédé est donc très proche de celui  
de l'exemple 5, mais avec une banque comportant  
seulement un ou deux inserts correspondant aux deux  
copies du gène présent chez un individu. La suite de ce  
10 procédé est identique à celle de l'exemple 5.

Ce travail peut être réalisé simultanément  
sur plusieurs inserts à la fois.

REVENDICATIONS

1) Procédé de clonage d'une séquence d'acide  
nucléique, caractérisé en ce qu'il comprend les étape  
5 suivantes :

- on prépare une banque d'ADN susceptible de  
contenir ledit fragment,

- on crible ladite banque en utilisant de  
façon combinatoire au moins 10 et de préférence 60 à 70  
10 enzymes de restriction, pour isoler par tout moyen  
approprié le clone contenant ledit fragment.

2) Procédé selon la revendication 1,  
caractérisé en ce que la préparation de la banque d'ADN  
15 susceptible de contenir le fragment d'acide nucléique,  
consiste à insérer chacun des fragments d'ADN d'un  
échantillon dans un vecteur substantiellement dépourvu  
de tout site de clivage par des enzymes de restriction à  
l'exception :

20 - d'un site pour la construction de la  
banque, et

- éventuellement de deux autres sites  
identiques et différent du premier site et le flanquant,  
utile pour sous-cloner la séquence d'acide nucléique une  
25 fois celle-ci identifiée et clonée.

3) Procédé de clonage d'un fragment d'acide  
nucléique selon l'une des revendications 1 à 2,  
caractérisé en ce qu'il comprend les étape suivantes :

30 (a) on prépare une banque d'ADN susceptible  
de contenir ledit fragment d'acide nucléique, consistant  
à insérer chacun des fragments d'ADN d'un échantillon  
dans un vecteur substantiellement dépourvu de tout site  
de clivage par des enzymes de restriction à l'exception

35 :

- d'un site pour la construction de la banque, et

5 - éventuellement de deux autres sites identiques, différents du premier site et le flanquant, utiles pour sous-cloner le fragment d'acide nucléique une fois celui-ci à identifiée et clonée.

10 (b) On réalise la digestion en parallèle de la banque par plusieurs enzymes de restriction, au moins 10 et de préférence environ 60 à 70 enzymes de restriction de façon obtenir autant de banques mono-digérés que d'enzymes utilisées.

15 (c) On transfecte indépendamment les banques mono-digérées dans des hôtes cellulaires appropriés de façon à obtenir des lots correspondant d'hôtes cellulaires.

20 (d) On teste par tout moyen approprié chacun des lots obtenus à l'étape (c) pour évaluer l'intégrité du fragment d'acide nucléique à cloner et ainsi établir sa CEM.

4) Procédé selon la revendication 3, caractérisé en ce qu'il comprend les étapes suivantes :

25 (e) On reprend la banque totale de l'étape (a) et on la digère sensiblement simultanément par les enzymes qui n'affectent pas l'intégrité du fragment d'acide nucléique à cloner.

30 (f) On isole le clone résistant contenant le fragment d'acide nucléique à cloner par tout moyen approprié et éventuellement on le sous-clone en utilisant les deux sites ménagés dans le vecteur à cet effet.

(g) On séquence éventuellement le fragment d'acide nucléique cloné.

5) Procédé selon les revendications 3 ou 4, caractérisé en ce qu'il comprend entre les étape (a) et (b) l'étape suivante :

5 (a') on vérifie la présence du fragment d'acide nucléique à cloner dans la banque en transfectant dans un hôte cellulaire qui n'a pas ledit fragment et en testant par tout moyen approprié la présence du fragment dans ledit hôte.

10 6) Procédé selon les revendications 4 et 5, caractérisé en ce qu'il comprend entre les étape (e) et (f) l'étape suivante :

15 (e') On transforme la banque multi-digérée de l'étape (e) dans des hôtes compétents, de façon à vérifier la nature des fragments clonés.

20 7) Procédé selon l'une quelconque des revendications 1 à 6, caractérisé en ce la banque d'ADN contient de 1 à  $10^8$  et de préférence de l'ordre de  $10^5$  à  $4.10^6$  fragments différents de l'ordre de 0,1 kb à 5 kb chacun, et de préférence de l'ordre de 1 à 2 kb.

25 8) Procédé selon l'une quelconque des revendications 3 à 7, caractérisé en ce que les tests réalisés aux étapes (d) et/ou (a') pour vérifier l'intégrité de la séquence d'acide nucléique à cloner sont tout moyen de mise en évidence soit de la séquence elle-même, comme une sonde, soit de la protéine codée par ladite séquence, tel qu'un ligand comme un anticorps, ou encore de l'activité de cette protéine, 30 comme un marqueur enzymatique, qui peuvent être détectées par tout moyen connu de l'homme du métier, comme un marquage fluorescent ou radioactif.

35 9) Procédé de clonage d'un gène par banque d'expression selon l'une quelconque des revendications 1

à 7, caractérisé en ce qu'il comprend les étapes suivantes :

5 a) On prépare une banque d'ADNc susceptible de contenir ledit gène en insérant dans un vecteur substantiellement dépourvu de tout site de clivage par des enzymes de restriction à l'exception d'un site pour la construction de la banque et de deux autres sites identiques, différents du premier site et le flanquant, utiles pour sous-cloner le gène une fois celui-ci  
10 identifié et cloné.

b) On vérifie la présence du gène recherché dans la banque en transfectant dans une lignée cellulaire qui n'a pas l'activité ou le phénotype recherché et en mesurant sa restauration de façon à  
15 distinguer les cellules transfectées des cellules non transfectées.

c) On digère la banque indépendamment par au moins 10 et de préférence environ 60 à 70 enzymes de restriction.

20 d) On transfecte indépendamment les banques mono-digérées de l'étape (c).

e) On teste par tout moyen approprié chacun des lots obtenus à l'étape (d) pour la présence de l'activité associée au gène à cloner et on évalue  
25 l'intégrité de la séquence dudit gène afin d'établir la CEM de l'activité associée audit gène.

f) On reprend la banque totale de l'étape (a) et on la digère sensiblement simultanément par les environ 50 à 55 enzymes qui n'affectent pas l'activité  
30 mesurée à l'étape (e).

g) On transforme la banque multi-digérée dans des bactéries compétentes.

h) On sous-clone en utilisant l'enzyme correspondant au site de sous-clonage ménagés dans le  
35 vecteur, puis éventuellement on séquence le gène.

10) Procédé de clonage par homologie selon l'une quelconque des revendications 1 à 9, caractérisé en ce qu'il comprend les étapes suivantes :

- 5 (a) on prépare une banque d'ADN susceptible de contenir ladite séquence d'acide nucléique, consistant à insérer chacun des fragments d'ADN d'un échantillon dans un vecteur substantiellement dépourvu de tout site de clivage par des enzymes de restriction à l'exception :
- 10 - d'un site pour la construction de la banque, et
- éventuellement de deux autres sites identiques et différent du premier site et le flanquant, utile pour sous-cloner la séquence d'acide nucléique une
- 15 fois celle-ci identifiée et clonée.
- b) On digère la banque indépendamment par chacune d'au moins 10 et de préférence environ 60 à 70 enzymes de restriction.
- c) On transforme les produits de la
- 20 digestion de l'étape (b) dans des bactéries compétentes.
- d) On fait pousser les bactéries transformées en milieu sélectif, de façon à produire des grandes quantités des banques digérées débarrassées des produits clivés.
- 25 e) Eventuellement on clive séparément chacune de ces banques par l'enzyme correspondant au site de sous-clonage ménagé dans le vecteur et l'on dépose séparément chacun des produits de digestion dans un puits de gel d'agarose.
- 30 f) f) On fait migrer les produits de digestion de l'étape (e), puis l'on transfère sur membrane par exemple de nitro-cellulose, et l'on hybride avec une sonde spécifique du gène à cloner par homologie, ou bien l'on dépose directement les produits
- 35 de l'étape (d) sur une membrane de nitrocellulose.
- g) On analyse la CEM du signal.

e) On fait les multi-digestions correspondantes, de façon à ce que le seul vecteur résistant soit le vecteur portant le gène à cloner.

5                   11) Application du procédé selon la revendication 10 à l'identification :

- d'allèles de différentes souches d'animaux de la même espèce, ou de différents individus chez l'homme,

10                   - d'équivalents géniques présents chez différentes espèces,

- des épissages alternatifs d'un gène.

- de différents membres d'une famille génique.

15

12) Procédé de Southern Blot d'identification d'inserts mettant en oeuvre un procédé de clonage d'une séquence d'acide nucléique selon l'une quelconque des revendications 1 à 10, caractérisé en ce qu'il comprend les étapes suivantes :

20

(a) on prépare une banque d'ADN susceptible de contenir ledit insert, consistant à insérer chacun des fragments d'ADN d'un échantillon dans un vecteur substantiellement dépourvu de tout site de clivage par des enzymes de restriction à l'exception d'un premier site pour la construction de la banque, et de deux autres sites identiques et différent du premier site et le flanquant, utile pour sous-cloner la séquence d'acide nucléique une fois celle-ci identifiée et clonée.

25

30

b) On digère cette banque par chacune d'au moins 10 et de préférence environ 60 à 70 enzymes de restriction.

35

c) On transforme les banques monodigérés obtenues à l'étape (b) ci-dessus dans des bactéries compétentes, ou des hôtes équivalents.

d) On fait pousser les bactéries en milieu sélectif, de façon à produire des banques digérées débarassées des produits clivés.

5 e) On clive séparément chacune de ces banques par l'enzyme correspondant aux deux autres sites identiques et différent du premier site et le flanquant, et on dépose les produits de digestion dans les puits d'un gel d'agarose ou d'acrylamide.

10 f) On fait migrer ce gel et on le transfère sur une membrane par exemple de nitrocellulose.

g) On utilise les inserts à identifier comme sondes marquées, soit un par un, soit plusieurs par plusieurs.

15 h) On teste par tout moyen approprié chacun des lots obtenus à l'étape (g) pour les associer à une CEM.

20 13) Procédé d'étude du polymorphisme mettant en oeuvre un procédé de Southern Blot d'identification d'inserts selon la revendication 12, caractérisé en ce que :

- La banque d'ADN génomique de l'étape (a) est issue du sujet étudié.

25 - Les inserts utilisés comme sondes sont des marqueurs de polymorphisme.

14) Procédé d'étude du polymorphisme caractérisé en ce qu'il comprend les étapes suivantes :

30 a) On définit la CEM de chacun des marqueurs connus de façon à permettre leur identification par le procédé de clonage selon la revendication 3.

35 b) On constitue une banque d'ADN génomique du sujet étudié consistant à insérer chacun des fragments d'ADN d'un échantillon dans un vecteur substantiellement dépourvu de tout site de clivage par des enzymes de restriction à l'exception :

- d'un site pour la construction de la banque, et

5 - éventuellement de deux autres sites identiques et différent du premier site et le flanquant, utile pour sous-cloner la séquence d'acide nucléique une fois celle-ci identifiée et clonée.

c) On effectue la digestion de la banque par des batteries d'enzymes correspondant aux CEM attribuées aux marqueurs étudiés.

10 d) On transforme les banques multidigérées dans des bactéries compétentes.

e) On cultive ces bactéries en milieu liquide ou solide contenant l'agent sélectif du plasmide. Si l'allèle recherché existe, les bactéries poussent, par contre, si cet allèle n'existe pas, elles ne poussent pas, et l'on dispose ainsi d'un profil pour chaque allèle.

15 15) Mélange d'au moins 10 et de préférence de 60 à 70 enzymes de restriction susceptible d'être utilisé dans un procédé selon l'une quelconque des revendications 1 à 14.

25 16) Vecteurs susceptible d'être utilisé dans un procédé selon l'une quelconque des revendications 1 à 14, caractérisé en ce qu'il est substantiellement dépourvu de tout site de clivage par des enzymes de restriction à l'exception :

30 - d'un site pour la construction de la banque, et

- éventuellement, de deux autres sites identiques, différents du premier site et le flanquant, utiles pour sous-cloner une séquence d'acide nucléique clonée dans ledit vecteur.

35

17) Vecteur selon la revendication 16, caractérisé en ce qu'il porte un système qui élimine tout vecteur seul refermé sur lui-même.

5                   18) Banque ou groupe de banques d'ADN multi-digérée ou mono-digérées obtenus dans un procédé selon l'une quelconque des revendications 1 à 14.

10                   19) Les hôtes cellulaires ou bactériens contenant une banque ou un groupe de banques selon la revendication 18.

15                   20) Support portant une banque ou un groupe de banques selon la revendication 18 ou des hôtes cellulaires selon la revendication 19.

Fig.1

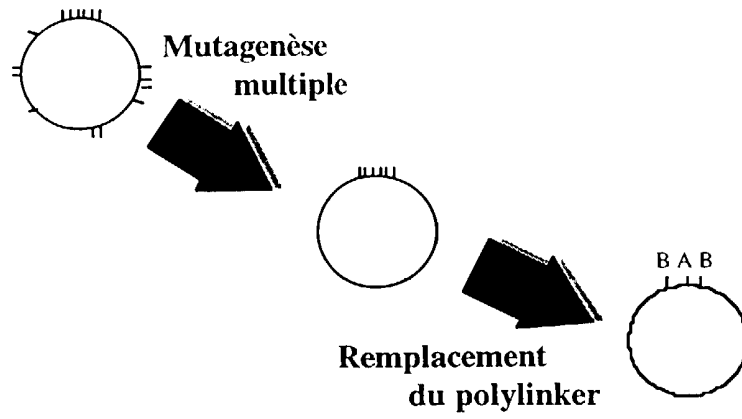


Fig.2

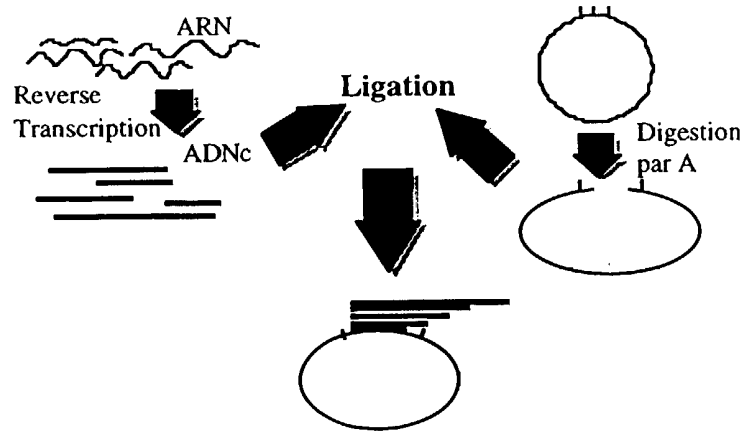


Fig.3

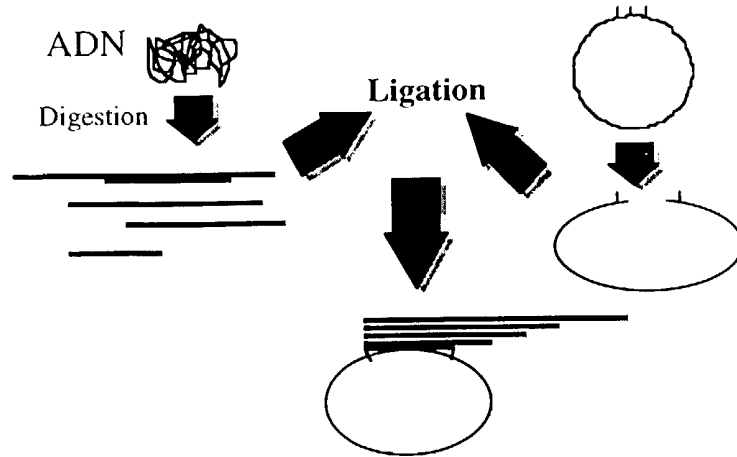


Fig.4

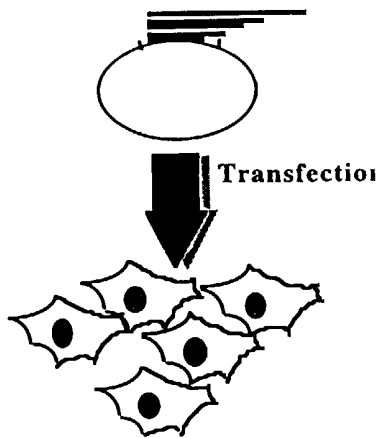


Fig.5

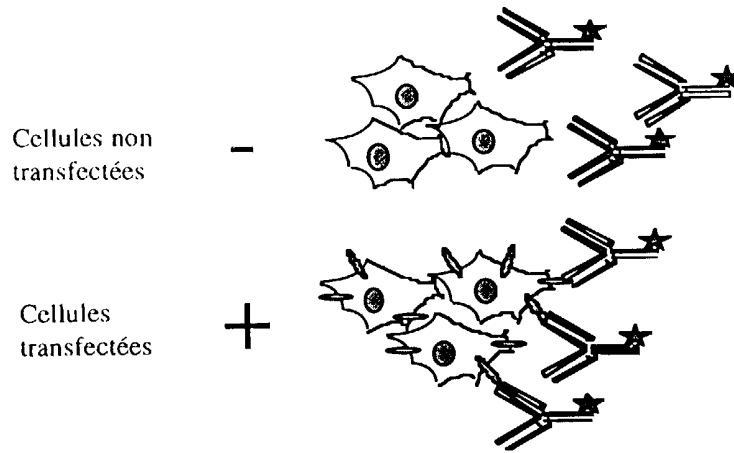


Fig.6

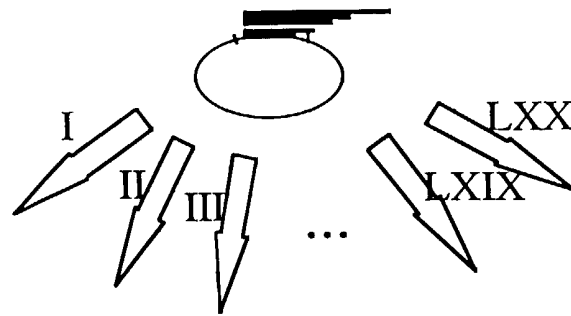


Fig.7

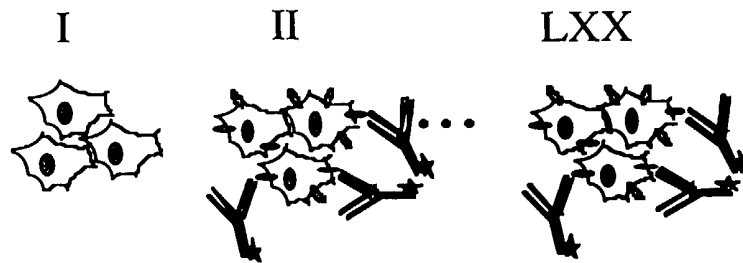


Fig.8

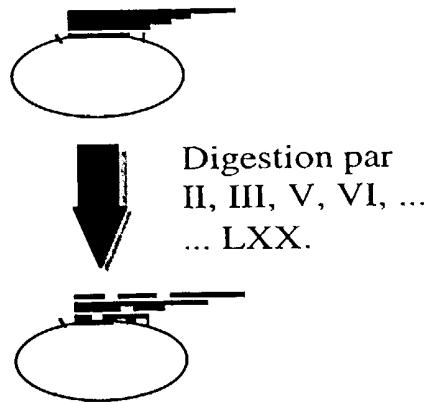


Fig.9

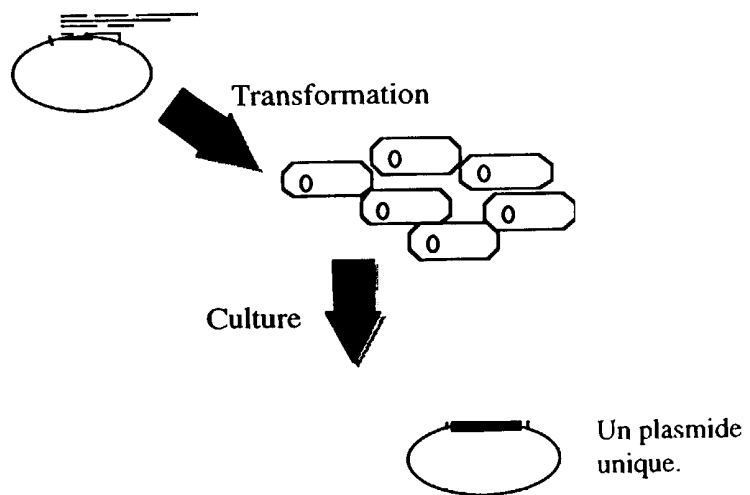
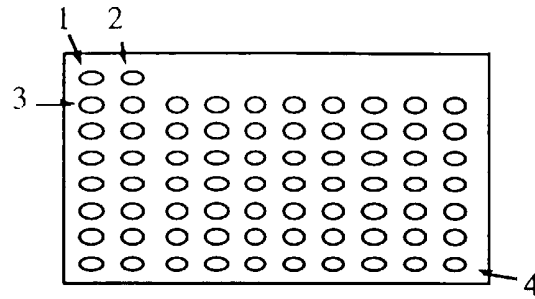
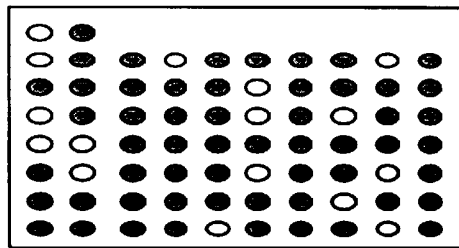


Fig.10



- 1: Extrait cellulaire de cellules non transfectées.
- 2: Extrait cellulaire de cellules transfectées par la banque totale.
- 3: Extrait cellulaire de cellules transfectées par la banque digérée par I.
- 4: Extrait cellulaire de cellules transfectées par la banque digérée par LXX.

Fig.11



CEM: I<sup>s</sup>II<sup>r</sup>III<sup>r</sup>IV<sup>s</sup>V<sup>r</sup>VI<sup>r</sup>...LXIX<sup>s</sup>LXX<sup>r</sup>.

Fig.12

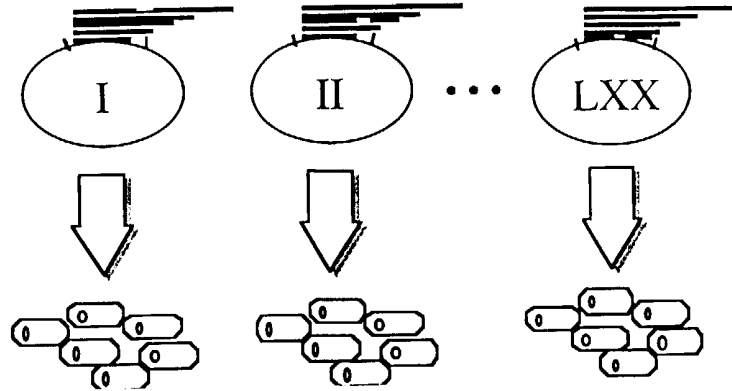


Fig.13

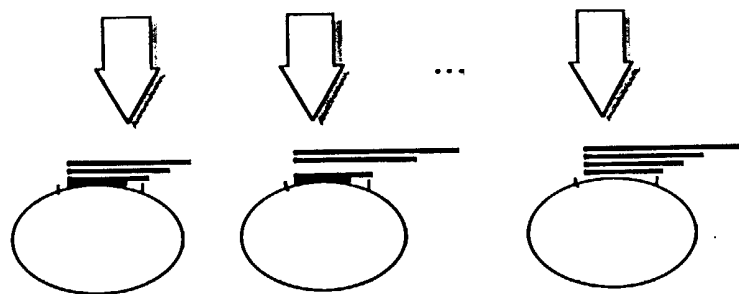


Fig.14

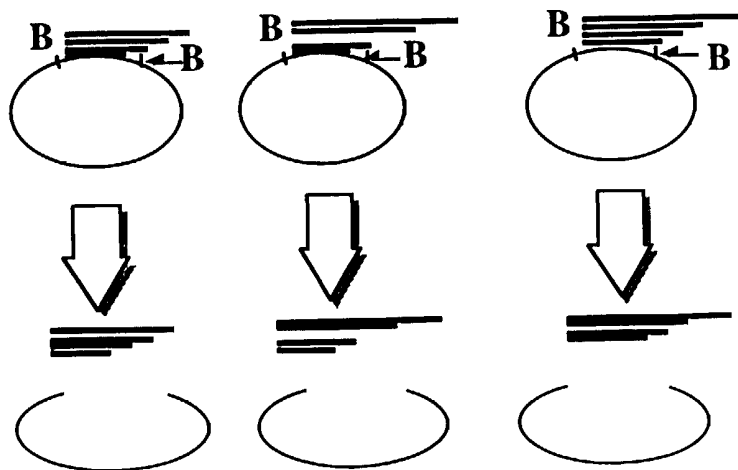
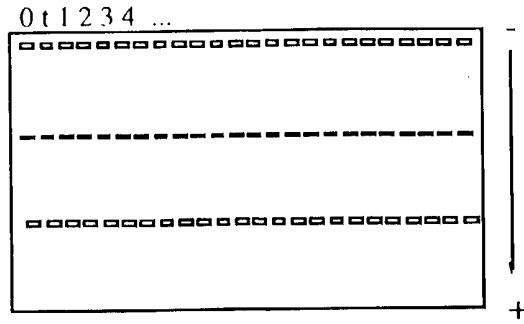
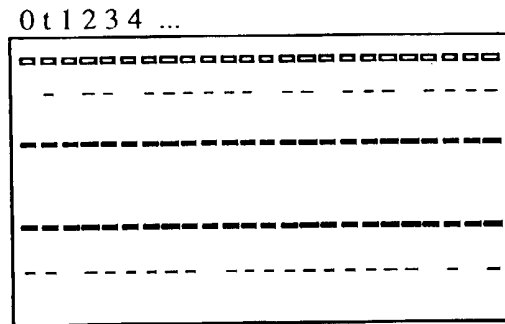


Fig.15



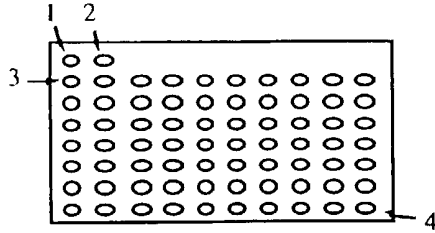
0: Plasmide seul (clivé B) t: Banque totale.  
1: Banque clivée par I. 2: Banque clivée par II.

Fig.16



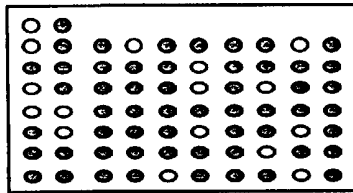
CEM: I<sup>s</sup>II<sup>r</sup>III<sup>r</sup>IV<sup>s</sup>V<sup>r</sup>VI<sup>r</sup>...LXIX<sup>s</sup>LXX<sup>r</sup>.

Fig.17



- 1: préparation d'ADN plasmidique de bactéries transformées par le plasmide seul.
- 2: préparation d'ADN plasmidique de bactéries transformées par la banque totale.
- 3: préparation d'ADN plasmidique de bactéries transformées par la banque digérée par I.
- 4: préparation d'ADN plasmidique de bactéries transformées par la banque digérée par LXX.

Fig.18



CEM: I<sup>r</sup>II<sup>r</sup>III<sup>r</sup>IV<sup>r</sup>V<sup>r</sup>VI<sup>r</sup>...LXIXLXX<sup>r</sup>.

Fig.19



CEM: A: I<sup>r</sup>II<sup>r</sup>III<sup>r</sup>IV<sup>r</sup>...  
 B: I<sup>r</sup>II<sup>r</sup>III<sup>r</sup>IV<sup>r</sup>S...  
 C: I<sup>r</sup>II<sup>r</sup>III<sup>r</sup>IV<sup>r</sup>...

INSTITUT NATIONAL  
de la  
PROPRIETE INDUSTRIELLE

RAPPORT DE RECHERCHE  
PRELIMINAIRE  
établi sur la base des dernières revendications  
déposées avant le commencement de la recherche

N° d'enregistrement  
national

FA 551255  
FR 9715319

DOCUMENTS CONSIDERES COMME PERTINENTS		Revendications concernées de la demande examinée
Catégorie	Citation du document avec indication, en cas de besoin, des parties pertinentes	
X	WO 95 04745 A (LIFE TECHNOLOGIES INC) 16 février 1995 * page 6 - page 7 * * revendications 1,4 *	18-20
X,D	JONES D.H. ET AL.: "Production of a vector to facilitate DNA mutagenesis and recombination" BIOTECHNIQUES, vol. 16, no. 4, 1994, pages 694-701, XP002076171 * figure 6 *	16,17
A	AMERSHAM LIFE SCIENCE CATALOGUE 1994 pages 164-165 XP002076172 * le document en entier *	15
A	US 5 252 724 A (KISHIMOTO TOSHIHIKO ET AL) 12 octobre 1993 * abrégé *	1-20
A	BONALDO FATIMA DE M ET AL: "NORMALIZATION AND SUBTRACTION: TWO APPROACHES TO FACILITATE GENE DISCOVERY" GENOME RESEARCH, vol. 6, no. 9, septembre 1996, pages 791-806, XP002039972 * le document en entier *	1-20
T	WO 98 28439 A (NAHAS NASRI ;GENSET (FR); DUMAS MILNE EDWARDS JEAN BAPTI (FR)) 2 juillet 1998 * abrégé *	18-20
Date d'achèvement de la recherche		Examineur
1 septembre 1998		Galli, I
<p>CATEGORIE DES DOCUMENTS CITES</p> <p>X : particulièrement pertinent à lui seul Y : particulièrement pertinent en combinaison avec un autre document de la même catégorie A : pertinent à l'encontre d'au moins une revendication ou arrière-plan technologique général O : divulgation non-écrite P : document intercalaire</p> <p>T : théorie ou principe à la base de l'invention E : document de brevet bénéficiant d'une date antérieure à la date de dépôt et qui n'a été publié qu'à cette date de dépôt ou qu'à une date postérieure. D : cité dans la demande L : cité pour d'autres raisons &amp; : membre de la même famille, document correspondant</p>		

1

EPO FORM 1503 03 B2 (P04C13)

DOMAINES TECHNIQUES  
RECHERCHES (Int.CL.6)

C12N  
C12Q