(19) RÉPUBLIQUE FRANÇAISE

INSTITUT NATIONAL
DE LA PROPRIÉTÉ INDUSTRIELLE

PARIS

(11) **N° de publication :** **2 913 694**

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

(21) **N° d'enregistrement national :** **08 51645**

(51) Int Cl[8] : ***C 12 N 15/29*** *(2006.01), C 12 N 15/82, 5/04, 5/10, C 07 K 14/415, A 01 H 5/00, G 01 N 33/53, C 12 Q 1/68*

(12) DEMANDE DE BREVET D'INVENTION A1

(22) **Date de dépôt :** 13.03.08.

(30) **Priorité :** 13.03.07 MY 20070394.

(43) **Date de mise à la disposition du public de la demande :** 19.09.08 Bulletin 08/38.

(56) **Liste des documents cités dans le rapport de recherche préliminaire :** *Ce dernier n'a pas été établi à la date de publication de la demande.*

(60) **Références à d'autres documents nationaux apparentés :**

(71) **Demandeur(s) :** *MALAYSIAN PALM OIL BOARD* — MY.

(72) **Inventeur(s) :** ABDULLAH SITI NOR AKMAR, CHEAH SUAN CHOO et ABD WAHAB NURNIWALIS.

(73) **Titulaire(s) :**

(74) **Mandataire(s) :** CABINET BEAU DE LOMENIE.

(54) ELEMENTS REGULATEURS D'EXPRESSION.

(57) L'invention concerne généralement un élément régulateur d'expression fonctionnel dans des plantes qui incluent le palmier à huile, et en particulier un élément régulateur d'expression fonctionnel sélectivement dans le tissu endosperme des plantes. Ledit élément régulateur d'expression comprend une séquence de nucléotides qui module spécifiquement l'expression d'une seconde molécule d'acide nucléique couplée de manière fonctionnelle audit élément régulateur d'expression.

FR 2 913 694 - A1

ARRIERE-PLAN DE L'INVENTION

DOMAINE DE L'INVENTION

La présente invention concerne d'une manière générale un élément régulateur d'expression qui est fonctionnel dans les plantes. Plus particulièrement, la présente invention fournit un élément régulateur d'expression sous forme d'un promoteur fonctionnel sélectivement dans le tissu de l'endosperme des palmiers à huile. De manière encore plus particulière, la présente invention concerne le promoteur associé avec un gène codant la glutéline dans le tissu de l'endosperme des palmiers à huile. La présente invention envisage aussi d'autres éléments régulateurs d'expression, comme des stimulateurs et/ou silenceurs, qui sont associés avec le promoteur du gène de glutéline ou qui sont positionnés à proximité de celui-ci ou qui modulent d'une autre façon l'activité ou la fonction du promoteur du gène de glutéline. Les éléments régulateurs d'expression de la présente invention ont une utilité, entre autres, pour faciliter l'expression spécifique de tissu de molécules d'acides nucléiques souhaitées en liaison fonctionnelle avec eux qui codent, entre autres, un ARN qui est éventuellement traduit en un peptide, un polypeptide ou une protéine. Au moins chez les palmiers à huile, l'expression est spécifique de tissu pour l'endosperme. De plus, la présente invention fournit en outre une nouvelle protéine glutéline isolée et les séquences génétiques la codant.

DESCRIPTION DE L'ETAT DE LA TECHNIQUE

Des détails bibliographiques concernant les publications citées dans cette description sont également regroupés à la fin de la description.

Une référence à tout état de la technique dans cette description n'est pas et ne doit pas être considérée comme une référence ou une forme quelconque de suggestion que cet état de la technique fait partie de la connaissance générale commune dans un pays quelconque.

Deux types différents d'huiles de réserve sont obtenus à partir des palmiers à huile, l'huile de palme provenant du mésocarpe et l'huile de noyau provenant du noyau. Les deux huiles sont différentes en ce qui concerne la composition en acides gras, les propriétés

physiques et les possibilités d'utilisation. L'huile de noyau est synthétisée dans l'endosperme 11 à 14 semaines après l'anthèse (SAA). Elle est riche en acide gras saturés à chaîne moyenne, c'est-à-dire 50 % d'acide laurique (C12 : 0) et 15 % d'acide myristique (C14 : 0), qui servent de charges importantes pour l'industrie oléochimique. Elle contient aussi 15 % d'acide linoléique (C18 : 2) et des niveaux inférieurs d'acide palmitique (C16 : 0), d'acide caprique (C10 : 0) et d'acide caprylique (C8 : 0), contribuant à raison d'environ 7 %, 4 % et 4 % respectivement. La pulpe du noyau des palmiers à huile, le sous-produit de l'extraction de l'huile, est un aliment intéressant pour les animaux. La synthèse de l'huile dans le mésocarpe débute plus tardivement vers 15 SAA et s'arrête quand le fruit mûrit à 20 SAA. L'huile de palme contient 44 % d'acide palmitique (16 : 0), 5 % d'acide stéarique (18 : 0), 39 % d'acide oléique (18 : 1) et 10 % d'acide linoléique (18 : 2). Les principales applications des produits d'huile de palme sont dans le domaine alimentaire, cependant des utilisations non alimentaires comme dans les cosmétiques et les biocarburants sont en progrès.

Etant des tissus de réserve, le mésocarpe et le noyau des palmiers à huile peuvent être la cible d'accumulation de produits modifiés par génie génétique sans effet néfaste pour la plante. Il serait très souhaitable de manipuler les voies biochimiques pour concevoir une synthèse d'huile dans le noyau sans affecter la production d'huile dans le mésocarpe. La composition en aminoacides des protéines du noyau peut aussi être modifiée pour augmenter leur valeur nutritionnelle. Le noyau serait aussi une cible appropriée pour accumuler des nutraceutiques de grande valeur et des produits pharmaceutiques de grande valeur par manipulation génétique. Pour atteindre ces buts, il est essentiel d'isoler les promoteurs de gènes spécifiques du noyau et de comprendre la régulation de l'expression génique dans le noyau.

Les protéines de réserve des graines sont synthétisées à des niveaux élevés dans le cotylédon et l'embryon des plantes dicotylédones ou dans l'endosperme des plantes monocotylédones et sont déposées dans des corps protéiques dans les cellules de ces tissus. L'expression de ces gènes est étroitement régulée, ce qui fournit un

système intéressant pour étudier les mécanismes moléculaires contrôlant l'expression génique spécifique des graines dans les plantes supérieures. Des promoteurs de gènes de réserve des graines ont été utilisés avec succès pour contrôler l'expression de gènes introduits pour modifier la composition des huiles dans les cultures de graines oléagineuses transgéniques. Ceux-ci incluent l'utilisation du promoteur du gène de napine pour le colza (Voelker *et al.*, *Science 257:* 72-73, 1992) et le promoteur du gène de β-conglicinine pour le soja (Kinney, *Journal of Food Lipids 3:* 273-292, 1996). Il existe aussi un intérêt important à utiliser des promoteurs provenant de gènes de protéines de réserve des graines pour cibler la production de nouveaux produits à haute valeur dans les graines. Le promoteur du gène de glutéline spécifique de l'endosperme du riz par exemple est utilisé pour produire des lignées de riz transgéniques qui augmentent le niveau de pro-vitamine A (Golden Rice) (Beyer *et al.*, *Proceedings of the American Society for Nutritional Sciences Symposium on Plant Breeding:* 506S-510S, 2002) et pour la production d'un vaccin sous-unités dans les graines du tabac transgénique (Tackaberry *et al.*, *Genome 46(3):* 521-526, 2003).

Les protéines de réserve des graines sont classées sur la base de différences dans leurs propriétés de solubilité. Les glutélines insolubles dans une solution salée sont trouvées dans le riz, le blé et d'autres céréales. Les glutélines sont les principales protéines de réserve des graines dans le riz, représentant plus de 80 % de la teneur en protéines totales de l'endosperme (Shewry, *Biol. Rev. 70:* 375-426, 1995). Typiquement, les protéines de réserve des graines incluant la glutéline sont codées par des familles multigéniques complexes. L'analyse des séquences a révélé que les glutélines du riz sont liées plus étroitement aux globulines 11S solubles dans les solutions salées trouvées largement dans les espèces dicotylédones (Muntz, *Plant Molecular Biology 38:* 77-99, 1998). Les glutélines du riz sont synthétisées sous forme de protéines précurseurs, qui codent une sous-unité protéique acide et une sous-unité protéique basique, plus un peptide leader pour le ciblage vers le réticulum endoplasmique. Les deux sous-unités protéiques se séparent après la traduction mais se réassocient ensuite via la formation de liaisons disulfure entre leurs

résidus cystéine. Plusieurs polypeptides se joignent pour former un oligomère ayant la conformation appropriée pour être déposé dans des corps protéiques (Takaiwa *et al.*, *Plant Mol. Biol. 17(4):* 875-885, 1991).

Selon la présente invention, un élément régulateur d'expression spécifique de tissu qui permet l'expression spécifique de tissu des molécules d'acides nucléiques associées à lui est identifié pour le tissu de l'endosperme d'un palmier à huile.

RESUME DE L'INVENTION

Dans toute cette description, sauf indication contraire, le mot "comprendre" ou les variantes comme "comprend" ou "comprenant" sera compris comme impliquant l'inclusion d'un élément ou entier ou d'un groupe d'éléments ou d'entiers indiqué mais pas l'exclusion de tout autre élément ou entier ou groupe d'éléments ou entiers.

La présente invention fournit un élément régulateur d'expression fonctionnel dans les plantes. Dans un mode de réalisation préféré, l'élément régulateur d'expression est un promoteur.

Le séquençage à une seule passe d'ADNc provenant de banques d'ADNc du palmier à huile, combiné avec le profilage de transcrits, a été utilisé pour identifier des gènes hautement exprimés spécifiquement dans le noyau des palmiers à huile. Une recherche dans des bases de données de séquences a identifié un marqueur de séquence exprimée (EST) comme codant la protéine de réserve des graines glutéline. Un clone génomique contenant le promoteur du gène a été obtenu et caractérisé totalement. La région promotrice a été identifiée et l'activité de ce promoteur a été analysée via un essai transitoire de gène rapporteur dans des coupes de tissu de palmier à huile. Le promoteur est fonctionnel sélectivement dans le tissu de l'endosperme.

La présente invention fournit donc, entre autres, un élément régulateur d'expression associé avec l'expression génique dans le tissu de l'endosperme des palmiers à huile. La présente invention concerne particulièrement un élément régulateur sous forme d'un promoteur, qui est au moins spécifique dans le tissu de l'endosperme des palmiers à huile. La présente invention s'étend cependant à d'autres éléments

régulateurs d'expression associés avec l'expression d'acides nucléiques comme des stimulateurs ou des silenceurs.

Ainsi, un aspect de la présente invention envisage une molécule d'acide nucléique isolée comprenant un élément régulateur d'expression spécifique de tissu, qui est spécifique de tissu dans le tissu de l'endosperme des palmiers à huile et qui facilite l'expression d'un ADN lié de manière fonctionnelle audit élément régulateur dans des cellules végétales.

Le terme "élément régulateur d'expression" tel qu'il est utilisé ici devrait être compris comme se référant à toute séquence nucléotidique qui facilite l'expression d'une seconde séquence d'acide nucléique liée de manière fonctionnelle à celle-ci. Dans un mode de réalisation particulier, l'élément régulateur d'expression est un promoteur. Cependant, comme indiqué ci-dessus, les stimulateurs et les silenceurs sont également envisagés par la présente invention. Le terme "spécifique de tissu" quand il est utilisé de manière adjective en combinaison avec "l'élément régulateur d'expression" devrait être compris comme se référant à un élément régulateur d'expression qui dirige l'expression d'une séquence d'acide nucléique dans un ou plusieurs tissus d'une plante, mais pas dans tous les tissus d'une plante. Il conviendrait de noter que, bien que les éléments régulateurs d'expression illustrés à titre d'exemples décrits ici soient spécifiques de l'endosperme dans les palmiers à huile, dans d'autres espèces végétales ils peuvent aussi être spécifiques de l'endosperme ou bien encore ils peuvent être spécifiques d'un ou plusieurs tissus différents ou peuvent avoir un spectre d'activité plus large, par exemple en facilitant l'expression dans de multiples tissus ou de tous les tissus de la plante.

L'élément régulateur d'expression décrit ici est illustré à titre d'exemple concernant le promoteur facilitant l'expression du gène codant la glutéline dans l'endosperme du palmier à huile, bien qu'il soit envisagé que les éléments régulateurs d'expression décrits ici soient actifs dans une plage d'espèces végétales. Il est proposé d'utiliser les éléments régulateurs d'expression pour faciliter l'expression d'un acide nucléique pour produire un ARN, un ARNm, un ARNf, un ARNmi, un ARNsi, un ARNsn, un ARNsnm, un ARNst, un ARNt, un ARNr, un ARNsno, des transcrits d'empreinte (« *imprinting transcripts* »), un ARN

antisens, des transcrits sens. L'ARNm peut aussi être en outre traduit en peptides, polypeptides ou protéines.

La présente invention envisage aussi des fragments de l'élément régulateur d'expression décrit ici, le fragment conservant l'activité régulatrice.

Un autre aspect de la présente invention concerne une construction d'ADN recombinant, comme un vecteur de clonage ou d'expression, comprenant une séquence nucléotidique définissant un élément régulateur d'expression qui est spécifique de tissu dans les palmiers à huile. De telles constructions peuvent comprendre éventuellement en outre une molécule d'acide nucléique liée de manière fonctionnelle à l'élément régulateur d'expression et/ou les moyens pour l'insertion d'une telle molécule d'acide nucléique.

La présente invention envisage particulièrement une construction d'ADN recombinant comprenant un vecteur d'expression qui est capable de faciliter l'expression d'un acide nucléique dans des cellules végétales, où le vecteur d'expression comprend un élément régulateur d'expression tel que décrit ici en liaison fonctionnelle avec une molécule d'acide nucléique hétérologue et où l'élément régulateur d'expression dirige, au moins dans les palmiers à huile, l'expression spécifique de la molécule d'acide nucléique hétérologue dans l'endosperme.

Le vecteur d'expression de la présente invention est particulièrement utile, entre autres, pour faciliter l'expression d'une molécule d'acide nucléique dans l'endosperme d'une plante. Une telle molécule d'acide nucléique exprimée peut former des produits qui modifient les qualités des protéines ou des huiles ; produire ou accumuler de nouvelles huiles ou de nouvelles protéines ; ou produire ou accumuler des nutraceutiques, des produits pharmaceutiques, des planticorps et/ou d'autres produits industriels dans l'endosperme de la plante transformée.

La présente invention s'étend en outre à une cellule isolée, un tissu ou un organe isolé comprenant un élément régulateur d'expression tel que décrit ici. De plus, la présente invention s'étend à des tissus, organes ou plantes entières régénérées dérivés de cellules,

tissus et organes et des propagules et leur descendance ainsi qu'à des graines et d'autres matériaux reproducteurs.

Un autre aspect de la présente invention envisage des procédés d'identification de partenaires de liaison, qui se lient ou s'associent d'une autre manière avec l'élément régulateur d'expression décrit ici, de tels procédés incluant des essais de mobilité électrophorétique (essais de déplacement sur gel et essais de retard sur gel) et des essais d'empreinte à la DNase.

Selon la présente invention, les procédés décrits ont une application particulière pour l'identification de facteurs de transcription qui se lient aux éléments régulateurs d'expression décrits ici.

Dans un aspect lié, la présente invention envisage un procédé pour moduler l'activité d'un élément régulateur d'expression, ledit procédé comprenant l'administration à une ou plusieurs cellules végétales comprenant l'élément régulateur d'expression d'un agent qui se lie à l'élément régulateur d'expression et qui module l'activité de l'élément régulateur d'expression. De plus, il est envisagé que la modulation de l'activité de l'élément régulateur d'expression modulera à son tour l'expression d'une molécule d'acide nucléique en liaison fonctionnelle avec l'élément régulateur d'expression.

Un autre aspect de la présente invention envisage une nouvelle protéine glutéline et de nouvelles séquences d'acides nucléiques incluant leurs transcrits d'ADN et d'ARN, codant celle-ci.

Un résumé des identificateurs de séquences utilisés dans toute la présente description est fourni dans le tableau 1.

TABLEAU 1

Résumé des identificateurs de séquences

SEQUENCE ID NO :	DESCRIPTION
1	Séquence de l'ORF complet du gène de glutéline
2	Séquence d'aminoacides de la glutéline
3	Séquence du promoteur de glutéline
4	Amorce KRA
5	Amorce KA1
6	Amorce KRB

7	Amorce KTC
8	Amorce KTD
9	Amorce GK4
10	Amorce GKN6
11	Amorce GAK5
12	Amorce GAK4
13	Amorce EK1
14	Amorce ANEK2

Une liste des abréviations utilisées ici est présentée dans le tableau 2.

TABLEAU 2

Abréviations

ABREVIATION	DESCRIPTION
2,4-D	Acide 2,4-dichlorophénoxyacétique
CaMV	Virus de la mosaïque du chou-fleur
EGFP	Protéine fluorescente verte stimulée
EST	Marqueur de séquence exprimée
GFP	Protéine fluorescente verte
GUS	β-glucuronidase
NOS	Nopaline synthase
OCS	Octopine synthase
PEG	Polyéthylèneglycol
scFv	Anticorps à fragment variable à une seule chaîne
SCSV	Virus du rabougrissement du trèfle
SAA	Semaines après l'anthèse

BREVE DESCRIPTION DES FIGURES

La figure 1 est une représentation graphique montrant les séquences nucléotidique et d'aminoacides déduite d'un clone d'ADNc de palmier à huile de longueur intégrale pour le gène de la glutéline (pOP-KT21). La séquence consensus pour le début de la traduction et les signaux de polyadénylation supposés sont soulignés.

La figure 2 est une représentation sous forme de tableau résumant les pourcentages d'identité aux niveaux nucléotidique et des aminoacides entre des séquences des différents gènes de glutéline du palmier à huile.

La figure 3 est une représentation graphique montrant un alignement de séquences multiples pour les séquences d'aminoacides déduites de trois gènes de glutéline du palmier à huile [pOP-KT21, pKT7 (AF261691) et pKT9 (AF193433)], du gène de glutéline du riz (P07730) et du gène de globuline 11S du sésame (AF240004). Les astérisques représentent les aminoacides identiques. Les doubles points indiquent la conservation de groupes forts et les points indiquent la conservation de groupes faibles. Les sites de clivage supposés pour le peptide signal et pour séparer les sous-unités protéiques acides et basiques sont indiqués par des flèches dirigées vers le bas. Les résidus cystéine conservés sont indiqués par ♦.

La figure 4 est une représentation sous forme de tableau montrant une comparaison entre les profils d'aminoacides de la séquence d'aminoacides déduite de pOP-KT21 et de la pulpe de noyau du palmier à huile. (A) est l'utilisation des aminoacides dans la séquence d'aminoacides déduite de pOP-KT21, (B) est la composition en aminoacides de la pulpe de noyau de palmier à huile.

La figure 5 est une représentation graphique montrant le diagramme d'expression des gènes de glutéline (sous-famille B) dans différents tissus du palmier à huile. Un « slot blot » contenant 1 µg d'ARN poly(A^+) provenant de différents tissus du palmier à huile (*E. guineensis*), comme indiqué, a été hybridé avec une sonde marquée avec ^{32}P préparée à partir de EST de pOP-KT21.

La figure 6 est une représentation graphique montrant le diagramme d'expression de gènes de glutéline (sous-famille B) dans différents tissus du palmier à huile. Un Northern Blot contenant 20 µg d'ARN total provenant de différents tissus du palmier à huile (*E. guineensis*), comme indiqué, a été hybridé avec une sonde marquée avec ^{32}P contenant la séquence d'ADNc de pOP-KT21. La membrane a été sondée de nouveau avec l'ADN ribosomique pour vérifier le chargement égal des échantillons d'ARN.

La figure 7 est une représentation graphique montrant le diagramme d'expression de pOP-KT21 dans différents tissus du palmier à huile. Un Northern Blot (7a) contenant 20 µg d'ARN total provenant de différents tissus du palmier à huile (*E. guineensis*) a été hybridé avec une sonde spécifique de gène marquée avec ^{32}P préparée à partir du 5'-UTR et des 14 bases provenant de l'extrémité 5' de la région codante de pOP-KT21. Les échantillons d'ARN provenant du noyau, 8, 9, 10, 13, 15 et 17 semaines après l'anthèse, du mésocarpe à 15 semaines après l'anthèse (M15), des racines (R), de jeunes plants germés (GS) et de feuilles (SL). Un gel coloré au bromure d'éthidium (7b) a été inclus pour montrer le chargement égal des échantillons d'ARN.

La figure 8 est une représentation photomicrographique montrant le diagramme d'expression spécifique de cellules de gènes de glutéline dans le tissu de l'endosperme des palmiers à huile. L'hybridation d'ARN *in situ* d'une coupe de noyau de fruits de palmier à huile avec des sondes d'ARN marquées avec DIG codant la glutéline de palmier à huile (pOP-KT21), et visualisation avec le Leica Q55OFW Imaging Microscope System. Panneau (a) : une sonde d'ARN sens a été utilisée sur une coupe épaisse de 5 µm (grossissement 33X). Panneau (b) : une sonde ARN antisens a été utilisée sur une coupe épaisse de 5 µm (grossissement 33X).

La figure 9 est une représentation photographique d'un gel d'agarose montrant les produits de la marche sur le génome utilisés pour isoler le promoteur du gène pOP-KT21. La réaction de PCR primaire a été réalisée au moyen de l'amorce spécifique de gène GK4 provenant de la séquence 5' de la région codant pOP-KT21, l'amorce AP1 et la banque de marche sur le génome *Eco* RV du palmier à huile comme matrice. Le produit de PCR primaire a été utilisé dans une réaction de PCR secondaire utilisant des amorces AP2 et l'amorce nichée spécifique de gène GKN6 provenant du 5'-UTR de pOP-KT21. Les fragments d'ADN amplifiés provenant des deux réactions ont ensuite été purifiés à partir du gel d'agarose et clonés dans le vecteur PCR II TOPO pour le séquençage. La piste 1 est le marqueur de masse moléculaire. La piste 2 montre le produit de PCR primaire de la marche sur le génome et la piste 3 est le produit de PCR secondaire.

La figure 10 est une représentation graphique montrant la séquence du promoteur du gène de glutéline de palmier à huile (pOP-KT21). Ce promoteur est trouvé dans le clone génomique du palmier à huile pKT21-Pa et a été isolé par marche sur le génome. Dans la séquence du promoteur, les éléments régulateurs d'expression supposés (motifs analogue à Skn-1 et AACA), la boîte TATA et l'adénine à l'extrémité 5' du produit 5' RACE (site de début de transcription probable) sont soulignés.

La figure 11 est une représentation graphique montrant un alignement de séquences d'ADN de pKT21-Pa (séquence génomique) et de pOP-KT21 (séquence d'ADNc). L'alignement de la séquence PKT21-Pa et de la séquence de pOP-KT21 (nucléotides 1-128) montrant clairement 100 % d'homologie dans la région de chevauchement.

La figure 12 est une représentation photographique d'un gel d'agarose montrant les résultats d'une analyse de restriction du vecteur de transformation chimérique GluP-EGFP. GluP-EGFP a été produit en introduisant le promoteur du gène de glutéline du palmier à huile (pOP-KT21) dans les sites *Bam* HI et *Xho* I trouvés dans la région de clonage multiple du vecteur pEGFP-1 (Clontech), qui contient GFP comme gène rapporteur. Pistes 1 et 3 : pEGFP-1 et GluP-EGFP digérés avec *Bam* HI, respectivement. Pistes 2 et 4 : pEGFP-1 et GluP-EGFP digérés avec *Xho* I, respectivement. Piste 5 : GluP-EGFP digéré avec *Bam* HI et *Xho* I. Piste M : marqueur d'échelle d'ADN (Fermentas).

La figure 13 est une représentation sous forme de tableau montrant des paramètres de bombardement optimisés pour une analyse d'essai transitoire utilisant des coupes de tissu de noyau de palmier à huile.

La figure 14 représente des représentations photographiques de tissus d'endosperme, de mésocarpe et de feuilles bombardés avec GluP-EGFP et CaMV-EGFP (témoin positif). Un essai de promoteur-gène rapporteur a été réalisé pour confirmer l'activité spécifique de l'endosperme du promoteur du gène de glutéline du palmier à huile (GluP) en comparant l'expression de la protéine fluorescente verte dans les tissus de l'endosperme, du mésocarpe et foliaire du palmier à huile bombardés avec GluP-EGFP (b, d et g, respectivement) et CaMV-EGFP (a, c et f, respectivement). Comme témoin interne, les tissus d et g ont

également été co-bombardés avec CaMV-GUS et l'expression de GUS a été détectée dans ces tissus comme cela est montré en e et h respectivement.

La figure 15 est une représentation sous forme de tableau montrant les nombres de points de GFP dans les tissus de feuilles, de mésocarpe et de noyau bombardés avec GluP-EGFP, CaMV-EGFP (témoin positif) et le témoin pEGFP (témoin négatif) ; moyenne de 3 expériences indépendantes.

DESCRIPTION DETAILLEE DE L'INVENTION

La présente invention est basée en partie sur un élément régulateur d'expression qui est spécifique de tissu dans les palmiers à huile. De manière appropriée, l'élément régulateur d'expression dirige ou facilite l'expression d'un gène codant la glutéline dans l'endosperme des palmiers à huile. L'isolement de cet élément régulateur d'expression permet entre autres l'expression de molécules d'acides nucléiques en liaison fonctionnelle avec l'élément régulateur d'expression, qui, au moins dans les palmiers à huile, sera spécifique de tissu dans l'endosperme. Dans d'autres espèces végétales l'expression peut aussi être spécifique de l'endosperme ou bien encore l'expression peut être spécifique d'un ou plusieurs tissus différents ou peut avoir un spectre d'activité plus large, par exemple en facilitant l'expression dans de multiples tissus ou dans tous les tissus de la plante.

Avant de décrire la présente invention en détail, il convient de comprendre que, sauf indication contraire, la terminologie utilisée est destinée à décrire des modes de réalisation particuliers seulement et n'est pas destinée à être limitative.

Telles qu'elles sont utilisées dans la présente description, les formes singulières "un" ou "une" et "le" ou "la" incluent les aspects pluriels sauf indication contraire. Ainsi, par exemple, une référence à "un élément régulateur d'expression" inclut un seul élément, ainsi que deux ou plusieurs éléments ou une combinaison d'éléments. Par exemple, "un élément régulateur d'expression" peut comprendre une seule région promotrice, de multiples régions promotrices, ou une ou plusieurs régions promotrices en même temps que d'autres éléments régulateurs d'expression comme des stimulateurs ou des silenceurs.

Ainsi, un aspect de la présente invention envisage une molécule d'acide nucléique isolée comprenant un élément régulateur d'expression spécifique de tissu, qui est spécifique de tissu dans le tissu de l'endosperme du palmier à huile et qui facilite l'expression d'un ADN lié de manière fonctionnelle audit élément régulateur dans des cellules végétales.

Le terme "élément régulateur d'expression" tel qu'il est utilisé ici devrait être compris comme se référant à toute séquence nucléotidique qui facilite l'expression d'une seconde séquence d'acide nucléique couplée de manière fonctionnelle à celle-ci. Dans un mode de réalisation particulier, l'élément régulateur d'expression est un promoteur. Cependant, comme indiqué ci-dessus, les stimulateurs et les silenceurs sont également envisagés par la présente invention. Le terme "spécifique de tissu" quand il est utilisé adjectivement en combinaison avec "l'élément régulateur d'expression" devrait être compris comme se référant à un élément régulateur d'expression qui dirige l'expression d'une séquence d'acide nucléique dans un ou plusieurs tissus, mais pas dans tous les tissus, d'une plante. Il conviendrait de noter que, bien que les éléments régulateurs d'expression illustrés à titre d'exemple qui sont décrits ici soient spécifiques de l'endosperme dans les palmiers à huile, dans d'autres espèces végétales ils peuvent aussi être spécifiques de l'endosperme ou bien encore ils peuvent être spécifiques d'un ou plusieurs tissus différents ou peuvent avoir un spectre d'activité plus large, par exemple en facilitant l'expression dans de multiples tissus de la plante ou dans tous les tissus de la plante.

Un "promoteur" est défini comme étant toute séquence d'acide nucléique qui facilite l'expression d'une seconde molécule d'acide nucléique liée de manière fonctionnelle à celle-ci.

Dans un mode de réalisation préféré de la présente invention, l'élément régulateur d'expression est un promoteur.

Cependant, la présente invention s'étend aussi à d'autres éléments régulateurs d'expression comme un stimulateur ou un silenceur.

Un "stimulateur" lie des facteurs de transcription connus comme étant des "protéines liant un stimulateur", qui peuvent aussi se

lier au promoteur d'un gène. La liaison d'un stimulateur et d'un promoteur à un facteur de transcription amène les deux régions en association, ce qui affecte ensuite l'activité du promoteur et augmente le taux de transcription de tout gène lié de manière fonctionnelle.

Les "silenceurs" sont des régions de contrôle de l'ADN qui, comme les stimulateurs, se lient à des facteurs de transcription. Cependant, contrairement aux stimulateurs, quand des facteurs de transcription se lient à un silenceur, l'expression du gène qu'ils contrôlent est réprimée.

L'élément régulateur d'expression de la présente invention est isolé de l'amont du gène codant la glutéline dans le palmier à huile. La glutéline est une protéine de réserve des graines qui s'accumule dans l'endosperme des plantes monocotylédones. "L'endosperme" est le nom donné au tissu triploïde riche en nutriments constitué par des cellules parenchymateuses, qui entoure l'embryon dans les graines végétales. Dans le palmier à huile, l'endosperme peut être appelé de manière interchangeable "noyau". La présente invention a identifié le promoteur du gène de glutéline dans le palmier à huile, et a identifié que ce promoteur conduit à une expression spécifique de l'endosperme d'une molécule d'acide nucléique couplée de manière fonctionnelle à celui-ci. La présente invention envisage entre autres l'utilisation des caractéristiques de ce promoteur pour exprimer spécifiquement des transgènes dans l'endosperme de plantes transgéniques. Il est envisagé que la re-direction du flux métabolique en direction de la production d'un produit hétérologue dans un tissu de réserve, comme l'endosperme, devrait être possible avec peu de coût métabolique pour la plante en termes de croissance et de vigueur.

Tel qu'il est utilisé ici, le terme "palmier à huile" devrait être compris comme désignant les palmiers à huile comme les espèces *E. guineensis* et *E. oleifera* incluant les variétés modifiées ou les variants génétiques de ceux-ci.

L'élément régulateur d'expression spécifique de tissu décrit ici est illustré à titre d'exemple concernant le promoteur facilitant l'expression du gène codant la glutéline dans l'endosperme du palmier à huile, bien qu'il soit envisagé que les éléments régulateurs d'expression décrits ici soient actifs dans une gamme d'espèces végétales en

facilitant l'expression d'une gamme de molécules d'acides nucléiques liées de manière fonctionnelle. Les exemples de telles molécules d'acides nucléiques incluent les séquences d'ADN qui sont transcrites en ARN codant comme l'ARNm qui est traduit ensuite en peptides, polypeptides ou protéines ; et un ADN qui est transcrit en un ARN ne codant pas une protéine (ARNnc), qui inclut, mais qui n'est pas limité à, l'ARNf, l'ARNmi, l'ARNsi, l'ARNsn, L'ARNsnm, l'ARNst, l'ARNt, l'ARNr, l'ARNsno, les transcrits d'empreinte, l'ARN sens, l'ARN antisens et analogues.

Plus particulièrement, la présente invention concerne une molécule d'acide nucléique ou un dérivé ou homologue de celle-ci comprenant une séquence nucléotidique codant un élément régulateur d'expression, où la séquence nucléotidique telle est présentée dans SEQ ID NO : 3 ou une séquence nucléotidique ayant au moins environ 50 % d'identité avec SEQ ID NO : 3 après l'alignement optimal ou une séquence nucléotidique capable de s'hybrider à SEQ ID NO : 3 ou sa forme complémentaire dans des conditions de faible stringence.

Le terme "similarité" tel qu'il est utilisé ici inclut l'identité exacte entre des séquences comparées au niveau nucléotidique ou au niveau des aminoacides. Quand il y a une non-identité au niveau nucléotidique, la "similarité" inclut les différences entre les séquences qui conduisent à des aminoacides différents qui sont néanmoins apparentés entre eux aux niveaux structural, fonctionnel, biochimique et/ou conformationnel. Quand il y a une non-identité au niveau des aminoacides , la "similarité" inclut des aminoacides qui sont néanmoins liés entre eux aux niveaux structural, fonctionnel, biochimique et/ou conformationnel. Dans un mode de réalisation particulièrement préféré, des comparaisons des séquences nucléotidiques et d'aminoacides sont faites au niveau de l'identité plutôt qu'au niveau de la similarité.

Les termes utilisés pour décrire les relations de séquence entre deux ou plusieurs polynucléotides ou polypeptides incluent "séquence de référence", "fenêtre de comparaison", "similarité de séquences", "identité de séquences", "pourcentage de similarité de séquences", "pourcentage d'identité de séquences", "sensiblement similaires" et "sensiblement identiques". Une "séquence de référence" a une longueur d'au moins 12 mais fréquemment de 15 à 18 et souvent

d'au moins 25 ou plus, comme 30 unités monomères, y compris les nucléotides et les résidus d'aminoacide. Du fait que deux polynucléotides peuvent comprendre chacun (1) une séquence (c'est-à-dire seulement une partie de la séquence polynucléotidique complète) qui est similaire entre les deux polynucléotides, et (2) une séquence qui est divergente entre les deux polynucléotides, les comparaisons de séquences entre deux (ou plusieurs) polynucléotides sont typiquement réalisées en comparant les séquences des deux polynucléotides sur une "fenêtre de comparaison" pour identifier et comparer des régions locales de similarité de séquence. Une "fenêtre de comparaison" désigne un segment conceptuel typiquement de 12 résidus contigus qui est comparé à une séquence de référence. La fenêtre de comparaison peut comprendre des additions ou des délétions (c'est-à-dire des vides) d'environ 20 % ou moins par rapport à la séquence de référence (qui ne comprend ni addition ni délétion) pour l'alignement optimal des deux séquences. L'alignement optimal de séquences pour aligner une fenêtre de comparaison peut être réalisé par des mises en œuvre informatiques d'algorithmes (GAP, BESTFIT, FASTA et TFASTA dans le Wisconsin Genetics Software Package Release 7.0, Genetics Computer Group, 575 Science Drive Madison, W1, USA) ou par inspection du meilleur alignement (c'est-à-dire conduisant à la plus grande homologie en pourcentage sur la fenêtre de comparaison) produite par l'un quelconque des différents procédés choisis. Il est possible aussi de se référer à la famille de programmes BLAST, décrite par exemple par Altschul *et al.* (*Nucl. Acids Res. 25:* 3389-3402, 1997). Une discussion détaillée de l'analyse de séquences peut être trouvée dans Unit 19.3 d'Ausubel *et al.* ("Current Protocols in Molecular Biology", John Wiley & Sons Inc., 1994-1998, chapitre 15, 1998).

Les termes "similarité de séquences" et "identité de séquences" tels qu'ils sont utilisés ici désignent l'étendue selon laquelle les séquences sont identiques ou fonctionnellement ou structuralement similaires sur une base de nucléotide par nucléotide ou sur une base d'aminoacide par aminoacide sur une fenêtre de comparaison. Ainsi, un "pourcentage d'identité de séquences", par exemple, calculé en comparant deux séquences alignées de manière optimale sur la fenêtre de comparaison, en déterminant le nombre de positions auxquelles

survient la base d'acide nucléique identique (par exemple A, T, C, G, I) ou le résidu d'aminoacide identique (par exemple Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys et Met) dans les deux séquences pour produire le nombre de positions appariées, divisant le nombre de positions appariées par le nombre total de positions dans la fenêtre de comparaison (c'est-à-dire la taille de la fenêtre) et multipliant le résultat par 100 pour donner le pourcentage d'identité de séquences. Pour les besoins de la présente invention, "l'identité de séquences" sera comprise comme signifiant le "pourcentage d'appariement" calculé par le programme d'ordinateur DNASIS (version 2.5 pour Windows ; disponible auprès de Hitachi Software engineering Co., Ltd., South San Francisco, Californie, USA) en utilisant des défauts standard tels qu'ils sont utilisés dans le manuel de référence accompagnant le logiciel. Des commentaires similaires s'appliquent à la similarité de séquences. Le terme "similarité" est particulièrement utile pour décrire des comparaisons de séquences d'aminoacides. Le terme "identité" est particulièrement utile pour décrire des comparaisons de séquences nucléotidiques.

Une référence à plus de 50 % d'identité ou de similarité inclut des identités en pourcentage et des similarités en pourcentage supérieures à 50 % comme 51 %, 52 %, 53 %, 54 %, 55 %, 56 %, 57 %, 58 %, 59 %, 60 %, 61 %, 62 %, 63 %, 64 %, 65 %, 66 %, 67 %, 68 %, 69 %, 70 %, 71 %, 72 %, 73 %, 74 %, 75 %, 76 %, 77 %, 78 %, 79 %, 80 %, 81 %, 82 %, 83 %, 84 %, 85 %, 86 %, 87 %, 88 %, 89 %, 90 %, 91 %, 92 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % et 100 %.

De préférence, l'élément régulateur d'expression comprend une molécule d'acide nucléique isolée comprenant la séquence d'acide nucléique présentée dans SEQ ID NO : 3.

La présente invention envisage aussi des mutants et des dérivés de l'élément régulateur d'expression ou d'un ADN comprenant l'élément régulateur d'expression, et en particulier les mutants ou fragments qui conservent une activité. Par exemple, un élément régulateur d'expression peut comprendre un promoteur et un élément stimulateur, qui peuvent être séparés en des fragments séparés, qui seraient situés l'un et l'autre dans le cadre de la présente invention. De

plus, l'élément régulateur d'expression de la présente invention peut être soumis à une analyse de délétion pour déterminer le fragment fonctionnel minimal nécessaire pour diriger l'expression génique dans des cellules parenchymateuses, ou dans le tissu de l'endosperme. A titre d'alternative, de tels mutants de délétion peuvent avoir une activité augmentée ou diminuée par délétion de régions stimulateurs ou silenceurs dans l'élément régulateur d'expression. De tels éléments régulateurs d'expression sont également envisagés par la présente invention.

Un autre aspect de la présente invention concerne des constructions d'ADN recombinant comprenant une séquence nucléotidique comprenant l'élément régulateur d'expression ou un fragment de celui comme décrit ici.

Les "constructions d'ADN recombinant" envisagées par la présente invention incluent toute molécule d'acide nucléique isolée comprenant l'élément régulateur d'expression. Les constructions d'ADN recombinant particulièrement utiles incluent les vecteurs comme les vecteurs de clonage et les vecteurs d'expression.

Le terme "vecteurs" tel qu'il est utilisé ici désigne une construction utilisée dans le but de multiplier, propager, transférer un acide nucléique d'un organisme à un autre (par exemple vecteurs de clonage) ou d'exprimer un acide nucléique (qui peut éventuellement coder une protéine) dans une cellule (vecteurs d'expression). Ainsi que l'homme du métier le déterminera aisément, dans certains cas, selon la molécule d'acide nucléique d'intérêt, un vecteur peut comprendre à la fois un vecteur de clonage et un vecteur d'expression. Par exemple, un vecteur de clonage avec un marqueur sélectionnable exprimé serait un vecteur de clonage concernant l'acide nucléique cloné, mais un vecteur d'expression concernant le gène marqueur sélectionnable. Les vecteurs particulièrement utiles incluent : les vecteurs plasmidiques, les vecteurs viraux, les vecteurs de type transposon, les cosmides et les chromosomes artificiels (incluant les chromosomes artificiels bactériens et les chromosomes artificiels de levure). D'autres vecteurs apparaîtront aisément à l'homme du métier, et la présente invention n'est absolument pas limitée par les types de vecteur énumérés ci-dessus.

Ainsi, un autre aspect de la présente invention envisage un vecteur comprenant un acide nucléique de la présente invention comme défini précédemment.

Dans un mode de réalisation préféré, le vecteur est un vecteur d'expression qui est capable de faciliter l'expression d'un acide nucléique dans des cellules végétales, où le vecteur d'expression comprend un élément régulateur d'expression comme décrit ici en liaison fonctionnelle avec une molécule d'acide nucléique hétérologue et où l'élément régulateur d'expression, au moins dans les palmiers à huile, dirige l'expression spécifique de la molécule d'acide nucléique hétérologue dans l'endosperme.

La mise en place d'une molécule d'acide nucléique sous le contrôle fonctionnel d'un élément régulateur d'expression, comme un promoteur, signifie le positionnement de la molécule de telle manière que l'expression est contrôlée par l'élément régulateur d'expression. Par exemple, les promoteurs sont généralement positionnés en 5' (en amont) des gènes qu'ils contrôlent. Dans la construction de combinaisons promoteur/gène structural hétérologues, on préfère généralement positionner le promoteur à une distance du site de début de la transcription du gène qui est approximativement la même que la distance entre ce promoteur et le gène qu'il contrôle dans son agencement naturel, c'est-à-dire le gène duquel le promoteur est dérivé. Comme on le sait dans la technique, une certaine variation dans cette distance peut être admise sans perte de fonction du promoteur. De manière similaire, le positionnement préféré d'un élément de séquence régulatrice par rapport à un gène hétérologue destiné à être placé sous son contrôle est défini par le positionnement de l'élément dans son agencement naturel, c'est-à-dire les gènes duquel il est dérivé. Là encore, comme on le sait dans la technique, une certaine variation dans cette distance peut également survenir.

La construction ou le vecteur contient de préférence des éléments régulateurs d'expression supplémentaires pour la transcription efficace, par exemple une séquence de terminaison de la transcription.

Le terme "terminateur" désigne une séquence d'ADN à la fin d'une unité de transcription qui signale la terminaison de la transcription. Les terminateurs sont des séquences d'ADN non traduites

en 3' contenant généralement un signal de polyadénylation qui facilite l'addition de séquences de polyadénylates à l'extrémité 3' d'un transcrit primaire. Des terminateurs actifs dans les cellules végétales sont connus et décrits dans la littérature. Ils peuvent être isolés à partir de bactéries, de champignons, de virus, d'animaux et/ou de plantes ou peuvent être synthétisés *de novo*.

Le terminateur peut être toute séquence terminatrice qui est fonctionnelle dans les cellules, les tissus ou les organes dans lesquels il est envisagé d'être utilisé.

Des exemples de terminateurs particulièrement appropriés pour être utilisés dans les gènes synthétiques de la présente invention incluent le signal de polyadénylation de SV40, le signal de polyadénylation de HSV TK, le terminateur CYCI, le terminateur ADH, le terminateur SPA, le terminateur du gène de nopaline synthase (NOS) de *Agrobacterium tumefaciens*, le terminateur du gène 35S du virus de la mosaïque du chou-fleur (CaMV), le terminateur du gène de zéine de *Zea mays*, les séquences terminatrices de gènes de la petite sous-unité du gène Rubisco (SSU), les terminateurs de séquence génique du virus du rabougrissement du trèfle (SCSV), tout terminateur de *E. coli* indépendant de *rho*, ou le terminateur de *lacZ* alpha, parmi d'autres.

Dans un mode de réalisation particulièrement préféré, le terminateur est le terminateur de l'octopine synthase (OCS) ou le terminateur de nopaline synthase (NOS) qui est actif dans les cellules, les tissus ou les organes végétaux.

L'homme du métier connaîtra d'autres séquences terminatrices qui peuvent être appropriées pour être utilisées dans la mise en œuvre de l'invention. De telles séquences peuvent être utilisées aisément sans expérimentation excessive.

Les constructions décrites ci-dessus sont capables d'être modifiées encore. Par exemple, par l'inclusion de séquences nucléotidiques marqueurs codant une enzyme marqueur détectable ou un analogue fonctionnel ou dérivé fonctionnel de celle-ci, pour faciliter la détection du gène synthétique dans une cellule, un tissu ou un organe où il est exprimé. Selon ce mode de réalisation, les séquences nucléotidiques marqueurs seront présentes dans un format pouvant être traduit et seront exprimées.

L'homme du métier saura comment produire les constructions décrites ici et les conditions pour obtenir leur expression. En particulier, l'homme du métier connaîtra les manipulations génétiques nécessaires pour réaliser la présente invention qui peuvent nécessiter la propagation d'une construction génétique décrite ici ou un dérivé de celle-ci dans une cellule procaryote comme une cellule de *E. coli* ou d'*Agrobacterium* sp. ou une cellule végétale ou une cellule animale.

Pour produire une construction génétique, un acide nucléique est inséré dans un vecteur approprié ou une molécule d'épisome appropriée, comme un vecteur bactériophage, un vecteur viral ou un plasmide, un cosmide ou un vecteur de type chromosome artificiel qui est capable d'être maintenu et/ou répliqué et/ou exprimé dans la cellule hôte, le tissu hôte ou l'organe hôte dans lequel il est ensuite introduit.

Ainsi, un autre aspect de l'invention fournit une construction génétique qui comprend, entre autres, un élément régulateur d'expression tel que décrit ici et éventuellement une ou plusieurs origines de réplication et/ou une ou plusieurs séquences de gènes marqueurs sélectionnables.

Typiquement, une origine de réplication ou un gène marqueur sélectionnable approprié pour être utilisé dans des bactéries est séparé physiquement de ces séquences génétiques contenues dans la construction génétique qui sont destinées à être exprimées ou transférées dans une cellule végétale, ou intégrées dans le génome d'une cellule végétale.

Tel qu'il est utilisé ici, le terme "gène marqueur sélectionnable" inclut tout gène qui confère un phénotype à une cellule dans laquelle il est exprimé pour faciliter l'identification et/ou la sélection de cellules qui sont transfectées ou transformées avec une construction génétique de l'invention ou un dérivé de celle-ci.

Les gènes marqueurs sélectionnables appropriés envisagés ici incluent le gène de résistance à l'ampicilline (Amp^r), le gène de résistance à la tétracycline (Tc^r), le gène de résistance à la kanamycine bactérien (Kan^r), le gène de résistance à la zéocine (la Zéocine est un médicament de la famille de la bléomycine qui est une marque déposée de InVitrogen Corporation), le gène *AURI-C* qui confère une résistance à l'antibiotique auréobasidine A, le gène de résistance à la

phosphinothricine, le gène de la néomycine phosphotransférase (*npt* II), le gène de résistance à l'hygromycine, le gène de la β-glucuronidase (GUS), le gène de la chloramphénicol acétyltransférase (CAT), le gène codant la protéine fluorescente verte ou le gène de la luciférase, parmi d'autres.

De préférence, le gène marqueur sélectionnable est le gène *npt* II ou le gène Kan^r ou le gène codant la protéine fluorescente verte (GFP).

L'homme du métier connaîtra d'autres gènes marqueurs sélectionnables utiles dans la mise en œuvre de la présente invention et la présente invention n'est pas limitée par la nature du gène marqueur sélectionnable.

La présente invention s'étend à toutes les constructions génétiques essentiellement telles que décrites ici, qui incluent d'autres séquences génétiques destinées au maintien et/ou à la réplication de ladite construction génétique dans les procaryotes ou les eucaryotes et/ou l'intégration de ladite construction génétique ou d'une partie de celle-ci dans le génome d'une cellule ou organisme eucaryote.

Un autre aspect de la présente invention envisage un procédé pour exprimer une molécule d'acide nucléique dans une plante transgénique, ledit procédé comprenant la transformation d'une plante, d'une cellule végétale, d'un tissu végétal ou d'un organe végétal, avec le vecteur décrit ici et la régénération de la plante transformée.

De préférence, la molécule d'acide nucléique est exprimée spécifiquement dans l'endosperme de la plante transformée ou du tissu transformé.

Des procédés standard peuvent être utilisés pour introduire les constructions dans la cellule, le tissu ou l'organe, par exemple la transfection ou transformation médiée par des liposomes, la transformation de cellules avec des particules virales atténuées ou des cellules bactériennes, l'appariement cellulaire, des procédures de transformation ou de transfection connues de l'homme du métier.

Des moyens particulièrement utiles pour introduire un ADN recombinant dans un tissu végétal ou des cellules végétales incluent, mais ne sont pas limités à, le bombardement par des microparticules d'un explant tissulaire ou de cellules, la transformation avec $CaCl_2$ et

des variations de celles-ci, l'absorption directe d'ADN dans des protoplastes, l'absorption médiée par le PEG dans des protoplastes, l'électroporation, la micro-injection d'ADN, l'infiltration sous vide d'un tissu avec un acide nucléique et le transfert à médiation par l'ADN T de *Agrobacterium* dans le tissu végétal.

Pour le bombardement de cellules avec des microparticules, une microparticule est propulsée dans une cellule pour produire une cellule transformée. Toute méthodologie de transformation cellulaire balistique appropriée et tout appareil correspondant peuvent être utilisés dans la mise en œuvre de la présente invention. Des exemples d'appareils et de procédures sont décrits par Stomp *et al.* (brevet U.S. n° 5 122 466) et Sanford et Wolf (brevet U.S. n° 4 945 050). Quand on utilise des procédures de transformation balistique, la construction génétique peut incorporer un plasmide capable de se répliquer dans la cellule destinée à être transformée.

Des exemples de microparticules appropriées pour être utilisées dans de tels systèmes incluent les sphères d'or de 1 à 5 µm. La construction d'ADN peut être déposée sur la microparticule par toute technique appropriée, par exemple par précipitation.

Pour la transformation à médiation par *Agrobacterium*, la construction elle-même peut comprendre des séquences nécessaires pour faciliter son propre transfert dans le tissu végétal, par exemple elle peut comprendre une région codant les gènes *vir* du plasmide Ti. Typiquement, ces vecteurs sont construits par recombinaison homologue entre un plasmide Ti de *Agrobacterium* désarmé (c'est-à-dire un plasmide dont les régions tumorigènes de l'ADN T ont subi une inactivation (« knock-out »)) et une construction contenant le gène d'intérêt et une région d'homologie avec le plasmide Ti désarmé. Un vecteur de ce type est appelé "vecteur co-intégré". Cependant, la région *vir* peut aussi être présente sur une construction séparée du gène destiné à être transféré, connue comme étant un système de vecteur binaire. Dans ce cas, les gènes destinés à être transférés (typiquement au moins le gène d'intérêt et un marqueur sélectionnable) sont flanqués par les régions bordant l'ADN T provenant d'un plasmide Ti sur une construction. De plus, dans le donneur *Agrobacterium*, une seconde construction doit être présente, qui comprend la région *vir* du

plasmide Ti, et qui fournit le mécanisme de transfert en *trans*. Typiquement, cette construction codant *vir* n'a pas de région d'ADN T, et aucune partie de cette construction "auxiliaire" n'est transférée. En plus de ces caractéristiques essentielles, il apparaîtra aisément à l'homme du métier que les constructions décrites ci-dessus peuvent inclure en outre d'autres caractéristiques telles que les suivantes, sans être limitées à celles-ci : des origines de réplication bactérienne ; des gènes de transfert bactérien comme *mob* ; des gènes marqueurs sélectionnables et analogues.

Dans un autre mode de réalisation de la présente invention, les constructions génétiques décrites ici sont adaptées pour l'intégration dans le génome d'une cellule dans laquelle elles sont exprimées. L'homme du métier saura que, pour obtenir l'intégration d'une séquence génétique ou d'une construction génétique dans le génome d'une cellule hôte, certaines séquences génétiques supplémentaires peuvent être nécessaires. Dans le cas des plantes, les séquences bordant à gauche et à droite l'ADN T du plasmide Ti de *Agrobacterium tumefaciens* seront généralement nécessaires.

Des plantes peuvent être régénérées à partir de cellules ou tissus ou organes végétaux transformés sur des milieux de régénération contenant des hormones ou sans hormones et les plantes régénérées peuvent revêtir différentes formes, comme des chimères de cellules transformées et de cellules non transformées ; des transformants clonaux (par exemple toutes les cellules transformées pour contenir la cassette d'expression) ; des greffes de tissus transformés et non transformés (par exemple un ensemble de racines transformées greffées à un scion non transformé dans les espèces d'agrumes). Les plantes transformées peuvent être propagées par différents moyens, par exemple par propagation clonale ou des techniques de reproduction classique. Par exemple, une première génération (ou T1) de plantes transformées peut être autofécondée pour donner une seconde génération homozygote (ou T2) de plantes transformées, et les plantes T2 peuvent être propagées encore par des techniques de reproduction classiques.

Les procédés envisagés par cet aspect de la présente invention, bien qu'ils soient applicables à toute espèce végétale, sont

particulièrement utiles pour les plantes choisies dans les espèces consistant en : *Elaeis guineensis* Jacq., *Elaeis oleifera* et leurs hybrides ; le maïs, le riz, le soja, le tabac, le coton, la luzerne, le blé, *Arabidopsis*, le colza, le canola et le cocotier.

Dans un mode de réalisation particulièrement préféré, les procédés de la présente invention sont utilisés pour transformer *Elaeis* spp. avec un acide nucléique qui, quand il est exprimé, réalise :

(i) une modification des qualités des protéines ou des huiles ;

(ii) la production ou l'accumulation de nouvelles huiles ou de nouvelles protéines ; ou

(iii) la production ou l'accumulation de nutraceutiques, de produits pharmaceutiques et/ou d'autres produits industriels ;

(iv) la production de planticorps.

dans l'endosperme de la plante transformée. Il est envisagé que l'expression d'un transgène sous le contrôle fonctionnel de l'élément régulateur d'expression de la présente invention sera sensiblement limitée à l'endosperme de la plante de sorte que le métabolisme dans d'autres parties de la plante, incluant l'accumulation d'huile dans le mésocarpe, ne sera pas affecté.

Il conviendrait aussi de noter que les procédés, les éléments régulateurs d'expression et les constructions de la présente invention peuvent être appliqués de manière équivalente pour l'expression d'un acide nucléique d'une plante différente du palmier à huile (*Elaeis* spp.). Ainsi, la présente invention n'est absolument pas limitée aux seules plantes *Elaeis* spp., et les procédés, les éléments régulateurs d'expression et les constructions de la présente invention sont applicables à toute plante.

Un procédé particulièrement utile pour modifier la teneur en huile de l'endosperme comprend l'expression d'une molécule d'acide nucléique qui affecte la biosynthèse des acides gras ou l'accumulation des acides gras. Des exemples de tels acides nucléiques incluent les acides nucléiques codant des enzymes de biosynthèse des acides gras et des acides nucléiques antisens (incluant ADN et ARNi) à des gènes de biosynthèse des acides gras.

Des exemples d'accumulation d'huiles de réserve incluent l'augmentation des niveaux d'acides gras à chaîne moyenne incluant l'acide laurique et l'acide myristique ou l'acide oléique C18 : 1 pour des applications dans l'industrie oléochimique. Ceci impliquerait une manipulation génétique de gènes codant des enzymes dans la voie de biosynthèse des acides gras végétaux comme l'acyl-ACP thioestérase (fatA et fatB), l'acétyl-CoA carboxylase, la β-cétoacyl-ACP synthase (KASI, KAS II, KASIII et KAS IV). Un gène ou une combinaison de gènes peuvent être manipulés pour produire une huile ayant la composition en acides gras souhaitée.

Des exemples d'accumulation de nutraceutiques ou de produits pharmaceutiques qui sont envisagés par la présente invention incluent :

(i) l'accumulation de pro-vitamine A ;

(ii) l'accumulation ou la biosynthèse de composés riches en fer ;

(iii) la biosynthèse ou l'accumulation de peptides antigéniques destinés à être utilisés comme vaccins oraux.

La pulpe de noyau de palmier à huile est communément utilisée comme aliment pour animaux. Ainsi, la présente invention envisage un procédé de production d'une plante génétiquement modifiée ayant une composition en aminoacides modifiée ou une composition en huile modifiée dans le noyau, de sorte que le noyau a plus grande valeur nutritionnelle que le noyau de palmier à huile de type sauvage.

Un exemple de modification de la valeur nutritionnelle de la pulpe du noyau de palmier à huile est l'expression d'une protéine dans l'endosperme qui contient des niveaux accrus d'un ou plusieurs des aminoacides "essentiels" pour les mammifères, et qui contient des niveaux équilibrés de tous les aminoacides essentiels pour les mammifères.

Les procaryotes comme *E. coli* peuvent produire les squelettes carbonés des 20 aminoacides et peuvent transaminer ces squelettes carbonés avec de l'azote provenant de la glutamine ou du glutamate pour réaliser les structures des aminoacides. Les êtres humains ne peuvent pas synthétiser les chaînes carbonées ramifiées

trouvées dans les aminoacides ramifiés ou les systèmes cycliques trouvés dans la phénylalanine et les aminoacides aromatiques ; les êtres humains ne peuvent pas non plus incorporer du soufre dans les structures à liaison covalente. De ce fait, les 10 aminoacides appelés aminoacides essentiels (voir le tableau 3) doivent être fournis par l'alimentation. Néanmoins, il conviendrait de reconnaître que, selon la composition de l'alimentation et l'état physiologique d'un individu, l'un ou l'autre des aminoacides non essentiels peut aussi devenir un composant élémentaire nécessaire. Par exemple, l'arginine est seulement normalement considérée comme étant un aminoacide essentiel pendant le développement précoce des enfant car le cycle de l'urée en fournit suffisamment pour les adultes.

Pour prendre un type d'exemple différent, la cystéine et la tyrosine sont considérées comme étant non essentielles mais sont formées à partir des aminoacides essentiels méthionine et phénylalanine, respectivement. Si suffisamment de cystéine et de tyrosine sont présentes dans l'alimentation, les besoins en méthionine et en phénylalanine sont nettement réduits ; inversement, si la méthionine et la phénylalanine sont présentes seulement en quantités limitées, la cystéine et la tyrosine peuvent devenir des composants alimentaires essentiels. Finalement, il conviendrait de reconnaître que si les α-cétoacides correspondants aux squelettes carbonés des aminoacides essentiels sont fournis dans l'alimentation, les aminotransférases dans l'organisme convertiront les cétoacides en leurs aminoacides respectifs, en fournissant largement les besoins de base.

Contrairement aux graisses et aux glucides, l'azote n'a pas de dépôt de réserve déterminé dans l'organisme. Comme la demi-vie de nombreuses protéines est courte (de l'ordre de quelques heures), des quantités alimentaires insuffisantes de même un aminoacide peuvent limiter rapidement la synthèse et réduire le niveau corporel de nombreuses protéines essentielles. Le résultat de la synthèse limitée et des taux normaux de dégradation des protéines est que l'équilibre de l'absorption d'azote et de l'excrétion d'azote est rapidement modifié de manière significative. Les adultes sains normaux sont généralement équilibrés en ce qui concerne l'azote, l'absorption et l'excrétion étant bien appariées. Les jeunes enfants en phase de croissance, les adultes

qui récupèrent d'une maladie grave et les femmes enceintes sont souvent en équilibre d'azote positif. Leur absorption d'azote dépasse leur perte car une synthèse de protéines nette se déroule. Quand une plus grande quantité d'azote est excrétée que celle qui est incorporée dans l'organisme, un individu est en équilibre de l'azote négatif. Des quantités insuffisantes de même un aminoacide essentiel sont adéquates pour faire passer un individu autrement normal en un individu ayant un équilibre d'azote négatif.

La valeur biologique des protéines alimentaires est liée à la mesure dans laquelle elles fournissent tous les aminoacides essentiels nécessaires. Les protéines d'origine animale ont généralement une grande valeur biologique ; les protéines végétales ont une large gamme de valeurs allant de sensiblement aucune valeur à une valeur relativement élevée. En général, les protéines végétales sont déficientes en lysine, méthionine et tryptophane et sont bien moins concentrées et moins digestibles que les protéines animales. L'absence de lysine dans les protéines de céréales de basse qualité utilisées comme principales sources alimentaires dans de nombreux pays sous-développés, conduit à une inaptitude à synthétiser les protéines (du fait du manque d'aminoacides essentiels) et finalement à un syndrome connu comme étant le kwashiorkor, commun parmi les enfants dans ces pays.

Ainsi, la présente invention envisage un procédé pour augmenter la valeur nutritionnelle de la pulpe du noyau du palmier à huile, par expression d'un ou plusieurs acides nucléiques sous le contrôle de l'élément régulateur d'expression décrit ici, où l'acide nucléique code une protéine ayant des niveaux accrus d'un ou plusieurs aminoacides essentiels par rapport à la protéine glutéline de palmier à huile de type sauvage.

Dans des modes de réalisation préférés, l'acide nucléique code une protéine ayant des niveaux accrus d'un ou plusieurs acides aminés parmi la lysine, la méthionine et le tryptophane par rapport à la protéine glutéline de palmier à huile de type sauvage.

TABLEAU 3

Aminoacides essentiels et aminoacides non essentiels

Non essentiels	**Essentiels**
Alanine	Arginine*
Asparagine	Histidine
Aspartate	Isoleucine
Cystéine	Leucine
Glutamate	Lysine
Glutamine	Méthionine*
Glycine	Phénylalanine*
Proline	Thréonine
Sérine	Tryptophane
Tyrosine	Valine

*Les aminoacides arginine, méthionine et phénylalanine sont considérés comme étant essentiels pour des raisons qui ne sont pas liées directement au manque de synthèse. L'arginine est synthétisée par les cellules de mammifère mais a un taux qui est insuffisant pour répondre aux besoins de croissance de l'organisme et la majorité qui est synthétisée est clivée pour former de l'urée. La méthionine est nécessaire en grandes quantités pour produire la cystéine si ce dernier aminoacide n'est pas fourni de manière adéquate dans l'alimentation. De manière similaire, la phénylalanine est nécessaire en grandes quantités pour former la tyrosine si cette dernière n'est pas fournie de manière adéquate dans l'alimentation.

Dans un autre mode de réalisation préféré, la molécule d'acide nucléique liée de manière fonctionnelle à l'élément régulateur d'expression code un planticorps. Les planticorps sont des anticorps humains ou d'autres animaux produits par et dans des plantes transgéniques. Les planticorps sont produits dans des plantes transgéniques par l'expression d'un acide nucléique codant un anticorps humanisé.

Tel qu'il est utilisé ici, le terme "planticorps" devrait être compris comme englobant tout anticorps complet, incluant les anticorps à une seule chaîne comme les anticorps à fragment variable à une seule chaîne (scFv), les anticorps à chaînes multiples et leurs

monomères, qui sont produits dans une plante. Les monomères d'anticorps envisagés par la présente invention incluent par exemple les chaînes légères et lourdes d'anticorps de mammifère. Le "planticorps" produit par la présente invention n'est pas nécessairement produit sous une forme fonctionnelle par la plante pour être englobé par la présente invention. Par exemple, l'invention envisage spécifiquement la production d'une chaîne ou monomère d'un anticorps multimérique dans une plante, qui peut ensuite être ultérieurement complexé avec d'autres chaînes ou monomères pour produire l'anticorps multimérique fonctionnel. La présente invention envisage aussi la transformation de deux ou plusieurs plantes avec différentes constructions codant des chaînes d'anticorps distinctes mais complémentaires, par exemple quand les plantes sont croisées, la descendance peut hériter de chacun des gènes codant des chaînes d'anticorps provenant de chaque parent puis être capable de produire un anticorps/planticorps fonctionnel.

La production d'un anticorps dans une plante sous forme de planticorps peut être bien plus efficace en termes de production d'anticorps qu'une culture de cellules de mammifère. Par exemple, une culture de cellules de mammifère de 10 000 litres donne typiquement 1 ou 2 kg d'anticorps utilisables. Au contraire, des planticorps ont été produits à des rendements d'environ 1,5 kg d'anticorps de qualité pharmaceutique par acre (4047 m^2) de maïs.

Les planticorps peuvent aussi réduire le risque de contamination, car les plantes ne sont pas sensibles aux maladies humaines comme le sont les cultures de cellules de mammifère. Bien que l'on doive s'assurer que les planticorps soient dépourvus de pesticides et d'autres types de contaminants, le criblage pour les virus et les toxines bactériennes peut être sensiblement réduit.

Plusieurs espèces végétales ont été utilisées dans la technique pour produire des planticorps, et certains exemples sont indiqués dans le tableau 4.

TABLEAU 4

Planticorps thérapeutiques et diagnostiques

Application et spécificité	Séquences signal	Nom ou type d'anticorps	Plante
Caries dentaires ; streptococciques	Peptides signal IgG murins	Guy's 13 (IgA sécrétoire)	*Nicotiana tabaccum*
Diagnostic : anti- IgG humaine	Peptides signal IgG murins	C5-1 (IgG)	Luzerne
Traitement du cancer ; antigène carcinoembryonnaire	Peptide signal IgG murin ; KDEL	ScFvT84.66 (SvFv)	*Triticum aestivum*
Traitement du cancer ; antigène carcinoembryonnaire	Peptide signal IgG murin ; KDEL	ScFvT84.66 (ScFv)	Riz
Traitement du cancer ; antigène carcinoembryonnaire	Leader de TMV ; peptides signal IgG murins ; KDEL	T84.66 (IgG)	*Nicotiana tabaccum* (expression transitoire avec infiltration de *Agrobacterium*)
Traitement du lymphome à cellules B ; vaccin idiotypique	α-amylase du riz	38C13 (scFv)	*Nicotiana Benthamiana*
Cancer du côlon ; antigène de surface	Peptide signal IgG murin ; KDEL	CO17-1A (IgG)	*Nicotiana Benthamiana*
Virus de l'herpès simplex 2	Peptide signal d'extensine du tabac	Anti-HSV-2 (IgG)	Soja

(Daniell *et al., Trends in Plant Science 6(5)* : 219-226, 2001)

Des procédés pour réaliser l'expression d'un gène codant un anticorps dans une plante sont bien connus de l'homme du métier. Cependant, une description des étapes typiquement réalisées pour produire un planticorps est présentée ci-dessous.

(i) Identification et caractérisation d'un anticorps monoclonal humain cible au moyen de systèmes murins

transgéniques, d'une banque de présentation sur phage ou par humanisation d'anticorps de souris.

(ii) Insertion des gènes d'anticorps dans un vecteur d'ADN qui permet l'expression dans une plante entière ou une partie spécifique de la plante par une ou plusieurs techniques bien établies, comme la délivrance biolistique, la transformation à médiation par *Agrobacterium*, l'électroporation et analogues.

(iii) Régénération de plantes entières à partir de cellules végétales transformées.

(iv) Purification d'un anticorps monoclonal humain à partir d'un tissu végétal.

Typiquement, la production d'un planticorps utilise la procédure décrite ci-dessus, cependant, il conviendrait de comprendre que la présente invention n'est absolument pas limitée à un procédé quelconque pour la production de planticorps.

Ainsi, la présente invention envisage un procédé pour la production spécifique d'un planticorps, dans l'endosperme d'une plante, ledit procédé comprenant la transformation d'une plante avec une construction d'ADN recombinant comprenant un acide nucléique codant un anticorps en liaison fonctionnelle avec l'élément régulateur d'expression décrit ici.

D'autres molécules d'acide nucléique pour lesquelles l'expression spécifique de tissu facilitée par les éléments régulateurs d'expression décrits ici serait souhaitable, seront aisément déterminées par l'homme du métier. Cependant, des exemples de molécules d'acide nucléique pour lesquelles il serait souhaitable d'avoir une expression spécifique dans l'endosperme incluent les acides nucléiques codant une ou plusieurs enzymes biosynthétiques, ou des transcripts antisens de celles-ci qui provoqueraient l'accumulation ou la réduction de la quantité d'un métabolite dans l'endosperme de la plante transformée. Certains exemples de tels gènes codant des enzymes biosynthétiques incluent, mais absolument pas à titre de limitation :

(i) les enzymes biosynthétiques de la vanilline, incluant la 3-déshydroshikimate déshydratase, la catéchol-o-méthyltransférase, l'arylaldéhyde déshydrogénase, la

féruloyl-CoA synthétase, l'énoyl-CoA hydratase/aldolase ;

(ii) les enzymes biosynthétiques du sorbitol incluant la glucose/fructose oxydoréductase ;

(iii) les enzymes biosynthétiques de PHA, incluant la 3-kétothiolase, l'acétoacétyl-CoA réductase, la PHA synthase, l'énoyl hydratase, la protéine support 3-hydroxyacyl-acyle:CoA-transférase ;

(iv) les enzymes biosynthétiques de l'indigo incluant la tryptophanase, la L-tryptophane indole lyase, la naphtalène dioxygénase, le produit du gène *R. eutrophica bec* ;

(v) les enzymes biosynthétiques du fructane incluant les fructosyltransférases et le lévanesucrases ;

(vi) les enzymes biosynthétiques de l'acide lactique, incluant la lactate déshydrogénase ;

(vii) les enzymes biosynthétiques de l'acide adipique, incluant la 3-déshydroshikimate déshydratase, la protocatéchuate décarboxylase et la catéchol 1,2-dioxygénase ;

(viii) les enzymes biosynthétiques de l'acide pétrosélinique, incluant la 3-cétoacyl-ACP synthase ;

(ix) les enzymes biosynthétiques du 1,3-propanediol incluant la glycérol déshydratase, la 1,3-propanediol oxydoréductase, la glycérol-3-phosphate déshydrogénase et la glycérol-3-phosphatase ; et/ou

(x) les enzymes biosynthétiques du 2-phényléthanol incluant la L-aminoacide aromatique décarboxylase, la 2-phényléthylamine oxydase et l'alcool arylique déshydrogénase.

(xi) les enzymes biosynthétiques de pHBA incluant la 4-hydroxycinnamoyl-CoA hydratas/lyase (HCHL) et la chorismate pyruvate lyase (CPL).

La présente invention s'étend en outre à une cellule, un tissu ou un organe isolé comprenant l'élément régulateur d'expression tel qu'il est décrit ici. De plus, la présente invention s'étend en outre à des

tissus, organes ou organismes complets régénérés dérivés desdites cellules, tissus ou organes et des propagules et leur descendance ainsi que les graines et d'autres matériels reproducteurs.

De préférence, la cellule, le tissu, l'organe ou l'organisme est une cellule végétale, un tissu végétal, un organe végétal ou une plante.

Le terme "matériel reproducteur" tel qu'il est utilisé ici doit être compris comme étant toute graine végétale, tissu, bouture, explant, cellule ou pollen qui peut être utilisée pour :

(i) propager la plante via reproduction sexuée ;

(ii) propager la plante via propagation végétative ; ou

(iii) produire une culture de cellules végétales ou de tissu végétal de la plante et/ou régénérer une plante à partir de la culture.

Des exemples de matériel reproducteur envisagé par la présente invention incluent, mais ne sont absolument pas limités à : les graines, les fruits, les cônes, les boutures, les explants, les cellules, les anthères, le pollen, les fleurs (incluant leurs parties), les tubercules, les bulbes, le tissu calleux, le cal embryogène, les cultures en suspension et les embryons (incluant les embryons immatures).

De préférence, la cellule végétale transgénique ou la plante, le tissu, l'organe ou la partie de plante transgénique comprenant une ou plusieurs desdites cellules, provient des espèces consistant en : *Elaeis guineesis* Jacq., *Elaeis oleifera* et leurs hybrides, le maïs, le riz, le soja, le tabac, le coton, la luzerne, le blé, *Arabidopsis*, le colza, canola et le cocotier.

Un autre aspect de la présente invention envisage un procédé d'identification de partenaires de liaison qui se lient ou s'associent d'une autre manière avec l'élément régulateur d'expression décrit ci-dessus.

Des procédés particulièrement préférés pour déterminer la liaison entre une molécule d'ADN (comme les éléments régulateurs d'expression de la présente invention) et un partenaire de liaison incluent les essais de mobilité électrophorétique (essais de déplacement sur gel et essais de retard sur gel) et les essais d'empreintes à la DNase. Ces essais seront bien de l'homme du métier.

En bref, les essais de déplacement de mobilité électrophorétique comprennent la mise en contact d'un partenaire de liaison supposé avec la molécule d'acide nucléique d'intérêt (par exemple un élément régulateur d'expression de la présente invention), ledit contact étant réalisé dans des conditions telles et pendant une durée telle qu'elles permettent à la liaison de se produire. La liaison entre le partenaire de liaison supposé et la molécule d'acide nucléique est déterminée par la mobilité électrophorétique du complexe acide nucléique supposé/partenaire de liaison. Une réduction de la mobilité électrophorétique de la molécule d'acide nucléique d'intérêt quand elle est complexée avec le partenaire de liaison supposé indique une liaison entre le partenaire de liaison supposé et la molécule d'acide nucléique.

Les essais de mobilité électrophorétique sont utiles pour la détection d'une gamme d'agents liant l'ADN incluant d'autres acides nucléiques (par exemple ADN et ARN) et les protéines.

Les essais d'empreinte à la DNase utilisent l'activité de la DNase I. La DNase I en présence de Mg^{2+}, hydrolyse chaque brin d'ADN indépendamment d'une manière statistiquement aléatoire. Cette propriété de la nucléase est utile dans l'analyse des complexes ADN-protéine. Les régions d'ADN qui ont des protéines liées sont protégées contre l'attaque à la DNase I diluée de manière appropriée. De ce fait, après la digestion avec la DNase I, les régions de l'ADN protégées par une protéine de liaison peuvent être isolées et séquencées. Des détails complets de tels essais sont décrits par Galas et Schmitz (*Nucl. Acids Res. 5* : 3157-3170, 1978). Ainsi, ce procédé peut être utilisé pour identifier les agents de liaison des éléments régulateurs d'expression de la présente invention par mise en contact d'un agent de liaison supposé avec un acide nucléique comprenant une séquence de nucléotides codant un élément régulateur d'expression décrit ici pendant une durée et dans des conditions permettant la liaison ; la digestion du complexe supposé avec la DNase I diluée de manière appropriée pour digérer l'ADN non protégé par une protéine de liaison ; et l'identification d'un agent de liaison supposé comme agent de liaison de l'acide nucléique par la détection d'une région non digérée, c'est-à-dire protégée, de la molécule d'acide nucléique.

La présente invention s'étend en outre à d'autres procédés connus de l'homme du métier qui utilisent les molécules d'acide nucléique fournies par la présente invention, ou l'information de séquence dérivée de celles-ci, pour identifier des agents de liaison des éléments régulateurs d'expression décrits ici.

Selon la présente invention, les procédés décrits ont une application particulière pour l'identification de facteurs de transcription qui se lient aux éléments régulateurs d'expression décrits ici.

Le terme "facteur de transcription" tel qu'il est utilisé ici désigne tout agent qui régule l'expression d'un gène en se liant à une ou plusieurs régions dans l'élément régulateur d'expression de ce gène. Ainsi, un facteur de transcription peut se lier à une ou plusieurs régions promotrices, stimulatrices ou silenceurs pour un gène particulier. Les facteurs de transcription sont le plus souvent des protéines, mais peuvent être aussi des coenzymes, des vitamines ou d'autres molécules organiques. Il conviendrait de comprendre que la présente invention n'est absolument pas limitée par la nature de tout facteur de transcription lui-même, qui est identifié selon la présente invention.

La présente invention s'étend en outre à des facteurs de transcription et d'autres agents qui se lient à l'élément régulateur d'expression décrit ici, quand il est identifié selon la présente invention.

Dans un aspect lié, la présente invention envisage un procédé pour moduler l'activité d'un élément régulateur d'expression, ledit procédé comprenant l'administration à une ou plusieurs cellules végétales comprenant l'élément régulateur d'expression d'un agent qui se lie à l'élément régulateur d'expression et qui module l'activité de l'élément régulateur d'expression. De plus, il est envisagé que la modulation de l'activité de l'élément régulateur d'expression modulerait à son tour l'expression d'une molécule d'acide nucléique en liaison fonctionnelle avec l'élément régulateur d'expression.

Finalement, un autre aspect de la présente invention fournit une molécule d'acide nucléique isolée comprenant une séquence de nucléotides, où la séquence nucléotidique code une protéine glutéline comprenant la séquence d'aminoacides présentée dans SEQ ID NO : 2 ou une séquence d'aminoacides ayant plus de 92 % de similarité avec SEQ ID NO : 2 après alignement optimal ou une séquence nucléotidique

telle que présentée dans SEQ ID NO : 1 ou une séquence nucléotidique ayant plus de 95 % d'identité avec SEQ ID NO : 1 après un alignement optimal.

Une référence à plus de 92 % de similarité inclut des identités et des similarités en pourcentage supérieures à 92 %, comme 92,1 %, 92,2 %, 92,3 %, 92,4 %, 92,5 %, 92,6 %, 92,7 %, 92,8 %, 92,9 %, 93 %, 94 %, 95 %, 96 %, 97 %, 98 %, 99 % et 100 %.

Une référence à plus de 95 % d'identité inclut des identités et similarités en pourcentage supérieures à 95 % comme 95,1 %, 95,2 %, 95,3 %, 95,4 %, 95,5 %, 95,6 %, 95,7 %, 95,8 %, 95,9 %, 96 %, 97 %, 98 %, 99 % et 100 %.

Encore un autre aspect de la présente invention envisage une protéine glutéline isolée comprenant une séquence d'aminoacides telle que présentée dans SEQ ID NO : 2 ou une séquence d'aminoacides ayant plus de 96 % de similarité avec SEQ ID NO : 2 après un alignement optimal.

Ainsi, la présente invention envisage aussi des constructions d'ADN recombinant, comme décrit précédemment, comprenant la séquence d'aminoacides présentée dans SEQ ID NO : 2 ou une séquence d'aminoacides ayant plus de 96 % de similarité avec SEQ ID NO : 2 après alignement optimal ou une séquence nucléotidique telle que présentée dans SEQ ID NO : 1 ou une séquence nucléotidique ayant plus de 96,4 % d'identité avec SEQ ID NO : 1 après alignement optimal.

Les constructions mentionnées ci-dessus peuvent comprendre en outre un élément régulateur d'expression tel que décrit ici et peuvent comprendre aussi en outre une séquence d'acide nucléique. Par exemple, une construction peut comprendre un élément régulateur d'expression, une molécule d'acide nucléique codant une protéine glutéline telle que décrite ici et peut comprendre éventuellement une autre molécule d'acide nucléique codant une protéine, de sorte que quand elle est exprimée sous le contrôle de l'élément régulateur d'expression, la protéine glutéline est fusionnée avec la seconde protéine.

La présente invention est décrite en outre par les exemples non limitatifs suivants :

EXEMPLE 1

Isolement d'un clone d'ADNc

Procédés

Séquençage à une seule passe de clones d'ADNc

Une banque d'ADNc provenant d'un tissu de noyau de palmier à huile de 12 semaines (*E. guineensis*) a été construite avec une amorce oligo (dT) et Lambda Uni-Zap XR comme vecteur de clonage (Stratagene). Des plages phagiques provenant de la banque d'ADNc ont été évidées et excisées *in vivo*. Des clones d'ADNc prélevés au hasard ont été séquencés depuis l'extrémité 5' et les résultats ont été renvoyés pour une recherche d'homologie avec une séquence connue dans la banque de données publique via BLASTX.

5'-RACE

L'ADNc du premier brin a été synthétisé à partir d'un ARN de noyau de palmier à huile à 12 saa en utilisant l'amorce spécifique de séquence antisens KRA (5' CCT GCC TTC CAG CCA ATA AG 3', SEQ ID NO : 4) dans un mélange réactionnel de 20 µl contenant 5 µg d'ARN total, 5,0 µl de mélange de dNTP 2 mM, 2,0 µl de DTT 0,1M, 1 µl de transcriptase inverse Superscript à 200 U/µl (Gibco BRL) et 4 µl de tampon Superscript 5X à 42°C pendant 1 h. Les molécules d'ARN ont été hydrolysées dans 12,5 µl d'hydroxyde de sodium 0,15N et 1 µl de EDTA 0,5M, pH 8,0 et incubées à 68°C pendant 15 min. Une séquence de queue poly(dG) a été introduite avec la désoxynucléotidyle transférase terminale dans un mélange réactionnel de 20 µl contenant 10 mM de Tris-acétate pH 7,5, 10 mM d'acétate de magnésium, 50 mM d'acétate de potassium, 0,2 µl de dGTP 100 mM et 0,2 µl d'enzyme à 300 U/µl. L'ADNc du second brin a été synthétisé dans un mélange réactionnel de 50 µl contenant 5,0 µl de dNTP 2 mM, 50 pmol d'amorce d'ancrage KA1 (5' CTT CCC CCC CCC CCC C 3', SEQ ID NO : 5), 4,5 µl d'ADNc simple brin à queue dG, 2,6 U de polymérase Expand High Fidelity (Boehringer Mannheim) et 5 µl de tampon d'enzyme 10X contenant 1,5 mM de $MgCl_2$. Les conditions de la PCR étaient les suivantes : 1 cycle ; 95°C pendant 3 min, 43°C pendant 1 min, 72°C pendant 2,5 min, suivi par 4 cycles ; 95°C pendant 1 min, 43°C

pendant 1 min, 72°C pendant 2 min. Le produit de PCR a été purifié avec le kit de purification de PCR QIAquick (Qiagen), et élué avec 50 µl de H_2O. Une portion aliquote de 5 µl a été ajoutée à 50 µl d'un mélange de PCR secondaire contenant 5 µl de dNTP 2 mM, 50 pmol d'amorce nichée antisens KRB (5' GAT CGC GAC CAC AGG GTT C 3', SEQ ID NO : 6), 2,6 U de polymérase Expand High Fidelity et du tampon d'enzyme 1X contenant 1,5 mM de $MgCl_2$. Les conditions de PCR suivantes ont été utilisées : 1 cycle ; 95°C pendant 3 min, 51°C pendant 1 min, 72°C pendant 30 s, suivi par 9 cycles ; 95°C pendant 1 min, 51°C pendant 1 min, 72°C pendant 1 min et 30 s, et 80°C d'immersion pendant laquelle 50 pmol d'amorce d'ancrage KA1 ont été ajoutés suivi par 30 cycles ; 95°C pendant 1 min, 51°C pendant 1 min et 72°C pendant 2 min. Le produit de PCR a été purifié et cloné dans le vecteur PCRII-TOPO (Invitrogen).

PCR de bout en bout

Les étapes impliquées dans la production d'un ADNc double brin provenant de l'ADNc de noyau de 12 semaines sont identiques à la réaction 5'-RACE à ceci près que l'amorce oligo-dT a été utilisée pour remplacer les amorces spécifiques de séquence KRA et KRB. L'ADNc double brin qui a été obtenu a été utilisé comme matrice dans l'amplification par PCR au moyen d'une paire d'amorces EK1 (5' AAG CAA TAG CCT TCA GCG TTT C 3', SEQ ID NO : 13) et ANEK2 (5' GCC CTT GCT CGC AGG CGG AG 3', SEQ ID NO : 14). L'amorce EK1 a été conçue sur la base de l'information de séquence obtenue à partir de l'extrémité 5' du produit 5'-RACE tandis que l'amorce ANEK2 a été obtenue à partir de la région 3'-UTR de la séquence EST de pOP-KT21.

Résultats

L'un des EST de l'extrémité 3' de 1,1 kb provenant de la banque d'ADNc de noyau de palmier à huile prélevait le gène de glutéline provenant du palmier à huile et du riz comme étant les meilleurs candidats d'après les recherches dans les banques de données avec BLASTP. 5'-RACE a été réalisé pour déterminer la séquence au niveau de la région 5' et pour obtenir le site de début de la transcription en utilisant l'ARN total provenant de tissus de noyau à 12 saa. L'amorce

spécifique de gène KRA et l'amorce nichée spécifique de gène KRB conçues sur la base de la séquence EST ont été utilisées dans la synthèse de l'ADNc du premier brin et de l'étape d'amplification par PCR finale dans la réaction 5'-RACE, respectivement pour obtenir le produit d'intérêt spécifique. L'information de séquence a été utilisée pour produire un clone d'ADNc de longueur intégrale par PCR de bout en bout utilisant une paire d'amorces spécifiques de gène sens et antisens EK1 et ANEK2 et un ADNc double brin produit à partir d'un tissu de noyau de 12 semaines comme matrice. La séquence du clone d'ADNc de longueur intégrale appelé pOP-KT21, telle qu'elle est montrée sur la figure 1, contient un ORF de 1413 pb (SEQ ID NO : 1) précédé par 45 pb de 5'-UTR et suivi par 147 pb de 3'-UTR. Elle code 470 résidus d'aminoacides (SEQ ID NO : 2) avec une masse moléculaire prédite de 53,5 kDa.

EXEMPLE 2

Analyse de séquences

Procédés

L'ADN plasmidique pour le séquençage a été extrait au moyen du minikit plasmidique Qiagen. Le séquençage de l'ADN a été réalisé depuis les deux directions en utilisant un séquenceur automatisé ABI. L'outil d'analyse de séquence basé sur l'Internet Biology Workbench Version 3.2 à http://workbench.sdsc.edu a été utilisé pour l'analyse des séquences et la recherche de similarité entre la séquence nucléotidique et la séquence d'aminoacides avec CLUSTALW (Thompson *et al., Nucl. Acid Res. 22* : 4673-4680, 1994).

Résultats

Analyse de la séquence nucléotidique

Cha et Shah (*Plant Science* 160 : 913-923, 2001) ont décrit l'isolement de clones d'ADNc codant trois isoformes différentes de protéine de réserve des graines glutéline à partir du palmier à huile, pKT7 (AF261691), pKT8 et pKT9 (AF193433) où pKT7 et pKT9 étaient des clones d'ADNc de longueur intégrale. Sur la base de l'analyse de séquence les gènes de glutéline du palmier à huile ont été divisés en deux sous-familles, la sous-famille A (pKT7) et la sous-famille B (pKT8

et pKT9). Ils ont montré en outre par analyse de Southern que la sous-famille B consiste éventuellement en au moins quatre membres dans le génome du palmier à huile. Les séquences nucléotidiques et d'aminoacide déduites de pOP-KT21 ont été comparées avec celles de pKT9 et pKT7 (figure 2). Les séquences nucléotidiques de pOP-KT21 et pKT9 partagent une forte identité de séquence dans les régions codante et 3'-UTR avec des identités de séquence de 96,4 et 90,7, respectivement. Cependant, l'identité de séquence dans le 5'-UTR est inférieure à 78,4 % tandis que la longueur est légèrement plus courte pour pOP-KT21 qui est de 45 pb par rapport à pKT9, 51 pb et la variation de séquence qui apparaît n'était pas confinée à une seule position spécifique. Plusieurs articles suggèrent que les éléments de séquence dans les régions 5'-UTR de gènes végétaux peuvent jouer un rôle dans la régulation de l'expression génique (Dickey *et al., Plant Cell 3* : 475-484, 1998 ; Bonaventure et Ohlrogge, *Plant Physiology 128* : 223-235, 2002). L'identité de séquence d'aminoacides déduite entre pOP-KT21 et pKT9 est 96 %. La région qui code la sous-unité basique est plus conservée avec une identité de séquences d'aminoacides de 99,5 % entre pOP-KT21 et pKT9 par rapport à la région codant la sous-unité acide avec une identité de séquences de 93,7 %. D'autre part, l'homologie de séquences nucléotidiques entre pOP-KT21 et pKT7 (sous-famille A) est bien inférieure avec des identités de seulement 69,8 %, 48,1 % et 43,6 % pour la région codante, 3'-UTR et 5'-UTR, respectivement. Tandis que l'identité des séquences d'aminoacides déduite est 57,4 %.

Analyse de la région codante (aminoacides)

L'alignement des séquences d'aminoacides déduites des gènes du palmier à huile (pOP-KT21, pKT7 et pKT9) avec un gène de glutéline provenant du riz et un gène de globuline 11S provenant du sésame (*Sesamum inducum*) tel qu'il est indiqué dans la figure 3 présentait une homologie de séquence suggérant que toutes ces séquences proviennent de la même origine au niveau de l'évolution. Ceci est en accord avec l'observation pour les gènes de glutéline du riz (Takaiwa *et al.*, *Mol. Gen. Genet. 208* : 15-22, 1987) qui montrait une homologie de séquence avec les gènes de globuline 11S de

légumineuses. La position pour les sites de clivage pour le peptide leader de transit dans pOP-KT21 est identique à celle proposée par Cha et Shah (*supra*). Le peptide de transit de pOP-KT21 contient aussi la séquence de cœur central hydrophobe conservé FSLCLLLL trouvé dans pKT7 et pKT9 qui convient pour assurer la médiation du transfert de la protéine jusqu'à la lumière du réticulum endoplasmique. Il a été montré que la séquence spécifique de Asn et Gly ainsi que la structure secondaire sont importantes pour déterminer le site de clivage post-traductionnel entre les sous-unités acide et basique dans la protéine précurseur (Takaiwa *et al.*, 1987, *supra*). La région acide proposée de pOP-KT21 contient 43 résidus acides (Glu et Asp) parmi 285 résidus d'aminoacides. Tandis que la région basique contient 28 résidus basiques (Arg et Lys) parmi 185 résidus d'aminoacides.

Les positions des 5 résidus de cystéine dans pOP-KT21 sont en accord avec celle décrite par Cha et Shah (*supra*). Quatre de ces résidus de cystéine sont conservés dans les glutélines et dans toutes les globulines 11S dans les graines de légumes et non légumes (Katsube *et al.*, *Engineering Crop plants for Industrial End Uses*, Shewry *et al.* (Eds), Portland Press, Londres, 1998, p. 65-76). Dans pOP-KT21, les résidus cystéine conservés sont trouvés aux positions 24, 57 et 100 de la sous-unité acide proposée et la quatrième à la position 7 de la sous-unité basique proposée. Les deux premières cystéines sont impliquées dans la formation d'une liaison disulfure intrachaîne, tandis que les deux dernières forment une liaison disulfure interchaîne (Utsumi, *Adv. Food Nutr. Res. 36* : 89-208, 1992 ; Katsube *et al., supra*). La formation de liaisons disulfure entre les sous-unités acide et basique stabilise la structure des protéines de réserve des graines. Plusieurs polypeptides peuvent être joints pour former un oligomère ayant la conformation appropriée pour être déposé dans des corps protéiques (Muntz, *supra*).

La composition en aminoacides de la séquence protéique déduite de pOP-KT21 est donnée sur la figure 4a. Bien que la séquence de pKT7 provenant de la sous-famille A soit différente, le profil d'aminoacides est très similaire aux séquences protéiques de glutéline provenant de la sous-famille B. Il était intéressant de noter que le gâteau de noyau de palmier à huile a un profil d'aminoacides très similaire (figure 4b). De manière notable, l'arginine, la leucine, la

glycine et la valine sont les résidus particulièrement prédominants avec de faibles niveaux d'histidine et de tryptophase. Ceci peut suggérer que la glutéline est une protéine de réserve majeure dans le noyau du palmier à huile.

EXEMPLE 3

Analyse par Northern Blot

Procédés

L'extraction de l'ARN total a été réalisée selon Siti Nor Akmar *et al.* (*Asia Pacific Journal of Molecular Biology and Biotechnology 2(2)* : 113-118, 1994). Les analyses par slot blot et Northern Blot ont été réalisées comme décrit par Siti Nor Akmar (PhD Thesis, University of East Anglia, 1999). L'analyse a été réalisée en utilisant le EST de pOP-KT21 comme sonde et au moyen de conditions de lavage à haute stringence. L'analyse par Northern Blot a été réalisée aussi en utilisant la sonde spécifique de gène conçue sur la base de la séquence de la région 5'-UTR de pOP-KT21. Les deux amorces utilisées pour produire la sonde spécifique de gène à partir du 5'-UTR et de 14 bases à l'extrémité 5' de la région codante de pOP-KT21 étaient KTC (5' CAA TAG CCT TCA GCG TTT C 3', SEQ ID NO : 7) et KTD (5' GAG AAG GAC GAC ATG GCT 3', SEQ ID NO : 8).

Résultats

L'analyse slot blot été réalisée en utilisant le EST de pOP-KT21 comme sonde et l'ARN total provenant de différents tissus du palmier à huile (figure 5). Un haut niveau d'expression a été observé dans les tissus du noyau. Elle montre aussi que l'expression de la glutéline du palmier à huile est étroitement régulée quand l'expression n'est pas détectable non plus dans les jeunes plants germés et les fleurs aux frondes 10, 14 et 19. Les résultats de l'analyse par Northern Blot utilisant pOP-KT21 comme sonde (figure 6) ont confirmé que la taille des transcrits est d'environ 1,6 kb. Ils confirment en outre que ce gène est hautement exprimé seulement dans le noyau, sans expression détectable dans les jeunes feuilles, les fleurs, les jeunes plants germés et le mésocarpe de fruits immatures (8 saa) et de fruits murs (20 saa). Il est hautement exprimé dans tous les différents stades du

développement du noyau (10-15 saa), avec de plus hauts niveaux d'expression après 12 saa (figure 6). Les résultats obtenus d'après l'analyse de slot blot et l'analyse de Northern Blot utilisant pOP-KT21 comme sonde ont confirmé le profil d'expression pour la sous-famille B comme démontré par Cha et Shah (*Plant Science 160* : 913-923, 2001). La sonde pOP-KT21 a été utilisée en outre pour déterminer le site spécifique d'expression dans le noyau (figure 6). Il a été démontré pour la première fois que, dans le noyau, l'expression est confinée à l'endosperme et qu'aucune expression n'a été détectée dans l'embryon. Comme l'endosperme est le site de synthèse d'huile dans le noyau du palmier à huile, le promoteur de pOP-KT21 peut être utilisé pour réguler l'expression de transgènes pour modifier la composition de l'huile du noyau sans interférer avec la synthèse de l'huile du mésocarpe.

Comme la séquence au niveau de la région 5'-UTR de pOP-KT21 et pKT9 présente une plus grande variabilité que la région 3'-UTR, une sonde spécifique de gène a été préparée sur la base de la séquence 5'-UTR de pOP-KT21 et de 14 pb de la séquence codant l'extrémité 5' qui montre aussi un haut niveau de variabilité et utilisée dans l'analyse Northern Blot (figure 7) avec des conditions de lavage à haute stringence. Il a été démontré que pOP-KT21 contribuait au haut niveau d'expression dans le noyau et que son profil d'expression était bien corrélé avec le profil d'expression des gènes de glutéline appartenant à la sous-famille B. Ceci confirme que l'expression de pOP-KT21 est régulée par un promoteur spécifique du noyau.

EXEMPLE 4

Hybridation in situ

Procédés

Synthèse de sondes

Le EST de 1,1 kb de palmier à huile contenant une séquence partielle de pOP-KT21 a été cloné dans le plasmide TOPO-TA (Invitrogen). Un mélange réactionnel contenant 1 µg de clone plasmidique, 2 µl de mélange de marquage d'ADN DIG 10X (Roche), 2 µl de tampon de transcription 10X (400 mM de Tris-HCl pH 8,0, 100 mM de $MgCl_2$), 100 mM de dithiothréitol et 20 mM de spermidine),

2 µl d'ARN polymérase (SP6 ou T7) et H_2O stérile jusqu'à un volume final de 20 µl a été préparé et incubé pendant 2 h à 37°C. La réaction a été stoppée par addition de 0,8 µl de EDTA 0,2 M (pH 8,0). Le transcrit d'ARN marqué a été purifié avec le QIAquick PCR Purification Kit (Qiagen). La sonde a été diluée à 0,2-10 ng/µl dans le tampon d'hybridation.

Prétraitement de coupe in situ

La fixation, l'inclusion et la formation de coupes de tissu ont été réalisées sur la base des procédés de Weigel et Glazebrook (*ARABIDOPSIS : A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002). Les coupes de tissu ont été déparaffinées deux fois dans Histoclear pendant 10 min. La coupe a été hydratée par des séries d'éthanol : 2X éthanol à 100 % pendant 1-2 min, éthanol à 95% pendant 1-2 min, éthanol à 90% pendant 1-2 min, éthanol à 80 % pendant 1-2 min, éthanol à 60 % pendant 1-2 min, éthanol à 30 % pendant 1-2 min, et finalement H_2O pendant 1-2 min. Les lames ont été ensuite plongées dans 2X SSC à la température ambiante pendant 15-20 min puis transférées à une solution de protéinase K (1 µg/ml) dans Tris-HCl 100 mM pH 8,0, EDTA 50 mM à 37°C pendant 30 min. La protéinase K a été bloquée avec 2 mg/ml de glycine dans PBS à la température pendant 2 min suivie par un lavage deux fois dans PBS à la température ambiante pendant 2 min. Le tissu a été fixé dans du paraformaldéhyde à 4 % pH 7 à la température ambiante pendant 10 min puis lavé deux fois dans PBS à la température ambiante pendant 5 min. Les lames ont été transférées dans de l'anhydride acétique à 0,5 % (v/v) dans la triéthanolamine 0,1 M pH 8. Ceci a été suivi par un lavage deux fois dans PBS à la température ambiante pendant 5 min. Les lames ont été déshydratées en étant déplacées dans les solutions suivantes : éthanol à 30 % pendant 30 s, éthanol à 60 % pendant 30 min, éthanol à 80 % pendant 30 s, éthanol à 90 % pendant 30 s, éthanol à 95 % pendant 30 s et deux fois éthanol à 100 % pendant 30 s. Les lames ont été conservées dans un récipient contenant une petite quantité d'éthanol pendant plusieurs heures à 4°C.

Hybridation in situ

La solution d'hybridation (suffisamment pour 5 paires de lames) a été préparée comme suit : 100 µl de sels 10X *in situ*, 400 µl de formamide désionisé, 200 µl de sulfate de dextrane à 50 %, 20 µl de solution de denhardt 50X, 10 µl d'ARNt (100 mg/ml), 70 µl de H_2O, qui a été amenée au volume total de 800 µl. Pour chaque paire de lames, une sonde a été ajoutée à 40 µl de formamide à 50 %, chauffée à 80°C pendant 2 min, enroulée et maintenue sur la glace. La solution d'hybridation (160 µl) a été ajoutée à chaque paire de lames. La sonde a été appliquée en étalant 100 µl sur chaque lame, avec le côté de l'extrémité de la pipette de sorte que tout le tissu était exposé à la sonde sans former de bulles. Les lames ont été prises en sandwich lentement. Les lames ont été recouvertes de paraffine et une hybridation a été réalisée à 50-55°C pendant une nuit.

Lavage et détection de signaux

Les lames ont été rincées dans 0,2X SSC à 55°C pour les séparer. Le lavage dans 0,2 X SSC a été répété pendant 60 min supplémentaires en agitant doucement. Ceci a été suivi par un lavage deux fois dans NTE (Tris 10 mM pH 8,0, NaCl 0,5 M, EDTA 5 mM) à 37°C pendant 5 min et 1X dans NTE contenant de la RNase (20 µg/ml) à 37°C pendant 5 min. Après rinçage avec NTE, les lames ont été placées dans 0,2X SSC à 55°C en agitant doucement pendant 60 min suivi par 1X PBS à la température ambiante pendant 5 min. Les lames ont été incubées dans une solution de blocage de Boehringer dans Tris-HCl 100 mM pH 7,5 et NaCl 150 mM pendant 45 min. Le bloc a été remplacé par BSA 1,0 % dans Tris-HCl 100 mM pH 7,5, NaCl 150 mM, 0,3 % de Triton X-100 pendant 45 min. Un anticorps anti-DIG a été dilué (1:1250) dans du tampon A (Tris 0,1 M pH 7,5, NaCl 0,15 M, BSA 1 % (m/v) et Triton X-100 0,3 % (v/v)) et mélangé dans une boîte en plastique. Les lames ont été disposées en sandwich ensemble et la solution a été poussée par action capillaire et une incubation a été réalisée pendant 2 h à la température ambiante. Les lames ont ensuite été lavées quatre fois dans du tampon A pendant 20 min à chaque lavage, sous agitation douce puis lavées dans du tampon de détection deux fois. Un mélange NBT/BCIP frais a été appliqué et les lames ont

été disposées en sandwich pour attirer la solution par action capillaire. Les lames ont été placées dans une chambre humidifiée à l'obscurité totale pendant 1-3 jours. Les lames ont été rincées dans TE pour stopper la réaction. Les lames ont ensuite été observées au microscope optique.

Résultats

L'hybridation *in situ* a été réalisée sur des coupes de tissu de noyau de palmier à huile en utilisant des sondes d'ARN sens et antisens produites à partir de la séquence EST de 1,1 kb du palmier à huile de pOP-KT21. Les signaux d'hybridation obtenus avec la sonde antisens montraient une distribution uniforme de transcrits dans l'endosperme où des signaux d'hybridation ont été observés dans toutes les cellules parenchymateuses de l'endosperme (figure 8). L'endosperme est composé d'un type cellulaire homogène, c'est-à-dire les cellules parenchymateuses. Ces cellules jouent des rôles importants dans le stockage des protéines et des huiles. L'observation d'une coupe d'endosperme provenant de cocotiers matures sous un grossissement de 6 fois, par exemple, montrait que les cellules parenchymateuses étaient occupées par de grands globules d'huile et de petits granules de protéines (Brown, *The Plant Kingdom*, Blaidsdell Publishing Company, New York, 1935). Ainsi, les promoteurs de gènes de protéine de réserve des graines peuvent diriger une régulation spatiale précise de gènes impliqués dans la synthèse des protéines et la synthèse des huiles car les deux produits de réserve sont des produits dans le même type de cellules.

EXEMPLE 5

Isolement du promoteur

Procédés

L'isolement du promoteur spécifique du mésocarpe a été réalisé en utilisant le Universal Genome Walker Kit (Clontech). L'ADN total a été isolé et purifié à partir de feuilles du palmier à huile avec le DNeasy Plant Mini Kit (Qiagen). Des portions aliquotes contenant 2,5 µg d'ADN ont été digérées avec les enzymes de restriction *Dra* I, *Eco* RV, *Pvu* II et *Stu* I qui produisent des extrémités franches et ligaturées à

l'adaptateur de marche sur le génome pour créer des banques de marche sur le génome. Une PCR primaire a été réalisée avec des portions aliquotes de 12 µl de chaque banque avec l'amorce spécifique de gène antisens GK4 (5' GAC TGG ACC CAA ATT GAG CCT GGG ACA C 3', SEQ ID NO : 9) provenant de l'extrémité 5'-terminale de la région codante de pOP-KT21 et l'amorce API fournie avec le kit. Le produit de PCR a été dilué 50 fois et 1 µl a été utilisé dans une réaction de PCR secondaire avec l'amorce nichée spécifique de gène antisens GKN6 (5' GAA AGA GAG AAA CGC TGA AGG CTA TTG C 3', SEQ ID NO : 10) provenant de l'extrémité 5' dans le 5'-UTR de l'ADNc et l'amorce AP2 du kit. La PCR a été réalisée avec le mélange de polymérase *Tth* Advantage de Clontech et un cycleur thermique Perkin-Elmer 9600 en suivant les conditions des cycles recommandées dans le manuel du kit GenomeWalker. Les produits de PCR primaires et secondaires ont été analysés et purifiés à partir du gel d'agarose en utilisant un kit d'extraction de gel de Qiagen et ont été clonés dans le vecteur PCRII-TOPO (Invitrogen). Le clone recombinant a été séquencé avec des amorces directes et inverses de M13.

Résultats

Le kit Universal Genome Walker de Clontech a été utilisé pour cloner le promoteur ou la région régulatrice en amont 5' de pOP-KT21. Des portions aliquotes de banques de GenomeWalker de palmier à huile *Dra* I, *Eco* RV et *Stu* I ont été amplifiées avec une amorce spécifique de gène 28-mère (GK4) provenant de l'extrémité 5'-terminale de la région codante de pOP-KT21 pour la protéine mature et l'amorce AP1 (séquence adaptateur) fournie avec le kit. La région où GK4 a été obtenue présentait une variabilité de séquence légèrement plus élevée entre pOP-KT21 et pKT9 que le reste de la région codante. L'identité des séquences nucléotidiques entre pOP-KT21 et pKT9 dans cette région est de 82 % seulement. Les banques *Dra* I, *Eco* RV et *Stu* I produisaient des produits de PCR d'environ 1,2 kb, 1,5 kb et 0,45 kb respectivement. Comme le produit de la banque *Eco RV* était le plus grand, il a été choisi pour une réaction de PCR supplémentaire. 1 µl d'une dilution 1/50 du produit de PCR primaire a été utilisé dans une réaction de PCR de seconde opération. Dans cette réaction, une amorce

nichée spécifique de gène 28-mère (GKN6) et une amorce nichée provenant de la séquence adaptatrice (AP2) ont été utilisées. Cette réaction de PCR secondaire amplifiait spécifiquement des fragments contenant la séquence de pOP-KT21. Les fragments produits dans la réaction de PCR primaire du fait de la liaison non spécifique d'amorces ne seront pas amplifiés. La figure 9 montre le produit des réactions de PCR primaires et secondaires. La taille de la bande obtenue à partir de la réaction de PCR secondaire était d'environ 1,4 kb. Ceci est la taille prévue en utilisant la paire d'amorces nichées car la séquence du gène GKN6 est seulement d'environ 100 pb à l'intérieur de GK4. Les banques provenant de la réaction de PCR primaire et secondaire ont été purifiées à partir du gel d'agarose et clonées dans le vecteur PCR II TOPO (Invitrogen). Mis à part le fait qu'elles sont plus courtes d'environ 100 pb à l'extrémité 3', les séquences de tous les clones provenant de la réaction de PCR secondaire étaient identiques à celles du produit de PCR primaire. Un clone génomique provenant du produit de PCR primaire a été utilisé dans l'analyse subséquente.

Le clone consiste en un fragment génomique de 1479 pb contenant une région promotrice de 1347, SEQ ID NO : 3 (figure 10). La séquence nucléotidique de la région chevauchante de ce fragment génomique avec la séquence 5'-terminale de l'ADNc de pOP-KT21 est totalement homologue (figure 11), ce qui indique que c'est en réalité le promoteur pour pOP-KT21 et non pas de pKT9 qui a 15 nucléotides différents dans cette région. Le site de début de transcription et la boîte TATA ont été identifiés dans la région promotrice ainsi que les motifs AACA et analogues à Skn-1 qui sont importants dans la régulation de l'expression spécifique de l'endosperme du gène de glutéline du riz, *GluB-1* (figure 10).

EXEMPLE 6

Clonage du promoteur de pOP-KT21 dans pEGFP-1

Procédés

Deux amorces ont été utilisées pour cloner un fragment génomique de 1375 pb contenant la séquence promotrice de 1347 pb dans le site de clonage multiple de pEGFP-1 (Clontech), un vecteur sans promoteur avec GFP comme gène rapporteur pour produire GluP-EGFP.

La première amorce GAK5 (5' CCG CTC GAG TCA AAT TAT CAA AAT ATC 3', SEQ ID NO : 11) est une amorce sens provenant de l'extrémité 5' du promoteur avec un site *Xho* I introduit. La seconde amorce GAK4 (5' CGC GGA TCC AAG AGA GAA ACG CTG 3', SEQ ID NO : 12) est une amorce antisens de 17 bases en amont du site de début de la traduction avec un site *Bam* HI introduit.

Le mélange de réaction de PCR (50 µl) pour amplifier le promoteur de pOP-KT21 contenait 5,0 µl de dNTP 2 mM, 3,3 µl d'amorce GAK5 15 µM, 3,3 µl d'amorce GAK4 15 µM, 25 ng de plasmide pKT21-Pa, 5,0 µl de tampon d'enzyme 10X contenant 1,5 mM de $MgCl_2$ et 2,6U de polymérase Expand High Fidelity (Roche). Les conditions de la PCR étaient les suivantes : 1 cycle ; 94°C pendant 3 min, 20 cycles ; 94°C pendant 1 min, 58°C pendant 1 min et 72°C pendant 90 s suivi par 1 cycle ; 72°C pendant 10 min. Le produit de PCR a été purifié avec le kit de purification de PCR QIAquick (Qiagen). La ligature a été réalisée avec un rapport molaire 1:3 de vecteur:insert dans un volume réactionnel de 15,0 µl contenant 1,5 µl de tampon ligase 10X et 1,5 µl d'ADN ligase de T4 (1 U/µl) et incubation à 16°C O/N. Deux microlitres ont été utilisés pour transformer des cellules compétentes JM101 comme décrit dans Siti Nor Akmar (1999, *supra*). Les cellules ont été étalées sur une boîte de LB contenant 50 µg/ml d'ampicilline pour sélectionner les cellules transformées. L'ADN plasmidique a été préparé à partir des colonies bactériennes avec le QIAprep spin miniprep kit (Qiagen). L'analyse de restriction a été réalisée par digestion avec *Xho* I et *Bam* HI pour confirmer la taille de l'insert. Le séquençage de GluP-EGFP a été réalisé avec l'amorce de séquençage EGFP-N (Clontech).

Résultats

Le fragment génomique de 1375 pb contenant la région promotrice de 1347 pb a été cloné dans un vecteur de transformation sans promoteur pEGFP (Clontech) contenant GFP comme gène rapporteur et le vecteur de transformation chimérique produit a été appelé GluP-EGFP. Dans EGFP, l'intron cryptique, qui conduit à une protéine non fonctionnelle dans *Arabidopsis* (Haseloff et Amos. *Trends*

Genet. 11 : 328-329, 1995) a été retiré. Une mutation d'aminoacide (Phe-64 en Leu) a été introduite aussi pour augmenter les signaux de fluorescence. L'analyse de restriction de GluP-EGFP avec *Xho* I et *Bam* HI montrait un clonage réussi du fragment génomique de 1375 pb dans pEGFP (figure 12) et est soutenue par les données de séquence obtenues avec l'amorce EGFP-N (Clontech).

EXEMPLE 7

Procédé biolistique et expression transitoire pour l'analyse du promoteur

Procédés

Préparation de coupes tissulaires, paramètres de bombardement et détection de points de GFP dans les tissus bombardés

Le mésocarpe et l'endosperme de palmier à huile des fruits à 12 saa et des feuilles ont été stérilisés par immersion dans RBS pendant 15 min suivie par chlorox à 25 % pendant 15 min pour le mésocarpe et l'endosperme et chlorox à 10 % pour les feuilles. Les tissus ont ensuite été rincés plusieurs fois avec de l'eau distillée stérile avant la culture.

Pour 1 litre de milieu de culture, 4,41 g de sels MS et 30 g de saccharose ont été utilisés. Les milieux ont été ajustés à pH 5,8 avec NaOH 1M et solidifiés avec de la phytogélose à 0,8 % (m/v). Le mésocarpe, l'endosperme et les feuilles de palmier à huile ont été coupés en petites coupes d'approximativement 1 cm x 1 cm de diamètre. Les explants ont été placés sur les milieux de culture et maintenus à 28°C à l'obscurité pendant 24-48 h avant le bombardement. Les coupes de tissu de palmier à huile ont été bombardées avec le système de délivrance de particules Biolistic™, PDS-1000/He (Biorad USA). La chambre et le macrosupport ont été stérilisés avec de l'éthanol à 100 % avant l'utilisation.

60 µg de microsupport (particules d'or d'une taille de 1,0 µm) ont été placés dans 1 ml d'éthanol à 100 % dans un tube de microcentrifugeuse puis vortexés et centrifugés pendant 1-2 min à 12 000 tr/min et le surnageant a ensuite été retiré. Ce processus a été répété trois fois. 1 µl de H_2O stérile a été ajouté, le mélange a été soumis à une sonication pendant 5 s avant la centrifugation et le retrait subséquent du surnageant. Cette étape a été répétée trois fois. Les particules d'or ont finalement été remises en suspension dans 1 ml

d'eau distillée stérile et mis sous forme de portions aliquotes dans un volume de 50 µl.

Les plasmides ont été isolés au moyen d'un Qiagen Spin Miniprep Kit. 10 µg d'ADN ont été ajoutés à la portion aliquote de 50 µl d'or. Pour les expériences avec un témoin interne, 5 µg de plasmide pBI221 contenant le promoteur de CaMV avec GUS comme gène rapporteur ont été ajoutés à 10 µg de la construction de promoteur ayant GFP comme gène rapporteur avant le mélange avec la portion aliquote de particules d'or de 50 µl. 100 µl de $CaCl_2$ 2,5M, 40 µl de spermidine 0,1M ont ensuite été ajoutés tout en traitant au vortex. Le mélange a été centrifugé à 10 000 tr/min. Le surnageant a été retiré et le microsupport a été lavé avec de l'éthanol à 100 %. Ces étapes ont été répétées deux fois et finalement le microsupport a été remis en suspension dans 60 µl d'éthanol et maintenu à -20°C jusqu'à son utilisation.

Pour chaque bombardement, 8 µl de particules d'or recouvertes d'ADN ont été chargés sur un microsupport et mis à sécher. Les tissus de mésocarpe et d'endosperme ont été bombardés avec une pression d'hélium de 10700 x 10^3 Pa (1550 psi) à une distance de 9 cm entre le macrosupport et le tissu cible. Les feuilles (tissus témoins) ont été bombardées à 1100 avec une distance de 6 cm entre le macrosupport et le tissu cible. La pression du vide a été maintenue à 35 x 10^2 Pa (27 mm de Hg).

L'expression de GFP a été déterminée en comptant les points de fluorescence verte produits visualisés sous un microscope à fluorescence Leica muni d'un ensemble de filtres GFP.

Essai histochimique avec GUS

Un essai enzymatique avec GUS a été réalisé avec le X-Gluc (5-bromo-4-chloro-3-indoly-β-D-glucuronide) comme substrat, selon le processus décrit par Jefferson *et al.* (1987) avec certaines modifications. Après l'analyse de la présence de points fluorescents de GFP, les coupes de tissu d'endosperme, de mésocarpe et de feuilles transformées provenant de chaque co-bombardement ont été placées séparément dans une quantité suffisante de solution de coloration de GUS (tampon $NaPO_4$ 0,1 M, pH 7,0, EDTA 10 mM, pH 7,0, ferricyanure

de K 0,5 mM, pH 7,0, ferrocyanure de K 0,5 mM, pH 7,0, X-Gluc 10 mM, et Triton-X 0,1 %) sans réaliser aucun processus de préfixation. Le récipient du tissu a été recouvert pour empêcher l'évaporation du mélange réactionnel puis incubé à 37°C pendant 12 h. L'activité enzymatique utilise X-Gluc comme substrat pour produire des précipités bleus qui ont été observés sous forme de points bleus dans différents tissus sous le spectre blanc au moyen d'un microscope Nikon SMZ1000.

Résultats

Les paramètres du bombardement qui ont été optimisés pour l'analyse du promoteur dans les tissus de mésocarpe et de feuille de palmier à huile par un essai transitoire (Zubaidah et Siti Nor Akmar, *Journal of Oil Palm Research* Vol. 15 n° 2 : 62-69, 2003) ont été utilisés dans l'analyse du promoteur de pOP-KT21 du palmier à huile. De plus, différentes conditions pour réaliser l'analyse du promoteur par un effet transitoire du gène rapporteur dans des coupes de tissu de noyau de palmier à huile ont été examinées en utilisant une construction promoteur de CaMV constitutif-gène rapporteur de protéine fluorescente verte (GFP). On a trouvé que le nombre de points fluorescents augmentait de plus de 20 fois quand les coupes de tissu utilisées pour le bombardement étaient obtenues à partir de la couche externe de l'endosperme. Ceci a été réalisé en utilisant une pression d'hélium de 1550 psi et une distance entre le macrosupport et le tissu cible de 9 cm (figure 13). Ces conditions ont ensuite été utilisées dans l'analyse de l'activité du promoteur de pOP-KT21.

Le promoteur de pOP-KT21 a été cloné dans un vecteur de transformation sans promoteur pEGFP (Clontech) qui contient le gène de la protéine fluorescente verte (GFP) comme gène rapporteur pour produire une construction de vecteur appelée GluP-EGFP. Cette construction génique a été bombardée dans des coupes de tissu de noyau de mésocarpe et de feuilles de palmier à huile. La comparaison de l'expression de GFP a été faite avec des coupes de tissu bombardées avec une construction de gène de GFP contenant le promoteur de CaMV (CaMV-EGFP) et comme témoin négatif, des tissus bombardés avec une construction de vecteur sans promoteur (pEGFP) ont été analysés aussi.

L'expression transitoire de GFP observée sous forme de points fluorescents verts a été détectée dans les coupes de tissu de noyau bombardées avec GluP-EGFP mais l'expression du GFP n'a pas été détectée dans les tissus de mésocarpe et de feuilles bombardées avec cette construction génique (figures 14 et 15) bien que l'expression de *uidA* cobombardé (*GUS*) se soit produite sous la régulation du promoteur de CaMV (figure 14). Des points fluorescents verts ont pu être détectés dans des coupes de tissu de noyau de mésocarpe et de feuilles bombardées avec CaMV-EGFP (figures 14 et 15). Des points fluorescents n'ont été trouvés dans aucune des coupes de tissu bombardées avec le vecteur sans promoteur pEGFP (figure 15). Sur la base de la comparaison de l'activité du promoteur de CaMV et du promoteur du gène de glutéline du palmier à huile (pOP-KT21) comme indiqué par l'expression transitoire de GFP, on peut conclure que le promoteur de pOP-KT21 du palmier à huile est un promoteur fonctionnel ayant une activité de promoteur spécifique du noyau.

L'homme du métier comprendra que l'invention décrite ici est susceptible de variations et de modifications différentes de celles décrites spécifiquement. Il est à noter aussi que l'invention inclut toutes ces variations et modifications. L'invention inclut aussi toutes les étapes, caractéristiques, compositions et composés cités ou indiqués dans cette description, individuellement ou collectivement, et toutes les combinaisons de ces étapes ou caractéristiques quelconques.

BIBLIOGRAPHIE

Altschul *et al., Nucl. Acids Res. 25* : 3389-3402, 1997

Ausubel *et al., Current Protocols in Molecular Biology*, John Wiley & Sons Inc, 1994-1998, chapitre 15, 1998 ;

Beyer *et al., Proceedings of the American Society for Nutritional Sciences Symposium on Plant Breeding* : 506S-510S, 2002 ;

Bonaventure and Ohlrogge, *Plant Physiol. 128* : 223-235, 2002 ;

Brown, *The Plant Kingdom,* Blaidsdell Publishing Company, New York, 1935 ;

Cha et Shah, *Plant Science 160* : 913-923, 2001 ;

Daniell *et al.*, *Trends in Plant Science 6(5)* : 219-226, 2001 ;

Dickey *et al., Plant Cell 3* : 475-484, 1998 ;

Galas et Schmitz, *Nucl. Acids Res. 5* : 3157-3170, 1978 ;

Haseloff et Amos, *Trends Genet. 11* : 328-329, 1995 ;

Katsube *et al.,* dans *Engineering Crop plants for Industrial End Uses*, Shewry *et al.* (Eds), Portland Press, Londres, 1998, p. 65-76 ;

Kinney, *Journal of Food Lipids 3* : 273-292, 1996 ;

Muntz, *Plant Molecular Biology 38* : 77-99, 1998 ;

Sanford et Wolf, brevet US n° 4 945 050 ;

Schmitz, *Nucl. Acids Res. 5* : 3157-3170, 1978 ;

Shewry, *Biol. Rev. 70* : 375-426, 1995 ;

Siew, *PORIM Bulletin 19* : 19-22, 1989 ;

Siti Nor Akmar *et al., Asia Pacific Journal of Molecular Biology and Biotechnology 2(2)* : 113-118, 1994 ;

Siti Nor Akmar, PhD Thesis, Univertisy of East Anglia, 1999 ;

Stomp *et al.*, brevet US n° 5 122 466 ;

Tackaberry *et al.*, *Genome 46(3)* : 521-526, 2003 ;

Takaiwa *et al.*, *Mol. Gen. Genet. 208* : 15-22, 1987 ;

Takaiwa *et al., Plant Mol. Biol. 17(4)* : 875-885, 1991 ;

Thompson *et al., Nucl. Acid Res.* 22 : 4673-4680, 1994 ;

Utsumi, *Adv. Food Nutr. Res. 36* : 89-208, 1992 ;

Voelker *et al.*, *Science 257* : 72-73, 1992 ;

Weigel et Glazebrook, *ARABIDOPSIS : A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 2002 ;

Zubaidah et Siti Nor Akmar, *Journal of Oil Palm Research Vol. 15 n° 2* : 62-69, 2003.

REVENDICATIONS

1. Molécule d'acide nucléique isolée comprenant une séquence de nucléotides définissant un élément régulateur d'expression caractérisée en ce que ledit élément régulateur d'expression module spécifiquement l'expression d'une seconde molécule d'acide nucléique couplée de manière fonctionnelle audit élément régulateur d'expression dans l'endosperme d'une plante.

2. Molécule d'acide nucléique isolée selon la revendication 1, caractérisée en ce que l'élément régulateur d'expression comprend un promoteur.

3. Molécule d'acide nucléique isolée selon la revendication 1, caractérisée en ce que l'élément régulateur d'expression comprend un stimulateur et/ou un silenceur.

4. Molécule d'acide nucléique isolée selon la revendication 2, caractérisée en ce que ledit acide nucléique code une région promotrice provenant d'un gène de glutéline d'une plante.

5. Molécule d'acide nucléique isolé selon la revendication 4, caractérisé en ce que ladite plante est un palmier à huile.

6. Molécule d'acide nucléique isolé selon la revendication 5, caractérisé en ce que ladite plante est du genre *Elaeis*.

7. Molécule d'acide nucléique isolé selon la revendication 6, caractérisé en ce que ladite plante est de l'espèce *Elaeis guineensis* ou *Elaeis oleifera*.

8. Molécule d'acide nucléique isolée selon l'une quelconque des revendications 1 à 7, caractérisée en ce que la séquence nucléotidique est telle que présentée dans SEQ ID NO : 3 ou une séquence nucléotidique ayant au moins environ 50 % d'identité avec SEQ ID NO : 3 après alignement optimal ou une séquence nucléotidique capable de s'hybrider à SEQ ID NO : 3 ou à sa forme complémentaire dans des conditions de faible stringence.

9. Molécule d'acide nucléique isolée selon la revendication 8, caractérisée en ce qu'elle comprend la séquence d'acide nucléique présentée dans SEQ ID NO : 3.

10. Mutant, dérivé ou fragment de la molécule d'acide nucléique selon l'une quelconque des revendications 1 à 9, caractérisé en ce que ledit fragment a une activité de régulation des gènes.

11. Fragment, partie ou portion de la molécule d'acide nucléique selon la revendication 10, caractérisé en ce que ladite activité de régulation des gènes est une activité promotrice.

12. Molécule d'acide nucléique isolée comprenant une séquence de nucléotides caractérisée en ce que la séquence nucléotidique code une protéine glutéline comprenant la séquence d'aminoacides présentée dans SEQ ID NO : 2 ou une séquence d'aminoacides ayant plus de 96 % de similarité avec SEQ ID NO : 2 après alignement optimal.

13. Molécule d'acide nucléique isolée selon la revendication 12 caractérisée en ce que la séquence nucléotidique code une séquence d'aminoacides présentée dans SEQ ID NO : 2.

14. Molécule d'acide nucléique isolée selon la revendication 12, caractérisée en ce que la séquence nucléotidique est telle que présentée dans SEQ ID NO : 1 ou une séquence nucléotidique ayant plus de 96,4 % d'identité avec SEQ ID NO : 1 après alignement optimal.

15. Molécule d'acide nucléique isolée selon la revendication 14, caractérisée en ce qu'elle comprend une séquence d'acide nucléique telle que présentée dans SEQ ID NO : 1.

16. Protéine glutéline isolée caractérisée en ce qu'elle comprend une séquence d'aminoacides présentée dans SEQ ID NO : 2 ou une séquence d'aminoacides ayant plus de 96 % de similarité avec SEQ ID NO : 2 après alignement optimal.

17. Protéine glutéline isolée selon la revendication 16, caractérisée en ce qu'elle comprend une séquence d'aminoacides telle que présentée dans SEQ ID NO : 2.

18. Construction d'ADN recombinant caractérisée en ce qu'elle comprend l'acide nucléique selon l'une quelconque des revendications 1 à 15.

19. Construction d'ADN recombinant selon la revendication 18 caractérisée en ce qu'elle comprend la molécule d'acide nucléique selon l'une quelconque des revendications 1 à 11.

20. Vecteur caractérisé en ce qu'il comprend la construction d'ADN recombinant selon la revendication 18 ou 19.

21. Vecteur selon la revendication 20, caractérisé en ce que ledit vecteur est un vecteur d'expression, ledit vecteur d'expression comprenant un élément régulateur d'expression défini par la molécule d'acide nucléique selon l'une quelconque des revendications 1 à 11 en liaison fonctionnelle avec une seconde séquence d'acide nucléique codant une molécule d'ARN codant une protéine ou ne codant pas de protéine.

22. Procédé pour exprimer le second acide nucléique tel que défini à la revendication 21, caractérisé en ce qu'il comprend la transformation d'une plante, d'une cellule végétale, d'un tissu végétal ou d'un organe végétal avec le vecteur selon la revendication 21 et la régénération de la plante transformée.

23. Procédé selon la revendication 22 caractérisé en ce que ledit second acide nucléique est exprimé spécifiquement dans l'endosperme.

24. Procédé selon la revendication 22, caractérisé en ce que ledit second acide nucléique est exprimé spécifiquement dans une ou plusieurs cellules parenchymateuses.

25. Procédé selon l'une quelconque des revendications 22 à 24, caractérisé en ce que la plante, la cellule végétale, le tissu végétal ou l'organe végétal est choisi parmi les espèces consistant en : *Elaeis guineesis* Jacq., *Elaeis oleifera* et leurs hybrides, le maïs, le riz, le soja, le tabac, le coton, la luzerne, le blé, *Arabidopsis*, le colza, le canola et le cocotier.

26. Procédé selon la revendication 25, caractérisé en ce que la plante, la cellule végétale, le tissu végétal ou l'organe végétal est un *Elaeis* spp. et ledit second acide nucléique, quand il est exprimé, réalise : une modification des qualités des protéines ou des huiles ; la production ou l'accumulation de nouvelles huiles ou de nouvelles protéines ; ou la production ou l'accumulation de nutraceutiques, de produits pharmaceutiques ou de planticorps dans l'endosperme de la plante transformée.

27. Procédé selon la revendication 26, caractérisé en ce que ledit second acide nucléique code une enzyme de biosynthèse des

acides gras ou code un transcrit antisens de celle-ci, et l'expression dudit seconde acide nucléique modifie la composition en acides gras de l'endosperme de la plante.

28. Procédé selon la revendication 26, caractérisé en ce que la provitamine A est accumulée dans l'endosperme de la plante.

29. Procédé selon la revendication 26, caractérisé en ce que des composés contenant du fer sont accumulés dans l'endosperme de la plante.

30. Procédé selon la revendication 26, caractérisé en ce qu'un vaccin oral est accumulé dans l'endosperme de la plante.

31. Procédé selon la revendication 26, caractérisé en ce que le procédé produit une plante ayant une composition en aminoacides modifiée ou une composition en huile modifiée dans le noyau, de sorte que le noyau a une plus grande valeur nutritionnelle quand il est utilisé comme aliment pour animaux que le noyau de palmier à huile de type sauvage.

32. Procédé selon la revendication 31, caractérisé en ce que le noyau a des niveaux accrus d'un ou plusieurs aminoacides essentiels pour un mammifère.

33. Cellule végétale modifiée génétiquement, caractérisée en ce qu'elle comprend la molécule d'acide nucléique selon l'une quelconque des revendications 1 à 15.

34. Cellule végétale modifiée génétiquement selon la revendication 33, caractérisée en ce que la plante comprend la molécule d'acide nucléique selon l'une quelconque des revendications 1 à 11.

35. Cellule végétale modifiée génétiquement selon la revendication 33, caractérisée en ce que la plante est produite selon le procédé selon l'une quelconque des revendications 22 à 32.

36. Plante, tissu végétal, organe végétal, partie d'une plante ou matériel reproducteur d'une plante incluant un matériel reproducteur végétatif et sexué, caractérisé en ce qu'il comprend une ou plusieurs cellules végétales modifiées génétiquement selon l'une quelconque des revendications 33 à 35.

37. Cellule, tissu, organe, plante ou partie d'une plante modifiée génétiquement selon la revendication 36 caractérisé en ce que

ladite cellule, tissu, organe, plante ou partie de celle-ci provient de la liste d'espèces consistant en : *Elaeis guineesis* Jacq., *Elaeis oleifera* et leurs hybrides, le maïs, le riz, le soja, le tabac, le coton, la luzerne, le blé, *Arabidopsis*, le colza, le canola et le cocotier.

38. Procédé pour identifier des partenaires de liaison qui se lient ou s'associent d'une autre manière avec un élément régulateur d'expression défini par la molécule d'acide nucléique selon l'une quelconque des revendications 1 à 11, caractérisé en ce qu'il comprend la mise en contact d'un partenaire de liaison supposé de ladite molécule d'acide nucléique, ledit contact étant dans des conditions telles et pendant une durée telle qu'elles permettent à une liaison de se produire ; où la liaison entre ledit partenaire de liaison supposé et ladite molécule d'acide nucléique est déterminée par une réduction de la mobilité électrophorétique de ladite molécule d'acide nucléique.

39. Procédé pour identifier des partenaires de liaison qui se lient ou s'associent d'une autre manière avec l'élément régulateur d'expression défini par la molécule d'acide nucléique selon l'une quelconque des revendications 1 à 11 caractérisé en ce qu'il comprend une empreinte à la DNase.

40. Procédé selon la revendication 38 ou 39, caractérisé en ce que ledit partenaire de liaison est un facteur de transcription.

41. Agent de liaison ou facteur de transcription, caractérisé en ce qu'il est identifié par le procédé selon l'une quelconque des revendications 38 à 40.

42. Procédé pour moduler l'activité de l'élément régulateur d'expression défini par la molécule d'acide nucléique selon l'une quelconque des revendications 1 à 11, caractérisé en ce qu'il comprend l'administration à une ou plusieurs cellules comprenant ledit élément régulateur d'expression d'un agent qui se lie audit élément régulateur d'expression et module l'activité dudit élément régulateur d'expression.

43. Procédé selon la revendication 42, caractérisé en ce que ledit agent est un facteur de transcription.

44. Procédé selon la revendication 42 ou 43, caractérisé en ce que la modulation de l'activité dudit élément régulateur d'expression module l'expression d'une seconde séquence d'acide nucléique liée de manière fonctionnelle audit élément régulateur d'expression.

```
                                                    M  S  S  F  S
aagcaatagccttcagcgtttctctctttcttaatttctaaagccatgtcgtccttctct   60
L  L  S  F  S  L  C  L  L  L  L  C  R  V  S  Q  A  Q  F  G
ttgctctcctttttctctttgcttgttgctcctgtgtcgtgtgtcccaggctcaatttggg  120
S  S  Q  E  S  P  F  Q  S  S  R  R  R  S  V  S  T  R  D  E
tccagtcaggagagcccatttcagagctcacgacgtcggtcggtatcaactcgggacgag  180
C  R  M  E  R  L  N  A  L  E  P  T  R  T  V  R  S  E  A  G
tgccggatggagaggctgaatgccctcgagccgacaaggactgtgaggtccgaagctggt  240
I  T  D  Y  F  D  E  D  N  E  Q  F  R  C  A  G  V  S  A  I
atcacggattactttgatgaggacaatgaacagttccggtgcgccggcgtgtccgcgatc  300
R  R  V  I  E  P  R  G  L  L  L  P  S  M  S  N  A  P  R  L
cgccgagtgatagaacccagaggccttctcctccctccatgtccaatgcccctcgcctc  360
V  Y  I  V  Q  G  R  G  L  V  G  L  V  M  P  G  C  P  E  T
gtctacatcgtccaaggaaggggtctcgttggacttgtaatgcctggctgccccgaaact  420
F  Q  S  F  Q  R  S  E  Q  Y  E  R  E  E  G  E  R  H  R  R
ttccaatccttccagcgatccgagcagtatgaacgcgaggaaggcgagcgacatcggaga  480
S  R  D  E  H  Q  K  V  Y  Q  F  Q  E  G  D  V  L  A  V  P
tccagagatgaacaccaaaaagtctaccagttccaagagggagacgtcctggctgtgcct  540
D  G  F  A  Y  W  C  Y  N  N  G  E  N  P  V  V  A  I  T  V
gatggatttgcttactggtgctacaacaatggggagaaccctgtggtcgcgatcaccgtc  600
L  D  T  G  N  D  A  N  Q  L  D  R  S  H  R  Q  F  L  L  A
ctcgataccggtaacgatgctaaccagctcgatcgcagccacagacaattcttattggct  660
G  R  Q  E  E  G  R  Q  R  Y  R  R  G  E  S  I  K  E  N  I
ggaaggcaggaggaaggccggcaacgatatcgccgtggggagagcatcaaggaaaacatc  720
L  R  G  F  S  T  E  L  L  A  A  A  F  G  V  N  M  E  L  A
ctgagagggttcagcaccgagttgctggcagcggcttttggcgttaacatggagctggca  780
R  K  L  Q  C  R  D  D  T  R  G  E  N  V  R  A  E  N  G  L
aggaagctccagtgcagggatgacacaaggggcgagaacgtgcgggcggagaacgggctt  840
Q  V  L  R  P  S  R  M  E  E  E  E  R  E  E  S  R  R  K  N
caggtgctgaggccctcaaggatggaggaagaagagagggaagagagtagaaggaaaaat  900
G  L  E  E  T  Y  C  S  M  K  I  K  Q  N  I  G  D  P  R  R
ggattggaggagacctattgctcgatgaagattaagcagaacattggggatccaaggcgt  960
A  D  V  F  N  P  R  G  G  R  I  T  T  L  N  S  E  K  L  P
gccgacgtcttcaacccaaggggcggaaggattactaccctcaacagcgagaagcttccg 1020
I  L  R  F  I  Q  M  S  A  E  R  V  V  L  Y  R  N  A  M  V
atccttaggttcatccaaatgagtgctgagagggtggttctctacaggaacgccatggta 1080
S  P  H  W  N  I  N  A  H  S  I  M  Y  C  T  G  G  R  G  R
tcaccccactggaacatcaatgccccacagcatcatgtattgtaccggaggacgaggccgt 1140
V  E  V  A  D  D  K  G  E  T  A  F  D  G  E  L  R  Q  G  Q
gtcgaggttgctgatgacaagggagaaactgcgttcgacggagagctccgccaaggtcag 1200
L  L  I  V  P  Q  N  Y  A  V  S  E  R  A  G  S  E  G  F  Q
ctcttgatcgtcccgcaaaactatgcggtgtcggaaagggcgggaagtgaaggcttccag 1260
F  V  S  I  K  T  S  D  R  A  M  V  S  A  I  V  G  K  T  S
ttcgtatcgatcaagaccagcgaccgggccatggtgagcgcaatcgttgggaagacgtcg 1320
A  L  R  G  M  P  E  E  V  L  M  N  S  Y  R  I  S  R  D  E
gctctccggggcatgccggaggaggtgctgatgaactcctaccgcatctcaagggacgaa 1380
A  R  R  V  K  L  N  R  G  D  E  V  A  I  F  T  P  R  R  E
gcaaggagggtcaaactcaacaggggggacgaggtggcgatcttcactcctaggagagag 1440
S  R  A  E  A  *
tcgagagctgaagcttgatgcgatcttagcagggagtgctgctaggagtagagagacttt 1500
agtcatcagcggtggtggataagtaaatagagcctgaggttgctcttcgttctctgcata 1560
tgagaatataataataaaatgaaatgtccgcctgcgagcaagggcgaattccagcacact 1620
ggcggccgttactagtggatccgagctcggtaccaagcttgatgcatagcttgagtatt  1679
```

Fig. 1 SÉQUENCE NUCLÉOTIDIQUE ET SÉQUENCE D'AMINOACIDES DÉDUITE D'UN CLONE D'ADNc DE PALMIER À HUILE DE LONGUEUR INTEGRALE (pOP-KT21) POUR UN GÈNE DE GLUTÉLINE

Fig. 2 REPRÉSENTATION SOUS FORME DE TABLEAU RÉSUMANT LES POURCENTAGES D'IDENTITÉ AU NIVEAU DES NUCLÉOTIDES ET AU NIVEAU DES AMINOACIDES ENTRE DES SÉQUENCES DES DIFFÉRENTS GÈNES DE GLUTÉLINE DU PALMIER À HUILE

ALIGNEMENT	POURCENTAGES D'IDENTITÉ (%)							
	5'-UTR	3'-UTR	NUCLÉOTIDES CODANTS			AMINOACIDES		
			SOUS-UNITÉ ACIDE	SOUS-UNITÉ BASIQUE	TOTAL	SOUS-UNITÉ ACIDE	SOUS-UNITÉ BASIQUE	TOTAL
pOP-KT21 ET pKT9	78.4	90.7	95.1	99.7	96.4	93.7	99.5	96
pOP-KT21 ET pKT7	43.6	48.1	66.7	74	69.8	49.7	70.4	57.4

2913694

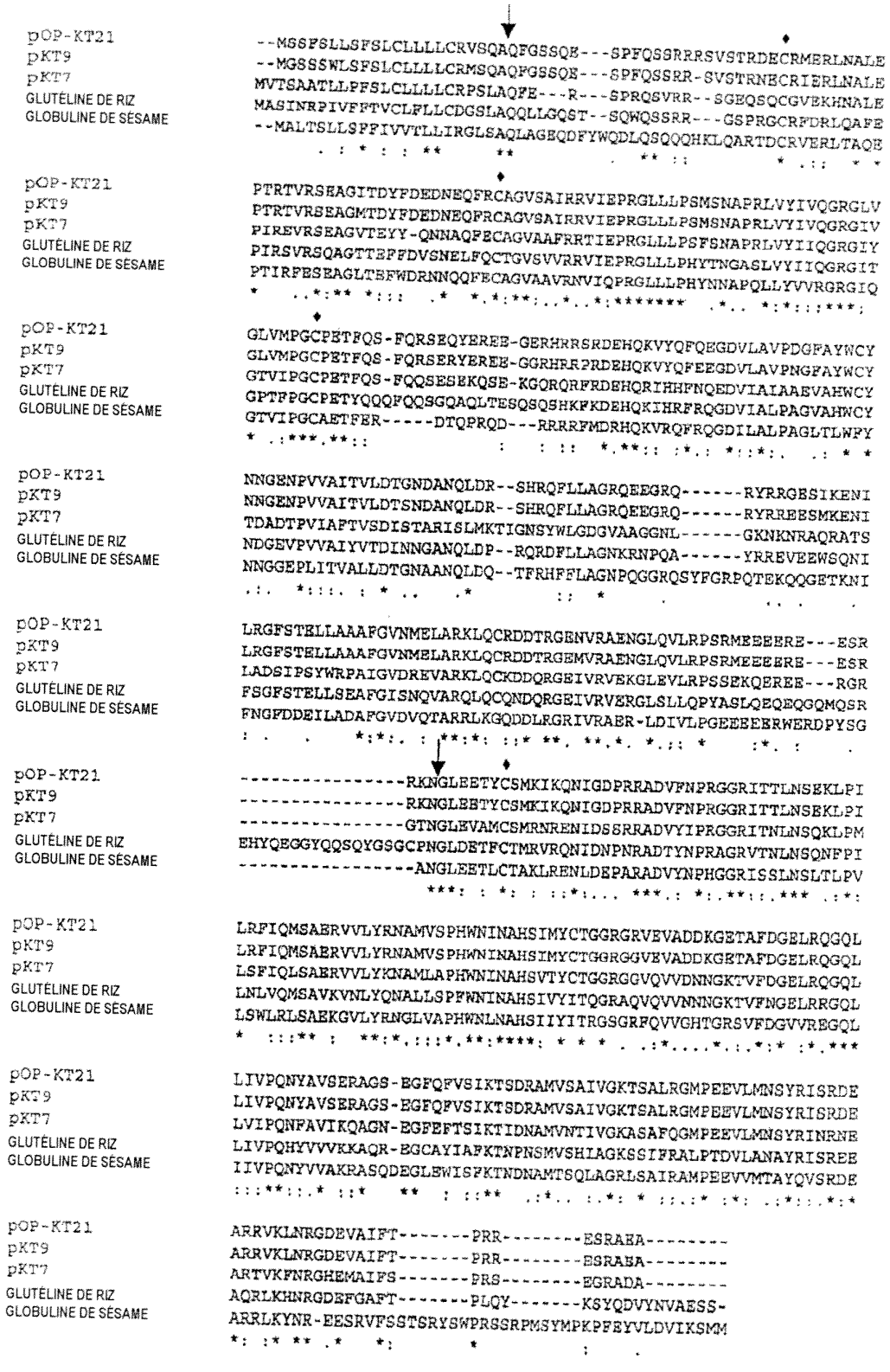


Fig. 3 COMPARAISON ENTRE LES SÉQUENCES D'AMINOACIDES DÉDUITES DE GÈNES DE GLUTÉLINE DE PALMIER À HUILE AVEC CELLE DU GÈNE DE GLUTÉLINE DU RIZ (P07730) ET DU GÈNE DE GLOBULINE DE SÉSAME 11S (AF240004)

Fig. 4 REPRÉSENTATION SOUS FORME DE TABLEAU MONTRANT UNE COMPARAISON ENTRE LES PROFILS D'AMINOACIDES DE LA SÉQUENCE D'AMINOACIDES DÉDUITE DE pOP-KT21 (A) ET LA COMPOSITION D'AMINOACIDES DE LA FARINE DE NOYAU DE PALMIER (B)

A. UTILISATION DES AMINO

AMINO TOTAL (LE NOMBRE D'AMINOACIDES) = 470
MASSE MOLÉCULAIRE = 53580.71

AMINOACIDE	NOMBRE	MOL (POUR-CENT)
Ala(A)	31	6.60
Arg(R)	59	12.55
Asn(N)	22	4.68
Asp(D)	20	4.26
Cys(C)	9	1.91
Gln(Q)	24	5.11
Glu(E)	46	9.79
Gly(G)	35	7.45
His(H)	5	1.06
Ile(I)	19	4.04
Leu(L)	38	8.09
Lys(K)	11	2.34
Met(M)	13	2.77
Phe(F)	19	4.04
Pro(P)	17	3.62
Ser(S)	39	8.30
Thr(T)	17	3.62
Trp(W)	2	0.43
Tyr(Y)	12	2.55
Val(V)	32	6.81
Total	470	99.99

AMINOACIDES	COMPOSITION EN AMINOACIDES DE LA FARINE DE NOYAU DE PALMIER (g/kg)
Arginine	23.6
Cystine	2.8
Glycine	8.1
Histidine	3.2
Leucine	11.9
Isoleucine	6.4
Lysine	5.4
Methionine	3.3
Phenylalanine	7.9
Threonine	6.1
Tryptophan	2.0
Tyrosine	4.7
Valine	8.2

B. COMPOSITION EN AMINOACIDES DE LA FARINE DE NOYAU DE PALMIER (g/kg) (BASE FRAÎCHE) (SIEW, 1989)

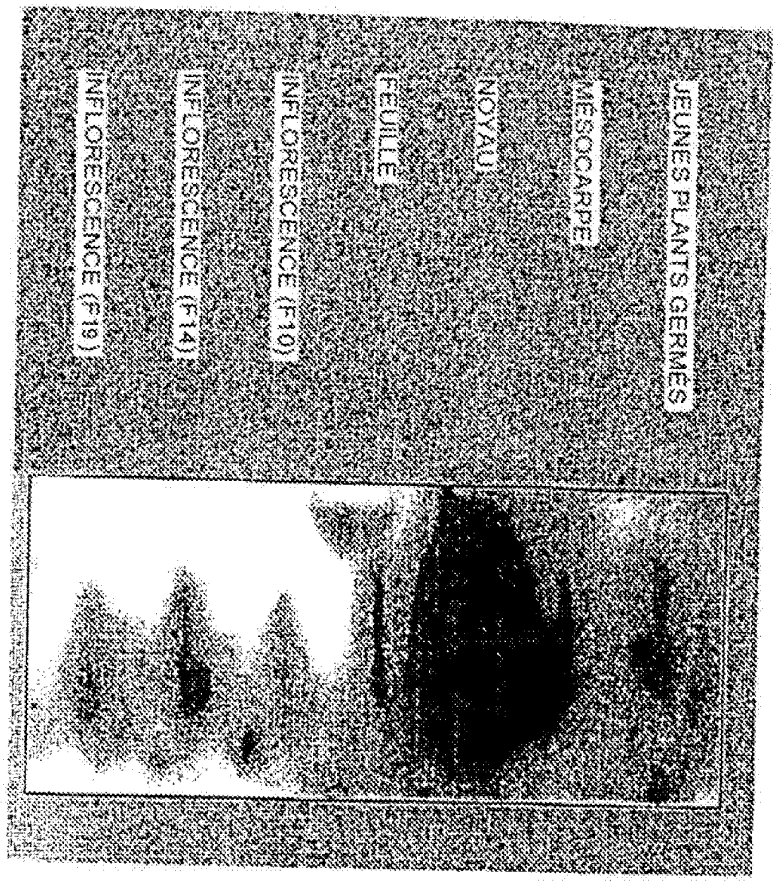


Fig. 5 DIAGRAMME D'EXPRESSION DU GÈNE DE GLUTÉLINE DU PALMIER À HUILE (SOUS-FAMILLE B) DANS DIFFÉRENTS TISSUS DE PALMIER À HUILE DÉTERMINÉ PAR ANALYSE PAR SLOT BLOT AVEC L'EST DE pOP-KT21 MARQUEE COMME SONDE.

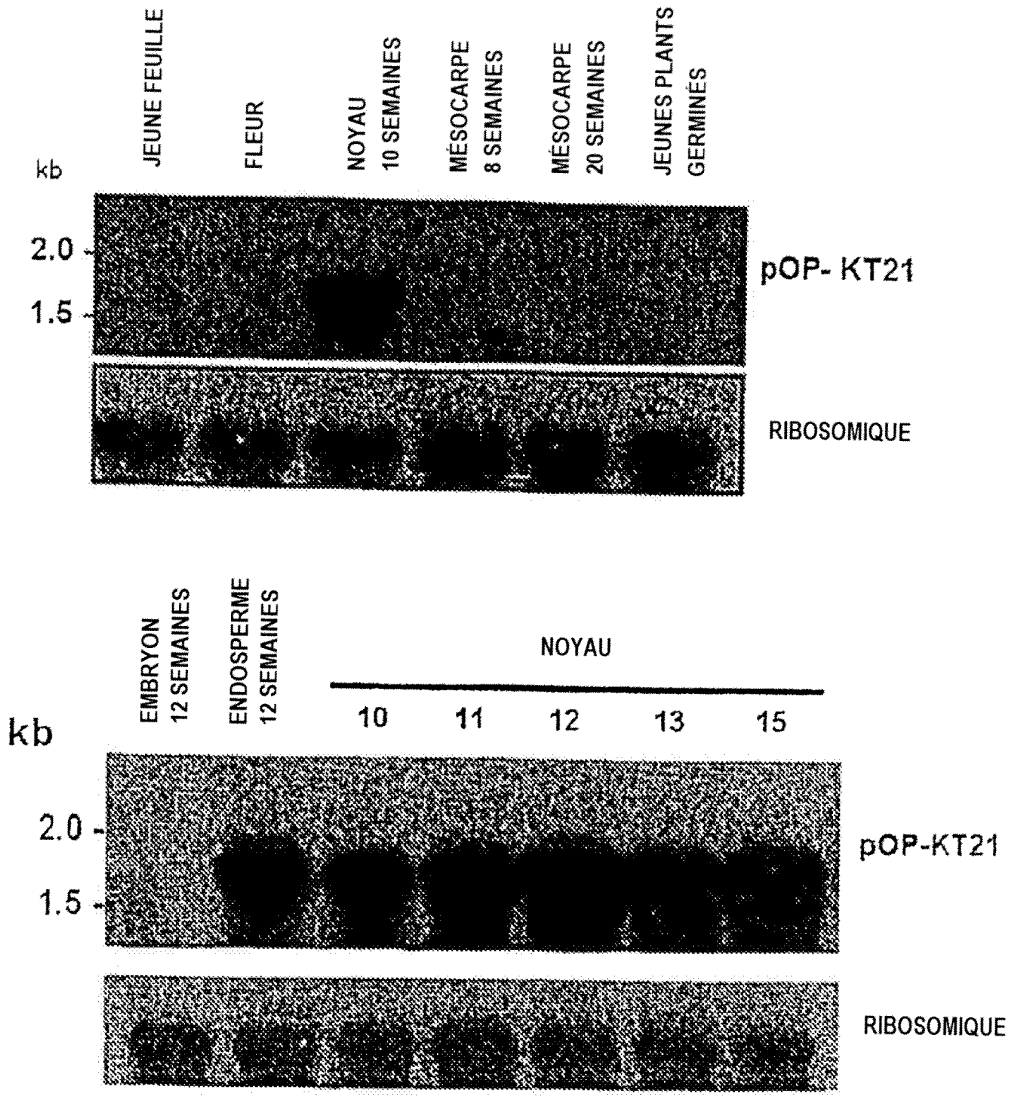


Fig. 6 DIAGRAMME D'EXPRESSION DU GÈNE DE GLUTÉLINE DU PALMIER À HUILE (SOUS-FAMILLE B) DANS DIFFÉRENTS TISSUS DU PALMIER À HUILE DÉTERMINÉ PAR ANALYSE PAR NORTHERN BLOT AU MOYEN D'UNE SONDE MARQUÉE CONTENANT LA SÉQUENCE D'ADNc DE pOP-KT21.

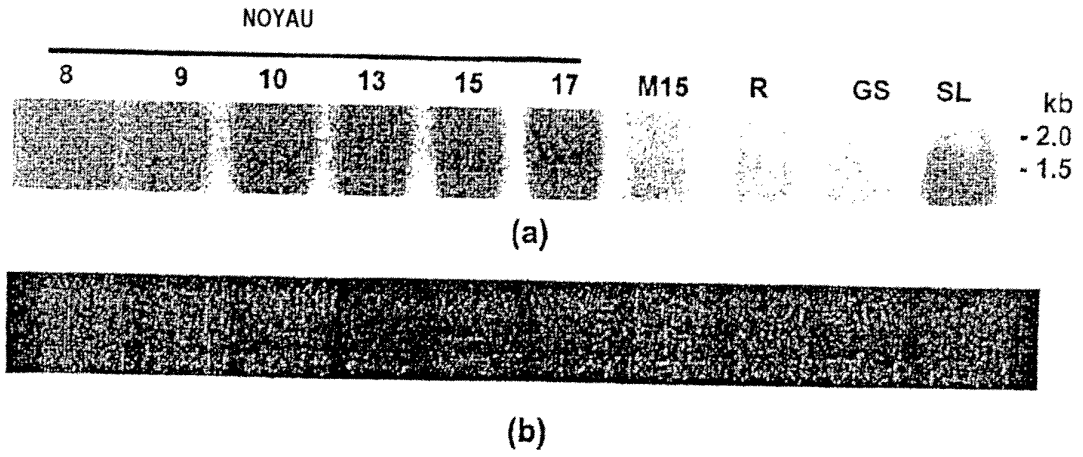
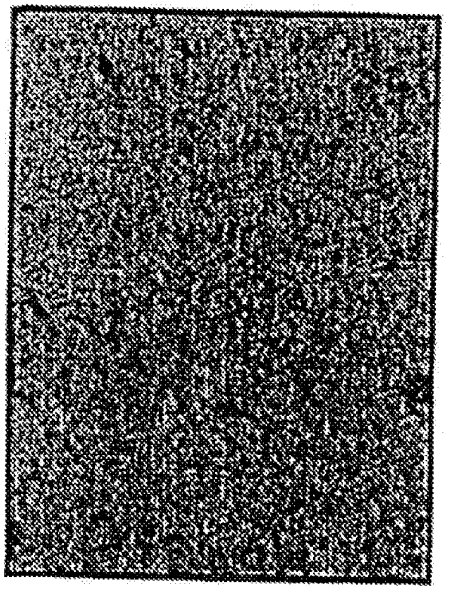
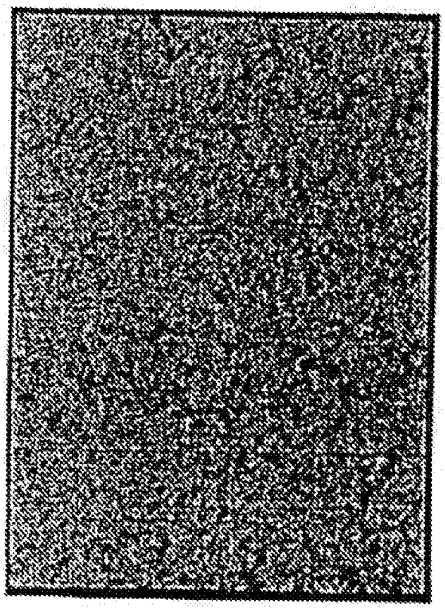


Fig. 7 DIAGRAMME D'EXPRESSION DE pOP-KT21 DANS DIFFÉRENTS TISSUS DU PALMIER À HUILE DÉTERMINÉ PAR ANALYSE PAR NORTHERN BLOT AU MOYEN D'UNE SONDE MARQUEE SPECIFIQUE DE GENE DE 5'-UTR ET DE LA RÉGION CODANTE EN 5' DE pOP-KT21.



PANNEAU (a)
SONDE ARN SENS



PANNEAU (b)
SONDE ARN ANTISENS

Fig. 8 DIAGRAMME D'EXPRESSION SPÉCIFIQUE DE CELLULE DU GÈNE DE GLUTÉLINE DU PALMIER À HUILE (SOUS-FAMILLE B) DANS LE TISSU D'ENDOSPERME DU PALMIER À HUILE DÉTERMINÉ PAR HYBRIDATION IN SITU D'ARN AU MOYEN D'UNE SONDE ARN SENS (PANNEAU A) ET ANTISENS (PANNEAU B) PRÉPARÉE AVEC L'EST DE pOP-KT21 DE PALMIER À HUILE.

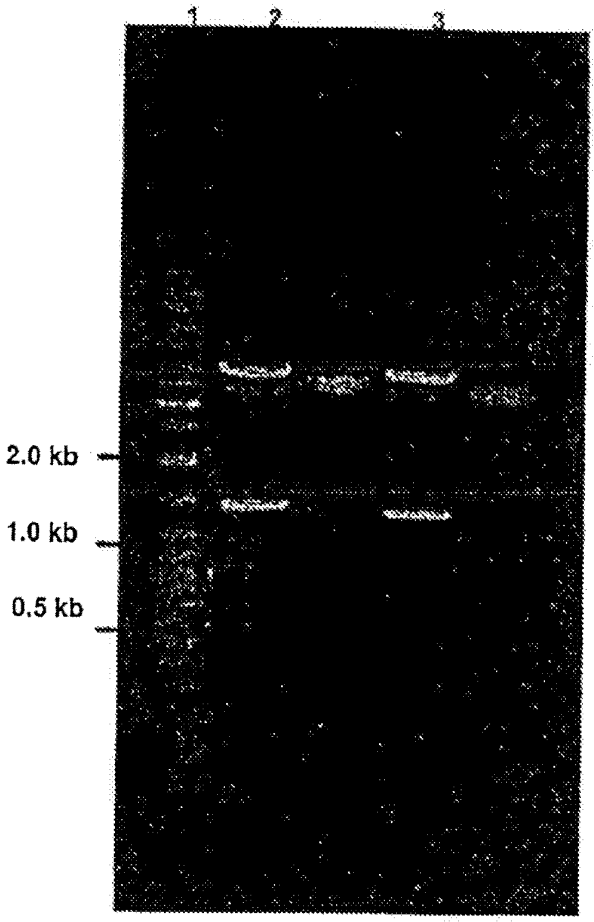


Fig. 9 PRODUIT DE LA RÉACTION DE MARCHE SUR LE GÉNOME POUR ISOLER LE PROMOTEUR DU GÈNE pOP-KT21

```
TCAAATTATC AAAATATCTC CAATAATTTT ATCCTGATGC TTTTCTTTTT  -1296
TTTTTAGTAA GAAGTGGTGG AAGTGGTATG GGTGTAGTGG TGACACGAAG  -1246
AAGCGGTGGC TGGAGGGTGT AAAAAATTGT GGGCATGGGG GGAGAGGGTG  -1196
GGTGTGCGCG GAGGAGCTAA GGGGGTGGTG GAGACGAGCG TGGAGGAGCC  -1146
GCAGGGGGGA GAGTGGTGGG GGTGGAGGGT AGAGGAGCCA TGGACAAGCG  -1096
AGGGGGTGGA GGAGCCACAA GTCGTGGGGG AGCAGTGTGC TTAAAGGAGC  -1046
TAGTGGGTGG TGGGGATGAG CATGGAAGAG CTGCGGGGGA TGGTGGGGGA  -996
                MOTIF ANALOGUE A Skn-1
TGAGGGAGGA GGAGTCATGG GCAGGCAAGG GGGTGGAGAA GTCACAGGTC  -946
AGGAGGGTGA TAAAGGAGCC ATGGGTAGGT GGTGGGGCAG AGGAGCTGCG  -896
GGCAAGCGGT AGAGATGGAC GAGCCATAGG TGATGCCGGG GGAGGGGTCG  -846
GGAAGGAGGG GGAGCGTTGG GCTATGGGTT TTATGTTTTT CTTTTTAAGG  -796
GATAATTTTA TCCAAAATTT TTATTGCAAT ACTGTCAAAA AAGTGGTAAT  -746
CTGATTAGCC ATTTCTTTCC GAGGCTATTT AAATTATCTC ATAGGAGGTA  -696
ATATAGATTG TCAAGACAAT TTGGATTGCC ACCCAAAAGC ATCCCAAATG  -646
CAGTAATCCA ATGGAGCCAC TTTAAATTGC TGACGATCTA AATAGATTAC  -596
CAGCTACCAA ATGCAACATA AAGGCTAACT CCCAAGCTTG CATTGCTTGC  -546
TAAATCTGTA GCACGTAGTC CTTGGACTCT TGATTTTGTT TGTCTTTTGA  -496
AAAATTTCTT TGGACTTCAG TATACTTCTC ATCTATCATA TTATACCTTA  -446
GGTCACTTAG CATTTTTAGA AGAAAGAGAT GTTATTTTGA AAATAAAAAA  -396
CTCTAACAGC ATACTAATTG AACAAGATCA TCAGAATACT TGGCAGGACC  -346
AAGAGGGATA GGAGCTAGCC CACCCAGCAA TATATATAAT TCACTGAAAT  -296
CATCTGATCT ACTCTTGCAT TCTGATGTGC TTGCTGCAAG CGGTTGGTAC  -246
GGTTGTCCTC TTCCTTCCAA TATTTACAGT CCATTCATCC AAGTGTATGA  -196
CACGTGGCCA CCTGTATTCT GTCCTCTCTG GACCAAGCTG CCACGCACCC  -146
ATCTCCATAT ATCTTATGCG GCAACACGAG AGCTCATGCA TGGCAAAGCA  -96
                                                MOTIF AACA
TGGGCGTCGT ATCTCACAAA AGTGCATCCA GATGGGTCCA AGCCTAAACA  -46
          BOÎTE TATA                    DÉBUT DE TRANSCRIPTION
ACTTTTTTGG CCTATAAATG CCCACCCCAT CTCGCCACCC TTGCTCAAGC  4

AATA                                                    8
```

Fig. 10 SÉQUENCE DU PROMOTEUR DU GÈNE DE GLUTÉLINE DU PALMIER À HUILE (pOP-KT21)

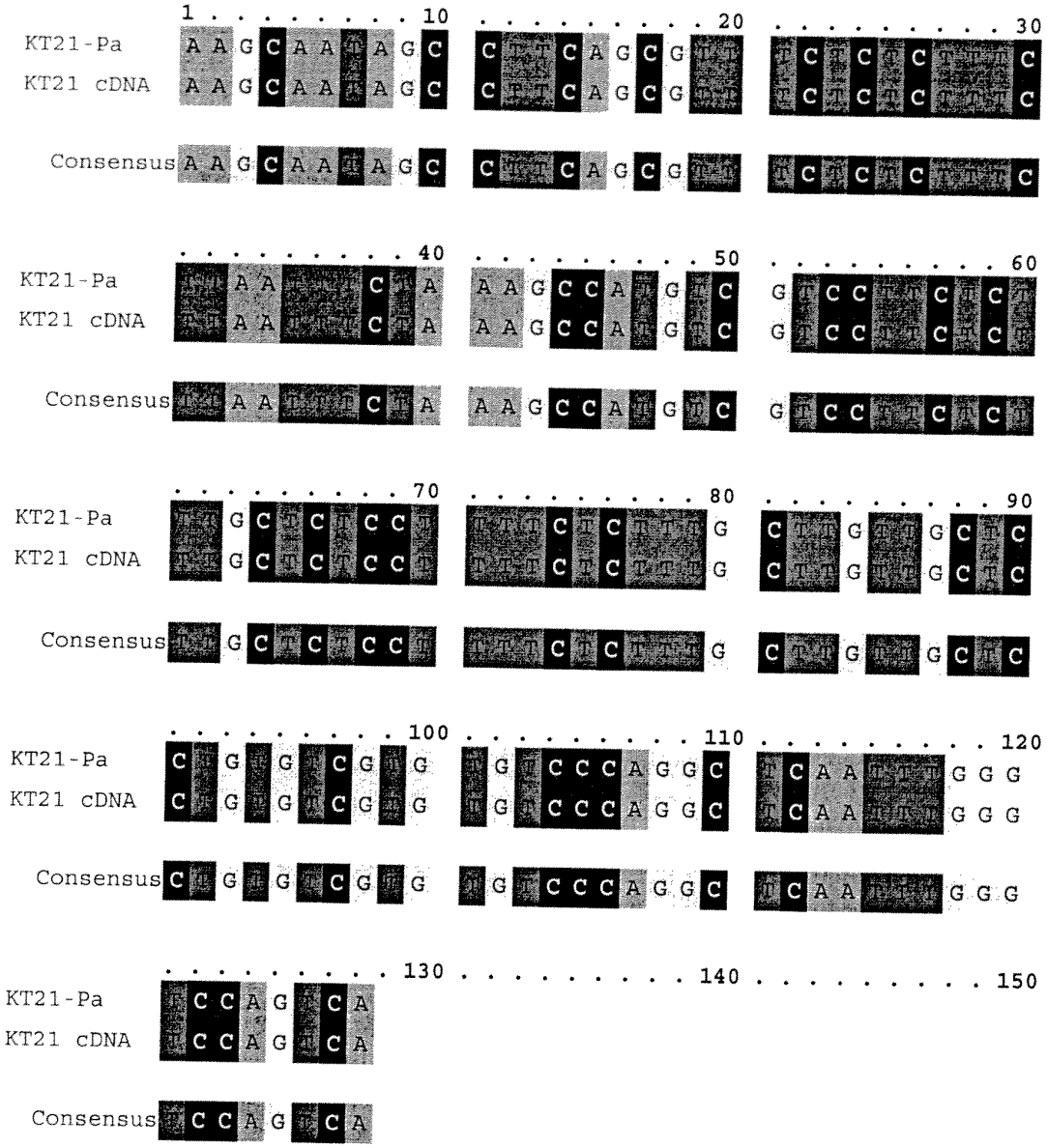


Fig. 11 ALIGNEMENT DE SEQUENCES D'ADN de Pkt21-Pa (SEQUENCE GENOMIQUE) ET DE pOP-KT21 (SEQUENCE D'ADNc)

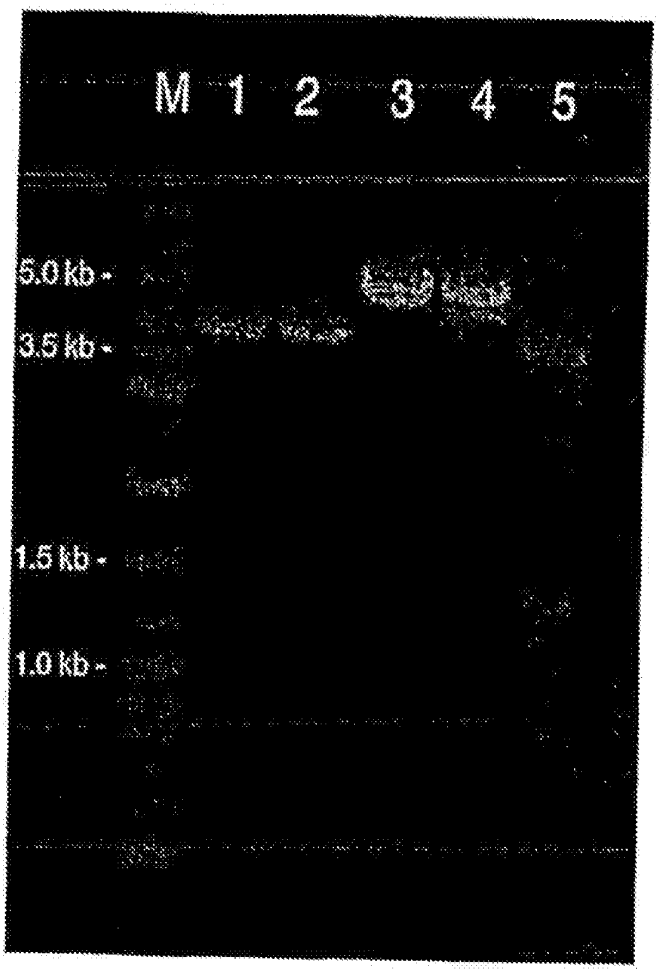


Fig. 12 ANALYSE DE RESTRICTION DU VECTEUR DE TRANSFORMATION CHIMÉRIQUE GluP-EGFP

Fig. 13 REPRÉSENTATION SOUS FORME DE TABLEAU MONTRANT LES PARAMÈTRES DE BOMBARDEMENT OPTIMISÉS POUR UNE ANALYSE D'ESSAI TRANSITOIRE UTILISANT DES COUPES DE TISSU DE NOYAU DE PALMIER À HUILE

COUPES DE TISSU	DISTANCE ENTRE LE MACROSUPPORT ET LES TISSUS CIBLES	PRESSION D'HÉLIUM	NOMBRE DE POINTS FLUORESCENTS
COUCHE EXTERNE DE L'ENDOSPERME	9 cm	1550 Psi	PLAQUE 1 – 88 PLAQUE 2 - 230
COUCHE INTERNE DE L'ENDOSPERME	9 cm	1550 Psi	PLAQUE 1 – 6 PLAQUE 2 - 7
COUCHE INTERNE DE L'ENDOSPERME	9 cm	1350 Psi	PLAQUE 1 – 1 PLAQUE 2 – 2
COUCHE INTERNE DE L'ENDOSPERME	9 cm	1880 Psi	PLAQUE 1 – 3 PLAQUE 2 - 6

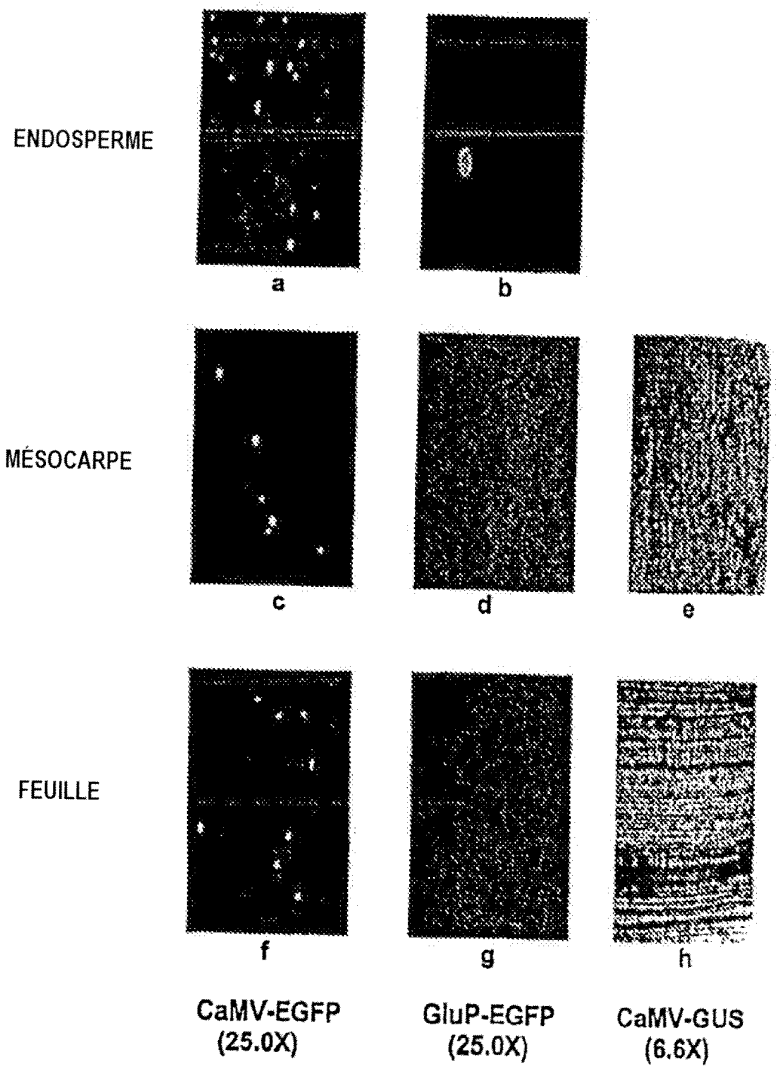


Fig. 14 ESSAI PROMOTEUR-GÈNE RAPPORTEUR POUR CONFIRMER L'ACTIVITÉ SPÉCIFIQUE DE L'ENDOSPERME DU PROMOTEUR DU GÈNE DE GLUTÉLINE DU PALMIER À HUILE (GluP) PAR COMPARAISON DE L'EXPRESSION DE LA PROTÉINE FLUORESCENTE VERTE DANS LES TISSUS DE L'ENDOSPERME, DU MÉSOCARPE ET FOLIAIRES DU PALMIER À HUILE BOMBARDÉS AVEC GluP-EGFP et CaMV-EGFP.

Fig. 15 REPRÉSENTATION SOUS FORME DE TABLEAU MONTRANT LES NOMBRES DE POINTS DE GFP DANS LES TISSUS FOLIAIRES, LE MÉSOCARPE ET LE NOYAU BOMBARDÉS AVEC GluP-EGFP, CaMV-EGFP et pEGFP

TISSUS/VECTEUR	GluP-EGFP	CaMV-EGFP	pEGFP
ENDOSPERME	7 ± 1	45 ± 2.5	0
MÉSOCARPE	0	71 ± 1.4	0
FEUILLES	0	170 ± 9.1	0

SEQUENCE LISTING

<110> MALAYSIAN PALM OIL BOARD
ABDULLAH, Siti Nor Akmar
CHEAH, Suan Choo
ABD. WAHAB, Nurniwalis

<120> Elements régulateurs d'expression

<130> 3J408770 0001 FR BN

<140> FR 08 51645
<141> 2008-03-13

<150> PI 20070394
<151> 2007-03-13

<160> 14

<170> PatentIn version 3.3

<210> 1
<211> 1413
<212> DNA
<213> Elaeis guineensis

<400> 1
atgtcgtcct tctctttgct ctccttttct ctttgcttgt tgctcctgtg tcgtgtgtcc 60

caggctcaat ttgggtccag tcaggagagc ccatttcaga gctcacgacg tcggtcggta 120

tcaactcggg acgagtgccg gatggagagg ctgaatgccc tcgagccgac aaggactgtg 180

aggtccgaag ctggtatcac ggattacttt gatgaggaca atgaacagtt ccggtgcgcc 240

ggcgtgtccg cgatccgccg agtgatagaa cccagaggcc ttctcctccc ctccatgtcc 300

aatgcccctc gcctcgtcta catcgtccaa ggaaggggtc tcgttggact tgtaatgcct 360

ggctgccccg aaactttcca atccttccag cgatccgagc agtatgaacg cgaggaaggc 420

gagcgacatc ggagatccag agatgaacac caaaaagtct accagttcca agagggagac 480

gtcctggctg tgcctgatgg atttgcttac tggtgctaca acaatgggga gaaccctgtg 540

gtcgcgatca ccgtcctcga taccggtaac gatgctaacc agctcgatcg cagccacaga 600

caattcttat tggctggaag gcaggaggaa ggccggcaac gatatcgccg tggggagagc 660

atcaaggaaa acatcctgag agggttcagc accgagttgc tggcagcggc ttttggcgtt 720

aacatggagc tggcaaggaa gctccagtgc agggatgaca caaggggcga gaacgtgcgg 780

gcggagaacg ggcttcaggt gctgaggccc tcaaggatgg aggaagaaga gagggaagag 840

agtagaagga aaaatggatt ggaggagacc tattgctcga tgaagattaa gcagaacatt 900

ggggatccaa ggcgtgccga cgtcttcaac ccaaggggcg gaaggattac taccctcaac 960

agcgagaagc ttccgatcct taggttcatc caaatgagtg ctgagagggt ggttctctac 1020

aggaacgcca tggtatcacc ccactggaac atcaatgccc acagcatcat gtattgtacc 1080

ggaggacgag gccgtgtcga ggttgctgat gacaagggag aaactgcgtt cgacggagag 1140

ctccgccaag gtcagctctt gatcgtcccg caaaactatg cggtgtcgga aagggcggga 1200

agtgaaggct tccagttcgt atcgatcaag accagcgacc gggccatggt gagcgcaatc 1260

gttgggaaga cgtcggctct ccggggcatg ccggaggagg tgctgatgaa ctcctaccgc 1320

atctcaaggg acgaagcaag gagggtcaaa ctcaacaggg gggacgaggt ggcgatcttc 1380

actcctagga gagagtcgag agctgaagct tga 1413

<210> 2
<211> 470
<212> PRT
<213> Elaeis guineensis

<400> 2

Met Ser Ser Phe Ser Leu Leu Ser Phe Ser Leu Cys Leu Leu Leu Leu
1 5 10 15

Cys Arg Val Ser Gln Ala Gln Phe Gly Ser Ser Gln Glu Ser Pro Phe
20 25 30

Gln Ser Ser Arg Arg Arg Ser Val Ser Thr Arg Asp Glu Cys Arg Met
35 40 45

Glu Arg Leu Asn Ala Leu Glu Pro Thr Arg Thr Val Arg Ser Glu Ala
50 55 60

Gly Ile Thr Asp Tyr Phe Asp Glu Asp Asn Glu Gln Phe Arg Cys Ala
65 70 75 80

Gly Val Ser Ala Ile Arg Arg Val Ile Glu Pro Arg Gly Leu Leu Leu
85 90 95

Pro Ser Met Ser Asn Ala Pro Arg Leu Val Tyr Ile Val Gln Gly Arg
100 105 110

Gly Leu Val Gly Leu Val Met Pro Gly Cys Pro Glu Thr Phe Gln Ser
115 120 125

Phe Gln Arg Ser Glu Gln Tyr Glu Arg Glu Glu Gly Glu Arg His Arg
130 135 140

Arg Ser Arg Asp Glu His Gln Lys Val Tyr Gln Phe Gln Glu Gly Asp
145 150 155 160

Val Leu Ala Val Pro Asp Gly Phe Ala Tyr Trp Cys Tyr Asn Asn Gly
165 170 175

Glu Asn Pro Val Val Ala Ile Thr Val Leu Asp Thr Gly Asn Asp Ala
180 185 190

```
Asn Gln Leu Asp Arg Ser His Arg Gln Phe Leu Leu Ala Gly Arg Gln
        195                 200                 205

Glu Glu Gly Arg Gln Arg Tyr Arg Arg Gly Glu Ser Ile Lys Glu Asn
    210                 215                 220

Ile Leu Arg Gly Phe Ser Thr Glu Leu Leu Ala Ala Ala Phe Gly Val
225                 230                 235                 240

Asn Met Glu Leu Ala Arg Lys Leu Gln Cys Arg Asp Asp Thr Arg Gly
                245                 250                 255

Glu Asn Val Arg Ala Glu Asn Gly Leu Gln Val Leu Arg Pro Ser Arg
            260                 265                 270

Met Glu Glu Glu Glu Arg Glu Glu Ser Arg Arg Lys Asn Gly Leu Glu
        275                 280                 285

Glu Thr Tyr Cys Ser Met Lys Ile Lys Gln Asn Ile Gly Asp Pro Arg
    290                 295                 300

Arg Ala Asp Val Phe Asn Pro Arg Gly Gly Arg Ile Thr Thr Leu Asn
305                 310                 315                 320

Ser Glu Lys Leu Pro Ile Leu Arg Phe Ile Gln Met Ser Ala Glu Arg
                325                 330                 335

Val Val Leu Tyr Arg Asn Ala Met Val Ser Pro His Trp Asn Ile Asn
            340                 345                 350

Ala His Ser Ile Met Tyr Cys Thr Gly Gly Arg Gly Arg Val Glu Val
        355                 360                 365

Ala Asp Asp Lys Gly Glu Thr Ala Phe Asp Gly Glu Leu Arg Gln Gly
    370                 375                 380

Gln Leu Leu Ile Val Pro Gln Asn Tyr Ala Val Ser Glu Arg Ala Gly
385                 390                 395                 400

Ser Glu Gly Phe Gln Phe Val Ser Ile Lys Thr Ser Asp Arg Ala Met
                405                 410                 415

Val Ser Ala Ile Val Gly Lys Thr Ser Ala Leu Arg Gly Met Pro Glu
            420                 425                 430

Glu Val Leu Met Asn Ser Tyr Arg Ile Ser Arg Asp Glu Ala Arg Arg
        435                 440                 445
```

Val Lys Leu Asn Arg Gly Asp Glu Val Ala Ile Phe Thr Pro Arg Arg
450 455 460

Glu Ser Arg Ala Glu Ala
465 470

<210> 3
<211> 1346
<212> DNA
<213> Elaeis guineensis

<400> 3
tcaaattatc aaaatatctc caataatttt atcctgatgc ttttcttttt tttttagtaa 60

gaagtggtgg aagtggtatg ggtgtagtgg tgacacgaag aagcggtggc tggagggtgt 120

aaaaaattgt gggcatgggg ggagagggtg ggtgtgcgcg gaggagctaa gggggtggtg 180

gagacgagcg tggaggagcc gcaggggga gagtggtggg ggtggagggt agaggagcca 240

tggacaagcg agggggtgga ggagccacaa gtcgtggggg agcagtgtgc ttaaaggagc 300

tagtgggtgg tggggatgag catggaagag ctgcgggga tggtggggga tgagggagga 360

ggagtcatgg gcaggcaagg gggtggagaa gtcacaggtc aggagggtga taaaggagcc 420

atgggtaggt ggtggggcag aggagctgcg ggcaagcggt agagatggac gagccatagg 480

tgatgccggg ggaggggtcg ggaaggaggg ggagcgttgg gctatgggtt ttatgttttt 540

ctttttaagg gataatttta tccaaaattt ttattgcaat actgtcaaaa aagtggtaat 600

ctgattagcc atttctttcc gaggctattt aaattatctc ataggaggta atatagattg 660

tcaagacaat ttggattgcc acccaaaagc atcccaaatg cagtaatcca atggagccac 720

tttaaattgc tgacgatcta aatagattac cagctaccaa atgcaacata aaggctaact 780

cccaagcttg cattgcttgc taaatctgta gcacgtagtc cttggactct tgattttgtt 840

tgtcttttga aaaatttctt tggacttcag tatacttctc atctatcata ttatacctta 900

ggtcacttag cattttaga agaaagagat gttattttga aaataaaaaa ctctaacagc 960

atactaattg aacaagatca tcagaatact tggcaggacc aagagggata ggagctagcc 1020

cacccagcaa tatatataat tcactgaaat catctgatct actcttgcat tctgatgtgc 1080

ttgctgcaag cggttggtac ggttgtcctc ttccttccaa tatttacagt ccattcatcc 1140

aagtgtatga cacgtggcca cctgtattct gtcctctctg gaccaagctg ccacgcaccc 1200

atctccatat atcttatgcg gcaacacgag agctcatgca tggcaaagca tgggcgtcgt 1260

atctcacaaa agtgcatcca gatgggtcca agcctaaaca acttttttgg cctataaatg 1320

cccaccccat ctcgccaccc ttgctc 1346

<210> 4
<211> 20
<212> DNA
<213> Artificial

<220>
<223> amorce antisens KRA

<400> 4
cctgccttcc agccaataag 20

<210> 5
<211> 16
<212> DNA
<213> Artificial

<220>
<223> amorce d'ancrage KA1

<400> 5
cttccccccc cccccc 16

<210> 6
<211> 19
<212> DNA
<213> Artificial

<220>
<223> amorce nichée antisens KRB

<400> 6
gatcgcgacc acagggttc 19

<210> 7
<211> 19
<212> DNA
<213> Artificial

<220>
<223> amorce KTC

<400> 7
caatagcctt cagcgtttc 19

<210> 8
<211> 18
<212> DNA
<213> Artificial

<220>
<223> amorce KTD

<400> 8
gagaaggacg acatggct 18

<210> 9
<211> 28
<212> DNA
<213> Artificial

<220>
<223> amorce antisens GK4

<400> 9

gactggaccc aaattgagcc tgggacac 28

<210> 10
<211> 28
<212> DNA
<213> Artificial

<220>
<223> amorce nichée antisens GKN6

<400> 10
gaaagagaga aacgctgaag gctattgc 28

<210> 11
<211> 27
<212> DNA
<213> Artificial

<220>
<223> amorce GAK5

<400> 11
ccgctcgagt caaattatca aaatatc 27

<210> 12
<211> 24
<212> DNA
<213> Artificial

<220>
<223> amorce GAK4

<400> 12
cgcggatcca agagagaaac gctg 24

<210> 13
<211> 22
<212> DNA
<213> Artificial

<220>
<223> amorce EK1

<400> 13
aagcaatagc cttcagcgtt tc 22

<210> 14
<211> 20
<212> DNA
<213> Artificial

<220>
<223> amorce ANEK2

<400> 14
gcccttgctc gcaggcggag 20