

GENOMIQUE ET NUTRITION HUMAINE Biodisponibilité des acides gras et apports nutritionnels conseillés

Oléagineux, Corps Gras, Lipides. Volume 9, Numéro 2, 135-8, Mars - Juin 2002, Dossier : Génomique et filière oléagineuse - Journées Chevreul de l' AFECG, Paris, 21-22 novembre 2001

Auteur(s) : Nicole COMBE, Département de biochimie et de nutrition, Itegr, Université Bordeaux I, avenue des Facultés, 33405 Talence Cedex , France.

Author(s) : Nicole COMBE

Résumé : Les recommandations en terme d'apports en lipides évoluent en même temps que la connaissance de leurs effets sur la santé. La troisième édition des apports nutritionnels conseillés (ANC), pour la population française, a été publiée en 2001. L'examen de l'état des lieux, au travers de diverses données récentes d'enquêtes alimentaires, permet de souligner les écarts importants par rapport aux ANC. Il y a lieu de diminuer la consommation d'acides gras saturés [de 18 à 8 % de l'apport énergétique total quotidien (AETQ)] et d'augmenter celle des acides gras monoinsaturés (de 15 à 20 %). Il ressort également que l'apport en 18:3 n-3 est insuffisant (0,3 % au lieu de 0,8 % AETQ recommandé). Les stratégies à envisager pour optimiser l'apport en acides gras polyinsaturés doivent prendre en compte un certain nombre de critères, dont celui de leur biodisponibilité. Il est rappelé que la biodisponibilité d'un acide gras ne dépend pas uniquement de sa teneur au sein des triglycérides ingérés mais également de sa répartition entre les positions interne (sn2) et externes (sn1 + sn3) du squelette glycérol. À partir de corps gras « structurés » ou « inter-estérifiés », des études ont montré que l'athérogénicité de l'acide palmitique est plus importante lorsqu'il est situé en position sn2 des triglycérides alimentaires. C'est en effet cette position qui, d'une part, lui assure le coefficient d'absorption intestinale maximum et, d'autre part lui permet de se maintenir à cette même position dans les lipides plasmatiques, ce qui contribuerait à augmenter son pouvoir athérogène. S'agissant des acides gras insaturés, la biodisponibilité de l'acide linoléique comme précurseur d'acides gras essentiels serait corrélée à ses proportions en sn2 dans les triglycérides ; celle des acides gras polyinsaturés à longue chaîne est fonction de leur digestibilité qui est médiocre lorsqu'ils sont situés en positions sn1 et sn3, comparativement à la position sn2. Dans le cas de l'acide alpha-linolénique, dont le métabolisme oxydatif in vivo est très rapide, c'est son maintien en position interne des triglycérides plasmatiques qui pourrait lui assurer la protection contre une dégradation oxydative précoce, préjudiciable à sa fonction de précurseur d'acides gras essentiels. Sur cette base sont évoquées les capacités de l'huile de colza supérieures à celles de l'huile de soja, avec le maintien sur la position sn2 des triglycérides lymphatiques de proportions plus élevées en acide alpha-linolénique.

Summary : Population reference intake (PRI) of lipids continues to evolve as a better understanding of their health effects is gained. The last PRI for French people was published in 2001. Average linoleic acid supply ($4.4 \pm 1.8\%$ of total energy) is in accordance with the recommended value (4%), whereas alpha-linolenic acid supply is too low ($0.34 \pm 0.1\%$ versus 0.8%). Attaining the goal would require a 2-3 fold increase in alpha-linolenic acid intake. Strategies for improving supply of polyunsaturated fatty acids have to take account of their bioavailability. Several aspects of

bioavailability are examined in this report. It emphasizes the role of the stereospecific structure of dietary triglycerides, based on published literature. Enhanced absorption of fatty acids in the sn2 position of triglycerides has been shown. In animals, the atherogenic potential of interesterified fats was increased as the level of palmitic acid in the sn2 position was also increased. Moreover, the position (sn1, sn2 or sn3) of linoleic acid in triglycerides might influence its metabolism by desaturation and elongation into essential fatty acids, including lipoprotein metabolism. This is based on a probable link between dietary fatty acids that modulate fatty acid stereospecificity in phospholipids via absorbed sn2 monoglycerides.

Mots-clés : apports en acides gras, biodisponibilité, structure glycéridique.

Keywords : dietary fats, bioavailability, triglyceride structure.

ARTICLE

Depuis 1996, la génomique végétale connaît un développement fulgurant dont le témoignage le plus marquant a été l'achèvement fin 2000 de la séquence du génome de l'espèce modèle *Arabidopsis thaliana* [1]. La génomique ne saurait se limiter au simple séquençage du génome d'une espèce modèle ; elle comprend de nombreux aspects. Elle commence par la réalisation de cartes génétiques détaillées, sur lesquelles il est possible de positionner de nombreux mutants ou des QTL déterminant des caractères complexes. Elle se développe avec l'établissement de collections d'EST (*Expressed Sequence Tags*) ou d'étiquettes qui sont des séquences partielles d'ADNc [2] et l'établissement de séquences génomiques. Une fois ce travail d'obtention de l'information réalisé, il reste à l'analyser et à l'interpréter : il s'agit de déterminer ce pour quoi code un gène ou une séquence, de déterminer les limites précises des gènes, ou encore de comprendre la logique de l'organisation des différentes séquences les unes par rapport aux autres. Ce travail d'annotation des séquences constitue la première étape de la phase post-génomique et c'est aussi la plus difficile et la plus coûteuse. Du fait que de nombreux gènes ont leur séquence largement conservée d'une espèce à l'autre, il est souvent possible de les identifier et de se faire une idée de leur fonction par la recherche d'homologie de séquences, le plus souvent au niveau des séquences traduites en protéines. Cela suppose d'énormes investissements en bio-informatique. Malgré cela, le nombre total de gènes chez *Arabidopsis* n'est pas définitivement établi. Les premières analyses de la séquence faisaient état de 25 000 gènes environ [1], mais les plus récentes suggèrent que leur nombre pourrait être plus élevé, de l'ordre de 29 000. Une fois le catalogue de gènes dressé, il reste à valider leur fonction présumée : par l'analyse des mutants, la production de protéines recombinantes et d'anticorps, l'analyse des fonctions biochimiques et la production de plantes transgéniques dans lesquelles le gène est exprimé de façon ectopique, ou encore réprimé. Il reste aussi à déterminer où et quand chaque gène s'exprime. Ces deux ou trois dernières années ont vu l'irruption de l'analyse globale de l'expression d'un génome avec le développement des puces à ADN (ou *microarrays*) qui vient compléter ou guider l'analyse individuelle par *northern blots* ou RT-PCR et par hybridation *in situ*. Enfin, il reste à repérer et à identifier la protéine correspondant au gène et à détecter ses éventuelles modifications post-traductionnelles : c'est là le rôle des approches de protéomique. Ces types d'approches sont tous en œuvre chez *Arabidopsis* dont le génome est pratiquement complet, mais également chez de très nombreuses espèces de grande culture comme le riz, dont le brouillon du génome devrait être disponible au public fin 2002 [3], le maïs, le soja, le cotonnier et beaucoup d'autres.

L'étude du métabolisme lipidique bénéficie bien sûr de ces avancées spectaculaires et l'objet de cette revue est de dresser un premier bilan de ce que nous avons appris, des surprises que nous avons eues, et de ce qu'il reste à découvrir.

Le métabolisme lipidique recouvre des facettes multiples, liées aux trois grands rôles que jouent les lipides dans la cellule végétale :

- constitution des membranes cellulaires et des organites ;
- constitution de réserves dont la finalité est de permettre la germination dans de bonnes conditions, le temps que les jeunes plantules deviennent autotrophes, et dont l'importance pour l'alimentation humaine ou animale et pour l'industrie en fait un objet d'étude essentiel ;
- enfin, et c'est là un rôle encore peu exploré chez les plantes, leur rôle dans les processus de signalisation en réponse à des modifications environnementales.

Dans chacun de ces trois registres, étudier le métabolisme lipidique consiste à examiner quatre aspects :

- la biosynthèse de chaque type de molécule lipidique ;
- sa dégradation ;
- son transport d'un point à un autre de la cellule ;
- et enfin, la régulation de chacune de ces opérations.

L'apport de la génomique à l'étude du métabolisme lipidique va donc consister à recenser tous les gènes qui interviennent dans le métabolisme lipidique ainsi que leurs gènes régulateurs, à identifier et à annoter leur fonction, à analyser leur expression et leur régulation et finalement à tenter de manipuler ou de modifier le métabolisme des huiles dans les plantes.

Exploitation des séquences d'EST

Depuis les premiers programmes d'EST végétales [2, 4, 5], la masse d'information disponible a augmenté de façon quasi-exponentielle à partir de 1996. Rien que dans les bases de données publiques, il y a déjà plus d'un million de séquences provenant d'ADNc de plantes. Le *tableau 1* récapitule la situation présente et montre que leur nombre a très sensiblement augmenté au cours des derniers mois.

Ces EST peuvent être comparées aux séquences de gènes et de protéines déjà connues, ce qui constitue une première indication de la fonction. Cette stratégie impose qu'au moins un gène ou une protéine de la famille ait déjà été caractérisé biochimiquement. Dans le domaine du métabolisme des lipides, nombreux sont les exemples qui illustrent l'intérêt d'une telle démarche. Un premier exemple est celui des acyls ACP thioestérases, enzymes impliquées dans le relargage d'acides gras du complexe acide gras synthase dans le chloroplaste. L'enzyme spécifique des acides gras en C12 a été purifiée et l'ADNc cloné à partir d'*Umbellularia*, espèce endémique californienne accumulant l'acide laurique. La publication de cette information [6] a immédiatement permis la caractérisation d'homologues chez *Arabidopsis* [7] et celle des protéines correspondantes [8]. Cependant, l'analyse

des spécificités de substrats a conduit à différencier plusieurs sous-familles de gènes dont les fonctions biochimiques sont distinctes. Le premier gène de désaturase déterminant une oméga 3 désaturase [9] et le gène *FAE1* [10], isolés initialement par clonage positionnel (dont une mutation était le point de départ) pour le premier et par étiquetage par un transposon pour le second, ont ainsi permis d'identifier parmi les EST d'*Arabidopsis*, puis des autres espèces, pratiquement tous les gènes des mêmes familles, révélant bien souvent des situations plus complexes que prévu, avec de multiples copies de gènes très voisins. Pour progresser, il faut alors caractériser les protéines recombinantes, leur activité enzymatique et leur spécificité de substrat, cartographier les gènes correspondants sur le génome et les associer à des QTL. C'est ainsi que plusieurs enzymes clés et leurs gènes, qui contrôlent la teneur en acide linoléique ou en acide érucique [11-13], ont été identifiées chez le colza. Cette vérification de fonction est importante, car l'homologie de séquence n'est pas suffisante pour assigner une fonction.

Ainsi, le gène *FAD5* présente des homologies avec des acyl-désaturases, mais seule l'analyse biochimique et phénotypique des mutants associés à la séquence de l'EST a permis de montrer que sa fonction est de coder une monogalactosyldiacylglycérol palmitate désaturase [14]. De même, les hydroxylases du ricin ont des séquences très voisines des désaturases et seule l'analyse biochimique permet de différencier les deux fonctions [15]. Parfois, le gène correspondant à l'enzyme étudiée n'a jamais été caractérisé chez les plantes : c'était le cas de la DAGAT, la diacylglycérol acyltransférase dont le gène a été cloné à peu près simultanément par homologie avec un gène de rat, par marche chromosomique et étiquetage par une insertion [16]. Maintenant que les collections d'EST sont très complètes, on peut se livrer à des analyses exhaustives de type *data-mining*. Une telle étude a été rapportée récemment pour *Arabidopsis* [14] : un minimum de 30 étapes enzymatiques sont requises au plan biochimique pour synthétiser des acides gras à partir d'acyl CoA, et 65 protéines ont pu être identifiées. Elles sont représentées par 2 600 séquences d'EST qui permettent ainsi souvent de reconstituer des ADNc pleine longueur, ou de les isoler et de définir ensuite la structure fine des gènes correspondants. Les mêmes études sont réalisées, souvent dans le secteur privé, pour des espèces oléagineuses comme le soja, le maïs ou le colza. C'est également une stratégie efficace pour repérer rapidement les gènes impliqués dans la biosynthèse des acides gras inhabituels [17].

L'intérêt des EST n'est pas limité à l'identification des gènes. La fréquence des EST reflète assez fidèlement le niveau d'expression des gènes. Il est donc possible, à partir de l'obtention d'un grand nombre d'EST sur des tissus, ou dans des situations physiologiques différentes, d'obtenir des indications sur les voies métaboliques prépondérantes dans ces conditions. Il est aussi possible de classer les EST en groupes de fonctions présumées. Pour que ces analyses soient valides, il faut que le séquençage des EST ait été réalisé sans biais majeur, ce qui n'est pas toujours le cas : pour des raisons de coût, il est souvent procédé à la soustraction des séquences redondantes. Une analyse exhaustive a ainsi été réalisée sur 28 000 ADNc préparés à partir de graines d'*Arabidopsis* [18]. Après soustraction des clones représentant les ARNm les plus abondants (protéines de réserve), 10 522 clones ont été séquencés et analysés. Le *tableau 2* donne un aperçu du type de résultats obtenus : 4,7 % des clones correspondent à des enzymes du métabolisme lipidique. Il est à noter que, dans cette expérience s'adressant à un organe spécialisé, près de 40 % des clones correspondent à des séquences nouvelles, non recensées dans les bases de données et que 38 % sont sans homologie significative avec quoi que ce soit, c'est-à-dire que leur fonction reste à découvrir.

L'analyse de la fréquence des EST a été particulièrement utile pour déterminer quelles sont les voies prédominantes pour convertir les assimilats photosynthétiques en huile [18]. Les EST correspondant à différents groupes de réaction ont été identifiées et leur fréquence évaluée (*figure*). Ainsi, les EST les plus importantes sont celles de l'aldolase, de la glycéraldéhyde-3-phosphate deshydrogénase, de l'énolase et de la pyruvate-kinase cytosolique montrant que la glycolyse cytoplasmique est sans doute très active. Les contributions de la synthèse et de la dégradation de l'amidon et celles du cycle des pentoses phosphates paraissent mineures. Au niveau des transferts de squelettes carbonés du cytoplasme vers le chloroplaste, ce sont le translocateur de glucose-6-phosphate et celui de phosphoénol pyruvate qui paraissent jouer le rôle majeur. Par exemple, aucune EST n'a été trouvée pour le transporteur de pyruvate. L'observation d'une fréquence élevée d'EST de pyruvate deshydrogénase, d'acétyl CoA carboxylase, des sous-unités de l'acide gras synthase et de la stéaroyl-ACP est en accord avec la forte activité de synthèse d'acides gras. Ces données, que l'on qualifie parfois de *Northern numérique*, sont cependant à prendre avec précaution car elles ne reflètent que l'abondance relative des ARNm et non pas nécessairement celle des protéines, ni le niveau de leur activité spécifique.

Exploitation des séquences génomiques

Bien souvent, les EST donnent une vision incomplète de l'équipement génique. En effet, on ne trouve une EST que si le gène est exprimé à un niveau suffisant dans les conditions où la banque d'ADNc a été réalisée. La disponibilité d'une séquence complète ou quasi complète permet maintenant de réaliser des catalogues exhaustifs. Cette opération n'est pour le moment possible que chez *Arabidopsis*, seul génome de plante séquencé avec un niveau de saturation et de fiabilité suffisant. Ce type d'analyse révèle l'existence de familles multigéniques. Celles-ci correspondent souvent à l'existence d'isoenzymes compartimentées dans différents organites (chloroplaste, mitochondrie) ou territoires cellulaires (cytoplasme, réticulum endoplasmique, Golgi, vacuole, paroi, etc.). Ces familles multigéniques sont aussi liées à la nature souvent polyploïde cryptique des espèces végétales. Ainsi, le maïs et *Arabidopsis*, pourtant considérés comme des diploïdes vrais ont pratiquement tous leurs gènes en double exemplaire, ce qui a sans doute permis une spécialisation fonctionnelle de chaque isoforme, qui reste souvent à découvrir. Le *tableau 3* donne quelques exemples du nombre de membres dans plusieurs familles multigéniques impliquées dans le métabolisme lipidique.

Notre laboratoire a plus particulièrement travaillé sur la famille des gènes de LPAAT (acide lysophosphatidique acyltransférase), l'enzyme responsable de l'insertion d'acide gras en position sn-2 du squelette glycérol. Chez le colza, c'est l'enzyme limitante pour la biosynthèse de la triérucine, puisque la LPAAT de colza impliquée n'accepte pas les acides gras à chaîne longue comme substrat [19]. Les premiers gènes de LPAAT ont été isolés soit par complémentation d'un mutant déficient d'*E. coli*, soit par purification biochimique [20, 21]. Le fait que les premiers gènes isolés n'aient pas les propriétés d'expression attendues a induit à rechercher d'autres gènes par *data-mining*. Dix gènes ont maintenant été repérés et localisés dans le génome d'*Arabidopsis* et davantage chez le colza. L'analyse phylogénétique de leur séquence révèle qu'ils se répartissent en trois classes, l'une de type procaryote, correspondant sans doute aux isoenzymes localisées dans le chloroplaste et la mitochondrie, l'autre aux enzymes localisées dans les microsomes et le plasmaleme et la troisième à des protéines dont la fonction réelle est inconnue. Dans cette situation, la génomique nous donne

accès à des catalogues de gènes, elle nous permet de les classer par famille et nous fournit les outils qui vont permettre d'en préciser les fonctions. Le travail en cours consiste à déterminer par RT-PCR dans quels tissus et à quelle étape du développement chacun de ces gènes est exprimé, à vérifier que chacun de ces gènes code pour une protéine qui a effectivement une activité LPAAT et à en définir les spécificités de substrat.

Outils d'analyse de l'expression et de la fonction des gènes

La génomique permet d'élaborer des catalogues de gènes. Elle a également permis la mise en œuvre de moyens d'analyse globale de l'expression des gènes et d'étude de leur fonction.

Une première approche évoquée plus haut consiste à réaliser des *Northern numériques*. Cependant, ce type de méthode est lourd et ne permet pas de réaliser un grand nombre d'analyses comparatives. Celles-ci sont devenues possibles avec la réalisation des puces à ADN. Les premières puces utilisées comportaient un nombre limité de séquences d'EST sur les lames de verre soumises à hybridation. La seule étude publiée à ce jour s'intéressant au métabolisme lipidique compare les niveaux d'expression de 2 600 séquences exprimées au niveau des graines et de la feuille [22]. Cette étude a révélé que 25 % des gènes étudiés sont deux fois plus exprimés dans la graine que dans la feuille et que 10 % y sont dix fois plus exprimés dans la feuille. De façon générale, les gènes des protéines du métabolisme lipidique sont légèrement sur-exprimés dans la graine par rapport à la feuille, avec quelques gènes, tels ceux des oléosines, d'une lipoxygénase et d'une lipase, nettement sur-exprimés. Parmi les gènes sur-exprimés, on note ainsi 28 ADNc codant des facteurs de transcription ou des protéines kinases dont les rôles biologiques sont souvent inconnus. Il est probable que quelques-uns des gènes ainsi mis en évidence régulent des gènes ou des enzymes impliqués dans le métabolisme lipidique.

On est loin, avec 2 600 gènes, d'une analyse globale et il faut des différences dans les niveaux d'expression relative supérieures à un facteur 2 pour qu'elles puissent être prises en compte dans l'analyse. On n'est pas sûr non plus que ces puces discriminent correctement les membres d'une même famille multigénique. Enfin, ces informations sont encore trop grossières pour repérer les gènes régulateurs du métabolisme lipidique.

Les nouvelles générations de puces, et en particulier les puces pan-génomiques à oligonucléotides spécifiques de chaque gène, devraient pallier ces inconvénients lorsqu'elles seront disponibles. En effet, pour définir des oligonucléotides spécifiques de chaque gène, il est indispensable de réannoter le génome de façon beaucoup plus fiable que ne le permet l'annotation automatique et grossière proposée actuellement.

En matière d'analyse fonctionnelle, les mutants continuent à jouer un très grand rôle. C'est grâce à eux et à l'analyse de leur composition en huile qu'il a été possible d'identifier les gènes de désaturase ou d'elongase [9, 10], enzymes membranaires qu'il est pratiquement impossible de purifier par des méthodes standard. Avec la constitution systématique de collections de mutants d'insertion, une nouvelle stratégie devient possible. Pour n'importe quel gène de fonction inconnue et pour lequel on dispose d'une séquence, il devient possible d'aller rechercher la lignée qui contient une insertion dans ce gène et d'en faire une analyse phénotypique détaillée. Cette recherche peut être faite directement par des cribles PCR, ou mieux maintenant, par simple recherche dans des bases de données de séquences flanquant les insertions (FST - *Flanking Sequence Tags*) [23]. Plusieurs groupes

ont en effet entrepris de séquencer systématiquement les FST associées à chaque lignée et de les mettre en accès public dans des conditions préservant leur propriété intellectuelle. Pour *Arabidopsis*, le simple alignement d'une séquence d'un gène avec une FST permet immédiatement de savoir qu'il y a une lignée mutante pour ce gène dont il sera possible d'analyser en détail le phénotype, permettant ainsi de substituer une analyse ciblée à un tri aléatoire sur l'ensemble de la population.

CONCLUSION

Cette revue a tenté de montrer comment les développements de la génomique ont permis de progresser dans la connaissance du métabolisme lipidique. Les données disponibles portent principalement sur *Arabidopsis*, mais elles permettent d'ores et déjà de s'intéresser aux nombreuses espèces à graines oléagineuses. La génomique a surtout révélé des situations beaucoup plus complexes que prévu, avec de nombreuses familles multigéniques qu'il est indispensable de caractériser de façon détaillée du point de vue de la fonction spécialisée de chaque gène. Ce n'est qu'à ce prix que l'on pourra comprendre les problèmes de compensation d'un gène par un autre et de compartimentation des synthèses ou des dégradations.

Il est clair que, si la grande majorité des gènes impliqués dans la biosynthèse des acides gras et des lipides membranaires et de réserve est maintenant identifiée, de nombreux détails sont encore manquants. La situation des processus de dégradation et de ré-allocation d'une membrane à l'autre est moins bien connue et constitue encore un enjeu important. Le rôle des lipides dans les processus de signalisation cellulaire, les enzymes impliquées et les gènes qui les contrôlent sont encore presque inconnus chez les plantes.

Enfin, une grande question reste encore sans réponse et devrait faire l'objet de développements importants dans le futur : quels sont les facteurs de transcription qui contrôlent l'activation des voies de biosynthèse des lipides des graines et peut-on les manipuler pour tenter d'augmenter la production globale d'acide gras par rapport aux protéines sans poser des problèmes physiologiques au reste de la plante ? Il est nécessaire de mettre en jeu des stratégies pour tenter de répondre à ces questions. La génomique a donc induit l'entrée de l'étude du métabolisme des lipides dans une toute nouvelle phase dans laquelle, plus que l'identification des espèces moléculaires et des enzymes qui catalysent leur formation, c'est maintenant la recherche des éléments de régulation qui devrait recevoir la priorité des recherches fondamentales.

Remerciements

Les auteurs remercient particulièrement leurs collègues Michel Renard, René Lessire, Jean-Claude Kader, Philippe Guerche et Françoise Labalette pour les nombreuses et stimulantes discussions qui ont sous-tendu leur programme de recherche. Ils ont bénéficié pour ce travail du soutien du CNRS (UMR 5096), du ministère de la Recherche et de la Technologie, de l'Organisation nationale interprofessionnelle des oléagineux et du Centre technique interprofessionnel des oléagineux métropolitains.

REFERENCES

1. *Arabidopsis* genome initiative (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, 408 : 796-826. <http://www.arabidopsis.org/agi.html>.
2. COOKE R, RAYNAL M, LAUDIE M, *et al.* (1996). Further progress towards a catalogue of all *Arabidopsis* genes: analysis of a set of 5,000 non-redundant ESTs. *Plant J*, 9 : 101-24.
3. DELSENY M, SALSE J, COOKE R, *et al.* (2001). Rice genomics: present and future. *Plant Physiol Biochem*, 39 : 324-34.
4. HOFTE H, DESPREZ T, AMSELEM J, *et al.* (1993). An inventory of 1,152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J*, 4 : 1051-61.
5. NEWMAN T, de BRUIJN FJ, GREEN P, *et al.* (1994). Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol*, 106 : 1241-55.
6. VOELKER TA, WORRELL AC, ANDERSON L, *et al.* (1992). Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. *Science*, 257 : 72-4.
7. GRELLET F, COOKE R, RAYNAL M, LAUDIE M, DELSENY M (1993). *Arabidopsis* systematic cDNA sequencing reveals a gene with homology to *Umbellularia californica* C12:0ACP-thioesterase. *Plant Physiol Biochem*, 31 : 599-602.
8. JONES A, DAVIES HM, VOELKER TA (1995). Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. *Plant Cell*, 7 : 359-71.
9. ARONDEL V, LEMIEUX B, HWANG I, GIBSON S, GOODMAN HM, SOMERVILLE CR (1992). Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. *Science*, 258 : 1353-5.
10. JAMES DW, LIM E, KELLER J, PLOOY I, DOONER HK (1995). Directed tagging of the *Arabidopsis* Fatty Acid Elongation (*FAE1*) gene with the maize transposon *Activator*. *Plant Cell*, 7 : 309-19.
11. JOURDREN C, BARRET P, BRUNEL D, DELOURME R, RENARD M (1996). Specific molecular markers of the genes controlling the linolenic acid level in rapeseed. *Theor Applied Genetics*, 93 : 512-8.
12. JOURDREN C, BARRET P, HOVAIS R, FOISSET N, DELOURME R, RENARD M (1996). Identification of RAPD markers linked to the loci controlling erucic acid level in rapeseed. *Molecular Breeding*, 2 : 61-71.
13. BARRET P, DELOURME R, RENARD M, *et al.* (1998). A rapeseed *FAE1* gene is linked to the *E1* locus associated with variation in the content of erucic acid. *Theoretical and Applied Genetics*, 96 : 177-86.
14. MEKHEDOV S, de ILARDUYA OM, OHLROGGE J (2000). Toward a functional catalog of the plant genome. A survey of genes for lipid biosynthesis. *Plant Physiol*, 122 : 389-402.
15. VAN DER LOO FJ, BROUN P, TURNER S, SOMERVILLE CR (1995). An oleate 12 hydroxylase from *Ricinus communis* L. is a fatty acyl desaturase homologue. *Proc Nat Acad Sci USA*, 92 : 6743-7.

16. ROUTABOUL JM, BENNING C, BECHTOLD N, CABOCHE M, LEPINIEC L (1999). The TAG1 locus of *Arabidopsis* encodes for a diacylglycerol acyltransferase. *Plant Physiol Biochem*, 37 : 831-40.
17. VAN DER LOO FJ, TURNER S, SOMERVILLE CR (1995). Expressed sequence tags from developing castor seeds. *Plant Physiol*, 113 : 1009-13.
18. WHITE JA, TODD J, NEWMAN T, *et al.* (2000). A new set of *Arabidopsis* expressed sequence tags from developing seeds. The metabolic pathway from carbohydrates to seed oil. *Plant Physiol*, 124 : 1582-94.
19. BERNETH R, FRENTZEN M (1990). Utilization of erucoyl-CoA by acyltransferases from developing seeds of *Brassica napus* (L.) involved in triacylglycerol biosynthesis. *Plant Sci*, 67 : 21-8.
20. BROWN AP, COLEMAN J, TOMMEY AM, WATSON MD, SLABAS AR (1994). Isolation and characterisation of a maize cDNA that complements a 1-acyl sn-glycerol-3-phosphate acyltransferase mutant of *Escherichia coli* and encodes a protein which has similarities to other acyltransferases. *Plant Mol Biol*, 26 : 211-23.
21. BOURGIS F, KADER JC, BARRET P, *et al.* (1999). A plastidial lysophosphatidic acid acyltransferase from oilseed rape. *Plant Physiol*, 120 : 913-22.
22. GIRKE T, TODD J, RUUSKA S, WHITE J, BENNING C, OHLROGGE J (2000). Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol*, 124 : 1570-81.
23. SAMSON F, BRUNAUD V, BALZERGUE S, *et al.* (2002). FLAGdb/FST: a database of mapped flanking insertion sites (FSTs) of *Arabidopsis thaliana* T-DNA transformants. *Nucleic Acids Res*, 30 : 94-7.

Illustrations

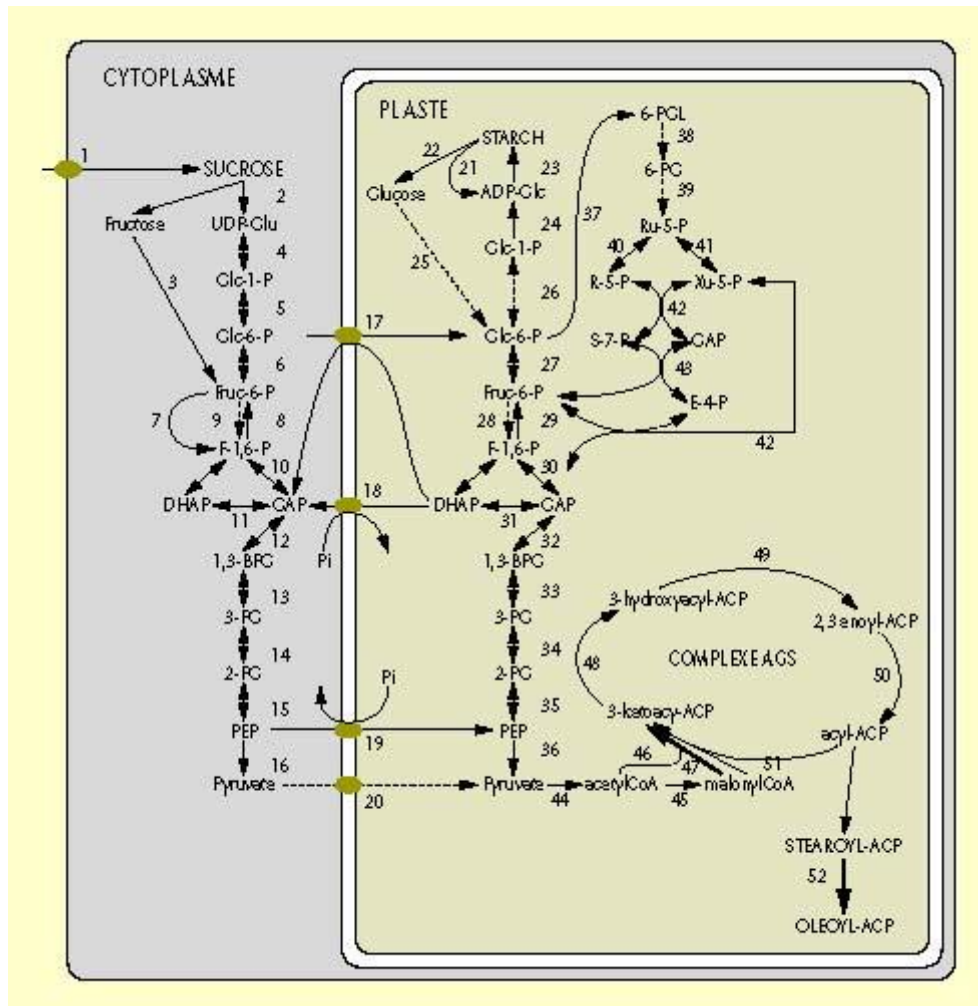


Figure. Transformation des assimilats photosynthétiques en huiles : métabolisme des carbohydrates et biosynthèse des acides gras dans l'embryon immature d'Arabidopsis (modifié à partir de White et al. [18]). Conversion du saccharose par glycolyse cytoplasmique (1-16) ; transfert vers le plaste (17-20) ; synthèse et dégradation de l'amidon (21-26) ; glycolyse chloroplastique (27-36) ; cycle des pentoses phosphates (37-43) ; complexe pyruvate déshydrogénase (44) ; synthèse des acides gras (45-52).

Tableau 1. *Etat des collections publiques d'EST de plantes cultivées.*

EST publiques	12.11.2001	21.03.2002
Soja	208 198	241 930
Tomate	141 687	148 350
Luzerne (<i>M. truncatula</i>)	137 588	138 957
<i>Arabidopsis thaliana</i>	113 330	113 330
Maïs	108 342	150 398
Riz	99 678	104 589
Bé	68 678	73 386
EST privées		
Toutes espèces confondues	> 1 500 000	
Maïs	> 500 000	

Tableau 2. *Analyse des EST de graines immatures d'Arabidopsis thaliana (d'après White et al. [18]).*

Classe fonctionnelle	%
Métabolisme des acides aminés	2,6
Métabolisme des sucres	6,7
Pari cellulaire	2,5
Développement	3,0
Métabolisme lipidique	4,7
Protéinases, ubiquitine	3,1
Ribosome, traduction protéines	3,1
Protéines de réserve	14,4
Transduction de signal	4,0
Homologie non-significative	24,3
Fonction non identifiée	13,5

Tableau 3. *Estimation des tailles (en nombre de gènes) des familles de gènes codant des enzymes du métabolisme lipidique chez Arabidopsis et le riz (d'après Mekhedov et al. [14]).*

Protéine	Estimation de la taille de la famille	
	<i>Arabidopsis</i>	Riz
Acyl Carrier Protein	8	6
Stearoyl ACP desaturase	4	4-9
Long chain acyl CoA synthetase	18-25	5-18
Phospholipase C	10-11	3-5
Phospholipase D	17-20	5-9
Keto-acyl CoA synthase	21-26	4-8
Lyso-phosphatidic acid acyltransferase	4-10	2-4
Oleosine	6-7	5-6
Acyl CoA oxydase	5	2-4
3-ketoacyl CoA thiolase	7	5-7